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

Tina Hviid Rydberg
Nuffield College, Oxford OX1 1NF, UK.
tina@imf.au.dk

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

November 3, 1998

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. Our initial attempts at using this framework will be based on generalisations of autologistic and autoregressive logarithmic models.

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: Autologistic, conditional independence, decomposition, durations, forecasting, logarithmic distribution, prediction decomposition, 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. This is for example 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 traded IBM prices at NYSE on four different days in 1995. In 1995 the restriction 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.

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

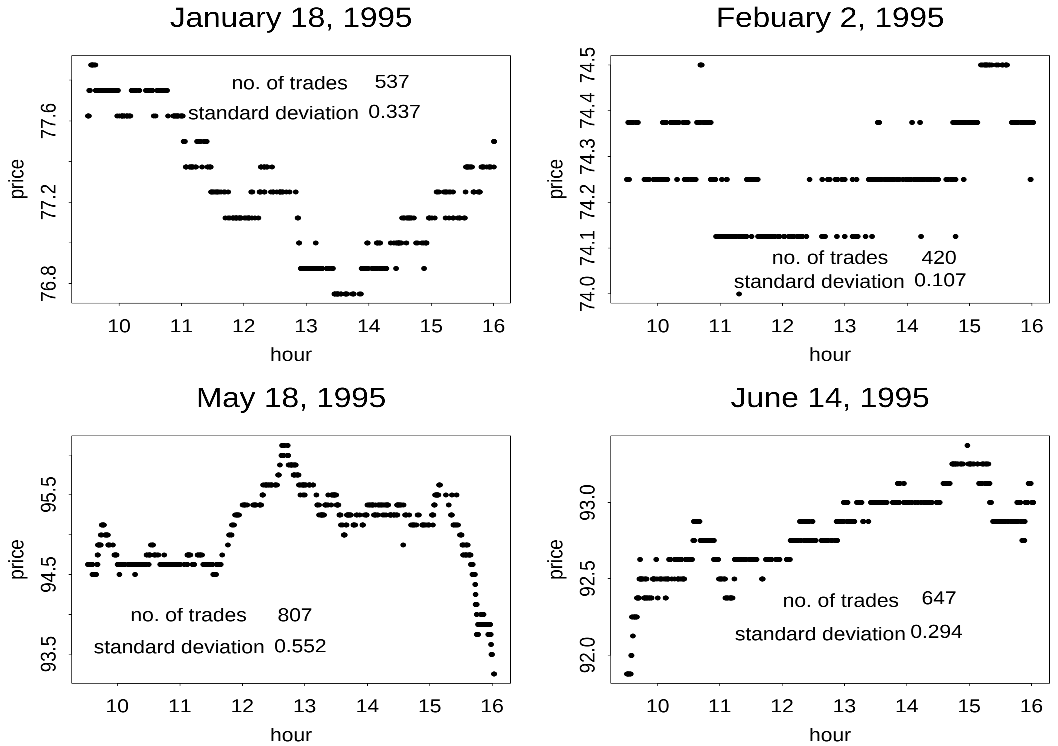

Figure 1: *Plot of all traded IBM prices at the New York stock exchange on four different days in 1995. A trade is represented as a dot.*

the price of the asset at time u, then they allowed

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

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 Cox process[1], that is a Poisson process with a random intensity. We should note immediately that many 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. 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 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.

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

[1]A special case of a Cox process is the influential autoregressive conditional duration (ACD) model advocated by Engle and Russell (1998).

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

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

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 $\Pr(Z_t|\mathcal{F}_{t-1})$. 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[2]. We will compare the Russell–Engle and Hasbrouck models to our suggestion towards the end of our paper.

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 looks at some important econometric issues such as multi-step ahead prediction, empirical model criticism and generalisations to allow for moving average style effects in the model. Section 5 places our suggestion in the context of the literature, as well as suggesting various extensions of the basic model construct. Section 6 concludes. In addition we have two Appendices. The first details the dataset we look at in this paper, while the second reviews the autologistic and logarithmic distributions we use in our empirical analysis as these maybe slightly unfamiliar to some readers.

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 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

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

[2]The econometrics of these types of models is dealt with in Manrique and Shephard (1997).

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 Basics

To start our empirical modelling we will work with the natural filtration of the price movements, building an initial empirical model for $\Pr(Z_t|\mathcal{F}_{t-1})$. In this draft of the paper we will not attempt to provide a complete model fit of the data, capturing all the aspects of the longer-term dependence in the structure of the data. Rather the focus will be on getting hold of the short run interdependency of the activities, directions and size random variables. In effect we will be modelling the marginal conditional likelihood

$$\Pr(Z_t|Z_{t-1}, \ldots, Z_{t-M}).$$

This might be called an order M (here we set $M = 20$) approximation of $\Pr(Z_t|\mathcal{F}_{t-1})$ (see Azzalini (1983)). Extending the filtration to capture the longer term movements is then straightforward, but involves the writing of specialist code (this is currently being written and preliminary results will be given in Section 4). At the moment we can fit all of our models using the software package STATA.

Some details of the econometric models we fit in this section is given in an Appendix to help readers who are unfamiliar with autologistic and logarithmic models.

3.2 The activity of prices

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 $\theta_t^A = x_t'\beta$ with x_t being potential combinations and subsets of $\mathcal{F}_{t-1}$. This model structure has some significant advantages. The corresponding log–likelihood is concave and so numerical optimisation is completely straightforward, allowing standard logistic regression software to be used to rapidly and reliably fit the model.

We use a general–to–specific model selection approach (see, for example, Hendry (1995)), estimating a complete model and then testing down insignificant lags. This results in an estimated model which suggests that

$$\theta_t^A = \text{Cont s.} + \kappa^R\left(S_{t-2} - A_{t-2}\right) + \sum_{i=1}^{20} \beta_i^A A_{t-i} + \sum_{j=1}^{2} \gamma_i^R D_{t-i}.$$

Figure 2 graphs the autocorrelation function of the activity variable together with the corresponding correlogram resulting from simulating for 200,000 time periods the activity from the fitted model (this requires a model for direction and size, which will be given later). We can see that our model captures the short-run dynamics of the activity variable in the model, but fails in the longer run. This is purely because we need to introduce a Moving Average (MA) structure into the autologistic model, since it shows quite a strong persistence in the autocorrelations. This is reflected in the estimated parameters which are given in Table 1. Here 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 in the lag structure. Our new software should deal with this problem.

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) and Barndorff-Nielsen and Shephard (1998)). 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.

