Autoregressive hidden Markov switching models of count data

Nicholas Taylor*
Cardiff Business School

January 9, 2002

Abstract

This paper introduces an alternative hidden Markov switching model. In particular, an autoregressive hidden Markov switching model of count data is formulated and applied to financial data. Through application of this new model, a theoretically-motivated representation of the dynamics of the number of orders placed per unit of time (referred to as *order-flow*) on the London Stock Exchange is provided. Using the economic arguments of Rock (1996), the suitability of a 2-state autoregressive hidden Markov switching model is demonstrated in this context. This model provides the best fit amongst competing discrete-valued time series models. Moreover, the parameters of this model are found to vary in a predictable manner according to whether the morning, lunch time, or afternoon trading sessions are considered.

Key Words: Discrete-valued, time series, Markov models.

JEL Classification Code: G14; C32.

*Cardiff Business School, Cardiff University, Cardiff CF10 3EU, United Kingdom, TEL: (01633) 431096.

1 Introduction

Discrete-valued time series models have existed in the statistical literature for over 20 years. The vast majority of these models modify the Gaussian autoregressive moving average (ARMA) model to incorporate discrete-valued data (see, eg., Al-Osh and Alzaid, 1991). More recently, an alternative class of discrete-valued time series models have been introduced and are commonly referred to in the statistics literature as *hidden Markov switching models*. These models assume the data are generated by a different discrete distribution depending on the value of an underlying unobservable state variable (see MacDonald and Zucchini, 1997, for a review). A related class of hidden Markov switching models has recently appeared in the econometrics literature that incorporates time-dependency amongst adjacent observations. In particular, Hamilton (1989, 1990) introduces continuous-valued models where the parameters of interest are allowed to vary according to the value of an underlying unobservable state variable. For example, a Gaussian-based autoregressive model is considered where the constant and the autoregressive parameters are subject to switching. The purpose of this paper is to introduce a new class of models that are essentially discrete-valued versions of these autoregressive switching models.

Despite the vast number of applications of existing discrete-valued time series models, few have been applied in the area of finance. Those that have been applied have concentrated on modeling trade-by-trade stock price movements in irregularly spaced transaction data (Engle and Russell, 1997, 1998, Rydberg and Shephard, 1998, 1999, 2000, and Engle, 2000). For instance, Rydberg and Shephard (1999) model the activity, direction and size decomposition of price movements for two stocks listed on the New York Stock Exchange using a multivariate compound Poisson process. Other aspects of financial markets have not been considered for the application of discrete-valued time series models. It is this apparent gap in the applied statistics literature that is addressed in this paper. Using a high frequency dataset covering 40 stocks listed on the London Stock Exchange (LSE), the number of orders placed per unit of time (referred to as *order-flow*) on a trading platform is modeled using existing discrete-valued models and the autoregressive hidden Markov switching model introduced in the paper.

The paper is organized as follows: The next section introduces the new model in the context of existing (and related) models. Section 3 contains the results of a Monte Carlo experiment designed to examine the robustness of selection criterion when applied to the new model. Section 4 outlines the economic issue addressed in this paper and an application is provided. Section 5 concludes.

2 The model

Several authors have modified the Gaussian ARMA model to incorporate discrete-valued data (see Jacobs and Lewis, 1978a–c, 1983, McKenzie, 1985a,b, 1986, 1987, 1988a,b, Al-Osh and Alzaid, 1987, 1988, 1991, Alzaid and Al-Osh, 1988, 1990, 1993, Du and Li, 1991, and Al-Osh and Aly, 1992, for theoretical papers in the area). The nature of the data necessitates use of the binomial thinning operator of Steutal and Van Harn (1979). The basic first-order autoregressive model with Poisson marginal introduced by Al-Osh and Alzaid (1987) has the following specification:

$$y_t = \alpha \circ y_{t-1} + \epsilon_t \tag{1}$$

where $\alpha \in [0,1]$, $\epsilon_t \sim \text{Poisson}(\lambda)$ and where the binomial thinning operator $\circ$ is defined as

$$\alpha \circ y_{t-1} = \sum_{k=1}^{y_{t-1}} B_k \tag{2}$$

where B_k is the binary outcome of a Bernoulli process with success probability, $\Pr[B_k = 1] = \alpha$. For consistency with other notation used in this paper, this particular model will henceforth be referred to as a *1-state autoregressive hidden Markov model.*

The model introduced in this paper augments the model given by (1) to allow state-dependent parameter values. That is, α and λ are each allowed to take M different values according to the value of some underlying unobservable state s_t taking a value $j \in M$, hence (1) becomes

$$y_t = \alpha_j \circ y_{t-1} + \epsilon_{j,t} \tag{3}$$

where $\alpha_j \in [0,1]$ and $\epsilon_{j,t} \sim \text{Poisson}(\lambda_j)$. This augmentation of (1) is comparable with the state-dependent augmentation of the basic Gaussian autoregressive models of Hamilton (1989, 1990). Following Hamilton, the unobservable state is assumed to evolve according to an irreducible homogeneous first-order Markov chain.[1] The transition matrix associated with this Markov chain $\mathbf{P}$ is given by

