Seoul Journal of Business
Volume 5, Number 1/2 (December 1999)

Buy-Sell Dependence and Classification Error in Market Microstructure Time-Series Models: A Markov Switching Regression Approach

Hyuk Choe
College of Business Administration
Seoul National University

Abstract

This paper conducts an empirical test of a market microstructure model using a new econometric approach. I treat the direction of a trade as a discrete latent variable following a stationary Markov chain. By overlaying a three-state Markov chain on a familiar market microstructure model, I can extract information on the directions of trades efficiently from time-series data. An analysis of 100 large and 100 small firms for the year 1990 yields several important results: (1) Order types (sale, cross, purchase) are serially correlated, and the mean transition probability matrix is very similar for large and small firms. (2) Information asymmetry is greater for smaller firms. (3) The per share order processing cost is greater for larger firms. (4) When trades are classified by the bid-ask test supplemented by the tick test, the estimated misclassification probabilities are typically small for sales and purchases, but they are often fairly large for crosses. (5) Buy-sell classification error results in systematic biases for regression coefficients.

I am grateful to Peter Algert, Charles Cao, Stephen Gray, Joel Hasbrouck, Frank Hatheway, Robert Korajczyk, William Kracaw, Joon-Haeng Lee, Ronald W. Masulis, Harold Mulherin, Dennis Sheehan, three anonymous referees, and session participants at the 1994 Western Finance Association Meetings (Santa Fe), the 1994 joint meetings of the Korean Finance Association and the Korean American Finance Association (Seoul) and the 1994 Financial Management Association Meetings (St. Louis), and seminar participants at Baruch College for helpful comments on earlier versions. The Institute of Management Research at Seoul National University provides financial support for this research.

1. Introduction

Central to many empirical studies of market microstructure is the determination of whether a trade is buyer- or seller-initiated. For example, time-series models of intraday stock price changes are concerned with the impacts of trades on prices. Since the effect of a trade on price changes depends on whether the trade is buyer- or seller-initiated, variables representing the direction of a trade play critical roles in these models. In addition, the estimated coefficients of the trade initiation and volume variables signed according to the direction of a trade often have direct implications for the severity of information asymmetry, the magnitude of order processing costs and the market maker's inventory considerations in setting prices. Important recent studies of time-series models that fall in this category include those of Ho and Macris (1984), Glosten and Harris (1988), Hasbrouck (1988, 1991a), Madhavan and Smidt (1991), Hausman, Lo and MacKinlay (1992), and Foster and Viswanathan (1993).

Publicly available transaction databases do not distinguish between buyer- and seller-initiated trades. Thus, researchers are forced to use a buy-sell classification scheme and use the classifications as exogenous data in a time-series model (or in an event study). This practice has two potential problems, which have not been properly addressed in previous studies. First, treating buy-sell classifications as exogenous data assumes that order types are serially independent.[1] However, Hasbrouck and Ho (1987) report that purchases tend to follow purchases and sales tend to follow sales. Thus, the current practice of ignoring order dependence underutilizes information contained in the sequence of order flows.

Second, this approach assumes that trades are perfectly classified into sales and purchases. However, misclassified trades, if they exist, cause biased and inconsistent coefficient estimates due to the well-known errors in variables problem. Furthermore, misclassified trades affect not only the coefficients

1) In this paper, the term 'order type' has the same meaning as the 'direction of trade.'

of the signed trade and the signed volume variables, but also the coefficients of other variables (such as the lagged price changes) measured without error. In particular, classification errors for very large trades could be potentially detrimental when a signed volume variable is a regressor because the magnitude of the measurement error is proportional to the trade volume.

This paper proposes a general approach that can effectively extend many market microstructure models by accounting for both serial dependence in order types and buy-sell classification errors in an intuitive manner. The main idea is that a Markov chain can be used as a general way of modeling the order dependence, where the state variable is the direction of a trade (sale, cross, purchase). In essence, the proposed approach overlays a Markov chain on a familiar market microstructure model. It is further assumed that the 'true' directions of trades are not directly observable.

A traditional classification indicator may or may not be used. If used, it is treated as a proxy for the direction of a trade with error. Then, the resulting econometric model is essentially a variant of the Cosslett and Lee (1985) model. Cosslett and Lee show that by imposing a specific probability structure on classification errors, it is possible to estimate the parameters of the time-series model simultaneously with the two layers of probabilities (that is, order type transition probabilities and misclassification probabilities). On the other hand, if a classification variable is not used, the model becomes the Hamilton model (1989, 1990). Hamilton shows that this type of model is identifiable with the Markov chain assumption alone.

In principle, the present approach can be used for any microstructure time-series model that requires a buy-sell indicator variable. In this paper, I choose the Madhavan and Smidt model as the base market microstructure model for several reasons. First, it is one of the most generally used in the literature and nests many previous models. Second, the model yields natural measures of information asymmetry and order processing costs. These measures are of interest not only to researchers but also to practitioners and policy makers. Third, Madhavan and Smidt examine only sixteen stocks obtained from a specialist firm.[2)] Thus, it is worthwhile to provide more extensive evidence on information asymmetry and order processing costs beyond the few stocks considered in their original study.

I analyze a sample of 100 large and 100 small firms for the year 1990. My findings confirm that order types are indeed serially correlated. The mean conditional probability that the next trade is the same type as the current trade is in the range between 0.52 and 0.55. The mean estimate is very similar across the three order types (sale, cross, purchase). The coefficient estimates also indicate that smaller firms exhibit a greater information asymmetry than larger firms. This evidence is intuitively appealing because more information is likely to be produced for larger firms. Finally, the estimated per share order processing cost is greater for larger firms (6.1 cents) than smaller firms (5.0 cents).

In addition to these results, the new approach, as a by-product, provides information that is useful in addressing two important questions: i) How accurate are popular buy-sell classification schemes? And ii) how large is the impact of classification errors on previously reported results?[3] I find that, when trades are classified by the bid-ask test supplemented by the tick test,[4] the estimated misclassification probabilities are typically small for sales and purchases (around 1% at the 75th percentile for large firms, around 3% for small firms) but are often large for crosses (around 2~5% at the 75th percentile for large firms, around 10~11% for small firms). A comparison of the coefficient estimates from the conventional approach with

2) One of their objectives is to understand the impacts of specialists' inventory positions on prices. The availability of inventory data limits their sample to one specialist firm.

3) Obviously, a better way to answer these questions is to examine a data set that allows researchers to directly observe sales and purchases. Such an approach is infeasible in most circumstances without acquiring proprietary data. Among publicly available U.S. databases, the NYSE TORQ (Transactions, Quotes, Order Processing) data set is the only one that provides detailed information on the directions of trades. However, it has problems: i) it covers only three months (including year-end) for a limited sample of 144 NYSE stocks, and ii) since it includes only orders submitted to the NYSE via the SuperDot system, it omits many large trades.

4) The bid-ask test classifies a given trade as being seller-initiated (buyer-initiated) if its transaction price is below (above) the midpoint of the bid and ask quotes that prevail at the time of the trade. The tick test classifies a trade as being seller-initiated (buyer-initiated) if its transaction price is smaller (greater) than the price of the previous trade.

those obtained from my three-state Markov model suggests that classification errors often result in systematic biases in the regression coefficients.

The rest of the paper is organized as follows: Section 2 presents a three-state Markov model of intraday stock price changes in which purchases, sales and crosses are imperfectly observed. Section 3 provides a brief description of the model estimation procedure. Section 4 explains how the time-series data are constructed from the original transaction database. Section 5 provides the empirical results for a sample of firms. Finally, Section 6 concludes this paper.

2. The Model

Since the construction of a buy-sell classification variable is an important building block in the present model, this section first explains the classification scheme used and then describes the time-series model.

2.1 A Modified Bid-ask Test

A buy-sell classification scheme commonly found in the literature is the bid-ask test which compares the price of a given trade to the specialist's quote prevailing at the time of the transaction. There are several variants of the bid-ask test. Among them, the simplest form is to treat a trade as seller-initiated (buyer-initiated) if the trade is made at the bid (ask). This is because on the NYSE or AMEX the specialist posts his bid and ask quotation prior to a trade, and honors it when an investor wants to buy or sell stocks. The most frequently used form is to classify a trade as being seller-initiated (buyer-initiated) if the price of the trade is below (above) the midpoint of the bid and ask.