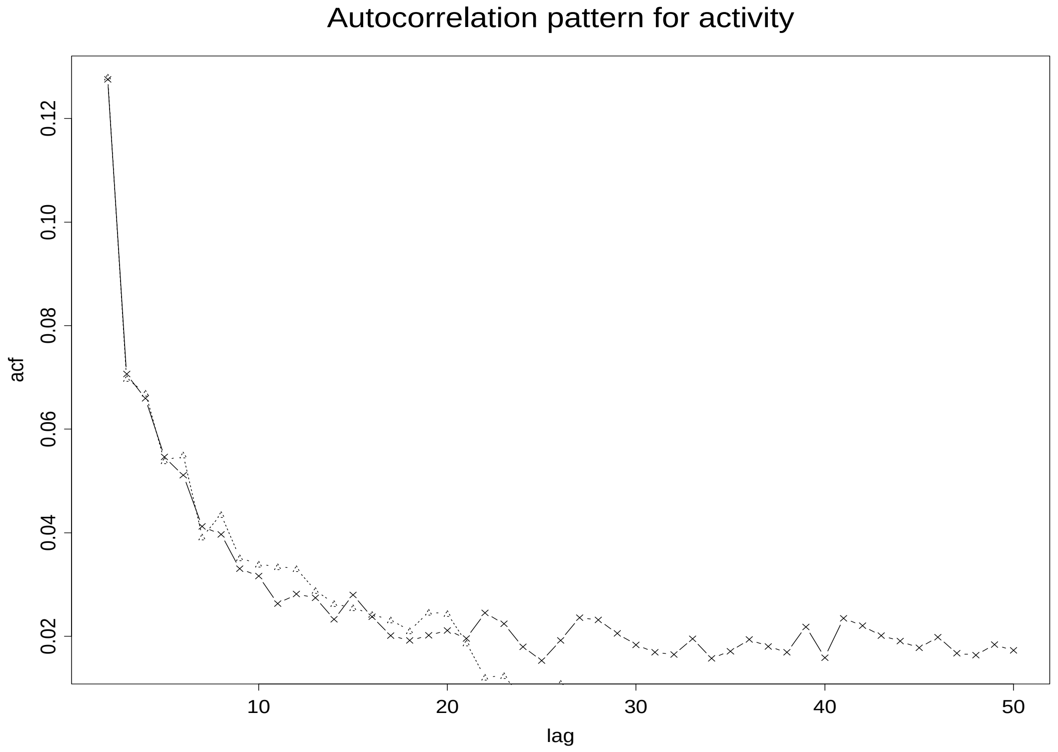

Figure 2: *Plot of the autocorrelation function for the observed activity and a simulated series based on the estimated model. The observed pattern is indicated by* $\times$*.*

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

$$T_t = \sup\{s < t : D_s \neq 0\}$$

then $M_t = D_{T_t}$ is the last price change prior to D_t which was 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

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

Furthermore, we find that only the last 8 lags are significant

$$\theta_t^D = \text{Const.} + \sum_{i=1}^{8} \gamma_i^D D_{t-i} + \sum_{j=1, j\neq 4}^{5} \delta_j^D ED_{t-j} + \sum_{k=1}^{5} \mu_k^D M_{t-k},$$

Variable	Coef.	StR. Err.	t–stat.	Variable	Coef.	StR. Err.	t–stat.
A_{t-1}	0.634	0.014	44.44	A_{t-9}	0.080	0.016	5.14
D_{t-1}	-0.103	0.013	-8.19	A_{t-10}	0.048	0.016	3.11
A_{t-2}	0.250	0.015	16.66	A_{t-11}	0.072	0.016	4.61
E_{t-2}	0.305	0.090	3.37	A_{t-12}	0.069	0.016	4.43
D_{t-2}	-0.049	0.013	-3.66	A_{t-13}	0.043	0.016	2.73
A_{t-3}	0.247	0.015	16.42	A_{t-14}	0.084	0.016	5.39
A_{t-4}	0.176	0.015	11.59	A_{t-15}	0.054	0.016	3.44
A_{t-5}	0.168	0.015	11.03	A_{t-16}	0.036	0.016	2.29
A_{t-6}	0.109	0.015	7.10	A_{t-17}	0.037	0.016	2.33
A_{t-7}	0.115	0.015	7.48	A_{t-18}	0.048	0.016	3.06
A_{t-8}	0.077	0.015	4.95	A_{t-19}	0.058	0.016	3.74
Const.	1.955	0.012	162.92	A_{t-20}	0.053	0.016	3.42

Table 1: *Estimation for the activity. Variable is the explanatory variable. Log Likelihood = -82862.959. 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.*

see Table 2. From the estimated parameters we see that the process D_t is strongly mean–reverting reflecting the observed directions. That we do in fact capture the features of the direction is seen from Figure 3. This figure shows the autocorrelation for the empirical data and the autocorrelation for a simulated series of $\{D_s\}$ generated from our fitted model, simulating it and then removing all the zero movement observations. This transform of the real data is also carried out and compared in Figure 3.

Exp. Var.	estimate	Std. Err.	t–stat.	Exp. Var.	estimate	Std. Err.	t–stat.
D_{t-1}	-2.194	0.045	-47.84	ED_{t-1}	0.712	0.186	3.83
D_{t-2}	-0.595	0.036	-16.31	ED_{t-2}	-0.488	0.156	-3.12
D_{t-3}	0.068	0.034	1.99	ED_{t-3}	-0.838	0.193	-4.34
D_{t-4}	0.330	0.033	9.91	ED_{t-5}	-0.717	0.190	-3.76
D_{t-5}	0.402	0.033	12.20	M_{t-1}	-0.274	0.020	-13.52
D_{t-6}	0.330	0.032	10.23	M_{t-2}	0.182	0.016	11.76
D_{t-7}	0.204	0.031	6.47	M_{t-3}	-0.099	0.015	-6.72
D_{t-8}	0.155	0.030	5.20	M_{t-4}	-0.031	0.015	-2.17
Const.	0.040	0.013	3.077	M_{t-5}	-0.034	0.014	-2.94

Table 2: *Estimation results for the direction of trade. After lag 8 the parameter was insignificant. The Log Likelihood = -18191.333. Variable is the explanatory variable. Std. Err. denotes the standard deviation, t–stat. denotes the t–statistic for the value being zero.*

An implication of this fitted model is that the dynamics generating the directions seems perfectly symmetrical, although the intercept indicates there are quite a lot more ups in the data than there are downs. Further, the model shows direction is quite a short memory process.