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{21} & \dots & pM1 \\ p_{12} & p_{22} & \dots & pM2 \\ \vdots & \vdots & \dots & \vdots \\ p_{1M} & p_{2M} & \dots & pMM \end{bmatrix}. \tag{4}$$

This matrix implies that $\Pr(s_t = j | s_{t-1} = i, s_{t-2} = k, \dots) = \Pr(s_t = j | s_{t-1} = i) = p_{ij}$ and $\mathbf{P}'\mathbf{1} = \mathbf{1}$. As the Markov chain is assumed to be irreducible then there exists a unique, strictly positive, stationary distribution, denoted by the vector $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_M]$.[2] For example, for $M = 2$ this vector is given by

$$\boldsymbol{\pi} \equiv \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} = \begin{bmatrix} (1-p_{22})/(2-p_{11}-p_{22}) \\ (1-p_{11})/(2-p_{11}-p_{22}) \end{bmatrix}. \tag{5}$$

This vector gives the unconditional probability of being in state 1 and 2, respectively. The model given by (3) and (4) will henceforth be referred to as an *M-state autoregressive hidden Markov switching model.*

To estimate the parameters in (3) and (4) (collectively denoted by $\boldsymbol{\theta}$) by (unconditional) maximum likelihood, the unconditional density of y_t is required. Before this can be derived we require the density of y_t conditional on the unobservable random variable s_t. This is achieved by augmenting the density given by Al-Osh and Alzaid (1987) for estimation of (1) by inclusion of state-dependent parameters,

$$f(y_t | s_t = j; \boldsymbol{\theta}) = \begin{cases} \frac{(\lambda_j/(1-\alpha_j))^{y_t}}{y_t!} e^{-\lambda_j/(1-\alpha_j)} & \text{if } t = 1, \\ e^{-\lambda_j} \sum_{i=0}^{\min(y_{t-1}, y_t)} \frac{\lambda_j^{y_t - i}}{y_t - i} \binom{y_{t-1}}{i} \alpha_j^i (1-\alpha_i)^{y_{t-1}-i} & \text{if } t = 2, \dots, T. \end{cases} \tag{6}$$

To obtain the unconditional density of y_t, we sum (over all M) the above conditional density

[1] Second-order chains can also be used, however, we describe first-order chains only for ease of exposition.

[2] This vector equals the eigenvector associated with the unit eigenvalue of $\mathbf{P}$ and under certain conditions is referred to as the vector of *ergodic probabilities*. See Hamilton (1994, p. 681)

multiplied by the unconditional probability of being in state j given by (4),

$$f(y_t;\boldsymbol{\theta}) = \sum_{j=1}^{M} f(y_t|s_t = j;\boldsymbol{\theta}) \Pr(s_t = j) = \sum_{j=1}^{M} f(y_t|s_t = j;\boldsymbol{\theta})\pi_j. \tag{7}$$

To obtain the (unconditional) maximum likelihood estimates of $\boldsymbol{\theta}$, the following log-likelihood is maximized:

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f(y_t;\boldsymbol{\theta}) \tag{8}$$

subject to the constraint that $\pi_1 + \pi_2 + \ldots \pi_M = 1$ and $\pi_j \geq 1$ for $j = 1, 2, \ldots, M$.[3]

To test the significance of the maximum likelihood parameter estimators we make use of the parametric bootstrap methodology of Efron and Tibshirani (1993). This methodology is used in favor of an asymptotic test because of a lack of evidence concerning the accuracy of asymptotic tests covering even restricted versions of the model introduced in this paper. In particular, by allowing non-zero values of α_j, the model given by (3) is in fact a more general version of the M-state non-autoregressive hidden Markov switching models described in MacDonald and Zucchini (1997). Though the consistency and asymptotic normality of the estimators of these models has been established (Ryden, 1994), the accuracy has not. Indeed, for this very reason the parametric bootstrap methodology is often used in the context of non-autoregressive hidden Markov switching models (see, eg., Albert, 1991).

Under the assumption that $M = 2$, the parametric bootstrap technique can be broken down into the following steps:

1. Having estimated the parameters of the model, $\widehat{\boldsymbol{\theta}} = (\hat{p}_{12}, \hat{p}_{21}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\lambda}_1, \hat{\lambda}_2)$, the first stage of the parametric bootstrap technique involves generating T realizations of the fitted hidden Markov model. The first value of the state s_1^* is calculated from the stationary distribution $\boldsymbol{\pi}$ as follows:

$$s_1^* = \begin{cases} 1 & \text{if} \quad 0 \leq u_1 \leq \hat{\pi}_1, \\ 2 & \text{otherwise,} \end{cases}$$

where u_1 is a drawing from a uniform $[0, 1]$ random number. The values of s_t^* for $t =$

[3]Details concerning the algorithm used to maximize (8), the computation time used to estimate the parameters of the model, and other estimation details can be obtained upon request.

$2, 3, \ldots, T$ are then calculated using the values of s_{t-1}^* with evolution determined by the transition probability matrix $\widehat{\mathbf{P}}$.

