Dynamics of trade–by–trade price movements: decomposition and models

Tina Hviid Rydberg
Nuffield College, Oxford OX1 1NF, UK.
tina.rydberg@nuf.ox.ac.uk

Neil Shephard
Nuffield College, Oxford OX1 1NF, UK.
neil.shephard@nuf.ox.ac.uk

February 2, 1999

Abstract

In this paper we introduce a decomposition of the joint distribution of price changes of assets recorded trade–by–trade. Our decomposition means that we can model the dynamics of price changes using quite simple and interpretable models which are easily extended in a great number of directions, including using durations and volume as explanatory variables. Thus we provide an econometric basis for empirical work on micro market structure using time series of transactions data.

We use maximum likelihood estimation and testing methods to assess the fit of the model to a year of IBM stock price data taken from the New York Stock Exchange.

Keywords: Activity, autologistic, conditional independence, decomposition, directions, durations, forecasting, GLARMA, logarithmic distribution, prediction decomposition, size, transactions data.

1 Introduction

In a recent paper Rydberg and Shephard (1998) proposed a continuous time framework for the evolution of the prices and times at which transactions are carried out on a common asset traded on a stock exchange trading only with fixed tick sizes. In almost all financial markets prices are restricted to such lattice points. For example this is the case for stocks traded at the New York Stock Exchange (NYSE) which can be seen from Figure 1 which displays the price of all NYSE based transaction on the IBM stock on four different days in 1995. In 1995 the tick size was 1/8th of a dollar, in 1997 it changed to 1/16th of a dollar. See O'Hara (1995, Ch. 1) for an introduction to markets and market making on the NYSE.

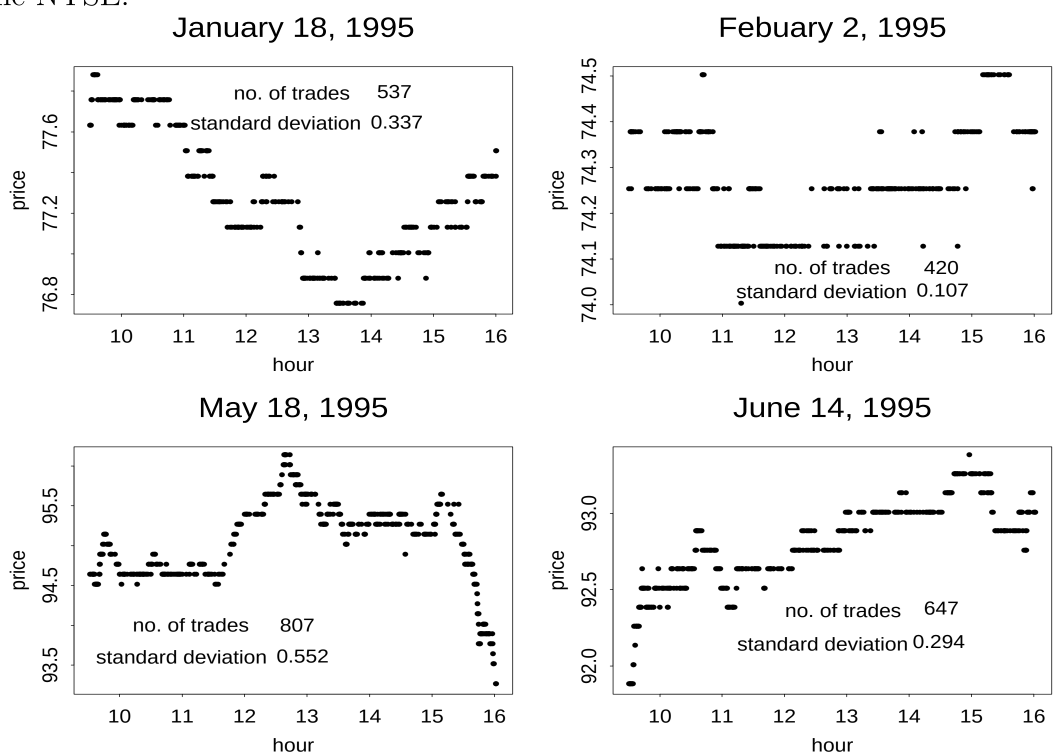

Figure 1: *Plot of all traded IBM prices at the New York stock exchange on four different days in 1995. A trade is represented as a dot.*

In order to model the discreteness of the price changes a compound process was suggested in Rydberg and Shephard (1998). It had the following structure. Let $p(u)$ denote the price of the asset at time u, then they allowed

$$p(u) = p(0) + \sum_{t=1}^{N(u)} Z_t, \tag{1}$$

where $N(u)$ is the number of trades recorded up until time u and Z_t is the price movement or change associated with the $t-th$ trade. Rydberg and Shephard (1998) modelled $N(u)$ as a counting process with new arrivals being generated by a Cox process[1], that is a Poisson process with a random intensity. We should note immediately that for the IBM dataset around 80% of the $\{Z_t\}$ take on the value zero, reflecting the fact that trades can occur without changing the price of the asset due to the bid-ask spread. Of the 33,184 non-zero $\{Z_t\}$ in 1995, 261 were greater than one tick in absolute value, with just 18 price movements being larger than 3 ticks in absolute value during the whole year. These figures are not atypical for NYSE trades – see for example Campbell, Lo, and MacKinlay (1997, pp. 107–128). In general, the dynamics of the Cox and price movements processes can be adapted to a wide class of filtrations involving just their own past or more extensive information sets. This is purely an issue of combining both the empirical evidence and a priori economic theory, reflecting both the purpose of the modelling exercise and the data generating mechanism.

Rydberg and Shephard (1998) were unspecific about the price innovation process beyond the use of simple descriptive Markov chains – putting off a full discussion of this issue to this paper. A special case (which assumed the $\{Z_t\}$ were independent over time and of the counting process) of this style of continuous time construction has been independently proposed by Rogers and Zane (1998), although that paper did not assess the empirical evidence behind such a construction. Some of the econometric issues which arise with unequally spaced financial data were discussed at length in Engle (1996).

In this paper we study the detailed joint distribution of price movements, returning to the problem of modelling using larger filtrations later in the paper. For expositional convenience we will assume the tick size of the market is one and so have rescaled our data accordingly. Then let $Z_t \in I$ be an integer process and $\mathcal{F}_t = \sigma(Z_s : s \leq t)$ be its natural sigma algebra or filtration (sequential information set). Hence $\mathcal{F}_{t-1}$ contains all the past integers from time 1 to $t-1$. We will first be interested in the joint distribution of the movements which are given by

$$\Pr(Z_1, ..., Z_n|\mathcal{F}_0) = \prod_{t=1}^{n} \Pr(Z_t|\mathcal{F}_{t-1}),$$

[1] An convenient example of a Cox process is the influential autoregressive conditional duration (ACD) model advocated by Engle and Russell (1998), which allows straightforward likelihood based econometric inference.

using a prediction decomposition. The problem will be specifying $\Pr(Z_t|\mathcal{F}_{t-1})$. Prior to the first draft of this paper, Russell and Engle (1998) have proposed using a conditional multinomial style model for specifying $\Pr(Z_t|\mathcal{F}_{t-1})$, in a sense generalising previous work of Hausman, Lo, and MacKinlay (1992) on multivariate probit models for transaction data. Also Hasbrouck (1996) has proposed a class of dynamic latent variable models for efficient prices and traders cost which uses economically motivated truncation to enforce prices to live on lattices. We will compare the Russell–Engle and Hasbrouck models to our suggestion towards the end of our paper. The general issue of discreteness of asset prices is discussed at length by Campbell, Lo, and MacKinlay (1997, pp. 98–144).

Other work which relates to this topic includes a paper by Meddahi, Renault, and Werker (1998) who combined GARCH models for price returns with autoregressive duration style models for the times between trades. In this context this work has the disadvantage that their price process does not live on the required lattice observed in the data. The paper by Darolles, Gourieroux, and Le Fol (1998), which we first saw after the circulation of the first draft of this paper, is much closer to the framework looked at in Rydberg and Shephard (1998) and Rogers and Zane (1998). The Darolles, Gourieroux, and Le Fol (1998) paper uses the structure (1) but assumes the Z_t process is Markov living on the points $-1, 0, 1$. This modelling assumption is combined with a reduction in the dataset by using quote data to allow them only to model buys from the market maker — which in turn makes the assumption that the price movements are at most one tick in absolute value more realistic.

Our paper has the following basic structure. In Section 2 we introduce our decomposition of the price movements. Section 3 will look at our initial empirical models for the activity, direction and size of price movements – taken together these three models yield an overall model of price movements. Section 4 places our suggestion in the context of the literature, as well as suggesting various extensions of the basic model construct. Section 5 concludes.

2 Decomposition of price movements

Potentially the distribution of $\Pr(Z_t|\mathcal{F}_{t-1})$ can be quite complicated. Our approach is to break down the pieces of Z_t into bits and then model these sequentially. Note that there is no loss of information in this decomposition.

To carry out our decomposition define the $t-th$ price move as

$$Z_t = A_t D_t S_t.$$

We will let A_t take on only two values: $0, 1$. When $A_t = 0$, we define for notational convenience (there is no loss in doing this), $D_t = S_t = 0$. Otherwise, when $A_t = 1$ we let D_t and S_t live on the structure:

$$D_t = -1, 1 \qquad \text{and} \qquad S_t = 1, 2, \ldots$$

Thus we have that if A_t is zero then Z_t must be zero. This means the price does not move or, in other words, is *In-Active*. If $A_t = 1$ then there are *Active* price movements. The non–zero price movement must be $Z_t = D_t S_t$. Likewise, if we assume $A_t = 1$, then D_t controls the *Direction* of the price move. If $D_t = 1$ the price moves upwards, else it moves downwards. Finally, S_t controls the *Size* of price movements. This suggests the decomposition of price movements into

$$\Pr(Z_t = 0|\mathcal{F}_{t-1}) = \Pr(A_t = 0|\mathcal{F}_{t-1})$$

while for $z_t \neq 0$

$$\begin{aligned} \Pr(Z_t &= z_t|\mathcal{F}_{t-1}) = \Pr(A_t = 1|\mathcal{F}_{t-1}) \\ &\times \left\{ \begin{array}{c} \Pr(S_t = z_t|\mathcal{F}_{t-1}, A_t = 1, D_t = 1) \Pr(D_t = 1|\mathcal{F}_{t-1}, A_t = 1) + \\ \Pr(S_t = -z_t|\mathcal{F}_{t-1}, A_t = 1, D_t = -1) \Pr(D_t = -1|\mathcal{F}_{t-1}, A_t = 1) \end{array} \right\}. \end{aligned}$$

The implication of this decomposition is that there are exactly three pieces of modelling to carry out

- $\Pr(A_t|\mathcal{F}_{t-1})$ — a binary process on $\{0, 1\}$ modelling *activity* (the price moves or not).