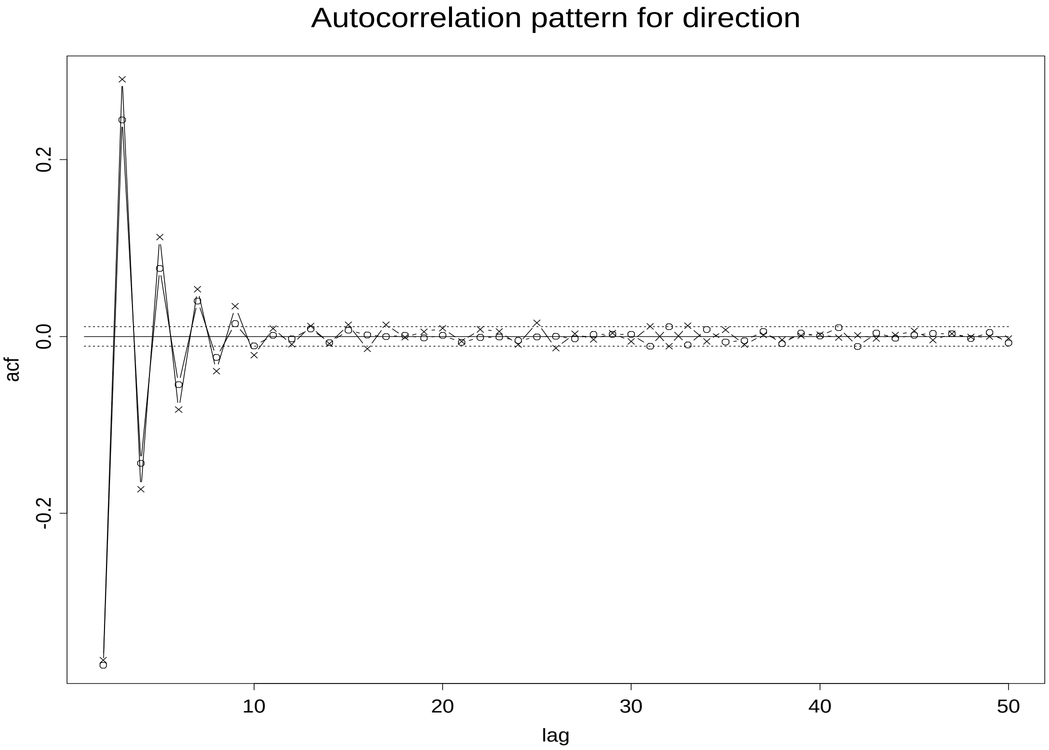

Figure 3: *Plot of the autocorrelation function for the observed direction and a simulated series based on the estimated model. The observed pattern is indicated by* $\times$*.*

The economic meaning of this fitted model is that the directions are mostly generated by bid/ask bounce. 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. Of course one of the potential uses of our model is that it could provide a probability calculus for the chance of price movements being reversed. That is to allow traders to work out if the movements in the market price are due to bid–ask bounce or to movements in the efficient price.

3.4 The size of price movements

This section is devoted to a discussion about how to model $\Pr(S_t|\mathcal{F}_{t-1}, A_t = 1, D_t)$. As we have noted above this is a process on the strictly positive integers. We will use the logarithmic distribution

$$\Pr(S_t = s_t|\mathcal{F}_{t-1}, A_t = 1, D_t) = \kappa_t \frac{p(\theta_t^S)^{s_t}}{s_t}, \qquad p(\theta_t^S) \in (0,1),$$

where

$$\kappa_t = \frac{-1}{\log\{1 - p(\theta_t^S)\}}, \qquad s_t = 1, 2, 3, \ldots,$$

as our first basic parametric model[3] and investigate different scenarios of how to include conditioning variables into the parameter p_t. The logarithmic distribution was selected as it is quite heavy tailed. Again we exploit the functional form

$$p(\theta_t^S) = \frac{\exp(\theta_t^S)}{1 + \exp(\theta_t^S)}$$

and model θ_t^S linearly in terms of components of $\mathcal{F}_{t-1}, A_t = 1, D_t$. We have chosen the logarithmic distribution because the observed density of sizes seems to have very heavy tails, see Figure 4.

We go from general to specific. In this setup we are able to test for asymmetry in the distribution of the size of the price changes. This can occur in one of two ways.

- $\Pr(S_t = s_t | \mathcal{F}_{t-1}, A_t = 1, D_t) = \Pr(S_t = s_t | \mathcal{F}_{t-1}, A_t = 1)$, so given the filtration, are the sizes independent of the direction of the price move. This is conditional symmetry.

- $\Pr(S_t = s_t | \mathcal{F}_{t-1}, A_t = 1, D_t) = \Pr(S_t = s_t | \mathcal{F}_{t-1}^S, A_t = 1)$, so the current sizes only depend on sizes of previous price movements. This is stochastic symmetry.

As can be seen from Table 3 we find that both the direction of the current price change and the preceding are significant, so rejecting both of the concepts of symmetry given above. We further find that we only need to include functions of the last 8 lags. This leads to a model

$$\begin{aligned} \theta_t^S &= \text{Const.} + \sum_{i=1}^{2} \gamma_i^S D_{t-i+1} + + \sum_{k\in\{2,4,5,6\}} \mu^S M_{t-k} + \sum_{j\in\{1,3,4,6,8\}} \beta_j^S \left(S_{t-j} - A_{t-j}\right) \\ &+ \sum_{k\in\{1,3,5,6,7\}} \delta^S S_{t-k} + \zeta_j^E D_{t-6} \left(S_{t-6} - A_{t-6}\right). \end{aligned}$$

The negative sign of the contemporaneous value of direction means that when the market goes down the down movements are typically larger than corresponding up movements. However, 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. In other words returns are very skewed.

Having discovered the asymmetry we can estimate a model for each of the scenarios, $D_t = 1$ and $D_t = -1$. The estimation results are found in Tables 4 and 5. It is quite interesting that when we consider the two instances $D_t = 1$ and $D_t = -1$ separately we find that we produce a more parsimonious model, but achieving an increase in likelihood of 178.4.

Even more interesting is the observation that the dynamic structure has changed dramatically. Almost all the significant coefficients have switched signs, reflecting the vast difference between up and down movements in the markets.

[3]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.

Variable	estimate	Std. Err.	t–stat.	Variable	estimate	Std. Err.	t–stat.
D_t	-0.449	0.074	-6.09	M_{t-5}	-0.268	0.065	-4.11
D_{t-1}	-0.631	0.100	-6.32	S_{t-5}	0.355	0.112	3.17
S_{t-1}	0.879	0.122	7.22	M_{t-6}	-0.158	0.064	-2.46
E_{t-1}	1.223	0.331	3.69	S_{t-6}	-0.374	0.152	-2.46
M_{t-2}	-0.191	0.070	-2.73	E_{t-6}	2.988	-0.458	5.46
S_{t-3}	0.264	0.128	2.06	ED_{t-6}	1.368	0.488	2.80
E_{t-3}	1.050	0.416	2.52	S_{t-7}	0.382	0.116	3.29
E_{t-4}	0.995	0.336	2.96	E_{t-8}	0.500	0.251	1.99
M_{t-4}	-0.166	0.067	-2.47				
Const.	-4.781	0.110	-43.52				

Table 3: *Estimation for the logarithmic distribution of the price Log Likelihood = -1631.9. changes. Variable is the explanatory variable. Std. Err. denotes the standard deviation, t–stat. denotes the t–statistic for the value being zero.*

Variable	estimate	Std. Err.	t–stat.	Variable	estimate	Std. Err.	t–stat.
S_{t-1}	1.549	0.171	9.06	M_{t-3}	0.373	0.108	3.44
M_{t-1}	1.173	0.121	9.66	ED_{t-3}	1.368	0.658	2.08
S_{t-2}	0.579	0.184	3.15	E_{t-6}	2.212	0.545	4.06
M_{t-2}	0.295	0.113	2.61	S_{t-7}	0.492	0.177	2.78
S_{t-3}	0.490	0.215	2.28				
Const.	-4.699	0.136	-34.65				

Table 4: *Estimation for the logarithmic distribution of the positive price changes. Log Likelihood = -592.0. Variable is the explanatory variable. Std. Err. denotes the standard deviation, t–stat. denotes the t–statistic for the value being zero.*

Judging from Figure 4 the logarithmic density is failing slightly. This can only be seen from the right panel, where the histogram is shown on a log–scale. In the centre of the distribution we have an extremely good fit, but we do not get enough simulated values which are strictly bigger than 3. It should be said here that values bigger than 3 accounts for only 0.01% (18 observations in total) of the price changes. The logarithmic distribution has only got one parameter which is probably the reason why we are not able to get both the centre and the tails correct. This can be improved if we introduce a distribution with both a location and a scale parameter. One suggestion of such a distribution which is discussed in Appendix B.4.3.