2. Having generated a series of T two-state values of $\{s_t^*\}$, the required realization of the two-state values of $\{y_t^*\}$ are generated by drawing from the autoregressive model given by (3) with coefficients $\{\hat{\alpha}_1, \hat{\lambda}_1\}$ if $s_t^* = 1$ and $\{\hat{\alpha}_2, \hat{\lambda}_2\}$, otherwise.

3. Steps 1 and 2 are then repeated B times to give B independent sequences of $\{y_t^*\}$. For each of these realizations, the parameter estimates are calculated using the same method used to calculate the original set of estimates $\widehat{\boldsymbol{\theta}}$ from the original observations.

4. Having generated B parameter vectors, $\widehat{\boldsymbol{\theta}}^*(1), \ldots, \widehat{\boldsymbol{\theta}}^*(B)$, the parameter bootstrap estimator of the variance-covariance matrix of $\widehat{\boldsymbol{\theta}}$ is given by

$$\frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\boldsymbol{\theta}}^*(b) - \widehat{\boldsymbol{\theta}}^*(.)\right)'\left(\widehat{\boldsymbol{\theta}}^*(b) - \widehat{\boldsymbol{\theta}}^*(.)\right),$$

where

$$\widehat{\boldsymbol{\theta}}^*(.) = \frac{1}{B}\sum_{b=1}^{B}\widehat{\boldsymbol{\theta}}^*(b).$$

The estimated models considered in this paper are based on $B = 100$. Before proceeding to an application of these models the performance of certain selection criteria is considered.

3 Model selection

Selection of the correct number of states in non-autoregressive hidden Markov switching models is an important issue. Studies by Ryden (1995) and Zhang and Stine (2001) both find that the Akaike Information Criterion (AIC) and the Schwarz Information Criterion (SIC) perform well in this respect. However, the performance of these criteria in the context of autoregressive hidden Markov switching models is unknown. For this reason a Monte Carlo simulation is conducted and the performance of these criteria is assessed.

Five data generating processes (DGP's) are considered. The first two are based on the 1-state autoregressive hidden Markov model (ie. the discrete-valued first-order autoregressive model of

Al-Osh and Alzaid, 1987) given by (1). As these DGP's are not dependent on an unobserved state they are referred to as 1-state autoregressive hidden Markov processes. The latter three DGP's are based on the 2-state autoregressive hidden Markov switching model given by (3) and (5). The parameter values used are based on their empirical counterparts found in the subsequent empirical section and are given below:

DGP A: $\alpha = 0, \quad \lambda = 7$

DGP B: $\alpha = 0.15, \quad \lambda = 5$

DGP C: $\alpha_1 = \alpha_2 = 0, \quad \lambda_1 = 4, \quad \lambda_2 = 12, \quad \pi_1 = \frac{2}{3}, \quad \pi_2 = \frac{1}{3}$

DGP D: $\alpha_1 = \alpha_2 = 0.15, \quad \lambda_1 = 3, \quad \lambda_2 = 11, \quad \pi_1 = \frac{2}{3}, \quad \pi_2 = \frac{1}{3}$

DGP E: $\alpha_1 = 0.15, \quad \alpha_2 = 0.30, \quad \lambda_1 = 3, \quad \lambda_2 = 10, \quad \pi_1 = \frac{2}{3}, \quad \pi_2 = \frac{1}{3}$

During each replication (R) of the data and for each DGP the following (and corresponding) models are estimated: a 1-state non-autoregressive hidden Markov model (denoted HM(1)), a 1-state autoregressive hidden Markov model (denoted HM(1)-AR(1)), a 2-state non-autoregressive hidden Markov switching model (denoted HM(2)), a 2-state autoregressive hidden Markov switching model with the same AR coefficients across states (denoted HM(1)-AR(1)), and a 2-state autoregressive hidden Markov switching model with different AR coefficients in each state (denoted HM(1)-AR(2)).

Upon estimation of the model (by unconditional maximum likelihood) the AIC and SIC are calculated. These criteria are then used to give an indication of the best model given each replication of the DGP. The proportion of times each model is selected for each DGP is chosen as the metric of importance. Moreover, a success is defined as selection of the most appropriate model given each DGP. For instance, the most appropriate model for DGP A is the HM(1) model. The results of conducting this experiment for $R = 100$ and $T = 500$ are given in Table I. For DGP A the success rates for the AIC and SIC are 88% and 96%, respectively. However, when an autoregressive term is introduced the AIC and SIC are less successful. In particular, these criteria select the most appropriate model only 41% and 16% of the time. When the 2-state hidden Markov switching models are considered, the success rates of the AIC are 50%, 77%, and

66%, for DGP's C, D and E, respectively. Despite these relatively low success rates, two other points should be considered in support of the AIC and SIC. First, on no occasion are the 1-state hidden Markov models selected for these DGP's. Second, when autoregressive hidden Markov switching processes are considered (DGP's D and E), autoregressive hidden Markov switching models (HM(1)-AR(1) or HM(1)-AR(2)) are selected 99% and 98% of the time when the AIC is used. One can conclude that, for these empirically calibrated DGP's, the AIC and SIC are largely successful.