- $\Pr(D_t|\mathcal{F}_{t-1}, A_t = 1)$ — another binary process on $\{-1, 1\}$ modelling the *direction* of the price moves.

- $\Pr(S_t|\mathcal{F}_{t-1}, A_t = 1, D_t)$ — a process on the strictly positive integers modelling the *size* of price moves.

Potentially each of these models has to be constructed separately — basing each on the complete history of the Z_t process. Although this sounds a difficult task we will see that our empirically based models will have very simple interpretable structures which does not immediately appear when we model the Z_t directly. It will be helpful to decompose the natural filtration $\mathcal{F}_t$ into its constituent parts — $\mathcal{F}_t^A = \sigma(A_s : s \leq t)$, $\mathcal{F}_t^D = \sigma(D_s : s \leq t)$ and $\mathcal{F}_t^S = \sigma(S_s : s \leq t)$. Of course $\mathcal{F}_t = \mathcal{F}_t^{A,D,S}$.

Finally, before we detail the modelling of activity, direction and size of the price movements we should note that although we can model these processes separately we are specifying a multivariate model. Hence in principle we cannot simulate a sequence of activities using just $\{\Pr(A_t|\mathcal{F}_{t-1})\}$ as we need all three models to simulate past values of Z_t. Thus we are not specifying a marginal model for the processes for activities, directions or sizes! An implication of this is that a structural break in any of the three processes $\{A_t|\mathcal{F}_{t-1}\}$, $\{D_t|\mathcal{F}_{t-1}, A_t = 1\}$ and $\{S_t|\mathcal{F}_{t-1}, A_t = 1, D_t\}$ will imply a structural break in the joint process.

3 Preliminary models for the components

3.1 The data

To start our empirical modelling we will work with the natural filtration of the price movements, building an initial empirical model for $\Pr(Z_t|\mathcal{F}_{t-1})$ via the construction of three models: those for activity, direction and size. The next section will extend this work to allow us to condition on a wider filtration. The trade data used in this paper is for the IBM stock recorded electronically at the New York Stock Exchange in 1995. We first construct a time series for each day on which the exchange was open, computing the price changes at each trade (rescaling the data to have a tick size of one). We then have deleted the first 15 minutes of every day. This is to avoid having to deal with the effects of the call auction which takes place in the morning to set off the trading. We also cut out all trades registered after 16.00 as this is the official closing of the exchange and our initial data analysis suggested the data was significantly different when it had a time stamp which was after 16.00.

For this paper we constructed a single series by concatenating each of the above series (whose overnight effects were removed by the action of removing the first 15 minutes of the day) for individual days. The length of the total data set when all exchanges in the US is considered 413,906, this is too much data to initially handle and therefore we have limited out analysis to the trades performed at NYSE (trades coded with an N). We have also deleted all trades which have an error code. This leaves us with a total of 173,146 observations to model. Of these 33,184 are non-zero (and so moved the price), which means the data we model for directions and size will only be 19% of the size of the activity series.

In our analysis we have not used quotes data. In Hausman, Lo, and MacKinlay (1992) the quotes where used to generate a trivariate variable describing if a trade where a bid, ask or indetermined. Our work could easily be extended in this direction, but we have decided to focus solely on the trades data.

3.2 The activity of prices

3.2.1 Autologistic model

Our initial parametric model for $\Pr(A_t|\mathcal{F}_{t-1})$ will be an autologistic model based on $\mathcal{F}_t = \mathcal{F}_t^{A,D,S}$. Recall for an autologistic we write

$$\Pr(A_t = 1|\mathcal{F}_{t-1}) = p(\theta_t^A), \qquad \text{where} \qquad p(\theta_t^A) = \frac{\exp(\theta_t^A)}{1 + \exp(\theta_t^A)}$$

and

$$\theta_t^A = x_t'\beta + g_t, \qquad \text{where} \qquad g_t = \sum_{j=1}^{p} \beta_j A_{t-j} \tag{2}$$

with x_t being potential combinations and subsets of $\mathcal{F}_{t-1}^{D,S,A}$. This model structure was introduced by Cox (1958) (see Cox and Snell (1989) for an exposition) and has some significant advantages[2]. The log–likelihood for the autologistic is concave and so numerical optimisation is completely straightforward, allowing standard logistic regression software to be used to rapidly and reliably fit the model (e.g. McCullagh and Nelder (1989)).

We use a general–to–specific model selection approach (see, for example, Hendry (1995)), estimating a complete model and then testing down insignificant lags. To start off we only allow 20 lags of all of the variables to enter the model. After the model is fitted

[2] A standard latent variable interpretation of these models writes

$$\Pr(Y_t = 1|X_t) = \Pr(\theta_t + U > 0) = p(\theta_t),$$

where U has a logistic distribution with parameters $(0, 1)$. Replacing the logistic distribution by a standard normal produces a probit models.

we will look at a portmanteau test to see its ability to capture the main features of the data. We use the normalised sequence for the fit of a binary process $\{A_t\}$ by constructing

$$u_t = \frac{A_t - p(\theta_t^A)}{\sqrt{p(\theta_t^A)\left\{1 - p(\theta_t^A)\right\}}},$$

for, if the model was correctly specified, the $\{u_t\}$ should be uncorrelated with zero mean and unit conditional (and unconditional) variance. The $\{u_t\}$ could then be used inside a Box-Pierce statistic as a measure of residual dependence.

The result of the model fitting and diagnostic checking are shown in Table 1. At lag two a variable $E_{t-2} = S_{t-2} - A_{t-2}$, which we will call excess price change, is significant. It indicates that if there is a large movement in the market followed by another trade, then there will be an increased probability of subsequent movements (of any size) in the price. That is large movements are associated with subsequent high volatility. The direction variables are negative, which suggests past falls in the prices tend to increase the chance that there will be a future movement in the market. This seems close to the famous leverage effect which is emphasised in the ARCH literature (e.g. Nelson (1991)).

Table 1 also indicates bid-ask bounce, for if the two lagged direction variables have opposite signs then the direction variable is damped down so reducing the chance of future price movements. However, this last effect will be clearer when we model the dynamics of the direction of price movements.

Finally, the Table shows the coefficients in front of the activity variables decay down — starting at 0.6 at lag one and falling to 0.2 at lag three. However, for longer lags the decay is quite slow and is not sufficiently well captured by our imposed artificial cutoff at lag 20. As a result the diagnostic checks on the residuals behave well at short lags, but poorly at longer lags as there is significant dependence at 1000s of lags in the activity variable. In order to model this parsimoniously we have to move away from autologistic models and into constructions which allow moving average type behaviour. This can be carried out by introducing GLARMA type models.

3.2.2 GLARMA binary model

We could generalise the autologistic structure (2) by allowing

$$g_t = \sum_{j=1}^{p} \gamma_j g_{t-j} + \sum_{j=1}^{q} \delta_j A_{t-j},$$

Variable	Coef.	StR. Err.	Variable	Coef.	StR. Err.
A_{t-1}	0.641	0.014	A_{t-9}	0.078	0.016
D_{t-1}	-0.105	0.013	A_{t-10}	0.049	0.016
A_{t-2}	0.244	0.015	A_{t-11}	0.066	0.016
E_{t-2}	0.289	0.092	A_{t-12}	0.069	0.016
D_{t-2}	-0.050	0.013	A_{t-13}	0.041	0.016
A_{t-3}	0.253	0.015	A_{t-14}	0.090	0.016
A_{t-4}	0.175	0.015	A_{t-15}	0.050	0.016
A_{t-5}	0.173	0.015	A_{t-16}	0.036	0.016
A_{t-6}	0.111	0.015	A_{t-17}	0.035	0.016
A_{t-7}	0.113	0.015	A_{t-18}	0.039	0.016
A_{t-8}	0.077	0.015	A_{t-19}	0.059	0.016
Const.	1.958	0.012	A_{t-20}	0.053	0.016
Q	$T\sum_{j=1}^{Q} r_j^2$			Log-like = -82313	
20	23.20	(31.41)			
100	698.4	(124.3)			
1500	5489	(1591)			

Table 1: *Estimation for the activity using an autologistic model. Variable is the explanatory variable. Std. Err. denotes the standard deviation, t–stat. denotes the t–statistic for the value being zero. E_{t-2} denotes the excess price changes. r_j denotes the series correlation coefficient at lag j for the the standardised residuals u_t. The figures in brackets are corresponding 95 percentage points on the χ^2_Q distribution to give a rough benchmark.*

but this is typically numerically unstable and so difficult to work with. Shephard (1994) has studied a number of alternatives, which are called generalized linear autoregressive moving average (GLARMA) models. The one we favour here puts

$$g_t = \sum_{j=1}^{p} \gamma_j g_{t-j} + \sigma v_t + \sigma \sum_{j=1}^{q} \delta_j v_{t-j}, \qquad \sigma > 0,$$

where

$$v_t = \frac{\left\{A_{t-1} - p(\theta^A_{t-1})\right\}}{\sqrt{p(\theta^A_{t-1})\left\{1 - p(\theta^A_{t-1})\right\}}}. \tag{3}$$

Importantly $\{v_t\}$ is a Martingale difference sequence with a unit conditional variance. This style of model is adopted in Russell and Engle (1998) in their multinomial construction.