The bid-ask test leaves many trades unclassified because many transaction prices fall exactly in the middle of the bid and ask. These quote-midpoint transactions arise for various reasons. First, the existence of hidden limit orders may be a cause. McInish and Wood (1995) report that many limit orders that are better than the standing quotes are not posted as quotes. Thus, market orders matched to these hidden limit orders are executed at the price within the posted bid and ask. Since price moves in a discrete fashion ($1/8), the chance of quote midpoint transactions would be nontrivial. Second, floor brokers' participation in trades may also contribute to the problem. Since floor brokers compete with the specialist, the situation is very similar to the case of hidden limit orders.[5] Third, this problem may be caused by the discrepancy between the actual sequence of trades and quotes and the recorded sequence in the database. For example, the transmission speed of trades is likely to lag that of quotes. If the specialist tends to shift his or her bid-ask quote centering on the previous transaction price, the missequenced trade looks like a quote midpoint transaction.

Faced with the nonclassification problem, the volume of an unclassified trade often enters the given time-series model with a zero value, while the volume of a classified trade is signed depending on the direction of the trade. Although some studies exclude unclassified trades from the sample, this practice is not desirable in the context of a time-series model because arbitrary exclusion of observations may alter the lag structure.

Lee and Ready (1991) present an algorithm to alleviate this problem. Following their recommendation, researchers often classify a quote midpoint trade as a sale (purchase) if it is a downtick (uptick) transaction. I call such a sequential procedure of supplementing the bid-ask test with the tick test the 'modified bid-ask test.' While the modified bid-ask test classifies more trades, numerous unclassified trades still remain. In addition, there is no guarantee that the additionally classified trades are correct. Since the modified bid-ask test is the most frequently used in the literature, it is used throughout this paper.

While previous studies tend to make an effort to reduce the number of unclassified trades, such an effort may lead to over-classification. The underlying assumption of this practice is that the specialist participates in all trades, and thus, transactions must be either buyer- or seller-initiated. In reality, the specialist on the NYSE participated in only 19.9% of trades measured by volume during 1990 (See NYSE Fact Book 1991). The majority of the NYSE transactions are agency-to-agency trades in which the

---

5) See Hasbrouck (1988, pp.240-241) for details.

distinction between buyer- and seller-initiation is blurred. Many agency-to-agency trades are initiated by both buyers and sellers. Each side contacts a broker and places instructions. Also, each side may or may not think that they have valuable information. Therefore, it is important to recognize that not all trades can be classified as pure sales or purchases.

2.2 A Markov Model of Market Microstructure

Consider a simple time-series model of intraday price changes:

$$\begin{aligned} r_t &= \phi_1 r_{t-1} + \phi_2 r_{t-2} + \alpha_0 i_t x_t + \alpha_1 i_{t-1} x_{t-1} + \alpha_2 i_{t-2} x_{t-2} \\ &\quad + \gamma_0 i_t + \gamma_1 i_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \quad t = 3, 4, \cdots, T, \end{aligned} \tag{1}$$

where r_t is the price change (or a measure of return) between time t-1 and time t, x_t is an appropriately transformed trade size (shares or dollar volume), and i_t is an indicator which takes -1 for a seller-initiated trade and 1 for a buyer-initiated trade.[6] Since not all trades can be classified as pure sales or purchases as discussed in the previous section, i_t may take 0.[7] Thus, there are three states: sale, cross, and purchase. This model does not include a constant term because the expected transaction-by-transaction return is likely to be close to zero. The number of lags in Equation (1) is arbitrary.[8]

Equation (1) is a prototype of many time-series models in the market microstructure literature, and the estimated coefficients have natural interpretations. For example, the γ coefficients capture the price impact of the information conveyed by a trade. Since the information content of a trade is presumably an increasing function of the trade size [Easley and O'Hara (1987)], the sum of the α coefficients is expected to be positive. The γ coefficients have an implication for both the information effect and the order processing cost. In Glosten and Harris (1988) and Madhavan and Smidt (1991), -γ_1 represents the per share order processing cost; therefore, the estimated γ_1 is expected to be

6) Examples include log transformation or Box-Cox transformation. This paper does not transform the trade size for the ease of interpretation.

7) I label i_t=0 'cross' for expositional convenience.

8) There is an alternative specification. For example, to avoid the arbitrariness of the state classification, one may want to give a restriction that enforces the α's to be positive.

negative.[9] Further, in Madhavan and Smidt, $-\gamma_1/\gamma_0$ is a measure of information asymmetry, and should lie between zero and one.[10] This is a unique feature of the Madhavan and Smidt model where the market maker updates his belief using the Bayesian updating rule. The ratio of $-\gamma_1/\gamma_0$ represents the specialist's weight on prior beliefs, and it is inversely related to the degree of information asymmetry in the market. In many other models, this ratio should be one.

While the base model is essentially the Madhavan and Smidt model, there are several differences. First, Madhavan and Smidt examine the effect of inventory adjustments on price dynamics. Since they find that the inventory effect is weak, I ignore it. Second, unlike the Madhavan and Smidt model, Equation (1) includes the lagged price change variables. This is intended to control for the autocorrelation in price changes induced i) by bid-ask bounce that is not fully captured by i_t variables, or ii) by unknown factors that are not specified in the model.[11] Finally, Equation (1) includes variables representing the lagged volume effect ($i_{t-1}x_{t-1}$ and $i_{t-2}x_{t-2}$). These lagged volume variables capture i) possible lagged price adjustments to a trade[12] and ii) the impact of the price continuity rule.[13]

I assume that the sequences of r_t and x_t are observed but the sequence of i_t is not directly observed. Thus, a proxy for the buy-sell indicator, j_t, is constructed based on the modified bid-ask test as described in the previous subsection. The observed indicator j_t can take a value from {-1, 0, 1} (that is, sale, nonclassification and purchase).

9) Since the present paper incorporates 'cross', the magnitude of the coefficient would be smaller than that reported in Madhavan-Smidt.

10) Thus, γ_0 is expected to be positive and greater than $-\gamma_1$.

11) The original Madhavan and Smidt model has an MA(1) error. Unfortunately, incorporating an MA error structure in my model is difficult with current technology, because the likelihood function depends on the entire history. I hope that the inclusion of the lagged price change variables alleviates the problem caused by this omission.

12) The specialist in Madhavan and Smidt makes price adjustments ('regret-free') in anticipation of the trade. Thus, the price instantly responds to the trade. But, if the specialist revises quotes in response to the information inferred from the trade, the lagged adjustment may arise. See Hasbrouck (1991a).

13) The NYSE obligates the specialist to make an 'orderly' market and to maintain a 'continuous' price path. This obligation may cause a delayed response. See Hasbrouck (1991a).

To make Equation (1) identifiable, it is necessary to impose two probability structures: i) a structure that depicts the stochastic process of the 'true' order type i_t, and ii) another structure that links j_t to i_t. Following Cosslett and Lee (1985), the true buy-sell arrival process is assumed as a first-order Markov process i_t with the stationary transition probability matrix P. The (i, j)-th element of P is

$$p_{i,j} = \Pr[i_t = j \mid i_{t-1} = i], \quad i, j \in \{-1, 0, 1\}, \tag{2}$$

where 'Pr' denotes conditional probabilities. The Markov chain assumption conveniently captures the known empirical regularity that purchases tend to follow purchases and sales tend to follow sales. It also incorporates the possibility that crosses may tend to follow crosses.

I assume a simple probability structure regarding j_t: the observed indicator j_t depends only on the current state i_t. That is, the (i, j)-th element of the classification probability matrix Q is

$$q_{i,j} = \Pr[j_t = j \mid i_t = i], \quad i, j \in \{-1, 0, 1\}, \tag{3}$$

Obviously, the conventional approach of using j_t rather than i_t is a special case where $q_{-1,-1}=1$, $q_{0,0}=1$, and $q_{1,1}=1$.

An econometric concern that arises from using j_t rather than i_t is the errors in variables problem. If there is a single signed trade or volume variable in the regression equation, the slope coefficient will be biased toward zero. However, the direction of the bias in the present model (especially, with lagged dependent variables) cannot be determined *a priori* - it depends on the covariance structure of all the regressors. Another noteworthy point is that, as Cosslett and Lee (1985) show, the errors in variables problem in this situation (where a discrete variable has measurement errors) cannot be solved by the use of the conventional instrumental variables method. The reason is that the measurement error $(j_t - i_t)x_t$ is correlated with the 'true' value $i_t x_t$. It is, therefore, almost impossible to choose an instrumental variable that is correlated with $i_t x_t$ but uncorrelated with $(j_t - i_t)x_t$.