4 An Application

This section contains a description of the economic issue addressed, motivates use of the model introduced in this paper, describes the data used, and contains the results obtained when estimating various hidden Markov models.

4.1 The economic issue

The dynamics of the number of orders placed (termed *order-flow*) on the London Stock Exchange (LSE) are examined in this paper. This issue has previously been investigated using data from other financial markets. For instance, Biais, Hillion, and Spatt (1995), Hamao and Hasbrouck (1995), and Harris and Hasbrouck (1996) find evidence of order-flow dependence for the Paris, Tokyo, and New York limit-order exchanges, respectively. Three alternative explanations are posited for this order-flow dependence: The first of these argues that traders indulge in strategic order splitting. In particular, it is possible that informationally motivated trading will have less market impact if a large order is split into several smaller orders. Alternatively, traders may imitate each others behaviour. This is likely to be the case when traders observe the orders placed by informed traders. Finally, traders may place orders in response to the release of an important piece of information. However, traders may react to information at different speeds. This would again imply serial dependence in order-flow.

The analysis undertaken in this paper innovates on previous empirical studies of order-flow dynamics in two distinct ways: First, the discreteness and serial dependence in the order-flow data is explicitly considered. This is achieved using the 1-state autoregressive hidden

Markov models described in Section II. Second, alternative models are considered which take into account the institutional features of the LSE. In particular, these features imply use of a 2-state autoregressive hidden Markov switching model. The first of these innovations is rather obvious but the second requires some explanation.

The motivation for use of a 2-state autoregressive hidden Markov switching model is the ability of traders to switch between the SETS and SEAQ trading platforms available on the LSE.[4] This motivation is based on the arguments of Rock (1996) who focuses on the level of the unobservable fundamental price of the stock with respect to the prices quoted by market makers and the prices placed in *competing* limit-orders. The adverse selection problem faced by market makers may result in obsolete orders and hence a transition in trading from SETS to SEAQ. In addition to the arguments of Rock (1996), the empirical findings of Biais, Hillion, and Spatt (1995) and the theoretical arguments of Parlour (1998) are used. In particular, both of these papers find that a thick order book results in a large number of trades while a thin order book results in a large number of orders. This is because when the order book is thick, the likely of a trade via a limit-order is small because of 'crowding out'. Thus, traders can only trade by placing market orders. Meanwhile, a thin order book offers traders the possibility that preferential prices will be obtained through the time-preferencing nature of limit-order markets.

The argument can be summarized as follows: Traders can trade immediately through a market maker on SEAQ or can hope to obtain a better price by placing a limit-order on SETS. For example, assume that a trader places a bid limit-order on SETS with a specified price of £9.80 and that this represents the best order currently offered. Meanwhile, assume that the best ask offered specifies a price of £10.20, that the (unobservable) fundamental price of the stock is £10.00, and that a market maker quotes bid and ask prices of £10.20 and £9.80, respectively. Consider the situation where market makers revise their assessment of the true fundamental price of the stock. If the new fundamental price is downwardly revised then the

[4] A fundamental change to the UK equity market occurred when the LSE introduced the Stock Exchange Electronic Trading Service (SETS) on October 20, 1997. The most important aspect of SETS was the introduction of an order-driven trading system alongside the existing quote-driven Stock Exchange Automatic Quotation (SEAQ) trading system. Under the current dual trading system offered by the LSE, market makers post quotes but these are only indicative. Competition for order flow is increased by allowing members to post orders on the electronic order book via a computer terminal. Under this order-driven system, members can post a variety of firm orders, including limit-orders and 'at best' (or market) orders. Limit-orders allow members to post orders with a specified price and volume. Only when a corresponding order is placed with a matching price will a trade occur.

revised quotes offered by the market maker may now result in obsolete best prices offered on SETS. For example, assume that the downward revised bid and ask quotes offered by the market maker are now £9.60 and £9.40. The new situation implies that the trader placing the original bid limit-order with price £9.80 can purchase the stock more cheaply via the market maker on SEAQ. The redundant bid orders on SETS are then deleted over a (short) period of time to avoid being hit by the ask orders still on the order book. As these bid orders disappear so to do the ask orders as they have very little chance of being executed because they contain prices far away from the fundamental price. Thus, after a period of time a thin order book is observed. At this point in time traders will start placing large numbers of orders on SETS with prices set around the new fundamental price.[5]

The above argument implies that order-flow takes one of two discrete distributions at each point in time. If the majority of traders are making use of SETS then order-flow on SETS is likely to be high. However, if SEAQ is the primary platform used by traders then order-flow (on SETS) is likely to be low. Moreover, the argument implies that this intensity will switch over time according to the state of the order book. The 'state' in the current context is defined as the relationship between the fundamental price of the stock and the range of prices placed on SETS. This follows from the arguments of Rock (1996) where trading platform transition occurs when the fundamental price lies outside the range of prices placed on SETS. As the fundamental price cannot be observed then the 'state' must necessarily be unobservable. Moreover, if the fundamental price follows a random walk then there is likely to be persistence in the process followed by the state.