GLARMA models have some of the properties of ARMA models. This can be seen by focusing in on $\{v_t\}$, which has zero mean and unit conditional and unconditional variance. The implication of this is that $\{g_t\}$ is a linear ARMA process driven by a weak white noise error term. Hence it is covariance stationary and invertible if this model obeys the usual stationarity and invertability constraints on the polynomials $1 - \sum_{j=1}^{p} \gamma_j L^j$

and $1 + \sum_{j=1}^{q} \delta_j L^j$. The implication is that the autocorrelation function of $\{g_t\}$ and the corresponding unconditional variance can be found using standard results on covariance stationary linear processes.

In our numerical work we have imposed that $\{g_t\}$ obeys the covariance stationarity and invertability constraints on the GLARMA. We impose these conditions by using the partial autocorrelations $\{\rho_j, j = 1, ..., p\}$ and the inverse partial autocorrelations $\{\bar{\rho}_j, j = 1, ..., q\}$ (see Barndorff-Nielsen and Schou (1973) and Jones (1987)). We numerically maximised the corresponding likelihood using analytic first derivatives and the BHHH algorithm[3], using the initial conditions that $g_0, ..., g_{-p+1}, v_0, ..., v_{-q+1}$ are all set to zero.

Using an AIC model selection rule we have chosen a GLARMA(3,1) model for the activity dataset. The estimated parameters and diagnostic statistics are given in Table 2. Notice that the likelihood for this model is much higher, and the diagnostics much better behaved, than for the previous models fitted for activity given in Table 1. However, the estimated coefficients for the lagged values of direction and (to a lesser extent) excess have not changed very much. Table 2 shows that both E_{t-1} and E_{t-2} are marginally significant. Both their corresponding parameter estimates and standard errors are not very stable across different GLARMA models. The direction variables are much more important in this context and these are estimated precisely and are not very sensitive to the parameterisation of the dynamics we use.

The estimated parameters suggest a great deal of memory in the activity series. Of course, we have imposed stationarity on this process and so it will be important to model the possibility that this series is non-stationary. Probably more realistically we need a more intricate model of activity which takes into account intra-day, intra-week and month effects on the series. Work on this topic is reported in the next section.

We do not investigate the impact of reporting rules. One such rule that might have an impact is the practice that it is the seller who reports on the transaction. On average this should not have an effect on the activity level. A procedure which does create longer spells of no price movements is that trades which occur closely in time might be pre–negotiated in order not to make the broker appear as if she is purchasing at one price and selling at

[3] A plain BHHH method was used mapping the real variables being maximised into partial autocorrelations and inverse partial autocorrelations using the transform $x/(1+|x|)$. No interventions or numerical problems were encountered.

Variable	Coef.	StR. Err.	
ρ_1	0.99988		
ρ_2	-0.649		
ρ_3	-0.289		
$\bar{\rho}_1$	-0.986		
D_{t-1}	-0.103	(.012)	
D_{t-2}	-0.0575	(.013)	
E_{t-1}	-0.198	(.085)	
E_{t-2}	0.231	(.095)	
Q	$T\sum_{j=1}^{Q} r_j^2$		Log-like = -81927
20	21.88	(31.41)	
100	92.9	(124.3)	
1500	1447	(1591)	

Table 2: *Estimation for the activity using a GLARMA model. Parameter estimates of the fitted GLARMA model, using a maximum likelihood criteria and the (3) error term. The figures in brackets are the standard errors on the regressors computed using the GLARMA model. Model order selected using AIC. r_j denotes the series correlation coefficient at lag j for the the standardised residuals u_t. The figures in brackets are corresponding 95 percentage points on the χ^2_Q distribution to give a rough benchmark.*

another.

3.3 The direction of price changes

An important feature of our decomposition is that we are now able to focus on a model of the directions of the price changes, given that the price has changed: $\Pr(D_t|\mathcal{F}_{t-1}, A_t = 1)$. This maybe helpful for high frequency traders who are interested in the movement of prices over very small periods of time.

Again we will use an autologistic model, but this time the outcome variable will live on the support $\{-1, 1\}$, rather than $\{0, 1\}$. To start out with we will let $\mathcal{F}_t = \mathcal{F}_t^{D,A,S}$. After testing out insignificant explanatory variables we end up with directions and excess–direction as the only information of significance, where excess–direction is given by

$$ED_t = D_t(S_t - A_t).$$

Furthermore, let

$$T_t = \sup_{s_1,\dots,s_k} \left\{ \begin{array}{rl} s_1 < t; & A_{s_1} = 1 \\ s_2 < s_1; & A_{s_2} = 1 \\ \vdots & \\ s_k < s_{k-1}; & A_{s_k} = 1 \end{array} \right\}.$$

The T_t vector is $k \times 1$ and contains the times at which the last k active prices occurred — trades which moved the price level. We call this concept of a time scale "activity time". We found that this concept of time is extremely significant statistically.

Then D_{T_t} will be a vector of the last k price changes different from 0. We refer to the $i-th$ element of this vector as $D_{T_t,i}$, which is the $i-th$ last price move that has been observed, standing at time t. A simple example of this is $D_{T_t,1}$ which is the sign of the last price movement different from 0.

This gives us an autologistic model for

$$\Pr(D_t = 1|\mathcal{F}_{t-1}, A_t = 1) = p(\theta_t^D), \qquad \text{where} \qquad p(\theta_t^D) = \frac{\exp(\theta_t^D)}{1+\exp(\theta_t^D)},$$

and so

$$\Pr(D_t = -1|\mathcal{F}_{t-1}, A_t = 1) = \frac{1}{1+\exp(\theta_t^D)}.$$

By going from general to specific we find that direction has only very short memory, see Table 3. From the estimated parameters we see that the process D_t is strongly mean–reverting (in activity time) reflecting the observed directions. An implication of this fitted model is that the dynamics generating the directions seems symmetrical, although there are more up directions than down ones — we will see down movements are typically bigger than up moves which compensates for this feature.

The lagged values of D_t and ED_t seem reasonably straightforward. If the price moved on the last trade then there is a large chance that this movement will be reverse if there is an active trade. If it moved by two ticks this probability of a reversal is reduced, although not by a great deal. On the other hand, if the last active trade was two periods ago the chance of a reversal is not very much different from evens.

The really interesting variables are the overwhelmingly significant, but quite modest in effect, variables $\{D_{T_t,i}\}$. These relate current directions to the last active trade — which appears at a random number of trades ago. That is they work in activity time not in trading time. They suggest that the sign of the last active trade has a persistent effect on the probability of an up or down. A simple example is that, at whatever lag in trading time, if the last price movement was down then there is a slightly higher probability of a reversal than a non-reversal.

An interesting effect, which is only marginally significant, is the $\sum_{i=11}^{20} D_{T_t,i}$ variable. If this variable is larger than zero then the series has tended to have a lot of up price movements and not many downs. So this is recording the presence of local trends in the price. It suggests this has a mildly positive effect on the direction process.

Our fitted model has some empirical failures. The diagnostic checks in Table 3 suggest we are slightly failing the check on serial correlation for this model. This failure should be put in some perspective. When we fit a model with just a constant (the directions are independently identically distributed Bernoulli — an implication of the model suggested by Rogers and Zane (1998)) the log-likelihood is -22966, and the Box-Pierce statistic at 20 lags is 9173. An alternative model is a simple autologistic in activity time — that is regressing just on $\{D_{T_t,i}, i = 1, 2, ..., 10\}$, then we have a log-likelihood of -20416 and a Box-Pierce of only 21. That model has reasonably good diagnostics but not an enormous amount of explanatory power. A simple alternative is to run a logistic regression using $\{D_{t-i}, i = 1, 2, ..., 20\}$, which is modelling directions using data ordered in transaction time, rather than activity time. This has a quite a high log-likelihood of -18419 but its Box-Pierce at 20 lags is 688. Hence this model has the opposite problem — being able to predict many of the directions but failing dramatically the diagnostics. It seems very hard to remove this model failure when we only use the concept of transaction time. The introduction of activity time seems essential for this type of process.

The economic meaning of this fitted model is that the directions are mostly generated by bid/ask bounce — see, for a review of empirical work on this topic, Campbell, Lo, and MacKinlay (1997, pp. 99-107). That is people buying shares from market makers have to pay higher prices for them than those selling them to the market makers (an elegant model of bid and ask dynamics is given by Hasbrouck (1996)). Sequences of no price movements are thought of as a series of consecutive buys (or sells) by the market makers. A price movement could reflect either a change in the efficient price or, more likely, a sell (or buy) by the market maker. As this buying and selling around the efficient price dominates in magnitude the actual large movements in the efficient price, it will automatically generate very strong negative autocorrelation in the directions sequences. That is changes in the traded price are almost certainly reversed.

A series of price changes in the same direction would, if the data were viewed at a lower frequency, be seen as large price movements. As we will see from the next section there are very few really big prices changes at this high frequency. One of the reasons is probably that market makers on the NYSE are encouraged to adjust the price smoothly.