3. Estimation Method

Cosslett and Lee (1985) provide a general recurrence algorithm for evaluating the likelihood function of a Markov switching regression model. Subsequent studies of Markov switching regression models adopt a similar algorithm. Among them, Hamilton (1990) provides an estimation method that is easy to implement. This paper closely follows Hamilton's suggestions for estimating the model.[14]

The log-likelihood of the model is the sum of the conditional log-likelihoods of all observations:

$$L(R_T, J_T) = \sum_{t=1}^{T} \ln f(r_t, j_t \mid R_{t-1}, J_{t-1}, X_t), \tag{4}$$

where $R_t \equiv \{r_t, r_{t-1}, \cdots, r_1\}$, $J_t \equiv \{j_t, j_{t-1}, \cdots j_1\}$, $X_t \equiv \{x_t, x_{t-1}, \cdots, x_1\}$ and f is a conditional joint density function.[15] Since trading volumes entering Equation (4) as conditioning variables are assumed exogenous, X_t is omitted throughout the paper for expositional simplicity.

3.1 The Basic Filter

The core of the algorithm that evaluates the likelihood function is a filtering procedure that takes conditional probability $\Pr[i_{t-1}, i_{t-2} \mid R_{t-1}, J_{t-1}]$ as the input and computes $\Pr[i_t, i_{t-1} \mid R_t, J_t]$ as the output. In general, if the model has r lags, the filtering procedure takes $\Pr[i_{t-1}, \cdots, i_{t-r} \mid R_{t-1}, J_{t-1}]$ as the input and computes $\Pr[i_t, \cdots, i_{t-r+1} \mid R_t, J_t]$ as the output. Since Equation (1) has two lags, the procedure for the case of r=2 is illustrated as follows.

14) Excellent descriptions of the algorithm [Hamilton (1989, 1990)] are available, but here I explain it briefly for completeness at the risk of redundancy.

15) Throughout this paper, 'f' represents a probability density of continuous variables (or a mixture of continuous and discrete variables), and 'Pr' represents a probability of discrete variables.

Step 1. Compute

$$\Pr[i_t, i_{t-1}, i_{t-2} \mid R_{t-1}, J_{t-1}] = \Pr[i_t \mid i_{t-1}] \Pr[i_{t-1}, i_{t-2} \mid R_{t-1}, J_{t-1}].$$

The simplification in the first term is due to the first-order Markov assumption. The Markov assumption implies that the realization of the state at t depends only on the state at $t-1$. The second term is the input of the filter. Since each of i_t, i_{t-1}, and i_{t-2} takes one of -1, 0 and 1, twenty-seven ($3 \times 3 \times 3$) possible combinations should be considered.

Step 2. Using the result from Step 1, compute

$$\begin{aligned} &f(r_t, j_t, i_t, i_{t-1}, i_{t-2} \mid R_{t-1}, J_{t-1}) \\ &= f(r_t \mid R_{t-1}, i_t, i_{t-1}, i_{t-2}) \Pr[j_t \mid i_t] \Pr[i_t, i_{t-1}, i_{t-2} \mid R_{t-1}, J_{t-1}]. \end{aligned}$$

The product of the first two terms on the RHS is the conditional likelihood of one observation. The first term is evaluated as

$$\begin{aligned} f(r_t \mid R_{t-1}, i_t, i_{t-1}, i_{t-2}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2}(r_t - \phi_1 r_{t-1} - \phi_2 r_{t-2} - \alpha_0 i_t x_t \\ - \alpha_1 i_{t-1} x_{t-1} - \alpha_2 i_{t-2} x_{t-2} - \gamma_0 i_t - \gamma_1 i_{t-1})^2\}. \end{aligned}$$

The simplification in the second term is due to the assumed error structure.

Step 3. Compute

$$f(r_t, j_t \mid R_{t-1}, J_{t-1}) = \sum_{i_t} \sum_{i_{t-1}} \sum_{i_{t-2}} f(r_t, j_t, i_t, i_{t-1}, i_{t-2} \mid R_{t-1,J_{t-1}}).$$

Step 4. Use the results from Step 2 and Step 3 to compute

$$\Pr[i_t, i_{t-1}, i_{t-2} \mid R_t, J_t] = \frac{f(r_t, j_t, i_t, i_{t-1}, i_{t-2} \mid R_{t-1}, J_{t-1})}{f(r_t, j_t \mid R_{t-1}, J_{t-1})}.$$

Step 5. Finally, compute the output of the filter

$$\Pr[i_t, i_{t-1} \mid R_t, J_t] = \sum_{i_{t-2}} \Pr[i_t, i_{t-1}, i_{t-2} \mid R_t, J_t].$$

The filter starts from t=3 (one plus the number of lags), using the unconditional probability $\Pr[i_2, i_1]$ as the initial input. This treatment of the initial input is based on the assumption that the initial state is drawn from the equilibrium distribution of the Markov chain. Alternatively, the nine possible values of $\Pr[i_2, i_1]$ can be treated as parameters to be estimated. The unconditional probability is

$$\Pr[i_2, i_1] = \Pr[i_2 \mid i_1]\Pr[i_1],$$

where $(\Pr[i_1=-1], \Pr[i_1=0], \Pr[i_1=1])'$ is the solution $(\pi_{-1}, \pi_0, \pi_1)'$ of the equation

$$\begin{bmatrix} 1-p_{-1,-1} & -p_{0,-1} & -p_{1,-1} \\ -p_{-1,0} & 1-p_{0,0} & -p_{1,0} \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \pi_{-1} \\ \pi_0 \\ \pi_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

The likelihood function of the model is computed as a by-product of the filtering procedure (Step 2). The filtering algorithm is repeated until the conditional log-likelihoods of all observations are obtained. Adding up the conditional log-likelihoods of all observations yields the log-likelihood of the model.

3.2 The Full-sample Smoother

After running through the basic filter, the full-sample smoother $\Pr[i_t, i_{t-1} \mid R_T, J_T] (t=2, \cdots, T)$ can be computed. This smoother is useful not only to gather inferences about the state i_t given the full sample, but also essential to update the parameter values during the iteration process. Unfortunately, the algorithm provided by Hamilton (1989, 1990) is computationally demanding: the number of calculations is proportional to the square of the number of observations.

Lee and Han (1993) and Kim (1994) provide an alternative algorithm that requires the number of calculations be proportional only to the number of observations. Their smoothing algorithm is a 'backward filtering' procedure that takes $\Pr[i_{t+1}, i_t, i_{t-1} \mid R_T, J_T]$ as the input and obtains $\Pr[i_t, i_{t-1}, i_{t-2} \mid R_T, J_T]$ as the output. Only one pass of the backward filter is required to obtain the full-sample smoother. The procedure that computes the smoother for a given time t is as follows:

Step 1. Compute

$$\Pr[i_{t+1}, i_t, i_{t-1}, i_{t-2} \mid R_t, J_t] = \Pr[i_{t+1} \mid i_t] \Pr[i_t, i_{t-1}, i_{t-2} \mid R_t, J_t],$$

where $\Pr[i_t, i_{t-1}, i_{t-2} \mid R_t, J_t]$ is from the basic filter (Step 4).

Step 2. Use the result from Step 1 to compute

$$\Pr[i_{t+1}, i_t, i_{t-1} \mid R_t, J_t] = \sum_{i_{t-2}} \Pr[i_{t+1}, i_t, i_{t-1}, i_{t-2} \mid R_t, J_t].$$

Step 3. Using the results from the previous two steps, compute

$$\Pr[i_{t-2} \mid i_{t+1}, i_t, i_{t-1}, R_t, J_t] = \frac{\Pr[i_{t+1}, i_t, i_{t-1}, i_{t-2} \mid R_t, J_t]}{\Pr[i_{t+1}, i_t, i_{t-1} \mid R_t, J_t]}.$$

Step 4. Compute the output of the backward filter

$$\Pr[i_t, i_{t-1}, i_{t-2} \mid R_T, J_T]$$
$$= \sum_{i_{t+1}} \Pr[i_{t-2} \mid i_{t+1}, i_t, i_{t-1}, R_T, J_T] \Pr[i_{t+1}, i_t, i_{t-1} \mid R_T, J_T].$$

As shown in Appendix A, the first component on the RHS is identical to the conditional probability $\Pr[i_{t-2} \mid i_{t+1}, i_t, i_{t-1}, R_t, J_t]$. Thus, to evaluate the first component on the RHS, we can use the result from Step 3. The second term is the input of the filter.