All of these features suggests use of a 2-state hidden Markov switching model. Moreover, the autoregressive version of this model as given by (3) and (5) is used to allow for first-order dependence in order-flow within each state. This latter feature will incorporate the effects of strategic order splitting and/or traders imitating each others behaviour.

[5]This argument rather subtly relies on another feature of the fundamental price. In particular, it has been assumed that the fundamental price on SEAQ more efficiently responds to information. For instance, when there is a change in the fundamental price this is more rapidly revealed on SEAQ via the market makers assessment of the fundamental price. Otherwise, there would be no reason for switching trading platforms. Such efficiency on SEAQ has been documented in a recent study by Ellul (2000). Using a Kalman smoothing technique to generate the fundamental stock price, he finds that price volatility on SEAQ is lower than on SETS. From this he concludes that the price on SEAQ more efficiently tracks the unobservable fundamental price.

4.2 Data description

Order-flow data covering 40 stocks over the period August 3rd, 1998 to October 31th, 1998 were obtained from the London Stock Exchange Data Service. The stocks and the sectors to which they belong are listed in Table A in the Appendix. These stocks are FTSE100 companies randomly selected from each of the major sectors. We use a frequency of 15 minutes which is sufficiently low to avoid use of stale information and high enough to capture short-run movements in the variable of interest.[6] The resulting sample consists of 1920 observations per stock.

4.3 Empirical results

Five models are estimated for the order-flow associated with each stock. The models considered are; the HM(1) model, the HM(1)-AR(1) model, the HM(2) model, the HM(1)-AR(1) model, and the HM(1)-AR(2) model. Space limitations preclude detailed presentation of model estimates for each stock. Instead, results are given for a typical stock and a summary of the estimates is supplied for all stocks.

It is apparent from the data that order-flow exhibits intraday periodicity.[7] This can be observed from Figure 1 which shows the mean intraday order-flow for all stocks in the sample. In particular, order-flow tends to be high during the morning period between 9.00 and 12.30, and the afternoon period between 14.30 and 16.30.[8] By contrast, the lunch time period between 12.30 and 14.30 tends to be characterized by low order-flow. For this reason each model is estimated separately for each of these time periods. On each occasion, the HM(1)-AR(2) model provides the best fit of the data according to the AIC and the SIC. The parameter estimates associated with this model for a typical stock, in this case 3I Group, are given below:

[6] The trading day starts at 9.00am and ends at 4.30pm. However, the way that the variables are constructed means that the variables observed at 9.00am are the same as the variables observed at 4.30pm on the previous day. To avoid this duplication we only consider variables observed between 9.15am and 4.30pm (inclusive) during each trading day.

[7] Such intraday periodicity is often found in financial data. See Baillie and Bollerslev (1989), Schwert (1990), Harvey and Huang (1991), Gallant, Rossi, and Tauchen (1992), and Bollerslev and Ghysels (1996), for empirical examples.

[8] The increase in order-flow observed at 14.30 is due to the opening of the major US markets at this time.

TIME PERIOD: 9.00 TO 12.15

$$y_t = \underset{(0.0482)}{0.2523} \circ y_{t-1} + \epsilon_{1,t}, \quad \text{with} \quad \hat{\lambda}_1 = \underset{(0.3627)}{5.3452} \quad \text{and} \quad \hat{p}_{12} = \underset{(0.0833)}{0.0454}, \tag{9}$$

$$y_t = \underset{(0.0303)}{0.4176} \circ y_{t-1} + \epsilon_{2,t}, \quad \text{with} \quad \hat{\lambda}_2 = \underset{(0.8941)}{15.1621} \quad \text{and} \quad \hat{p}_{21} = \underset{(0.1174)}{0.0913}. \tag{10}$$

TIME PERIOD: 12.30 TO 14.15

$$y_t = \underset{(0.0547)}{0.1382} \circ y_{t-1} + \epsilon_{1,t}, \quad \text{with} \quad \hat{\lambda}_1 = \underset{(0.1486)}{1.6300} \quad \text{and} \quad \hat{p}_{12} = \underset{(0.0986)}{0.0645}, \tag{11}$$

$$y_t = \underset{(0.0479)}{0.3011} \circ y_{t-1} + \epsilon_{2,t}, \quad \text{with} \quad \hat{\lambda}_2 = \underset{(0.4837)}{6.2829} \quad \text{and} \quad \hat{p}_{21} = \underset{(0.1606)}{0.1385}. \tag{12}$$

TIME PERIOD: 14.30 TO 16.30

$$y_t = \underset{(0.0631)}{0.1280} \circ y_{t-1} + \epsilon_{1,t}, \quad \text{with} \quad \hat{\lambda}_1 = \underset{(0.2388)}{2.9421} \quad \text{and} \quad \hat{p}_{12} = \underset{(0.0938)}{0.0543}, \tag{13}$$

$$y_t = \underset{(0.0444)}{0.3180} \circ y_{t-1} + \epsilon_{2,t}, \quad \text{with} \quad \hat{\lambda}_2 = \underset{(0.7897)}{10.3919} \quad \text{and} \quad \hat{p}_{21} = \underset{(0.1502)}{0.0997}. \tag{14}$$

where numbers in parentheses are the parametric bootstrap standard errors.