4 Econometric methods

4.1 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 $(s+1)$–

Variable	estimate	Std. Err.	t–stat.	Variable	estimate	Std. Err.	t–stat.
S_{t-1}	1.375	0.156	8.82	D_{t-3}	-0.583	0.175	-3.34
M_{t-1}	-0.463	0.135	-3.43	ED_{t-3}	-1.708	0.532	-3.21
D_{t-1}	-1.266	0.193	-6.57	M_{t-4}	-0.237	0.087	-2.71
ED_{t-1}	1.339	0.463	2.89	ED_{t-4}	-1.010	0.439	-2.30
M_{t-2}	-0.242	0.100	-2.40	M_{t-5}	-0.268	0.084	-3.20
D_{t-2}	-0.652	0.200	-3.26	S_{t-5}	0.344	0.150	-5.945
ED_{t-2}	-1.767	0.506	-3.49	E_{t-6}	1.363	0.134	3.06
Const.	-4.325	0.112	-38.625				

Table 5: *Estimation for the logarithmic distribution for the negative price changes. Log Likelihood = -861.5. Variable is the explanatory variable. Std. Err. denotes the standard deviation, t–stat. denotes the t–statistic for the value being zero.*

periods ahead price movements, (ii) predictions of the asset price levels $(s+1)$–periods ahead. We first of all deal with the former.

4.1.1 Predicting price movements

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

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

In our model Z_t gives 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.

4.1.2 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

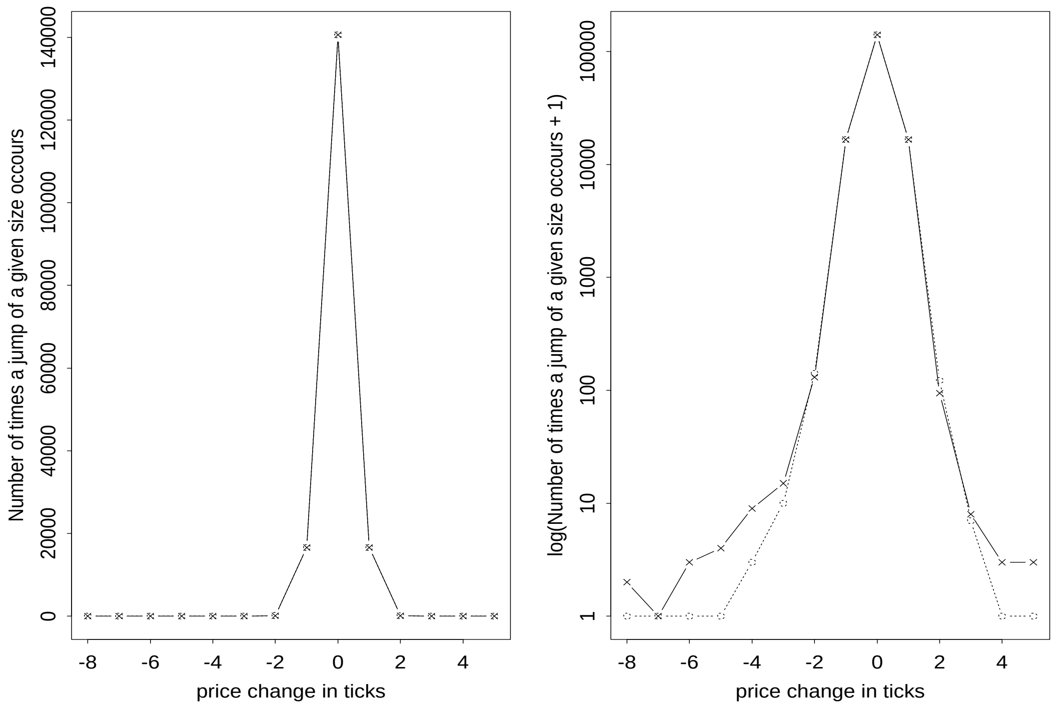

Figure 4: *Left panel shows the observed and the simulated histogram for price changes. Right panel shows the two histogram on a log scale. The observed data are denoted by solid lines and* $\times$ *as tick marks. It is only on the log–scale we are able to distinguish the two patterns.*

standard errors as

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

Figures 5, 6 and 7 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 5 two down movements have just been observed and this is seen to give an increased probability for moving one tick up. Figure 6 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 7, corresponds to "bid–ask bouncing". This has decreased the probability of moving away from the current price level compared to Figure 5. 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 6). In Table 6 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).

4.2 Empirical model criticism

The principle way we test the econometric specification of our models is to add lagged dependent variables and test their importance. This can be done either by score or

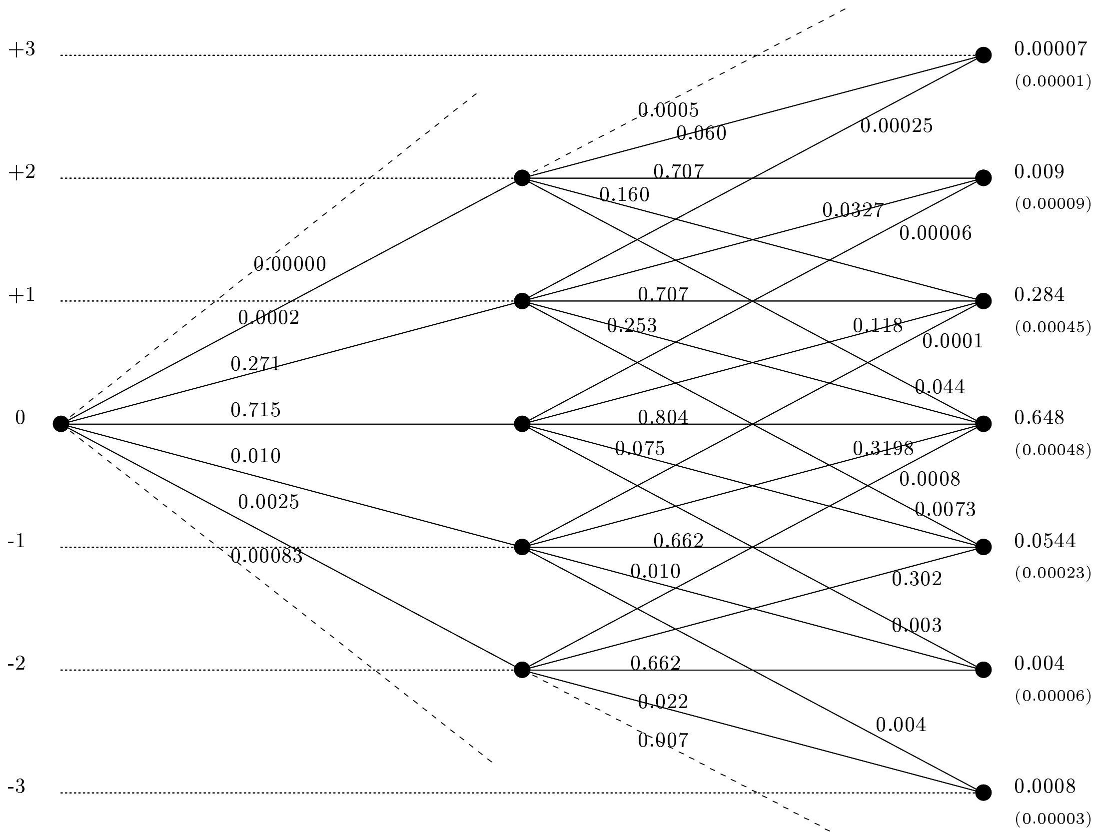

Figure 5: *History $\mathcal{F}^A = (-1, -1, 0, ..., 0)$, $\mathcal{F}^M = (-1, -1, -1, 1, -1)$. The numbers to the right are the simulated probabilities of ending up at this level after two trades with the given history. The numbers in brackets are the standard errors of the simulated values.*