Exp. Var.	estimate	Std. Err.	Exp. Var.	estimate	Std. Err.
D_{t-1}	-2.192	.043	ED_{t-1}	0.629	.180
D_{t-2}	-0.672	.033	ED_{t-2}	-0.506	.160
D_{t-4}	0.296	.030	ED_{t-3}	-0.837	.200
D_{t-5}	0.395	.033	ED_{t-5}	-0.625	.191
D_{t-6}	0.337	.034	$D_{T_t,1}$	-0.403	.038
D_{t-7}	0.249	.034	$D_{T_t,2}$	0.307	.036
D_{t-8}	0.233	.034	$D_{T_t,3}$	-0..069	.031
D_{t-9}	0.141	.034	$D_{T_t,5}$	-0.056	.027
D_{t-10}	0.073	.031	$D_{T_t,8}$	0.059	.027
D_{t-13}	-0.086	.030	$D_{T_t,10}$	-0.059	.027
D_{t-14}	-0.067	.029	$\sum_{j=11}^{20} D_{T_t,j}$	0.025	.010
$D_{(T_t,1)-1}$	0.315	.032	Const.	-0.053	.067
$D_{(T_t,3)-1}$	-0.087	.030			
$D_{(T_t,4)-1}$	-0.134	.033			
$D_{(T_t,5)-1}$	-0.105	.033			
$D_{(T_t,6)-1}$	-0.128	.033			
$D_{(T_t,7)-1}$	-0.154	.032			
$D_{(T_t,8)-1}$	-0.106	.030			
Q	$T\sum_{j=1}^{Q} r_j^2$			Log-like = -17947	
20	39.25	(31.41)			
100	101.4	(124.3)			
1500	1510	(1591)			

Table 3: *Estimation results for the direction of active trade using an autologistic model. Variable is the explanatory variable. The figures in brackets are the standard errors on the regressors computed using the autologistic model. r_j denotes the series correlation coefficient at lag j for the the standardised residuals u_t. The figures in brackets are corresponding 95 percentage points on the χ^2_Q distribution to give a rough benchmark.*

At this frequency this will appear as a sequence of price changes of one tick size, all in the same direction. Another reason could be that not all limit order traders adjust their orders continuously. As a result if the price rapidly changes then "stale" limit orders will smooth the observed prices.

3.4 The size of price movements

This section is devoted to constructing a model for $\Pr(S_t|\mathcal{F}_{t-1}, A_t = 1, D_t)$. As we have noted above this is a process on the strictly positive integers. Although the sample size is around 33,000, there are only 261 of these which are not one. Hence, we have to use quite simple models in this part of the paper as this dataset is not very informative about the dynamics of the size of price movements. We will use a negative binomial based GLARMA process for excess movements $S_t - 1$. Recall the negative binomial (NegBin) is a generalisation of the Poisson, allowing overdispersion (see, for example, Johnson, Kotz, and Kemp (1992, pp. 204-5)). The model will have

$$\Pr(S_t = s_t|\mathcal{F}_{t-1}, A_t = 1, D_t) = \frac{\Gamma(\alpha + s_t - 1)}{\Gamma(\alpha)(s_t - 1)!}\left(\frac{\alpha}{\mu_t + \alpha}\right)^{\alpha}\left(\frac{\mu_t}{\mu_t + \alpha}\right)^{s_t - 1},$$

implying

$$\begin{aligned} E(S_t|\mathcal{F}_{t-1}) &= 1 + \mu_t, \\ Var(S_t|\mathcal{F}_{t-1}) &= \alpha\left(\frac{\mu_t}{\mu_t + \alpha}\right)/\left(\frac{\alpha}{\mu_t + \alpha}\right)^2 = \mu_t + \frac{1}{\alpha}\mu_t^2. \end{aligned}$$

Notice that μ_t will be typically very small and so μ_t^2 is mostly tiny. As α, the overdispersion parameter, goes to infinity so NegBin approaches a conditional Poisson model. Here we allow $\mu_t = \exp(\theta_t^S)$ where

$$\theta_t^S = x_t'\beta + g_t, \qquad \text{and} \qquad g_t = \sum_{j=1}^{p}\gamma_j g_{t-j} + \sigma v_t + \sigma\sum_{j=1}^{q}\delta_j v_{t-j},$$

where x_t will include D_t and elements of $\mathcal{F}_{t-1}$. We define

$$v_t = \frac{\{(S_{t-1} - 1) - \mu_{t-1}\}}{\sqrt{\mu_{t-1} + \frac{1}{\alpha}\mu_t^2}}.$$

as our basic parametric model[4]. The NegBin distribution was selected as it is simple and familiar. The NegBin GLARMA structure has the advantage that g_t is covariance stationary under the usual conditions on the autoregressive and moving average polynomials — following the same line of argument as in the binary GLARMA case.

The model was fitted using a maximum likelihood estimator, selecting p, q using AIC. The resulting estimated model is detailed in Table 4. The most interesting feature is that the direction of the current price change and the preceding are significant and all have negative coefficients. This rejects, quite significantly, two hypotheses of symmetry. Firstly big price movements tend to be preceded by non–symmetric falls in the price (directions being negative). This is the familiar dynamic leverage effect (see Nelson (1991)) and so its presence is not surprising.

[4]In foreign exchange markets the tick size tends to be smaller compared to the bid–ask spread. However, many writers have observed that there is a tendency for prices to cluster on "natural numbers" such as integers. Our model would have to be altered to allow for such a characteristic, but from a methodological viewpoint this raises no new issues.

The second form of non-symmetry is a contemporaneous one and is simply dependent on the significance of D_t (not its lags). The negative sign associated with it suggest big falls are more common than big rises. This implies the unconditional distribution of returns should be skewed with a longer left hand tail (if D_t had a zero coefficient then the implied unconditional distribution of returns would be symmetric). This is counterbalanced by the fact that the average value of $S_t D_t$ is positive, which suggests the market trends upwards over time due to the predominance of small positive movements (over small negative movements), but tends to fall back sometimes with quite large falls.

The above non-symmetry of the unconditional distribution has not been found in previous work on this type of data. This is perhaps not surprising as the t-statistic on it is only around 8 and so will not be found unless it is very directly tested. It is, however, important. Our results suggest excess movements (movements of more than one tick) downwards are, on average, around twice as large as excess movements upwards.

Variable	Coef.	StR. Err.	Variable	Coef.	StR. Err.
Const	-5.140	(.138)	σ	0.180	
ρ_1	0.998		ρ_2	-0.343	
$\bar{\rho}_1$	-0.816		D_t	-0.320	(.070)
D_{t-1}	-0.394	(.095)	D_{t-3}	-0.299	(.112)
$D_{(T_t,5)}$	-0.121	(.067)	$D_{(T_t,6)}$	-0.206	(.064)
α	0.0713				
Q	$T\sum_{j=1}^{Q} r_j^2$			Log-like = -1478	
20	31.88	(31.41)	$E(v_t) = -.001$		
100	85.39	(124.3)	$Var(v_t) = 1.207$		
1500	1743	(1591)			

Table 4: *Estimation for the NegBin based GLARMA(3,1) model of the excess price movements. Variable is the explanatory variable. Std. Err. denotes the standard deviation, t–stat. denotes the t–statistic for the value being zero. The figures in brackets are the standard errors on the regressors computed using the GLARMA model. Model order selected using AIC.* r_j *denotes the series correlation coefficient at lag* j *for the the standardised residuals* u_t*. The figures in brackets are corresponding 95 percentage points on the* χ^2_Q *distribution to give a rough benchmark.*

Judging from Table 4 the NegBin based GLARMA model is failing slightly as the variance of the residuals is slightly bigger than one. This could be due to misspecification in the construction of $\{\mu_t\}$, or a mild distributional failure in our choice of the NegBin model. In the next section we will condition on a wider information set and so may hope to remove this problem by using a more subtle version of $\{\mu_t\}$.

To put the performance of this model structure in context we can compare its fit to a constant which gives a log-likelihood of -1628 and a Box-Pierce statistic (using 20 lags) of 1762. When we take out the GLARMA structure completely, leaving just explanatory variables, the log-likelihood is -1602 and the Box-Pierce is 2065. The other extreme is where we drop all the explanatory variables and leave simply the GLARMA(2,1) model. This has a log-likelihood of -1502 and Box-Pierce (using 20 lags) of 16.6. Thus we can see that it is the time series modelling aspect of this particular model which dominates the fitting of this series.

The NegBin model was not our first choice. We did try out more familiar models such as the Poisson which did not have thick enough tails. We also tried the logarithmic distribution as a possible model. This is a one parameter distribution and therefore not quite as flexible as the NegBin.

The above observation seems to rule out the use of trade–by–trade data to estimate a meaningful extremal tail index of returns, unless quite complicated adjustments are made, at least at this frequency. At aggregated levels not only the sizes matters, but also the directions as sequences of small moves in the same direction sum to a large price change.

4 Comments

4.1 Predictive distributions

4.1.1 Multi-step prediction

A crucial use of our model structure is to produce multi–step ahead predictions of asset price movements. This can be expressed in two basic ways: (i) predictions of the $(s+1)$–periods ahead price movements, (ii) predictions of the asset price levels $(s+1)$–periods ahead. We first of all deal with the former.

4.1.2 Predicting price movements

The object of interest is $\Pr(Z_{t+s}|\mathcal{F}_{t-1})$, which is simply

$$\begin{aligned}\Pr(Z_{t+s}|\mathcal{F}_{t-1}) &= \sum_{Z_t} \cdots \sum_{Z_{t+s-1}} \Pr(Z_t, ..., Z_{t+s}|\mathcal{F}_{t-1}) \\ &= \sum_{Z_t} \cdots \sum_{Z_{t+s-1}} \prod_{j=0}^{s} \Pr(Z_{t+j}|\mathcal{F}_{t-1+j}).\end{aligned}$$

In our model Z_t lives on the integers which makes complete enumeration of these quantities impossible. We can respond to this in two ways — by using simulation or by truncating the state space of Z_t. For small values of s the latter probably is the most effective, while for long horizons simulations would seem perfectly satisfactory for most purposes.

For s very large the multi-step ahead forecast distribution will approach the unconditional distribution of our fitted model. Although this is of little economic meaning it can be a useful diagnostic check on the fitted model.

4.1.3 Predicting price levels

Computing analytically predicted price levels can be carried out using similar arguments to those given above for any value of s.