The intraday periodicity in order-flow intensity is apparent from equations (9)–(14). In particular, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are higher for the morning and afternoon periods than the lunch time periods. For all periods, these parameter estimates indicate a large degree of heterogeneous order-flow intensity across states. This heterogeneity is most likely due to traders switching between SETS and SEAQ, with high intensity order-flow representing the state where traders make primary use of SETS. The switching probabilities are given by $\hat{p}_{12}$ and $\hat{p}_{21}$. These estimates indicate a large degree of persistence in the state value. Moreover, from these probabilities one can calculate the stationary distribution of the states using equation (5). The resulting probabilities associated with state 2 (ie. traders using SETS) for each time period are $\pi_2 = \{0.3321, 0.3177, 0.3526\}$. Thus it would appear that SETS is not the primary trading platform used by traders. This observation is backed up by anecdotal evidence that SETS has failed to attract a significant amount of trading activity. The final parameter estimates of interest are the autoregressive coefficients within each state. For each time period considered, these coefficients are significantly greater than zero. Moreover, the degree of serial dependence is greater when in state 2 (ie. traders use SETS) than when in state 1 (ie. traders use SEAQ). This suggests that some form of order splitting, or related behaviour, is more prevalent when traders make primary use of SETS.

A similar picture emerges when the cross-sectional means (and there associated standard errors) are calculated across all stocks. There results are given in Table II. Of the five models considered the HM(1)-AR(1) model and the HM(1)-AR(2) model provide the best fit of the data for each stock. Moreover, the latter of these models universally provides the best fit during the morning and afternoon periods when the AIC is used. Only during the (quiet) lunch time period does the former model seem to fit the data (approximately) as many times of the latter model. The parameters of the latter model indicate heterogeneous intensities across states, positive autocorrelation within states, and less use of SETS (ie. state 2) than SEAQ (ie. state 1).

5 Concluding remarks

The model introduced in the paper is capable of modeling count data where the intensity of the process is expected to switch across states and where positive first-order serial dependence is expected within each state. Such a model is particular useful when describing order-flow on the LSE. The particular design of the trading platforms available on the LSE necessitates use of a model that is capable of allowing for the above dynamics.

Future research is likely to concentrate on generalizing the model in two ways. First, the model considered in this paper is based on a Poisson marginal distribution. However, it is a trivial matter to augment the model by allowing alternative discrete marginal distributions such as the negative binomial or geometric distribution. This has already been achieved in the context of discrete-valued ARMA models (see McKenzie, 1986, Al-Osh and Aly, 1992). Second, only first-order dynamics were considered in this paper. However, higher order autoregressive dynamics with inclusion of moving average components can be incorporated into the model.

This can be achieved by augmenting the discrete-valued ARMA model of Alzaid and Al-Osh (1993) to a hidden Markov switching model context in a similar fashion to that achieved in this paper.

References

Albert, P. (1991). A two-state Markov mixture model for a time series of epileptic seizure counts. *Biometrics* 47, 1371–1381.

Al-Osh, M. and Aly, E. (1992). First-order autoregressive time series with negative binomial and geometric marginals. *Commun. Statist.-Theory Meth.* 21, 2483–2492.

Al-Osh, M. and Alzaid, A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* 8, 261–275.

——— (1988). Integer-valued moving average (INMA) process. *Statist. Papers* 29, 281–300.

——— (1991). Binomial autoregressive moving average models. *Stoch. Models* 7, 261-282.

Alzaid, A. and Al-Osh, M. (1988). First-order integer-valued autoregressive ((INAR(1)) process: Distributional and regression properties. *Statist. Neerl.* 42, 53–61.

——— (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *J. Appl. Prob.* 27, 314–324.

——— (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions. *Ann. Inst. Statist. Math.* 45, 223-232.

Baillie, R. and Bollerslev, T. (1989). The message in daily exchange rates: A conditional variance tale. *Journal of Business and Economic Statistics* 7, 297–305.

Biais, B., Hillion, P. and Spatt, C. (1995). An empirical analysis of the limit-order book and the order-flow in the Paris Bourse. *Journal of Finance* 50, 1655–1689.

Bollerslev, T. and Ghysels, E. (1996). Periodic autoregressive conditional heteroscedasticity. *Journal of Business and Economic Statistics* 14, 139–151.

Du, J. and Li, Y. (1991). The integer-valued autoregressive INAR(p) model. *J. Time Ser. Anal.* 12, 129–142.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Ellul, A. (2000). As you like it: An investigation of trading behaviour and price volatility on auction and dealership market architectures. Working paper, London School of Economics, London.

Engle, R. (2000). The econometrics of ultra-high frequency data. *Econometrica* 68, 45–67.

Engle, R. and Russell, J. (1997). Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal of Empirical Finance* 12, 187–212.

——— (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66, 1127–1162.

Gallant, A., Rossi, P. and Tauchen, G. (1992). Stock prices and volume. *Review of Financial Studies* 5, 199–242.

Hamao, Y. and Hasbrouck, J. (1995). Securities trading in the absence of dealers: Trades and quotes of the Tokyo Stock Exchange. *Review of Financial Studies* 8, 849–878.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.