likelihood ratio methods — we favour the latter given the speed with which we can fit our models and the better small sample behaviour of LR tests. This method suggests that currently the main weakness in our model is the model for activity which needs much more memory. This is remarked upon earlier. Our other models pass this style of test.

Clearly we could use other forms of portmanteau testing. An example (suggested, for example, in Russell and Engle (1998)) of testing the activity model would be to construct a normalised sequence for the fit of a binary process $\{Y_t\}$ by $\{p(\theta_t)\}$ using

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

for if the model was correctly specified the $\{u_t\}$ should be uncorrelated with zero mean and unit conditional variance. The $\{u_t\}$ could then be used inside a Box-Ljung statistic as a measure of residual dependence. Of course such a test is likely to have poor power as we would expect the $\{u_t\}$ to have had their linear dependence removed by the fitting of the model. Russell and Engle (1998) suggest using this structure in the case where we are checking the fit of Z_t (using a multivariate version of the Box-Ljung statistic).

We could test the validity of the model for $\{Z_t\}$ directly. Recall our models specify the process $\Pr(Z_t|\mathcal{F}_{t-1})$. Hence we can compute exactly $F_t = \Pr(Z_t \leq z_t|\mathcal{F}_{t-1})$. Unfortunately

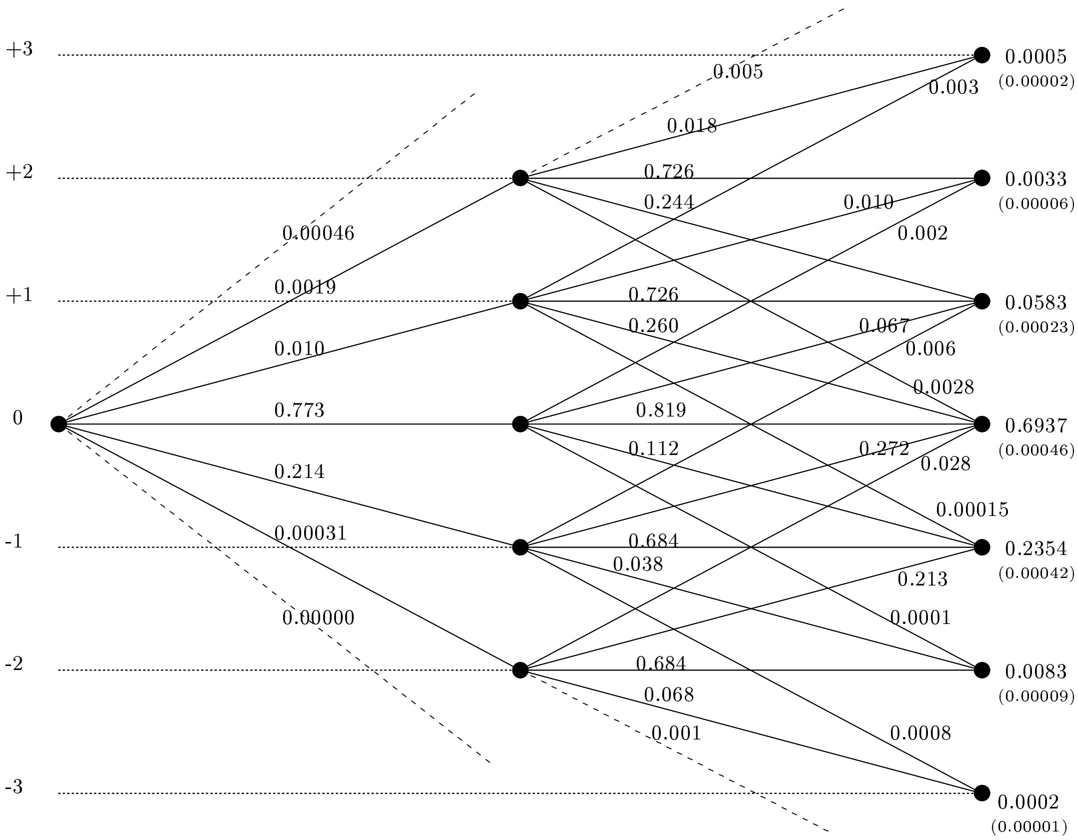

Figure 6: *History* $\mathcal{F}^A = (1, 1, 0, ..., 0)$, $\mathcal{F}^M = (1, 1, -1, 1, -1)$. *The numbers to the right are the simulated probabilities of ending up at this level after two trades with the given history. The numbers in brackets are the standard errors of the simulated values.*

F_t, by necessity, lives on discrete points. However, this lattice can be refined through randomisation. This risks the loss of information but will not introduce bias if carried out in the following way.

Define

$$v_t | Z_t = s_t, \mathcal{F}_{t-1} \overset{ind}{\sim} U \left\{ \Pr(Z_t \leq s_t - 1 | \mathcal{F}_{t-1}), \Pr(Z_t \leq s_t | \mathcal{F}_{t-1}) \right\}. \tag{1}$$

Then if the model is correct $\{v_t\}$ are uniformly, independently and identically distributed on $(0, 1)$. The same property will hold for the Kim, Shephard, and Chib (1998) reflected version $\left\{ u_t = 2 \left| v_t - \frac{1}{2} \right| \right\}$, which focuses on the dispersion of the model. We can now do standard diagnostic checks on both $\{v_t\}$ and $\{u_t\}$. In particular we can look at their QQ–plots and the associated correlograms and Box-Ljung statistics once a inverse normal distribution function transform has been made.

4.3 GLARMA type models of activity

In this subsection we will give some initial results on fitting simple autoregressive moving average models to the activity series. In this work we will abstract from all variables except lagged activity, returning to the role of other variables which appear in the filtration of Z_t in a later revision.