The calculations become intricate when s is large as there are many groups of price changes which achieve the same terminal price. Hence, in practical work the best way of proceeding is by the use of simulation. Hence given $\mathcal{F}_{t-1}$ we simulate the process N times and count the number of simulated prices which fall on particular lattice points. As our model is extremely easy to simulate from this can be carried out for very large values of N (in the simulations discussed below $N = 10,000$) even if s is large. We can also obtain standard errors as

$$\sqrt{\frac{p(1-p)}{N}}$$

Figures 2, 3 and 4 show the centre of the two–step ahead forecast distribution based on different histories – mapping out the probabilistic evolution of the forecast distribution. In Figure 2 two down movements have just been observed and this is seen to give an increased probability for moving one tick up. Figure 3 corresponds to the opposite observation, namely two up movements, here we have the opposite result that the probability of moving down is increased. The last of the 3 trees, Figure 4, corresponds to "bid–ask bouncing". This has decreased the probability of moving away from the current price level compared to Figure 2. From all 3 figures it is seen that the predominant behaviour is mean reversion of one tick size and that the last directions are important in determining how likely a price reversal is. This implies that, with the given history, when we have seen a movement of two one ticks down(up) after two trades the price will still be a least one down(up) with probability 0.990(0.991) and at least two down(up) with probability 0.707(0.756), (see Table 5). In Table 5 the 10–step ahead forecast is also given. In this case we get that when we have seen a movement of two one ticks down(up) after ten trades the price will still be a least two down(up) with probability 0.779(0.809).

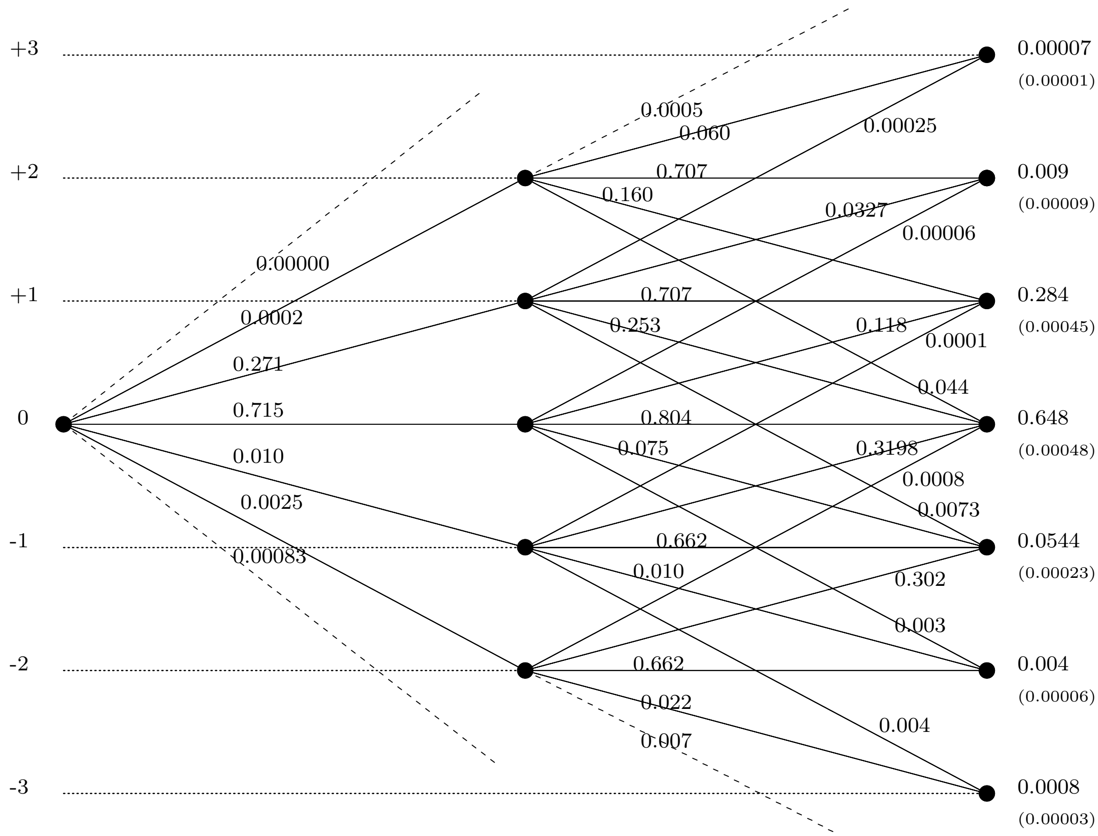

Figure 2: *History* $\mathcal{F}^A = (1, 1, 0, ..., 0)$, $\mathcal{F}^M = (-1, -1, -1, 1, -1)$. *The numbers to the right are the simulated probabilities of ending up at this level after two trades with the given history. The numbers in brackets are the standard errors of the simulated values.*

4.2 Previous works

4.2.1 Conditional multinomial models

In a recent highly stimulating paper Russell and Engle (1998) have suggested modelling price movements using a conditional multinomial distribution. Their paper can be viewed as a time series extension of a multivariate probit (Russell and Engle (1998) prefer to work with logistic functions rather than probit ones) analysis of transaction data proposed by Hausman, Lo, and MacKinlay (1992). Here we will discuss their work and its relationship to our own. We will initially abstract our discussion from the time series feature of the model and so we will write Y_t to denote the indicator for the movements which we will assume live only on $-2, -1, 0, 1, 2$. So if the movement is 1, then $Y_t = (0, 0, 0, 1, 0)'$, while

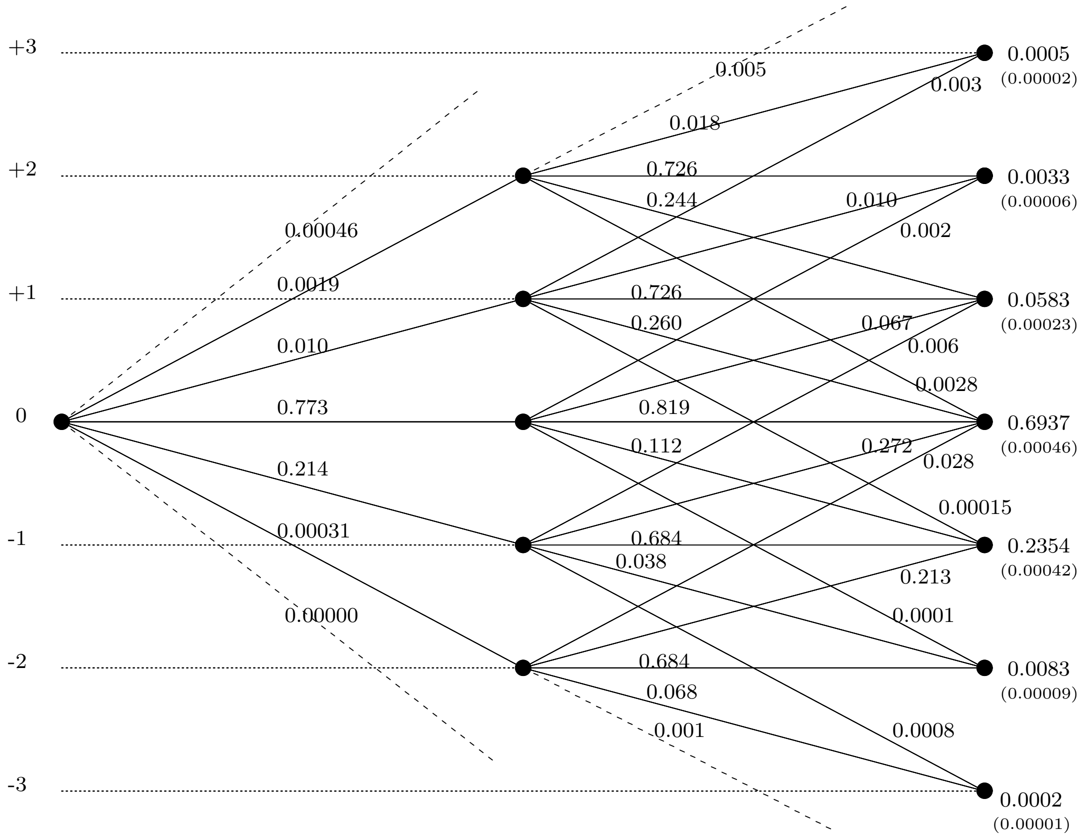

Figure 3: *History* $\mathcal{F}^A = (1, 1, 0, ..., 0)$, $\mathcal{F}^M = (1, 1, -1, 1, -1)$. *The numbers to the right are the simulated probabilities of ending up at this level after two trades with the given history. The numbers in brackets are the standard errors of the simulated values.*

if it were -1 then $Y_t = (0, 1, 0, 0, 0)'$. We suppose we use some regressors X_t to model the changing probabilities of these movements. In practice X_t will depend upon some features of the filtration of Y_t, $\mathcal{F}_t = \sigma(Y_s, s \leq t)$.

At this level, there is only one loss of generality (and information) compared to our decomposition — price movements have to live on a small finite grid (mainly due to parsimony). Next Russell and Engle (1998) use a multinomial logit structure (see e.g. McFadden (1984, Section 3.4)).

$$\Pr(Y_t = i | X_t) = p_i(\theta_t), \qquad i = -2, -1, 0, 1, 2,$$

where $\theta_t = (\theta_{-2t}, \theta_{-1t}, \theta_{0t}, \theta_{1t}, \theta_{2t})' = X_t\beta$ and

$$p_i(\theta_t) = \frac{\exp(\theta_{i,t})}{1 + \sum_{j=-2}^{2} \exp(\theta_{j,t})}, \qquad i = -2, -1, 0, 1, 2.$$

In practice this structure is not identified and so constraints are placed on $X_t\beta$. A typical