Step 5. Finally, compute the full-sample smoother

$$\Pr[i_t, i_{t-1} \mid R_T, J_T] = \sum_{i_{t-2}} \Pr[i_t, i_{t-1}, i_{t-2} \mid R_T, J_T].$$

The backward filter starts with $\Pr[i_T, i_{T-1}, i_{T-2} \mid R_T, J_T]$ as the initial input, which is obtained from Step 4 in the final stage of the basic filter. The filter runs through backward from t=T to t=3.

3.3 EM Algorithm

I use the EM algorithm to estimate the model.[16] Starting with

16) See Hamilton (1990) for the explanation of the EM algorithm.

arbitrary initial values of the parameters, I run through the basic filter and the smoother. The parameter values are updated by the EM rule. This is easy to implement since, as Hamilton (1990) shows, there is a closed form solution for the new parameter values. I repeat the basic filter and the smoother with the updated parameter values. Iterating this sequential procedure until convergence results in the final estimates. Hamilton notes that the EM algorithm is numerically stable compared to standard optimization algorithms, and thus, it is particularly suitable for a large scale application such as mine. A problem is that the EM algorithm leads to a local maximum. Since the likelihood function of the present model may have multiple maxima, it is necessary to try many different initial values to attain the global maximum.

4. The Data

All firms in the CRSP file (NYSE and AMEX) are grouped into five quintiles in increasing order on the basis of their market capitalizations at the end of 1989. Firms in quintiles 5 and 2 are sorted alphabetically by ticker symbol within each market capitalization quintile.[17] From each of these two quintiles (henceforth, large and small firms), the first 100 common stocks that have transaction records for at least 250 days (out of a total of 253 trading days) for the year 1990 in the ISSM transactions database are chosen.[18]

My intraday time-series data are constructed from the 1990 ISSM transactions database. Trades flagged by the ISSM as errors as well as non-standard delivery trades are eliminated.[19] All BBO-ineligible quotes are also eliminated, where BBO stands for 'Best Bid/Offer.'[20] Following Hasbrouck's (1991b) suggestion,

17) Initially, I attempted to obtain the sample from quintile 1 (the smallest quintile), but experienced difficulty in finding firms with sufficient time-series observations.
18) Partnerships, ADRs and closed-end funds are excluded from the sample.
19) Specifically, trades with the ISSM condition codes A, C, D, N, O, R,and Z are eliminated. Opening trades are also eliminated because they are usually made in batch auctions.
20) BBO-ineligible quotes are closing quotations, trading halts, pre-opening indications, and non-firm quotations. For details, see the ISSM manual (1993), p.14.

I cumulate those trades that occur within five seconds of each other without intervening quotes as a single observation to alleviate the problem of reporting fragmentation.[21] The 'prevailing' quote is matched to each 'clean' trade that survived the above screening process. Since Lee and Ready (1991) show that trades are often reported with a lag, the 'prevailing' quote at the time of a trade is defined as the quote reported at least five seconds prior to the trade.

The number of 'clean' trades varies widely from firm to firm. The number of trades ranges from 1,987 to 12,297 for small firms, and only one (CHL.B) is greater than 10,000. For large firms, the numbers are in the range of 5,333 to 154,810, and only five of them (AAL, AOC, BCE, BFB, CIN) are less than 10,000. Since the estimation of the model is time consuming, I select randomly twenty-five days for each firm if the number of trades exceeds 10,000.[22] Consequently, ninety-five large firms and one small firm are subject to this twenty-five day sampling scheme. Any trading day with fewer than five trades is eliminated because I am interested in intraday variation in stock price changes.

5. Empirical Results

The dependent variable r_t in my time-series model is the change in transaction prices (in dollars) at transaction time t. The volume variable x_t is the number of shares traded (in millions) at transaction time t. The estimates of the model, like those of any parametric models, are sensitive to outliers. Therefore, following Hausman, Lo and MacKinlay (1992), a value of x_t greater than the 99.5 percentile of x_t's for that stock is set to the 99.5 percentile.[23] Price changes and share volume are

---

21) For details, see Hasbrouck (1991b), p.581.

22) It is possible to devise an alternative sampling scheme. An example is to use the first 10,000 observations for actively traded firms. This procedure is more likely to draw observations from earlier months of the year, which makes cross-firm comparison difficult. My sampling procedure also has a problem. Some actively traded firms may have less data than the less actively traded firms. Despite this drawback, I believe, my sampling procedure preserves qualitative characteristics of the data.

23) The 99.5% cutoffs for trade size truncation range between 5,000 and 124,700 shares (mean=20,058 shares) for small firms, and between 9,700

adjusted for stock splits and stock dividends.

While I estimate the model using time-series data spanning twenty-five days at the minimum, the estimation procedure uses intraday data only to avoid a problem of econometric specification of overnight price changes. Thus, filtering and smoothing are done within each day[24] though the true values of the parameters are assumed to be constant over the year. The implicit assumption is that the evolution of the states on a given day is independent of the states on the previous day.

5.1 An Example: American Aluminum Company

To illustrate how the model works, I show the details of the model estimates for the first firm in the large firm sample (ticker symbol=AA). The number of observations used is 3,783 (25 days), and the 99.5% cutoff for trade size truncation is 34,500 shares.

Table 1 presents the maximum likelihood estimates of the three-state and two-state Markov models, as well as the least square estimates of the conventional model that assumes perfectly measured buy-sell indicators. Several test statistics are also presented. Standard errors are computed by inverting the information matrix of the parameters, where the information matrix is computed analytically.

Comparisons of the estimates among the three models suggest that the three-state model is quite reasonable. First, the estimated transition probability matrix shows that, as in Hasbrouck and Ho (1987), purchases tend to follow purchases and sales tend to follow sales. Further, crosses also tend to follow crosses. The point estimates together with the small standard errors indicate that every diagonal element in the transition probability matrix is significantly greater than the benchmark 1/3.[25] Second, the three-state model shows the greatest sensitivity ($\Sigma\alpha$) of price changes to order size. In addition, the negative autocorrelations ($\Sigma\phi$) in price changes are substantially reduced. That is, the model seems to extract

---

and 100,000 shares (mean=35,005 shares) for large firms.

24) Confining the analysis to intraday data also reduces the computation time significantly.

25) I do not mean that the unconditional probability is 1/3. 1/3 is an arbitrary benchmark.

Table 1. Model Estimates: the Case of American Aluminum Company[a]

The basic model: $r_t=\phi_1 r_{t-1}+\phi_2 r_{t-2}+\alpha_0 i_t x_t+\alpha_1 i_{t-1} x_{t-1}+\alpha_2 i_{t-2} x_{t-2}$
$+\gamma_0 i_t+\gamma_1 i_{t-1}+\varepsilon_t$, $\varepsilon_t \sim N(0,\sigma^2)$, $t=3,4,\cdots,T$

-	No error model[b]		Two-state model 2[c]		Three-state model 3[d]	
Parameter estimates:						
ϕ_1	-0.0214	(0.0140)	-0.0509	(0.0097)	0.0161	(0.0101)
ϕ_2	-0.0522	(0.0109)	-0.0315	(0.0099)	-0.0477	(0.0098)
α_0	1.0964	(0.2250)	1.0332	(0.1566)	1.0735	(0.1698)
α_1	1.3253	(0.2259)	0.4603	(0.2092)	1.2438	(0.2176)
α_2	0.3915	(0.2142)	0.9402	(0.2023)	0.5964	(0.2217)
γ_0	0.0719	(0.0011)	0.0671	(0.0011)	0.0745	(0.0012)
γ_1	-0.0510	(0.0014)	-0.0537	(0.0013)	-0.0554	(0.0015)
$p_{-1,-1}$			0.6249	(0.0127)	0.4565	(0.0145)
$p_{-1,0}$					0.1718	(0.0122)
$p_{0,-1}$					0.2413	(0.0161)
$p_{0,0}$					0.4700	(0.0196)
$p_{1,-1}$			0.3274	(0.0117)	0.3276	(0.0122)
$p_{1,0}$					0.1285	(0.0101)
$q_{-1,-1}$			0.7564	(0.0119)	0.9954	(0.0053)
$q_{-1,0}$			0.2395	(0.0118)	0.0000	(0.0047)
$q_{0,-1}$					0.0001	(0.0119)
$q_{0,0}$					0.9844	(0.0172)
$q_{1,-1}$			0.0113	(0.0028)	0.0134	(0.0035)
$q_{1,0}$			0.1987	(0.0102)	0.0064	(0.0049)
Log-likelihood:				1440.0		1921.3
Hypothesis testing:						
H_0:[e]	$\phi_1+\phi_2=0$		28.0	[0.0000]	4.1	[0.0416]
H_0:[f]	$\alpha_0+\alpha_1+\alpha_2=0$		63.4	[0.0000]	73.4	[0.0000]
H_0:[g]	$q_{-1,1}=0$, $q_{1,-1}=0$		19.9	[0.0000]	18.8	[0.0000]
H_0:[h]	$q_{-1,-1}=1$, $q_{0,0}=1$, $q_{1,1}=1$				13.4	[0.0010]

[a] Standard errors are in parentheses. P-values are in square brackets. All three models use the buy-sell indicator j_t (-1=sale, 0=nonclassification, 1=purchase) that is exogenously determined. A trade made at the price below (above) the quote midpoint is classified as seller (buyer) initiated. A midpoint trade is classified as a sell (buy) if it is a downtick (uptick) transaction.