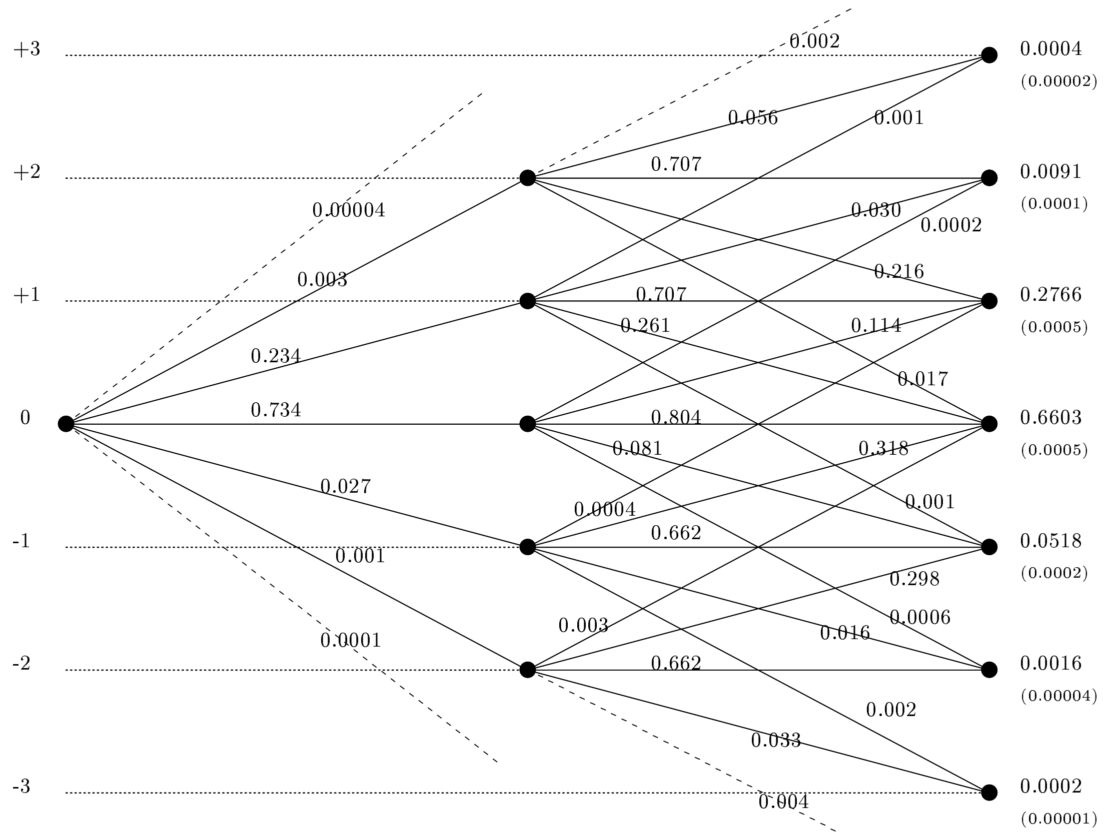

Figure 7: *History* $\mathcal{F}^A = (-1, 1, 0, ..., 0)$, $\mathcal{F}^M = (-1, 1, -1, 1, -1)$. *The numbers to the right are the simulated probabilities of ending up at this level after two trades with the given history. The numbers in brackets are the standard errors of the simulated values.*

Let us write abstractly the problem of modelling a logistic based binary time series Y_t as one of constructing

$$\Pr(Y_t|\mathcal{F}^y_{t-1}) = \frac{\exp(\theta_t)}{1+\exp(\theta_t)} = p_t.$$

The question is how to allow θ_t to depend upon previous values of Y_t. We have used autologistics which have

$$\theta_t = \alpha + g_t, \qquad \text{where} \qquad g_t = \sum_{j=1}^{p} \beta_j y_{t-j}.$$

The obvious generalisation is to write

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

but this is typically numerically unstable and so difficult to work with. Shephard (1994) has studied two alternatives, which are called generalized linear autoregressive moving average (GLARMA) models. The first puts

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

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 6: *The Table show the simulated probabilities for having moved x ticks after the given history.* $\mathcal{F}^A = (|?|, |?|, 0, ..., 0)$ *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

$$v_t = \frac{(y_{t-1} - p_{t-1})}{\sqrt{p_{t-1}(1 - p_{t-1})}}, \tag{2}$$

which are a Martingale difference sequence with a unit conditional variance. This style of model is adopted in Russell and Engle (1998) in their multinomial construction[4]. The main alternative to this model is where we construct

$$v_t = \frac{(y_{t-1} - p_{t-1})}{p_{t-1}(1 - p_{t-1})}. \tag{3}$$

At first sight this is harder to understand, but was derived by Shephard (1994) as a plausible function using topological arguments[5]. Again, v_t is a Martingale sequence but now its conditional variance is $1/\sqrt{p_{t-1}(1 - p_{t-1})}$ which is largest when p_t is close to one

[4]In our numerical work we have imposed that this model obeys the usual stationarity and invertibility constraints on the AR parameters $\{\gamma_j\}$ and the moving average parameters $\{\delta_j\}$. We have done that even though we do not know if this is necessary or sufficient for $\{Y_t\}$ to be weakly stationary. The imposition of these conditions are carried out by using the partial autocorrelations and the inverse partial autocorrelations (see Barndorff-Nielsen and Schou (1973) and Jones (1987)). We numerically maximised the corresponding likelihood using analytic first derivatives and the BHHH algorithm.

[5]The argument is that is we are modelling a function of the conditional mean $h(\mu_t)$, as being linear in transformations of past data, then we should use $h()$ to transform the data so that the mean and the data are on the same scale. Ideally this would lead to models of the form

$$h(\mu_t) = \alpha + \beta h(y_{t-1}).$$

A simple example of this is the log-GARCH model of Geweke (1986) who argued for modelling variances on the log-scale using log-squares of previous data.

For binary data using logistic transform, $\log\{p_t/(1 - p_t)\}$ we cannot use $h(y_{t-1})$ due to singularities.

or zero. We have fitted both types of models, but find that latter model does slightly worse than the first suggestion[6]. Hence we will focus on models using (2).

In our initial empirical work we have taken the activity series $\{A_t\}$ and modelled it simply as depending on its own past — ignoring the more complex dependencies we found on other feature of lagged $\{Z_t\}$ we found earlier. Table 7 gives the maximised likelihoods for each model we have fitted. We have only taken $p = 1, 2, ..., 5$ and $q = 0, 1$ at the moment[7]. This suggests an GLARMA(3,1) or GLARMA(4,1) model is the best fitting model. Notice that the likelihood for this model is much higher than for the previous models fitted for activity given in Table 1.

	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$q=0$	-83019.	-82905.	-82884.	-82864.	-82855.
$q=1$	-82842.	-82581.	-82526.	-82524.	-82524.

Table 7: *Fitted likelihood function for various values of p and q in the GLARMA structure, using (2).*

The estimated parameters of the model are given in Table 8. The table is presented in terms of partial autocorrelations and inverse partial autocorrelations as these are stable as we change the model in all cases except where the fit of the model changes dramatically. We can see that when we move to a GLARMA(2,1) model the estimated structure dramatically changes. This is because we can model separately the short term fall in the

Shephard (1994) argued that we might replace $h(y_{t-1})$ by

$$\begin{aligned} h(y_{t-1}) &\simeq h(p_{t-1}) + \frac{\partial h(p_{t-1})}{\partial p_{t-1}}(y_{t-1} - p_{t-1}) \\ &= h(p_{t-1}) + \frac{(y_{t-1} - p_{t-1})}{p_{t-1}(1 - p_{t-1})}. \end{aligned}$$

So the suggested model is

$$h(p_t) = \alpha + \beta h(p_{t-1}) + \beta \frac{(y_{t-1} - p_{t-1})}{p_{t-1}(1 - p_{t-1})}.$$

[6] *Fitted likelihood function for various values of p and q in the GLARMA structure, using (3). Compare to Table 7.*

	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$q=0$	-83045.	-82939.	-82920.	-82903.	-82895.
$q=1$	-82881.	-82624.	-82571.	-82571.	-82570.