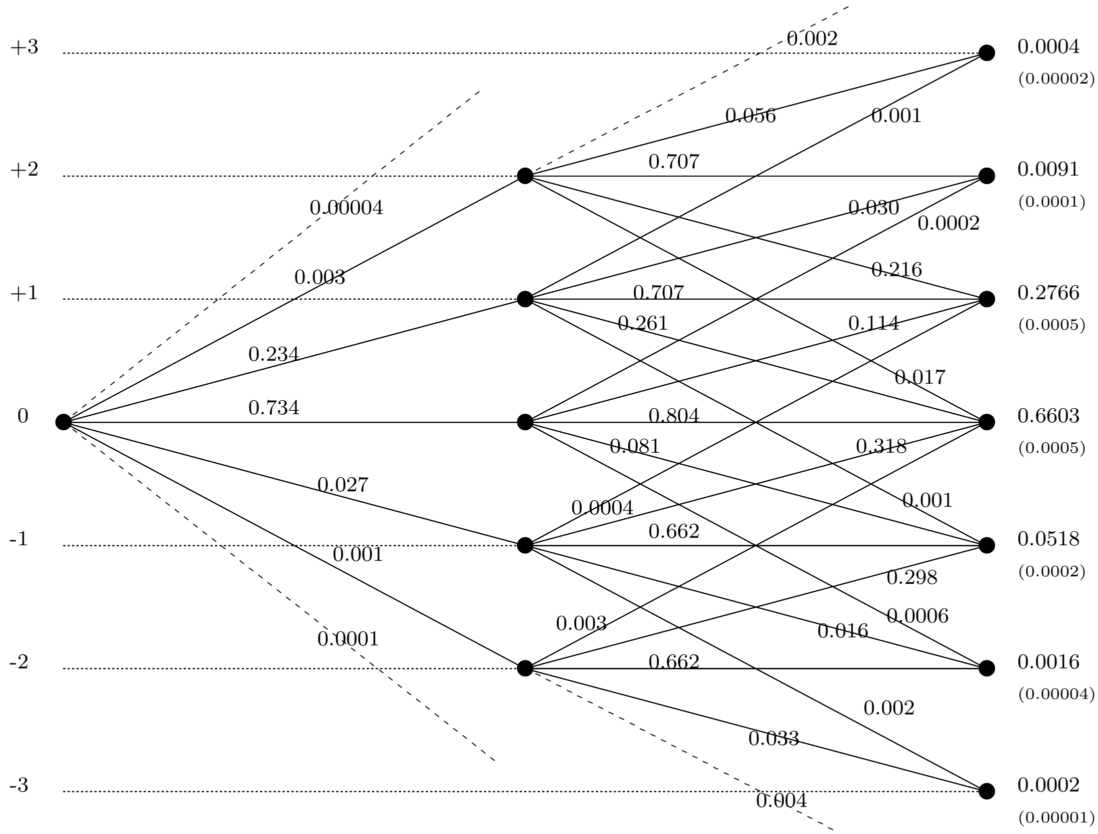

Figure 4: *History* $\mathcal{F}^A = (1, 1, 0, ..., 0)$, $\mathcal{F}^M = (-1, 1, -1, 1, -1)$. *The numbers to the right are the simulated probabilities of ending up at this level after two trades with the given history. The numbers in brackets are the standard errors of the simulated values.*

situation would be to define $\theta_{0t} = 0$ for all t, a solution followed by Russell and Engle (1998).

The important step in Russell and Engle (1998) is to define a vector generalised linear autoregressive moving average (VGLARMA) type structure on $\theta_t^* = (\theta_{-2t}, \theta_{-1t}, \theta_{1t}, \theta_{2t})'$, feeding in lagged values of $\{y_t\}$ using the variable given by (3). In particular if they define $v_t = (v_{1t}, v_{2t}, v_{3t}, v_{4t})'$ with

$$v_{it} = \frac{I(Y_t = i) - p_{it}}{\sqrt{p_{it}(1 - p_{it})}},$$

and

$$\theta_t^* = \alpha^* + g_t,$$

then model this system as

$$g_t = \sum_{j=1}^{p} \gamma_j g_{t-j} + \sigma v_t + \sigma \sum_{j=1}^{q} \delta_j v_{t-j},$$

tick moves	no. of trades	$\{-1,-1\}$	$\{-1,1\}$	$\{1,-1\}$	$\{1,1\}$
-3	2	0.00077	0.00017	0.00094	0.00020
	10	0.0027	0.0019	0.0025	0.0029
-2	2	0.00409	0.00164	0.00907	0.00834
	10	0.0108	0.0097	0.0104	0.0106
-1	2	0.05440	0.05176	0.24669	0.23538
	10	0.1512	0.1268	0.1982	0.1698
0	2	0.64761	0.66028	0.68549	0.69373
	10	0.6141	0.6115	0.6211	0.6104
1	2	0.28354	0.27661	0.05655	0.05833
	10	0.2006	0.2262	0.1505	0.1850
2	2	0.00894	0.00913	0.00103	0.00330
	10	0.0120	0.0140	0.0095	0.0117
3	2	0.00007	0.00036	0.00006	0.00050
	10	0.0023	0.0033	0.0024	0.0022

Table 5: *The Table show the simulated probabilities for having moved x ticks after the given history. $\mathcal{F}^A = (1, 1, 0, ..., 0)$ and $\mathcal{F}^M = (?, ?, -1, 1, -1)$. $?, ?$ is given in the top row of the Table. The estimated probabilities are based on $N = 10,000$ simulations.*

where α^* is a vector, while $\{\gamma_j\}$, σ and $\{\delta_j\}$ are 4×4 matrices. The only a priori constraint we might place on this structure is that σ should be lower triangular for identification.

Overall we can see that our analysis is quite closely related to that of Hausman, Lo, and MacKinlay (1992) and Russell and Engle (1998). Our goals are the same, although the technology that we use is very different. Our main advantages are: parsimony, easy interpretability via the decomposition and options for extensions.

4.2.2 Hasbrouck's truncation model

Hasbrouck (1996) introduced a dynamic model for the evolution of the bid and ask price of quotation data. Let μ_t denote the theoretical efficient price in the market and let α_t, σ_t represent the ask and bid costs respectively. Then Hasbrouck argued for a structure where the bid price is $Floor(\mu_t - \alpha_t)$ and the ask price is $Ceiling(\mu_t - \sigma_t)$. Here the $Floor$ function rounds down to the nearest tick and $Ceiling$ rounds up. Related papers include Bollerslev and Melvin (1994) and Harris (1994). Manrique and Shephard (1997) have studied the implied econometrics of this type of model.

The Hasbrouck (1996) bid/ask model is not immediately applicable to transaction data, but the principle of using a continuous time model which is then truncated in some way is potentially useful if combined (perhaps) with the Hausman, Lo, and MacKinlay (1992) static model of clustering.

4.3 Conditioning on signs

In some recent work, carried out independently from our own, Granger (1998) has emphasised the potential importance of modelling separately the direction (sign) and the size of stochastic processes. Typically he models these two variables independently, while we emphasise the sequential nature of our decomposition — which is empirically vital for our problem and more general. Abstracting from that detail, we can see that our analysis can be seen within his framework when we condition on the activity, A_t, being one.

4.4 Explanatory variables

4.4.1 Deterministic seasonality

It maybe that the activity, directions and size series are influenced by deterministic seasonal patterns for we know these patterns influence $N(u)$ the rate at which transactions occur in calendar time — see, for example, Rydberg and Shephard (1998). It is a straightforward task to include this information within our model, allowing seasonality to influence any or all of the sub-models for activity, direction and size. No new issues are raised by this and we will give empirical results on these problems in a later subsection.

4.4.2 Exogenous variables

Our modelling framework allows some very simple extensions which will be potentially enriching. Suppose in addition to the price movements $\{Z_t\}$ we have a sequence of other information sets such as volume and place of trade. Let us write these additional variables as $\{Y_t\}$, then we can do a prediction decomposition, using the extended filtration $\mathcal{F}_t^{z,y} = \sigma(Z_s, Y_s : s \leq t)$, to give

$$\begin{aligned} f(Z_1, Y_1, ..., Z_n, Y_n | \mathcal{F}_0^{z,y}) &= \prod_{t=1}^{n} f(Z_t, Y_t | \mathcal{F}_{t-1}^{z,y}) \\ &= \prod_{t=1}^{n} f(Y_t | \mathcal{F}_{t-1}^{z,y}) \Pr(Z_t | Y_t, \mathcal{F}_{t-1}^{z,y}) \\ &= \prod_{t=1}^{n} \Pr(Z_t | \mathcal{F}_{t-1}^{z,y}) f(Y_t | Z_t, \mathcal{F}_{t-1}^{z,y}) \end{aligned}$$

where the second stage of decomposition can be useful if we can find a sensible model for $f(Y_t|Z_t, \mathcal{F}_{t-1}^{z,y})$ and we can allow Y_t to enrich the decomposition of the price innovation process. The third stage decomposition can also be the focus on attention as it allows lagged information to improve our predictions of future price movements given the history of the Y_t process.

The above decomposition suggests that there are potentially two interesting densities for $\{Z_t\}$ to investigate further: $\Pr(Z_t|\mathcal{F}_{t-1}^{z,y})$ and $\Pr(Z_t|Y_t, \mathcal{F}_{t-1}^{z,r})$. The first is a pure forecast while the second allows contemporaneous explanatory variables to enter (see Engle, Hendry, and Richard (1983) and Hendry (1995, Ch. 5)). Tables 6 and 7 give results for the activity and direction series when we condition on lagged variables. The variables we use in this exercise are the logarithm of the volume traded and the logarithm of the number of seconds (plus one) elapsing before the trade. In addition we use dummy variables to denote the hour, day of the week and month of the year in which the trade takes place. Finally, we sometimes use two trending variables: "range" (a standardised version of the time index t) and "quadr" which is simply the square of "range". These trends are used as a parsimonious representation of the monthly seasonal pattern. Further, we have tried using the log of the actual price level of the IBM stock price, but this always tested out in our empirical work.

The empirical model for directions, reported in Table 7, is interesting as it completely tests out the effect of durations on the prediction of directions, while the influence of lagged volume is very small and almost all seasonal effects are irrelevant.

This is not the case when we look at the activity series, which is sensitive to many lags of durations. This is perhaps not surprising as the activity series is connected to volatility and so one would expected them to be influenced by other activity type series. The range variable is a trend variable, which we interpret as a monthly seasonal type variable rather than a typical trend as we only have a year of data. We used this range variable rather than a full set of monthly seasonal for reasons of parsimony.

An interesting feature for both the activity and direction series is that lagged volume and duration variables are sometimes statistically significant but not overwhelmingly so. Instead lagged data on previous price movements completely dominate the fit of these two models.