[b] The no error model assumes that the buy-sell classification variable i_t is measured without error; that is, $j_t=i_t$.

[c] The two-state model assumes that there are buyer- and seller-initiated trades only. It is a Markov switching regression model in

which the measured buy-sell indicator j_t is assumed imperfect. The unobserved 'true' indicator i_t (-1=sale, 1=purchase) follows a first order stationary Markov chain (Cosslett and Lee, 1985). The assumed probability structure is: $P_{i,j}=\Pr[i_t=j \mid i_{t-1}=i]$, $i,j \in \{-1,1\}$; $q_{i,j}=\Pr[j_t=j \mid i_t=i]$, $i \in \{-1,1\}$, $j \in \{-1,0,1\}$.

[d] The three-state model assumes that there are three states - buyer- and seller-initiated trades and crosses. It is a Markov switching regression model in which the measured buy-sell indicator j_t is assumed imperfect. The unobserved 'true' indicator i_t (-1=sale, 0=cross, 1=purchase) follows a first order stationary Markov chain. The assumed probability structure is: $p_{i,j}=\Pr[i_t=j \mid i_{t-1}=i]$, $i,j \in \{-1,0,1\}$; $q_{i,j}=\Pr[j_t=j \mid i_t=i]$, $i,j \in \{-1,0,1\}$.

[e] Wald test χ^2 and the corresponding p-value.

[f] Wald test χ^2 and the corresponding p-value.

[g] Modified Wald test χ^2 and the corresponding p-value. See Gourieroux, Holly and Monfort (1982).

[h] Modified Wald test χ^2 and the corresponding p-value. See Gourieroux, Holly and Monfort (1982).

information contained in trades more efficiently than the other two models do. Third, the likelihood value for the three-state Markov model is much greater than that for the two-state Markov model. The above three empirical observations are generally true for the other firms in the sample.

The probability of misclassification is very small for this firm. Consequently, it is interesting to test formally whether it is in fact zero. Since the probability under the null is on the boundary of the parameter space, conventional tests such as the likelihood ratio test, Lagrange multiplier test and Wald test, are not directly applicable. Thus, I use the p-value proposed by Kudo (1963), Gourieroux, Holly and Monfort (1982) and Shapiro (1985), among many others. Under the null hypothesis

$$H_0\text{: } q_{-1,1}=\text{(that is, } 1-q_{-1,-1}-q_{-1,0}=0) \text{ and } q_{1,-1}=0,$$

the standard Wald test statistic (χ^2-value) c is first calculated. Then, the p-value is

$$1-\{w+1/2\Pr[\chi_1^2<c]+(1/2-w)\Pr[\chi_2^2<c]\},$$

where $w \equiv (1/2\pi)\cos^{-1}\rho$ and ρ is the estimated correlation between $q_{-1,1}$ and $q_{1,-1}$.

A more general test of zero misclassification probability can be also conducted. Under the null hypothesis

H_0: $q_{-1,-1}=1$, $q_{0,0}=1$, and $q_{1,1}=1$,

the standard Wald test statistic c can be first calculated. Then, the modified p-value is

$$1-\{w_3+w_2\Pr[\chi_1^2<c]+w_1\Pr[\chi_2^2<c]+w_0\Pr[\chi_3^2<c]\},$$

where

$$\begin{aligned}
w_0 &\equiv (\pi^{-1}/4)(2\pi\cos^{-1}\rho_{12}-\cos^{-1}\rho_{13}-\cos^{-1}\rho_{23}),\\
w_1 &\equiv (\pi^{-1}/4)(3\pi\cos^{-1}\rho_{12.3}-\cos^{-1}\rho_{13.2}-\cos^{-1}\rho_{23.1}),\\
w_2 &\equiv 1/2-w_0,\quad w_3\equiv 1/2-w_1.
\end{aligned}$$

Here, ρ_{ij} represents the (i, j)-th element of the estimated correlation matrix of $(-q_{-1,-1},-q_{0,0},-q_{1,1})'$, and $\rho_{ij,k}\equiv(\rho_{ij}-\rho_{ik}\rho_{jk})/\{(1-\rho_{ik}^2)(1-\rho_{jk}^2)\}^{1/2}$. The results reported in Table 1 show that the null hypothesis of zero misclassification error is decisively rejected for the firm. More general results follow shortly.

5.2 The Estimates of the Transition Probability Matrix

For each of the 200 stocks in the sample, three-state Markov switching regression model is estimated. Table 2 reports the cross-sectional distributions of the estimated order type transition probabilities which measures the degree of persistence in buy-sell arrival sequences. Standard errors are computed from the cross-sectional distribution of the point estimates within each market capitalization group.

The probability estimates indicate that my Markov assumption is reasonable. All firms in the sample, except for one small firm, exhibit persistence in buy-sell arrivals - that is, $p_{-1,-1}>1/3$, and $p_{1,1}>1/3$. A measure of persistence in crosses, $p_{0,0}$, is also greater than 1/3 for all firms except for two large and eight small firms.

The mean transition probabilities are remarkably similar between large and small firms. Further, the estimated transition

Table 2. Estimates of Order Type Transition Probabilities for 100 Large and 100 Small Firms

Prob.[a]	Mean	StdErr[b]	Min.	10%	Median	90%	Max.
			A. Large Firms				
$p_{-1,-1}$	0.530	(0.071)	0.350	0.435	0.539	0.630	0.719
$p_{-1,0}$	0.099	(0.037)	0.031	0.053	0.101	0.150	0.255
$p_{-1,1}$	0.369	(0.059)	0.219	0.303	0.362	0.442	0.531
$p_{0,-1}$	0.226	(0.056)	0.089	0.153	0.232	0.288	0.359
$p_{0,0}$	0.553	(0.100)	0.285	0.447	0.537	0.701	0.821
$p_{0,1}$	0.220	(0.059)	0.083	0.143	0.221	0.292	0.377
$p_{1,-1}$	0.360	(0.054)	0.237	0.300	0.353	0.441	0.490
$p_{1,0}$	0.104	(0.041)	0.021	0.050	0.100	0.156	0.215
$p_{1,1}$	0.535	(0.075)	0.355	0.444	0.535	0.620	0.738
			B. Small Firms				
$p_{-1,-1}$	0.540	(0.087)	0.350	0.438	0.525	0.659	0.787
$p_{-1,0}$	0.088	(0.039)	0.006	0.030	0.091	0.143	0.169
$p_{-1,1}$	0.371	(0.066)	0.169	0.291	0.376	0.445	0.546
$p_{0,-1}$	0.237	(0.082)	0.028	0.129	0.248	0.344	0.458
$p_{0,0}$	0.519	(0.145)	0.242	0.349	0.491	0.720	0.812
$p_{0,1}$	0.242	(0.088)	0.051	0.128	0.251	0.353	0.419
$p_{1,-1}$	0.359	(0.061)	0.150	0.288	0.354	0.444	0.516
$p_{1,0}$	0.087	(0.043)	0.003	0.034	0.087	0.144	0.235
$p_{1,1}$	0.553	(0.077)	0.308	0.471	0.541	0.656	0.770

[a] The estimates are based on a three-state Markov switching regression model in which the measured buy-sell indicator j_t is assumed imperfect. The unobserved 'true' indicator i_t (-1=sale, 0=cross, 1=purchase) follows a first order stationary Markov chain: $p_{i,j}=\Pr[i_t=j \mid i_{t-1}=i]$, $i,j \in \{-1,0,1\}$.