	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$q=0$	2224.9	2108.1	2053.2	1928.3	1932.0
$q=1$	1785.8	1481.8	1442.1	1415.6	1470.5

Box-Ljung statistic using 1500 lags using the randomised residuals from the fit of the GLARMA model. If the model is correctly fitting the statistics should be around 1500 with a standard error of 55. Compare to Table 9.

[7] 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.

GLARMA(p,q)	Partial autocorrelations (PAC)					Inverse PAC
$p=1,q=0$	0.86900					
$p=2,q=0$	0.78736	0.38471				
$p=3,q=0$	0.77807	0.33820	0.15235			
$p=4,q=0$	0.77565	0.32580	0.10274	0.14321		
$p=5,q=0$	0.77927	0.32159	0.084344	0.11437	0.090818	
$p=1,q=1$	0.95605					-0.56540
$p=2,q=1$	0.99980	-0.74133				-0.98119
$p=3,q=1$	0.99987	-0.65661	-0.27944			-0.98563
$p=4,q=1$	0.99988	-0.65308	-0.26793	-0.036957		-0.98614
$p=5,q=1$	0.99989	-0.65238	-0.26506	-0.026453	-0.031731	-0.98659

Table 8: *Parameter estimates of the fitted GLARMA models, using a maximum likelihood criteria. The parameters are mapped into their corresponding partial autocorrelation and inverse partial autocorrelations as these are stable.*

autocorrelations of the series as well as the extremely dependent activity at long lags. It seems going much beyond a GLARMA(3,1) does not add much. When we have fitted move extensive moving average structure we typically do not achieve a very big fall in the likelihood function.

The estimated parameters suggest a great deal of memory in the activity series. The half life of a shock is over $6,000$ trades, which is around 10 days of physical time. 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 continues.

In Table 9 we also give the Box-Ljung statistics for the fitted model using the randomisation procedure, (1), discussed in the previous subsection. The statistic is computed using $1,500$ lags and is designed to pick up correlations at long lags. Figure 8 graphs all the cumulative correlograms ($\sum_{j=1}^{u} r_j$, where r_j are the serial correlation coefficients, drawn against u) for the fitted models. We can see that all the models with $q=0$ are rejected by this statistic, but when we take $q=1$ we can begin to fit the data. In particular all the models at least as complicated as an GLARMA(2,1) seem to broadly fit the data. Smaller models than this are overwhelmingly rejected.

	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$q=0$	2220.4	2081.1	2014.5	1885.2	1891.0
$q=1$	1791.7	1485.5	1436.9	1403.9	1465.6

Table 9: *Box-Ljung statistic using 1500 lags using the randomised residuals from the fit of the GLARMA model. If the model is correctly fitting the statistics should be around 1500 with a standard error of 55.*

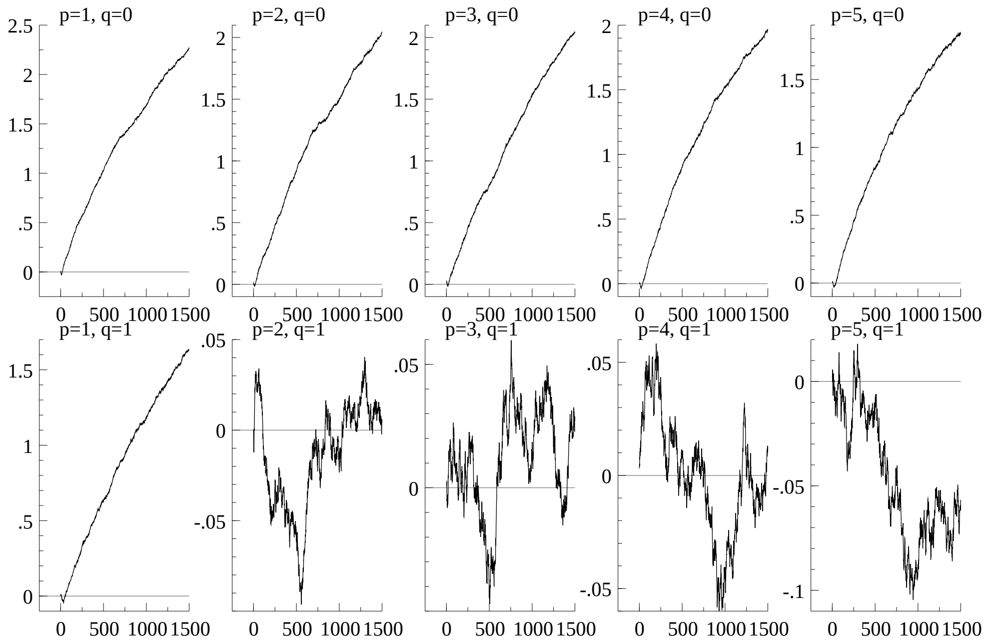

Figure 8: *Cumulative correlogram for the randomised residuals from the fit of the GLARMA(p,q) models. Graphs computed using 1500 lags. A well fitting model should deliver cumulative correlograms which are close to random walks.*

5 Comments

5.1 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

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

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. This can be seen from Figure 9 which show the observed and simulated autocorrelation functions. It it seen that the price mean–reverts very quickly which is indicated by a very large correlation at the first lag.

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

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

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.