In Tables 8 and 9 report the estimated activity and direction processes using contem-

Variable	estimate	Std. Err.	Variable	estimate	Std. Err.
Const.	-1.509	(.045)	Range	0.126	(.031)
ρ_1	0.9998		$\log(\text{dur})_{t-1}$	0.028	(.005)
ρ_2	-0.668		$\log(\text{dur})_{t-2}$	-0.018	(.005)
ρ_3	-0.252		$\log(\text{dur})_{t-3}$	-0.020	(.005)
$\bar{\rho}$	-0.987		$\log(\text{dur})_{t-4}$	-0.020	(.005)
E_{t-1}	-0.219	(.085)	$\log(\text{dur})_{t-5}$	-0.017	(.005)
E_{t-2}	0.235	(.095)	$\log(\text{dur})_{t-6}$	-0.016	(.005)
D_{t-1}	-0.101	(.012)	$\log(\text{dur})_{t-8}$	-0.009	(.005)
D_{t-2}	-0.058	(.013)	$\sum_{j=11}^{20} \log(\text{dur})_{t-j}$	-0.004	(.002)
			$\sum_{j=21}^{30} \log(\text{dur})_{t-j}$	-0.005	(.002)
Q	$T \sum_{j=1}^{Q} r_j^2$			Log-like = -81839	
20	19.06	(31.41)			
100	87.2	(124.3)			
1500	1437	(1591)			

Table 6: *Estimation for the activity including lagged* log(duration+1) *and the* log(volume) *(which tests out). Improvement in likelihood 88 compared to Table 2. Variable is the explanatory variable. Std. Err. denotes the standard deviation, t–stat. denotes the t–statistic for the value being zero.*

poraneous volumes and durations. For the activity series current durations have a very dramatic positive impact on activity. A smaller impact is made by volume. In addition hourly seasonal effects are now significant.

The quantitative effect of this quite large. Activity is effected positive by both volume and duration (see Table 8). In particular if the duration is high then this increases the chance that the price will move at the next trade, while if volume if high the same thing happens.

When we look at direction (see Table 9) we see that again both variables have significant impact. High durations reduce the chance that the price movement will be upwards, while high volume increases the chance of an up movement.

The estimated NegBin based GLARMA model for the excess variable $S_t - 1$ is reported in Table 10 using lagged durations and volume. The AIC measure selected a GLARMA(2,1) structure. Interesting the effect of lagged durations is modest, with its influence being played out over just two lags. Both of the estimated coefficients have t-statistics of only around 3. The lagged volume variables have a bigger impact, with longer lags than that given in the tables having positive but insignificant effects on the

Exp. Var.	estimate	Std. Err.	Exp. Var.	estimate	Std. Err.
D_{t-1}	-2.192	.043	ED_{t-1}	0.620	.180
D_{t-2}	-0.671	.033	ED_{t-2}	-0.506	.160
D_{t-4}	0.298	.030	ED_{t-3}	-0.851	.200
D_{t-5}	0.395	.033	ED_{t-5}	-0.626	.191
D_{t-6}	0.337	.034	$D_{T_t,1}$	-0.400	.038
D_{t-7}	0.248	.034	$D_{T_t,2}$	0.301	.036
D_{t-8}	0.232	.034	$D_{T_t,3}$	-0.071	.031
D_{t-9}	0.139	.034	$D_{T_t,5}$	-0.062	.027
D_{t-10}	0.072	.031	$\log(\text{vol})_{t-1}$	-0.030	.009
D_{t-13}	-0.083	.030	$\log(\text{vol})_{t-2}$	0.021	.008
D_{t-14}	-0.067	.029	$\sum_{j=11}^{20} D_{T_t,j}$	0.012	.005
$D_{(T_t,1)-1}$	0.312	.032	April	0.140	.050
$D_{(T_t,3)-1}$	-0.086	.030	Const.	-0.070	.039
$D_{(T_t,4)-1}$	-0.133	.033			
$D_{(T_t,5)-1}$	-0.103	.033			
$D_{(T_t,6)-1}$	-0.127	.033			
$D_{(T_t,7)-1}$	-0.152	.032			
$D_{(T_t,8)-1}$	-0.106	.030			
Q	$T\sum_{j=1}^{Q} r_j^2$			Log-like = -17938	
20	47.78	(31.41)			
100	110.4	(124.3)			
1500	1513	(1591)			

Table 7: *Estimation for the direction including lagged durations (which test out) and volume. Improvement in the log-likelihood for introducing these variables is 9. Variable is the explanatory variable. The figures in brackets are the standard errors on the regressors computed using the autologistic model.* r_j *denotes the series correlation coefficient at lag* j *for the the standardised residuals* u_t. *The figures in brackets are corresponding 95 percentage points on the* χ^2_Q *distribution to give a rough benchmark.*

excess variable. Importantly the range and quadr variable is taken as estimating the seasonal component of the process. This is significant, which is unsurprising given the size variable is, like the activity variable, a kind of volatility measure. Also the time series dependence in the GLARMA model has been reduced quite considerably by the presence of the explanatory variables.

Overall lagged explanatory variables improve the likelihood function by around 24, which is modest given we have include seven new explanatory variables in the fitted model. The diagnostic checks on the fitted model have improved, especially at short lags, but the model still suffers from slight over-dispersion suggesting some improvement could

Variable	estimate	Std. Err.	Variable	estimate	Std. Err.
Const.	-1.460	0.045	Range	0.157	0.028
ρ_1	0.9998		E_{t-1}	-0.215	0.086
ρ_2	-0.624		E_{t-2}	0.368	0.097
ρ_3	-0.245		D_{t-1}	-0.107	0.012
$\bar{\rho}$	-0.986		D_{t-2}	-0.048	0.013
10-11	-0.082	0.039	$\log(\text{vol})_t$	0.064	0.004
11-12	-0.154	0.043	$\log(\text{dur})_t$	0.368	0.005
12-13	-0.145	0.046	$\log(\text{dur})_{t-1}$	0.070	0.005
13-14	-0.173	0.047	$\log(\text{dur})_{t-2}$	-0.013	0.005
14-15	-0.128	0.044	$\log(\text{dur})_{t-3}$	-0.020	0.005
$\sum_{j=11}^{20} \log(\text{dur})_{t-j}$	-0.009	0.002	$\log(\text{dur})_{t-4}$	-0.024	0.005
$\sum_{j=21}^{30} \log(\text{dur})_{t-j}$	-0.009	0.002	$\log(\text{dur})_{t-5}$	-0.022	0.005
			$\log(\text{dur})_{t-6}$	-0.023	0.005
			$\log(\text{dur})_{t-8}$	-0.015	0.005
Q	$T\sum_{j=1}^{Q} r_j^2$			Log-like = -79022	
20	39.07	(31.41)			
100	115.0	(124.3)			
1500	1490	(1591)			

Table 8: *Estimation for the activity including contemporaneous and lagged* log(duration+1) *and the* log(volume)*. Improvement in likelihood 3895 compared to Table 1. Variable is the explanatory variable. Std. Err. denotes the standard deviation, t–stat. denotes the t–statistic for the value being zero.*

be gained from fitting a more complicated model than the NegBin structure we have used.

Table 11 shows the fitted model for the size variable using contemporaneous volumes and durations, in addition to lagged data and seasonal effects. The table shows that contemporaneous explanatory variables have a very significant effects on the excess variables $S_t - 1$. The volume variable has a very large positive impact on the chance that an active variable moves the price by more than one tick. The t-statistic on current volume is around 20, which is by far the largest of any of the significant variables we have found for the excess variable. Interesting the presence of current volume reduces the impact of lagged volume and removes the need to have daily seasonals and the quadratic trend (monthly seasonal). All that remains of these deterministic seasonals is the range variable, which we should interprete as saying big moves occur towards the end of the year. However, this effect is not very significant.

The contemporaneous duration variable also has a positive impact while at one lag the effect is reversed. We do not understand this effect.

Exp. Var.	estimate	Std. Err.	Exp. Var.	estimate	Std. Err.
D_{t-1}	-2.173	.043	ED_{t-1}	0.638	.181
D_{t-2}	-0.663	.033	ED_{t-2}	-0.490	.161
D_{t-4}	0.298	.030	ED_{t-3}	-0.830	.199
D_{t-5}	0.395	.033	ED_{t-5}	-0.607	.192
D_{t-6}	0.334	.034	$D_{T_t,1}$	-0.412	.038
D_{t-7}	0.241	.034	$D_{T_t,2}$	0.301	.036
D_{t-8}	0.225	.034	$D_{T_t,3}$	-0.073	.031
D_{t-9}	0.135	..034	$D_{T_t,5}$	-0.062	.027
D_{t-10}	0.071	.031	$\log(\text{vol})_t$	0.087	.008
D_{t-13}	-0.092	.030	$\log(\text{vol})_{t-1}$	-0.039	.009
D_{t-14}	-0.069	.029	$\log(\text{dur})_t$	-0.110	.010
$D_{(T_t,1)-1}$	0.316	.032	$\sum_{j=11}^{20} D_{T_t,j}$	0.012	.005
$D_{(T_t,3)-1}$	-0.087	.030	April	0.150	.050
$D_{(T_t,4)-1}$	-0.135	.033	Const.	0.082	.039
$D_{(T_t,5)-1}$	-0.103	.033			
$D_{(T_t,6)-1}$	-0.126	.033			
$D_{(T_t,7)-1}$	-0.155	.032			
$D_{(T_t,8)-1}$	-0.106	.030			
Q	$T\sum_{j=1}^{Q} r_j^2$			Log-like = -17837	
20	47.76	(31.41)			
100	112.3	(124.3)			
1500	1540	(1591)			

Table 9: *Estimation for the direction including lagged and contemporaneous durations and volume. Improvement in the log-likelihood for introducing the contemporaneous variables is 101. Variable is the explanatory variable. The figures in brackets are the standard errors on the regressors computed using the autologistic model.* r_j *denotes the series correlation coefficient at lag* j *for the the standardised residuals* u_t*. The figures in brackets are corresponding 95 percentage points on the* χ^2_Q *distribution to give a rough benchmark.*

The presence of these new explanatory variables cleans up the serial dependence structure in the data, for now the Box-Pierce statistics are satisfactory. Further the amount of over-dispersion in the fitted model is modest.