Hamilton, J. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics* 45, 39–70.

Hamilton, J. (1994). *Time Series Analysis.* Princeton University Press, Princeton.

Harris, L. and Hasbrouck, J. (1996). Market vs. limit-orders: The SuperDot evidence on order submission strategy. *Journal of Financial and Quantitative Analysis* 31, 213–231.

Harvey, C. and Huang, C. (1991). Volatility in the foreign currency futures market. *Review of Financial Studies* 4, 543–569.

Jacobs, P. and Lewis, P. (1978a). Discrete time series generated by mixtures I: Correlation and runs properties. *J. R. Statist. Soc. B* 40, 94–105.

——— (1978b). Discrete time series generated by mixtures II: Asymptotic properties. *J. R. Statist. Soc. B* 40, 222–228.

——— (1978c). Discrete time series generated by mixtures III: Autoregressive processes (DAR(p)). Technical report NPS55-78-022, Naval Postgraduate School, Monterey, California.

——— (1983). Stationary discrete autoregressive moving average time series generated by mixtures. *J. Time Ser. Anal.* 4, 19–36.

MacDonald, I. and Zucchini, W. (1997). *Hidden Markov Models and Other Models for Discrete-valued Time Series.* Chapman and Hall, New York.

McKenzie, E. (1985a). Contribution to the discussion of Lawrance and Lewis (1985). *J. R. Statist. Soc. B* 47, 187–188.

——— (1985b). Some simple models for discrete variate time series. *Water Resour. Bull.* 21, 645–650.

——— (1986). Autoregressive moving average processes with negative-binomial and geometric marginal distributions. *Adv. Appl. Prob.* 18, 679–705.

——— (1987). Innovation distributions for gamma and negative binomial autoregressions. *Scand. J. Statist.* 14, 79–85.

——— (1988a). The distributional structure of finite moving-average processes. *J. of Appl. Prob.* 25, 313–321.

——— (1988b). Some ARMA models for dependent sequences of Poisson counts. *Adv. Appl. Prob.* 20, 822–835.

Parlour, C. (1998). Price dynamics in limit-order markets. *Review of Financial Studies* 11, 789–816.

Rock, K. (1996). The specialist's order book and price anomalies. *Review of Financial Studies* 9, 1–20.

Ryden, T. (1994). Consistent and asymptotic normal parameter estimates for hidden Markov models. *Ann. Statist.* 22, 1884–1895.

Ryden, T. (1995). Estimating the order of Hidden Markov models. *Statistics* 26, 345–354.

Schwert, W. (1990). Indexes of US stock prices from 1802 to 1987. *Journal of Business* 63, 399–426.

Steutal, F. and van Harn, K. (1979). Central limit theorems for finite Walsh-Fourier transforms of weakly stationary time series. *J. Time Ser. Anal.* 6, 261–267.

Zhang, J. and Stone, R. (2001). Autocovariance structure of Markov regime switching models and model selection. *J. Time Ser. Anal.* 22, 107–124.

APPENDIX: Stock Sample Details

Table A: The sub-sample of FTSE100 stocks

This table gives the stocks examined and their respective sectors.

Stock	Sector
3I Group	Investment Trusts
Abbey National	Banks
Alliance and Leicester	Banks
Allied Domecq	Alcoholic Beverages
Asda Group	Food Retailers
Associated British Foods	Food Manufacturers
Bank of Scotland	Banks
British Gas	Oil and Gas
BOC Group	Chemicals
Boots	General Retailers
British Energy	Electricity
British Land	Properties
Cable and Wireless	Telecommunications
Cadbury Schweppes	Food Manufacturers
Carlton Communications	Media
CGU	Insurance
EMI Group	Media
General Electric	Electricity
Great Universal Stores	General Retailers
Halifax	Banks
Hays	Support Services
Imperial Chemical Industries	Chemicals
Ladbroke Group	General Retailers
Land Securities	Properties
Legal and General	Insurance
Lucasvarity	Engineering
Marks and Spencer	General Retailers
National Power	Electricity
National Westminster Bank	Banks
Norwich Union	Insurance
Orange	Telecommunications
Pearson	Media
Peninsular and Orient Steam	Transport
Powergen	Electricity
Prudential Corporation	Insurance
Railtrack Group	Transport
Reckitt and Coleman	Household Goods
Reed International	Media
Reuters Group	Media
Rolls-Royce	Engineering