[b] Standard errors are computed from the cross-sectional distribution of the point estimates for 100 firms in the corresponding market capitalization group.

probabilities appear to be almost symmetric between purchases and sales. The mean of $p_{-1,-1}$ (the probability that the next trade is seller-initiated conditional on the current trade being seller-initiated) is 0.53 for large firms and 0.54 for small firms. The mean of $p_{1,1}$ (the probability that the next trade is buyer-initiated conditional on the current trade being buyer-initiated) is 0.54 for large firms and 0.55 for small firms. Crosses also tend to follow crosses. The average $p_{0,0}$ is 0.55 for large firms and 0.52 for small firms. Thus, a common practice of treating the arrival of a cross as an independent random event may not be appropriate.

5.3 Information Asymmetry and Order Processing Cost

Table 3 reports summary statistics of the coefficient estimates. Since $-\gamma_1$ is a measure of order processing cost, its distribution is provided rather than that of γ_1. Further, since $-\gamma_1/\gamma_0$ is a measure of information asymmetry in Madhavan and Smidt, its distribution is also provided.

Table 3. Estimates of the Model Coefficients for 100 Large and 100 Small Firms

Coef.[a]	Mean	StdErr[b]	Min.	10%	Median	90%	Max.	% (p-val<0.05)[c]
				A. Large Firms				
$\Sigma\phi$	-0.074	(0.013)	-0.385	-0.251	-0.060	0.069	0.338	72%
$\Sigma\alpha$	2.665	(0.190)	0.373	0.865	2.216	5.114	11.363	93%
γ_0	0.079	(0.002)	0.058	0.064	0.074	0.107	0.153	100%
$-\gamma_1$	0.061	(0.001)	0.038	0.051	0.058	0.068	0.128	100%
$-\gamma_1/\gamma_0$	0.778	(0.011)	0.502	0.655	0.764	0.940	1.006	-
				B. Small Firms				
$\Sigma\phi$	-0.269	(0.021)	-0.846	-0.538	-0.290	-0.011	0.051	94%
$\Sigma\alpha$	4.269	(0.473)	-3.148	0.268	3.135	11.267	20.268	59%
γ_0	0.077	(0.002)	0.032	0.063	0.069	0.101	0.146	100%
$-\gamma_1$	0.050	(0.001)	0.019	0.033	0.047	0.064	0.129	100%
$-\gamma_1/\gamma_0$	0.678	(0.022)	0.218	0.397	0.655	0.976	1.028	-

[a] The estimates are based on a three-state Markov switching regression model in which the measured buy-sell indicator j_t is assumed imperfect. The unobserved 'true' indicator i_t (-1=sale, 0=cross, 1=purchase) follows a first order stationary Markov chain: $p_{i,j}=\Pr[i_t=j \mid i_{t-1}=i]$, $i,j\in\{-1,0,1\}$. The assumed classification error probability structure is: $q_{i,j}=\Pr[j_t=j \mid i_t=i]$, $i,j\in\{-1,0,1\}$.

[b] Standard errors are computed from the cross-sectional distribution of the point estimates for 100 firms in the corresponding market capitalization group.

[c] For each firm, p-values for $\Sigma\phi$ and $\Sigma\alpha$ are obtained from the Wald test for H_0: the sum of coefficients = 0. P-values for γ_0 and $-\gamma_1$ are based on the t-test.

The signs and the magnitudes of the coefficients are generally in line with those reported in previous studies, indicating that the biases induced by the errors in variables problem are not too serious to overturn the results established in the literature. For example, both γ_0 and $-\gamma_1$ are always positive and γ_0 is in general greater than $-\gamma_1$. Consequently, the ratio of $-\gamma_1/\gamma_0$ is typically less than one. The Madhavan and Smidt theory suggests that, in the absence of information asymmetry, the ratio should be one. Thus, firms experiencing a greater information asymmetry are likely to have a smaller $-\gamma_1/\gamma_0$.

An important finding is that small firms experience a greater information asymmetry than large firms. The average ratio for small firms (0.68) is less than that for large firms (0.78).[26] Since the standard errors are very small (0.02 for small firms and 0.01 for large firms), the difference in the ratio between the small and large firms is highly significant.

The observed difference between large and small firms is consistent with a common intuition that more information is generated for larger firms. Because news agencies and security analysts tend to follow larger firms more closely, more public information is likely to be produced for larger firms. In addition, firm size is highly correlated with trading frequency and trading share volume,[27] private information is likely to be revealed more quickly via trading for larger firms.

Order processing cost per share, $-\gamma_1$, is significantly greater for larger firms. The average order processing cost is 6.1 cents for large firms and 5.0 cents for small firms. The standard errors are very small (about 0.1 cent for both large and small firms). The evidence is consistent with the folklore that large firms subsidize small firms, but a clearer test requires a detailed analysis of individual specialist's behavior.

The sums of ϕ's are predominantly negative indicating the existence of unexplained negative autocorrelations in price changes. The sums of α's are mostly positive, which is consistent with information effects.

26) The average estimate provided by Madhavan and Smidt is 0.76, which is closer to that for my large firms.

27) Recall that only one among my small firms has the number of trades greater than 10,000 but ninety-five among the large firms in the sample has the number of trades greater than 10,000.

5.4 Misclassification Probabilities

Table 4 reports the estimates of the misclassification probabilities conditional on the order type.[28] Since the left tails of the distribution are uninteresting (very close to zero), only the right tails (from the median) together with the cross-sectional

Table 4. Estimates of Misclassification Probabilities for 100 Large and 100 Small Firms

Prob.[a]	Mean	StdErr[b]	Median	75%	90%	95%	Max.	% (p-val<0.05)[c]
				A. Large Firms				
$q_{-1,0}$	0.015	(0.023)	0.003	0.018	0.051	0.070	0.095	73%
$q_{-1,1}$	0.010	(0.012)	0.006	0.013	0.022	0.027	0.088	
$q_{0,-1}$	0.033	(0.054)	0.008	0.045	0.098	0.137	0.283	
$q_{0,1}$	0.027	(0.059)	0.004	0.024	0.066	0.167	0.352	
$q_{1,-1}$	0.010	(0.013)	0.006	0.012	0.025	0.031	0.096	
$q_{1,0}$	0.014	(0.025)	0.002	0.016	0.046	0.055	0.170	
				B. Small Firms				
$q_{-1,0}$	0.017	(0.025)	0.004	0.028	0.047	0.065	0.121	59%
$q_{-1,1}$	0.019	(0.019)	0.014	0.024	0.033	0.047	0.140	
$q_{0,-1}$	0.097	(0.174)	0.014	0.110	0.317	0.454	0.978	
$q_{0,1}$	0.086	(0.162)	0.011	0.097	0.247	0.368	0.797	
$q_{1,-1}$	0.023	(0.028)	0.017	0.028	0.049	0.059	0.217	
$q_{1,0}$	0.020	(0.034)	0.004	0.031	0.049	0.083	0.219	

[a] The estimates are based on a three-state Markov switching regression model in which the measured buy-sell indicator j_t is assumed imperfect. The unobserved 'true' indicator i_t (-1=sale, 0=cross, 1=purchase) follows a first order stationary Markov chain: $p_{i,j}=\Pr[i_t=j \mid i_{t-1}=i]$, $i,j\in\{-1,0,1\}$. The assumed classification error probability structure is: $q_{i,j}=\Pr[j_t=j \mid i_t=i]$, $i,j\{\in-1,0,1\}$.

[b] Standard errors are computed from the cross-sectional distribution of the point estimates for 100 firms in the corresponding market capitalization group.

[c] Modified Wald test χ^2 (Gourieroux, Holly and Monfort, 1982) for H_0: $q_{-1,-1}=1$, $q_{0,0}=1$, and $q_{1,1}=1$.

28) While I do not report the result separately, I find that large firms are more likely to be unclassified than small firms.

mean and the standard error are reported.

The results for large firms are generally consistent with the result for the American Aluminum stock presented in the previous section. The probability of misclassification for large firms is generally small except for a few firms, although a formal statistical test (a modified Wald test as explained in a previous section) rejects the null of zero error at the 5% confidence level for 73 out of 100 firms. In particular, the median of $q_{-1,1}$ (the probability that a seller-initiated trade is classified as buyer-initiated) is only 0.006, and 90% of the large firms have $q_{-1,1}$ less than 0.022. The median of $q_{1,-1}$ (the probability that a buyer-initiated trade is classified as seller-initiated) is 0.006, which is almost identical to $q_{-1,1}$. These probabilities are slightly greater for small firms (the medians are 0.014 and 0.017 respectively) than for large firms, but a formal test rejects the null of zero error only for 59 out of 100 small firms. A greater frequency of classification error arises for crosses particularly for small firms. Both the distributions of $q_{0,-1}$ and $q_{0,1}$ are highly skewed to the right.