4.5 Roll's model of bid–ask bounce

In an insightful paper, Roll (1984) proposed a simple measure for the effective bid–ask spread. The measure given for the spread was simply

$$2 \times \sqrt{-\text{Cov}(Z_t, Z_{t-1})}$$

Variable	Coef.	StR. Err.	Variable	Coef.	StR. Err.
Const	-5.546	(.152)	σ	0.175	
Tues	0.683	(.184)	Wednes	0.489	(.222)
Range	0.391	(.124)	Quadr	-0.342	(.114)
ρ_1	0.996		ρ_2	-0.264	
$\bar{\rho}_1$	-0.805		α	0.077	
D_t	-0.347	(.073)	D_{t-1}	-0.488	(.100)
D_{t-3}	-0.328	(.116)			
$D_{(T_t,6)}$	-0.183	(.068)			
$\log(\text{dur})_{t-1}$	-0.163	(.068)			
$\log(\text{vol})_{t-1}$	0.144	(.052)	$\log(\text{vol})_{t-2}$	0.110	(.045)
Q	$T\sum_{j=1}^{Q} r_j^2$			Log-like = -1454	
20	27.05	(31.41)	$E(v_t) = -.002$		
100	119.2	(124.3)	$Var(v_t) = 1.166$		
1500	1393	(1591)			

Table 10: *Estimation for the NegBin based GLARMA(2,1) model of the excess price movements ($S_t - 1$) including lagged* log(duration+1) *and the* log(volume). *Variable is the explanatory variable. Std. Err. denotes the standard deviation, t–stat. denotes the t–statistic for the value being zero. The figures in brackets are the standard errors on the regressors computed using the GLARMA model. Model order selected using AIC. r_j denotes the series correlation coefficient at lag j for the the standardised residuals u_t. The figures in brackets are corresponding 95 percentage points on the χ^2_Q distribution to give a rough benchmark.*

where Cov denotes the unconditional covariance. This was based on the observation that market efficiency should guarantee that the covariance between efficient prices should be 0 and that the actual observed covariance is due only to the bid–ask spread. The model proposed in Roll's paper is probably too simplistic to tell the whole story about serial correlation but what is important to note is that a large amount of the first–order serial covariance is due to bid–ask effects something which is easily captured by our model for the directions D_t.

Roll's measure of spread is an unconditional one, but our analysis suggests a generalization. We argue that one could infer spread from

$$2\sqrt{-\text{Cov}(Z_t, Z_{t-1}|\mathcal{F}_{t-2})}$$

a conditional correlation. As $\mathcal{F}_{t-2}$ varies, so does the implied spread. In particular it may widen if we have observed a sequence of active price movements.

Variable	Coef.	StR. Err.	Variable	Coef.	StR. Err.
Const	-5.710	(.120)	σ	0.149	
Range	0.315	(.118)	$\bar{\rho}_1$	-0.830	
ρ_1	0.996		ρ_2	-0.411	
D_t	-0.291	(.070)	D_{t-1}	-0.296	(.095)
$\sum_{j=1}^{6} D_{T_t,j}$	-0.195	(.061)			
$\log(\text{dur})_t$	0.179	(.066)	$\log(\text{dur})_{t-1}$	-0.196	(.064)
$\log(\text{vol})_t$	0.622	(.045)	α	0.161	
$\sum_{j=1}^{2} \log(\text{vol})_{t-j}$	0.081	(.027)	$\sum_{j=3}^{5} \log(\text{vol})_{t-j}$	0.062	(.023)
Q	$T\sum_{j=1}^{Q} r_j^2$			Log-like = -1355	
20	19.21	(31.41)	$E(v_t) = -.001$		
100	111.6	(124.3)	$Var(v_t) = 1.171$		
1500	1587	(1591)			

Table 11: *Estimation for the NegBin based GLARMA(2,1) model of the excess price movements ($S_t - 1$) including contemporaneous and lagged* log(duration+1) *and the* log(volume). *Variable is the explanatory variable. Std. Err. denotes the standard deviation, t–stat. denotes the t–statistic for the value being zero. The figures in brackets are the standard errors on the regressors computed using the GLARMA model. Model order selected using AIC. r_j denotes the series correlation coefficient at lag j for the the standardised residuals u_t. The figures in brackets are corresponding 95 percentage points on the χ^2_Q distribution to give a rough benchmark.*

5 Conclusions

In this paper we have proposed a decomposition of the price movements of trade–by–trade datasets. The decomposition means we have to model sequentially price activity, direction of moves and size of moves. Each modelling exercise is straightforward and interpretable. A number of extensions of the modelling framework are possible, including the use of relevant weakly exogenous variables.

When combined with a good model for the times between trades this analysis provides a complete model for the evolution of prices in real time.

Interesting open issues include: (i) modelling of two or more asset prices simultaneously, (ii) using trade data on the same stock but collected on different exchanges. Work continues on both of these issues.

6 Acknowledgements

Tina Rydberg thanks the Danish National Research Council for their financial support through a post-doctoral fellowship. Neil Shephard thanks the EU for financial support through their grant on "Econometric inference using simulation techniques". Both authors are grateful for support from The Centre for Analytical Finance, Aarhus, Denmark. The authors are grateful for the comments from the participants at the conference on "Econometrics and Financial Time Series" at the Isaac Newton Institute, Cambridge University, 12-16 October 1998 where the details of our decomposition were first presented. Also thanks to the participants at a seminar at Nuffield College. We thank Joel Hasbrouck and Richard Spady for their comments on an earlier draft of this paper and Rob Engle and Jeff Russell for various helpful conversations on these issues. Our greatest debt is to Jurgen Doornik who wrote a graphical front-end which could be used in conjunction with our Ox software (Doornik (1998)) for the fitting of GLARMA models for binary and count data. This allowed us to explore the empirical aspects of this paper with much more speed than would otherwise have been possible.

References

Barndorff-Nielsen, O. E. and G. Schou (1973). On the reparameterization of autoregressive models by partial autocorrelations. *J. Multivariate Analysis 3*, 408–419.

Bollerslev, T. and M. Melvin (1994). Bid-ask spreads in the foreign exchange market: an empirical analysis. *Journal of International Economics 36*, 355–72.

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

Cox, D. R. (1958). The regression analysis of binary sequences (with discussion). *J. Royal Statistical Society B 20*, 215–42.

Cox, D. R. and E. J. Snell (1989). *The Analysis of Binary Data* (2 ed.). London: Chapman & Hall.

Darolles, S., C. Gourieroux, and G. Le Fol (1998). Intra-day transaction price dynamics. Unpublished paper: CERMSEM-Paris I Univ. and CREST.

Doornik, J. A. (1998). *Ox: Object Oriented Matrix Programming, 2.0.* London: Timberlake Consultants Press.

Engle, R. F. (1996). The econometrics of ultra-high frequency data. Unpublished paper: Department of Economics, UC. San Diego.

Engle, R. F., D. F. Hendry, and J. F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304.

Engle, R. F. and J. R. Russell (1998). Forecasting transaction rates: the autoregressive conditional duration model. *Econometrica 66*, 1127–1162.

Granger, C. W. J. (1998). Comonotonicity. Lecture to Econometrics and Financial Time Series Workshop, Isaac Newton Institute for Mathematical Sciences, Cambridge University.

Harris, L. E. (1994). Minimum price variation, discrete bid-ask spreads and quotation sizes. *Review of Financial Studies 7*, 149–78.

Hasbrouck, J. (1996). The dynamics of discrete bid and ask quotes. Unpublished paper: Stern Business School, New York University.

Hausman, J., A. W. Lo, and A. C. MacKinlay (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics 31*, 319–30.

Hendry, D. F. (1995). *Econometric Methodology.* Oxford: Oxford University Press.

Johnson, N. L., S. Kotz, and A. W. Kemp (1992). *Univariate Discrete Distributions* (2 ed.). New York: John Wiley.

Jones, M. C. (1987). Randomly choosing parameters for the stationary and invertibility region of autoregressive-moving average models. *Applied Statistics 36*, 134–138.

Manrique, A. and N. Shephard (1997). Likelihood analysis of a discrete bid/ask price model for a common stock quoted on the NYSE. Unpublished paper: Nuffield College, Oxford.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models.* London: Chapman & Hall. 2nd Edition.

McFadden, D. L. (1984). Qualitative response models. In Z. Griliches and M. Intriligator (Eds.), *The Handbook of Econometrics, Volume 2*, pp. 1395–1457. North-Holland.

Meddahi, N., E. Renault, and B. Werker (1998). Modelling high frequency data in continuous time. Unpublished paper: CIRANO, CRDE, Montreal University.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset pricing: a new approach. *Econometrica 59*, 347–370.

O'Hara, M. (1995). *Market Microstructure Theory.* Oxford: Blackwell Publishers.

Rogers, L. C. G. and O. Zane (1998). Designing and estimating models of high frequency data. Unpublished paper: Department of Mathematics, University of Bath. Presented at Workshop on Mathematical Finance, University of Bremen, Germany, February.

Roll, R. (1984). A simple implicit measure of the efficient bid–ask spread in an efficient market. *J. Finance 39*, 1127–1139.

Russell, J. R. and R. F. Engle (1998). Econometric analysis of discrete-valued, irregularly-spaced financial transactions data using a new autoregressive conditional multinomial models. Unpublished paper: Graduate School of Business, University of Chicago. Presented at Second international conference on high frequency data in finance, Zurich, Switzerland, April.

Rydberg, T. H. and N. Shephard (1998). A modelling framework for the prices and times of trades made on the NYSE. In preparation: Nuffield College, Oxford. Presented at Workshop on Mathematical Finance, University of Bremen, Germany, February.

Shephard, N. (1994). Autoregressive based generalized linear models. Unpublished paper: Nuffield College, Oxford. Presented at the Econometric Society World Congress, Tokyo, 1995.