Table I: Model selection performance

This table contains the proportion of times the Akaike and the Schwarz Information Criterion select the correct model given one of five data generating processes. The data generating processes considered are: A. 1-state non-autoregressive hidden Markov model with $\{\lambda = 7, \alpha = 0\}$, B. 1-state autoregressive hidden Markov model with $\{\lambda = 5, \alpha = 0.15\}$, C. 2-state non-autoregressive hidden Markov switching model with $\{\lambda_1 = 4, \lambda_2 = 12, p_{12} = 0.4, p_{21} = 0.2, \alpha_1 = 0, \alpha_2 = 0\}$, D. 2-state autoregressive hidden Markov switching model with $\{\lambda_1 = 3, \lambda_2 = 11, p_{12} = 0.4, p_{21} = 0.2, \alpha_1 = \alpha_2 = 0.15\}$, and E. 2-state autoregressive hidden Markov switching model with $\{\lambda_1 = 3, \lambda_2 = 10, p_{12} = 0.4, p_{21} = 0.2, \alpha_1 = 0.15, \alpha_2 = 0.30\}$. The models considered under these data generating processes are: 1-state non-autoregressive hidden Markov model (HM(1)), 1-state autoregressive hidden Markov model (HM(1)-AR(1)), 2-state non-autoregressive hidden Markov switching model (HM(2)), 2-state autoregressive hidden Markov switching model with the same AR coefficients across states (HM(1)-AR(1)), and 2-state autoregressive hidden Markov switching model with different AR coefficients in each state (HM(1)-AR(2)).

	Data Generating Process				
Estimated Model	A	B	C	D	E
Panel A: Akaike Information Criterion					
HM(1)	0.88	0.48	0.00	0.00	0.00
HM(1)-AR(1)	0.11	0.41	0.00	0.00	0.00
HM(2)	0.00	0.05	0.50	0.01	0.02
HM(2)-AR(1)	0.01	0.02	0.26	0.77	0.32
HM(2)-AR(2)	0.00	0.04	0.24	0.22	0.66
Panel B: Schwarz Information Criterion					
HM(1)	0.96	0.82	0.00	0.00	0.00
HM(1)-AR(1)	0.04	0.16	0.00	0.00	0.00
HM(2)	0.00	0.01	0.73	0.14	0.08
HM(2)-AR(1)	0.00	0.00	0.23	0.78	0.60
HM(2)-AR(2)	0.00	0.01	0.04	0.08	0.32

Table II: Model estimates

This table gives the cross-sectional mean parameter values across all 40 stocks. The models estimated are: I. 1-state non-autoregressive hidden Markov model, II. 1-state autoregressive hidden Markov model, III. 2-state non-autoregressive hidden Markov switching model, IV. 2-state autoregressive hidden Markov switching model with the same AR coefficients across states, and V. 2-state autoregressive hidden Markov switching model with different AR coefficients in each state. The cross-sectional mean Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) are also given along with the number of times these criteria select the models estimated across the 40 stocks considered (#AIC* and #SIC*, respectively). Numbers in parentheses are the cross-sectional standard errors of the parameter estimates.

	Model				
Parameter	I	II	III	IV	V
Panel A: Sample period: 9.00am to 12.15pm					
λ	9.56 (2.30)	7.97 (1.76)			
α		0.20 (0.02)			
λ_1			6.02 (1.10)	5.22 (1.71)	4.68 (1.14)
λ_2			12.05 (4.32)	10.93 (3.41)	14.04 (2.34)
α_1				0.26 (0.07)	0.19 (0.02)
α_2				0.26 (0.07)	0.37 (0.04)
π_1			0.49 (0.24)	0.50 (0.24)	0.70 (0.04)
π_2			0.51 (0.24)	0.50 (0.24)	0.30 (0.04)
AIC	8.65	8.30	8.93	7.61	6.70
SIC	8.66	8.31	8.95	7.64	6.73
#AIC*	0	0	0	0	40
#SIC*	0	0	0	0	40

Table II: Model estimates (continued)

	Model				
Parameter	I	II	III	IV	V
Panel B: Sample period: 12.30pm to 2.15pm					
λ	6.50 (1.40)	5.42 (1.09)			
α		0.16 (0.03)			
λ_1			3.88 (1.09)	2.96 (0.94)	2.94 (0.73)
λ_2			11.88 (2.36)	10.60 (1.89)	9.84 (1.72)
α_1				0.16 (0.04)	0.15 (0.04)
α_2				0.16 (0.04)	0.29 (0.08)
π_1			0.66 (0.09)	0.67 (0.08)	0.67 (0.05)
π_2			0.34 (0.09)	0.33 (0.08)	0.33 (0.05)
AIC	7.02	6.85	5.88	5.79	5.75
SIC	7.02	6.87	5.91	5.83	5.79
#AIC*	0	0	0	10	30
#SIC*	0	0	0	23	17

Table II: Model estimates (continued)

Parameter	Model I	Model II	Model III	Model IV	Model V
Panel C: Sample period: 2.30pm to 4.30pm					
λ	12.06	10.00			
	(2.80)	(2.26)			
α		0.17			
		(0.03)			
λ_1			6.67	6.35	6.16
			(0.85)	(1.22)	(1.55)
λ_2			10.13	9.53	15.80
			(4.33)	(3.77)	(2.65)
α_1				0.28	0.10
				(0.12)	(0.04)
α_2				0.28	0.30
				(0.12)	(0.07)
π_1			0.33	0.35	0.62
			(0.21)	(0.19)	(0.05)
π_2			0.67	0.65	0.38
			(0.21)	(0.19)	(0.05)
AIC	9.10	8.85	11.39	8.82	7.00
SIC	9.11	8.86	11.41	8.85	7.04
#AIC*	0	0	0	0	40
#SIC*	0	0	0	2	38

Figure 1: Intraday order-flow periodicity