At this point, it would be useful to compare my estimates of misclassification probabilities with those in existing studies. Keim and Madhavan (1996) examine upstairs-negotiated trades obtained from a passive investment management firm. Their data contain information about the direction of each trade, and consist of 4,688 seller-initiated blocks and 937 buyer-initiated blocks. They report a much greater error frequency: 6.5% for seller-initiated trades and 20.1% for buyer-initiated trades. There are at least two explanations for the observed discrepancy:[29] i) They apply the tick test rather than a modified bid-ask test used in this study. ii) More importantly, their tick test compares the block price to the previous day's closing price (rather than the price immediately before the block), which may lead to many classification errors.

The estimated probabilities of misclassification are generally small, except for some relatively large firms. Thus, it is conceivable that conventional models may suffer from the errors in variables problem at least for some firms. I now turn to examining biases in the estimated coefficients caused by the

29) More than 60% of their trades comprise of NASDAQ trades, which also hinders a direct comparison between their results and mine.

errors in variables problem for a conventional model. My three-state Markov switching regression model is used as the benchmark.

Table 5 reports summary statistics of the biases in the coefficient estimates from the conventional model in which the measured buy-sell indicator j_t is assumed to be perfect. In general, the conventional model underestimates order processing cost, $-\gamma_1$, and overestimates the degree of information asymmetry (that is, underestimates $-\gamma_1/\gamma_0$). Information effect (the sums of α's) is also underestimated, which is a natural consequence of buy-sell misclassification.

The biases in the measured order processing costs appear economically small. For example, while the mean order processing cost per share obtained from my three-state Markov

Table 5. Biases in the Coefficient Estimates When the Measured Buy-sell Indicators are Assumed Perfect - 100 Large and 100 Small Firms

Coef.[a]	Mean	StdErr[b]	Min.	10%	Median	90%	Max.	% (overestimated)
				A. Large Firms				
$\Sigma\phi$	-0.042	(0.003)	-0.152	-0.075	-0.040	-0.012	0.006	4%
$\Sigma\alpha$	-0.308	(0.080)	-5.867	-0.753	-0.152	0.170	1.261	21%
γ_0	-0.003	(0.000)	-0.025	-0.010	-0.002	-0.000	0.001	7%
$-\gamma_1$	-0.005	(0.000)	-0.030	-0.009	-0.004	-0.002	0.000	1%
$-\gamma_1/\gamma_0$	-0.039	(0.002)	-0.111	0.076	-0.032	-0.007	0.005	2%
				B. Small Firms				
$\Sigma\phi$	-0.058	(0.007)	-0.290	-0.148	-0.025	0.001	0.019	12%
$\Sigma\alpha$	-0.095	(0.122)	-3.626	-1.424	-0.064	1.101	4.355	43%
γ_0	-0.006	(0.001)	-0.057	-0.011	-0.005	-0.001	0.001	1%
$-\gamma_1$	-0.008	(0.001)	-0.071	-0.017	-0.006	-0.009	0.001	6%
$-\gamma_1/\gamma_0$	-0.062	(0.006)	-0.241	-0.147	-0.048	0.009	0.062	18%

[a] The estimates are based on a three-state Markov switching regression model in which the measured buy-sell indicator j_t is assumed imperfect. The unobserved 'true' indicator i_t (-1=sale, 0=cross, 1=purchase) follows a first order stationary Markov chain: $p_{i,j}=\Pr[i_t=j \mid i_{t-1}=i]$, $i,j\in\{-1,0,1\}$. The assumed classification error probability structure is: $q_{i,j}=\Pr[j_t=j \mid i_t=i]$, $i,j\{-1,0,1\}$.

[b] Standard errors are computed from the cross-sectional distribution of the point estimates for 100 firms in the corresponding market capitalization group.

model is 6.1 cents (5.0 cents) for large (small) firms, the mean bias is only 0.5 cent (0.8 cent). For example, for a \$20 stock, the magnitude of the 0.5 cent bias corresponds to 0.025% of the stock price.

The sums of the estimated ϕ's are also downward biased. This result is consistent with the well-known empirical fact that a large portion of the negative autocorrelations in price changes is due to the bid-ask bounce. A correctly classified indicator variable (as in the three-state Markov model) should be able to capture such a bid-ask bounce effect.

5.5 Caveats and Extensions

In the introduction, I have emphasized the generality of the Markov model. This subsection discusses some avenues for future extensions. First, I assume that $q_{i,j}$'s are constant for a given stock. It may be more realistic to assume that misclassification probabilities are decreasing functions of the order size because larger trades are less likely to execute against limit orders which often occur within quotes. One way of handling this issue is to assume that classification probabilities are probit functions of order size:

$$q_{i,j}^{t} = \Phi(a_{i,j} + b_{i,j}x_t), \quad i \neq j,$$

where Φ is the cumulative density function of the unit normal distribution. The coefficients $b_{i,j}$'s are expected to have negative signs. The filtering and smoothing algorithms presented previously work without modification.[30] However, updating $a_{i,j}$ and $b_{i,j}$ using the EM algorithm is not as simple as before. It now requires a numerical optimization at each updating stage, which is computationally demanding.

Second, the assumption of normality for the error term may be objectionable,[31] because price moves in \$1/8 increments.

30) In this case, $\Pr[j_t \mid i_t]$ in Step 2 of the basic filter is no longer a constant. But, given the values of $a_{i,j}$ and $b_{i,j}$, it can be calculated from the assumed probit function.

31) It should be emphasized that in Equation (1) I do not assume the normality of r_t. This price change variable is in fact a mixture of normal distributions because the terms containing the state variable shift the conditional mean of r_t up and down. Therefore, the usual fat-tail problem is naturally taken care of.

Ignoring price discreteness may result in inefficient parameter estimates. However, the loss of efficiency is not an important concern here because most of the statistical inferences are based on the cross-sectional distribution of firm level estimates.

Perhaps, the most direct way of solving this discreteness problem is to impose an ordered probit model on Equation (1). In market microstructure literature, Hausman, Lo and MacKinlay (1992) introduce an ordered probit model to accommodate the price discreteness. The filtering and smoothing procedures in Section 2 are still valid but with some modifications. The dependent variable in Equation (1) should be r_t* which is an unobservable 'true' price change. The 'observed' price change r_t takes one of a few possible discrete values (for example, from -5 ticks to +5 ticks), each of which corresponds to a range of r_t* values:

$$r_t = s_m \quad \text{if} \quad r_t^* \in (a_{m-1}, a_m], \quad m = 1, 2, \cdots, M,$$

where $a_0=-\infty$, $a_M=+\infty$, and s_m is a discrete value which is a signed multiple of the minimum tick size (for example, -5/8, -4/8, ⋯, 5/8). Then, the likelihood function in Step 2 of the basic filter becomes

$$f(r_t \mid R_{t-1}, i_t, i_{t-1}, i_{t-2}) = \prod_{m=1}^{M} (\Phi(\frac{a_m - \delta_t' b}{\sigma}) - \Phi(\frac{a_{m-1} - \delta_t' b}{\sigma}))^{Y_m},$$

where $d_t'\ b=\phi_1 r_{t-1}+\phi_2 r_{t-2}+\alpha_0 i_t x_t+\alpha_1 i_{t-1} x_{t-1}+\alpha_2 i_{t-2} x_{t-2}+\gamma_0 it+\gamma_1 i_{t-1}$ and Y_m is an indicator that takes one if the realized r_t is s_m, and zero otherwise. Although it is straightforward to run through the filter and smoother given the input parameter values, updating the parameter values is difficult because there is no known closed-form solution for the new parameter values. It needs a numerical optimization procedure like the Newton-Raphson method in every EM step, which is time consuming. Since it is not clear whether the benefit outweighs the cost, I leave the actual implementation to a future study.[32]

32) Another problem in employing an ordered probit model is that the coefficient estimates are not unique. In other words, they are identifiable only up to an additive and multiplicative constants. To solve this identification problem, a normalization rule like $\sigma=1$ is commonly used. Thus, it is difficult to make economic interpretations of coefficients.

Finally, there are many ways to modify the base model. For example, it is easy to allow the coefficients of the signed volume variables to switch as the buy-sell state changes, which may capture the asymmetry between sales and purchases. Extending the proposed framework to vector autoregression as in Hasbrouck (1991a) is also conceptually straightforward. It is also possible to incorporate a nonlinear relationship between trade size and price changes by employing a Box-Cox transformation [Hausman, Lo and MacKinlay (1992)] with additional computational burden. In addition, a more complicated error structure, such as AR(r), can be imposed without much difficulty.

6. Summary

This paper tests an existing market microstructure time-series model using a new econometric approach. The proposed approach takes into account both serial dependence in buy-sell order types and possible buy-sell classification errors intuitively in a coherent framework.

The key intuition is that the direction of a trade (sale, cross and purchase) can be viewed as a discrete latent variable which follows a stationary Markov chain. By overlaying a three-state Markov chain on an existing market microstructure time-series model, I am able to extract information on the order type more efficiently from the entire sequence of data. In the model, the traditional buy-sell classification variable is treated as data with error.

An analysis of 100 large and an equal number of small firms for the year 1990 shows that the proposed model is not only viable, but is also useful in understanding the price formation process. I find that purchases tend to follow purchases and sales tend to follow sales. Similarly, crosses also tend to follow crosses, indicating that it is not desirable to treat the arrival of a cross as an independent random event. The mean estimate of the transition probability matrix is virtually invariant across large and small firm groups. The mean estimates of the transition probability matrix for large firms and small firms are:[33]

$$\begin{array}{l} \\ \\ \text{sale} \\ \text{cross} \\ \text{purchase} \end{array} \begin{array}{c} \text{Large Firms} \\ \begin{array}{ccc} \text{sale} & \text{cross} & \text{purchase} \end{array} \\ \begin{bmatrix} 0.53 & 0.10 & 0.37 \\ 0.23 & 0.55 & 0.22 \\ 0.36 & 0.10 & 0.54 \end{bmatrix} \end{array} \quad \begin{array}{c} \text{Small Firms} \\ \begin{array}{ccc} \text{sale} & \text{cross} & \text{purchase} \end{array} \\ \begin{bmatrix} 0.54 & 0.09 & 0.37 \\ 0.24 & 0.52 & 0.24 \\ 0.36 & 0.09 & 0.55 \end{bmatrix} \end{array}$$

I further find that information asymmetry is greater for small firms than for large firms, but per share order processing cost is greater for large firms (6.1 cents) than for small firms (5.0 cents). In addition, as a by-product, my analysis provides useful information on the accuracy of a popular classification scheme, and on the effect of classification errors on regression coefficients.

33) I am not aware of any study that reports an estimate of the transition probability matrix. Hasbrouck and Ho (1987) examine the autocorrelation in the buy-sell indicator, but they do not estimate the transition probability matrix directly. I hope the presented mean estimates will be useful for future studies as a benchmark or as an input to simulation work.

Appendix A

This appendix shows that the following equality holds for a general r-lag (r=0,1,2,$\cdots$) Markov switching model:

$$\Pr[i_{t-r} \mid i_{t+1}, i_t, \cdots, i_{t-r+1}, R_T, J_T] = \Pr[i_{t-r} \mid i_{t+1}, i_t, \cdots, i_{t-r+1}, R_t, J_t] \tag{A.1}$$

This equality is the key element of the backward filtering algorithm in computing the full-sample smoother.

Proof:

Define $R_T\text{-}R_t \equiv \{r_T, r_{T\text{-}1}, \cdots, r_{t+1}\}$ and $J_T\text{-}J_t \equiv \{j_T, j_{T\text{-}1}, \cdots j_{t+1}\}$. Then, the LHS of Equation (A.1) can be rewritten as

$$\begin{aligned} &\Pr[i_{t-r} \mid i_{t+1}, \cdots, i_{t-r+1}, R_T, J_T] \\ &= \Pr[i_{t-r} \mid i_{t+1}, i_t, \cdots, i_{t-r+1}, R_t, J_t, R_T - R_t, J_T - J_t] \\ &= \frac{\Pr[i_{t-r}, R_T - R_t, J_T - J_t \mid i_{t+1}, i_t, \cdots, i_{t-r+1}, R_t, J_t]}{\Pr[R_T - R_t, J_T - J_t \mid i_{t+1}, i_t, \cdots, i_{t-r+1}, R_t, J_t]}. \end{aligned} \tag{A.2}$$

The numerator in (A.2) can be written as the product of two conditional probabilities:

$$\begin{aligned} &\Pr[i_{t-r} \mid i_{t+1}, i_t, \cdots, i_{t-r+1}, R_t, J_t] \\ &\Pr[R_T - R_t, J_T - J_t \mid i_{t+1}, i_t, \cdots, i_{t-r+1}, R_t, J_t], \end{aligned} \tag{A.3}$$

Since $i_{t\text{-}r}$ does not contain information on $\{R_T\text{-}R_t, J_T\text{-}J_t\}$ beyond the information contained in $\{i_{t+1}, i_t, \cdots, i_{t\text{-}r+1}, R_t, J_t\}$, the second term in (A.3) becomes

$$\begin{aligned} &\Pr[R_T - R_t, J_T - J_t \mid i_{t+1}, i_t, \cdots, i_{t-r}, R_t, J_t] \\ &\Pr[R_T - R_t, J_T - J_t \mid i_{t+1}, i_t, \cdots, i_{t-r+1}, R_t, J_t], \end{aligned}$$

which is identical to the denominator in (A.2). Thus, Equation (A.1) is obtained.

References

Cosslett, Stephan R. and Lung-Fei Lee, 1985, Serial correlation in discrete variable models, *Journal of Econometrics* 27, 79-97.

Easley, David and Maureen O'Hara, 1987, Prices, trade size and information in securities markets, *Journal of Financial Economics* 19, 69-90.

Foster, F. Douglas and S. Viswanathan, 1993, Variations in trading volume, return volatility, and trading costs: Evidence on recent price formation models, *Journal of Finance* 48, 187-211.

Glosten, Lawrence and Lawrence Harris, 1988, Estimating the components of the bid-ask spread, *Journal of Financial Economics* 21, 123-142.

Gourieroux, Christian, Albert Holly and Alain Monfort, 1982, Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters, *Econometrica* 50, 63-80.

Hamilton, James D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357-384.

Hamilton, James D., 1990, Analysis of time series subject to changes in regime, *Journal of Econometrics* 45, 39-70.

Hasbrouck, Joel, 1988, Trades, quotes, inventories, and information, *Journal of Financial Economics* 22, 229-252.

Hasbrouck, Joel, 1991a, Measuring the information content of stock trades, *Journal of Finance* 46, 179-207.

Hasbrouck, Joel, 1991b, The summary informativeness of stock trades: An econometric analysis, *Review of Financial Studies* 4, 571-595.

Hasbrouck, Joel and Thomas Ho, 1987, Order arrival, quote behavior, and the return-generating process, *Journal of Finance* 42, 1035-1048.

Hausman, Jerry A., Andrew W. Lo and A. Craig MacKinlay, 1992, An ordered probit analysis of transaction stock prices, *Journal of Financial Economics* 31, 319-379.

Ho, Thomas and R. Macris, 1984, Dealer bid-ask quotes and transaction prices: An empirical study of some AMEX options, *Journal of Finance* 39, 23-45.

Keim, Donald B. and Ananth Madhavan, 1996, The upstairs market for large-block transactions: Analysis and measurement of price effects, *Review of Financial Studies* 9, 1-36.

Kim, Chang-Jin, 1994, Dynamic linear models with Markov-switching, *Journal of Econometrcs* 60, 1-22.

Kudo, Akio, 1963, A multivariate analogue of the one-sided test, *Biometrika* 50, 403-418.

Lee, Charles M. C. and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733-746.

Lee, Joon-Haeng and Wan-Sun Han, 1993, The analysis of the stock returns dynamics with multi-state Markov switching model (in Korean), *Korean Security Association Journal* 15, 561-595.

Madhavan, Ananth and Seymour Smidt, 1991, A Bayesian model of intraday specialist pricing, *Journal of Financial Economics* 30, 99-134.

McInish, Thomas H. and Robert A. Wood, 1995, Hidden limit orders on the NYSE, *Journal of Portfolio Management* 21, 19-26.

Shapiro, Alexander, 1985, Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints, *Biometrika* 72, 133-144.