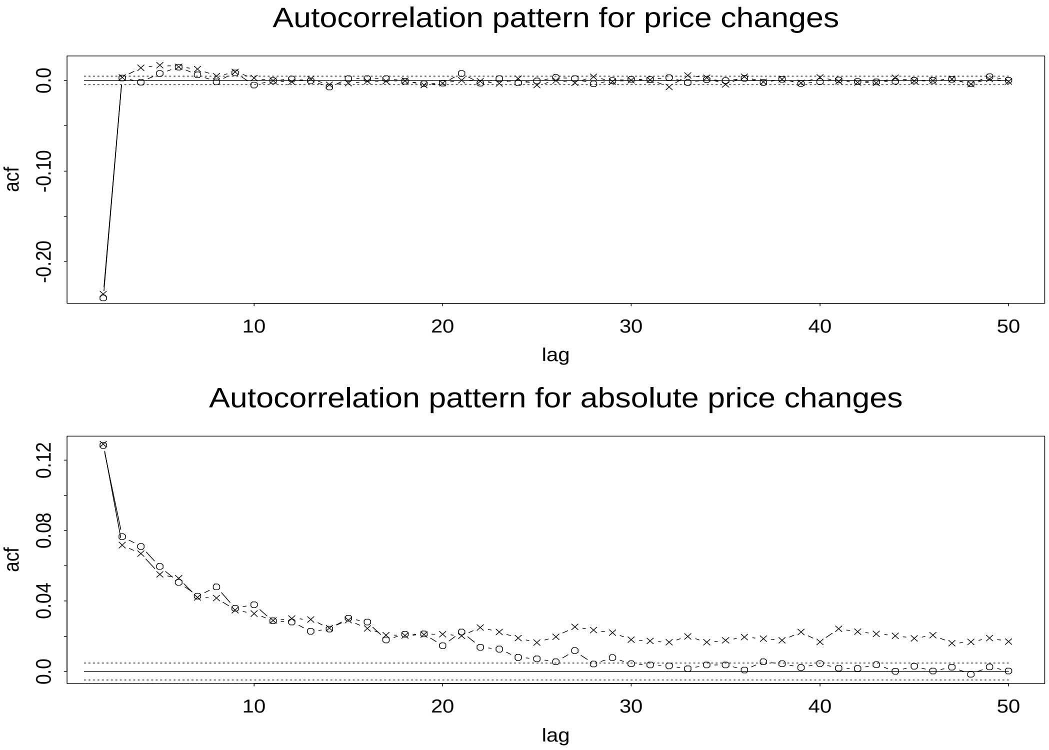

Figure 9: *The observed and simulated autocorrelation functions for the price changes and the absolute price changes. The observed values are denoted by* $\times$*.*

5.2 Explanatory variables

5.2.1 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.

Of course the decomposition of the joint process of $\{Z_t, Y_t\}$ implies we should think about when it is possible to safely (efficiently) ignore the process for $f(Y_t|\mathcal{F}_{t-1}^{z,y})$ and to simply model the conditional process $\Pr(Z_t|Y_t, \mathcal{F}_{t-1}^{z,y})$. For this to happen we will need that the $f(Y_t|\mathcal{F}^{z,y}_{t-1})$ process is weakly exogenous — which is discussed extensively in Engle, Hendry, and Richard (1983) and Hendry (1995, Ch. 5).

The above decomposition suggests that there are potentially two interesting densities for $\{Z_t\}$ to investigate further. $\Pr(Z_t|\mathcal{F}^{z,y}_{t-1})$ and $\Pr(Z_t|Y_t, \mathcal{F}^{z,r}_{t-1})$. The first is a pure forecast while the second allows contemporaneous log(durations) to enter. In Tables 10, 11, 12 and 13 the estimated parameters for the two models can be found for the activity variable and the direction variable. We have also allowed dependence on log10(volume). From Tables 10 and 11 where we have gone from general to specific, we see that the forecast density for activity depends only on the last 4 lags of log10(volume) while it depends on the last 9 lags of log(durations). The first lag of log10(volume) is however very significant. For the contemporaneous density we see that again we only have dependence on the contemporaneous volume and the first four lags, put we need more than 20 lags of durations.

For the densities for the direction variable the picture is different, see Table 12 and 13 here we have no dependence on the durations in the pure forecast density. In the contemporaneous density for the directions we find that the contemporaneous duration is very significant but there is no dependence on lagged durations. For both the forecast and the contemporaneous density we find that the first two lags of volume are significant and of course in the contemporaneous case the contemporaneous volume is also significant.

5.2.2 Trading durations and volume

An important example of the above is to formally test to see if the times between trades and volume are independently distributed from the price process. Write the times between trades and volume as Y_t. We can think of this in stages:

1. $\Pr(Z_t|Y_t, \mathcal{F}^{z,y}_{t-1}) = \Pr(Z_t|\mathcal{F}^{z,y}_{t-1})$, the current value of the price innovation is conditionally independent of the volume and time between this trade and the last.

2. $\Pr(Z_t|\mathcal{F}^{z,y}_{t-1}) = \Pr(Z_t|\mathcal{F}_{t-1})$, the distributions of the price movements do not depend upon the current and previous values of the volume and times between trades.

The combination of these two conditions is enough to show independence between the times and price innovation processes. We start by just conditioning on the current value of Y_t in both the model of activity and the model of direction.

The quantitative effect of this quite large. Activity is effected positive by both volume and duration (see Tables 10 and 11). In particular if the duration is high then this increases the chance the price will move at the next trade, while if volume if high the same thing happens.

When we look at direction (see Tables 12 and 13) 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.

5.2.3 Deterministic seasonality

Although the Cox process governing the arrivals of trades should deal with most of the clustering in trades and so movements caused by cyclical and deterministic seasonal patterns, it maybe possible that the actual time of trades could effect the dynamics of the

Variable	estimate	Std. Err.	t–stat.	Variable	estimate	Std. Err.	t–stat.
A_{t-1}	0.626	0.015	42.64	$\log(\text{dur})_{t-1}$	0.011	0.005	2.20
D_{t-1}	-0.086	0.013	-6.6960]	$\log(\text{dur})_{t-2}$	-0.031	0.005	-6.43
A_{t-2}	0.273	0.015	17.63	$\log(\text{dur})_{t-3}$	-0.025	0.005	-5.12
E_{t-2}	0.437	0.092	4.73	$\log(\text{dur})_{t-4}$	-0.021	0.005	-4.34
D_{t-2}	-0.036	0.014	-2.6509]	$\log(\text{dur})_{t-5}$	-0.024	0.005	-4.93
A_{t-3}	0.251	0.016	16.09	$\log(\text{dur})_{t-6}$	-0.021	0.005	-4.25
A_{t-4}	0.172	0.016	11.03	$\log(\text{dur})_{t-8}$	-0.017	0.005	-3.59
A_{t-5}	0.165	0.016	10.54	$\log(\text{dur})_{t-9}$	-0.013	0.005	-2.76
A_{t-6}	0.102	0.016	6.42	$\log10(\text{vol})_{t-1}$	-0.373	0.010	-36.69
A_{t-7}	0.099	0.016	6.37	$\log10(\text{vol})_{t-2}$	-0.041	0.010	-4.04
A_{t-8}	0.081	0.016	5.10	$\log10(\text{vol})_{t-4}$	0.023	0.010	2.32
A_{t-9}	0.075	0.016	4.70				
A_{t-10}	0.038	0.016	2.38				
A_{t-11}	0.066	0.016	4.24				
A_{t-12}	0.064	0.016	4.10				
A_{t-13}	0.036	0.016	2.29				
A_{t-14}	0.077	0.016	4.93				
A_{t-15}	0.049	0.016	3.14				
A_{t-16}	0.031	0.016	1.96				
A_{t-17}	0.031	0.016	1.99				
A_{t-18}	0.043	0.016	2.70				
A_{t-19}	0.053	0.016	3.38				
A_{t-20}	0.047	0.016	3.03				
Const.	0.295	0.065	4.54				

Table 10: *Estimation for the activity including the lags of* log*(duration+1) and the* $\log_{10}$*(volume). Log Likelihood = -82052.0. Improvement in likelihood from Table 1 is 810.96. Variable is the explanatory variable. Std. Err. denotes the standard deviation, t–stat. denotes the t–statistic for the value being zero.*

price movement process. However, 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.

5.3 Directions

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.

Variable	estimate	Std. Err.	t–stat.	Variable	estimate	Std. Err.	t–stat.
A_{t-1}	0.760	0.015	50.76	$\log(\text{dur})_t$	0.348	0.005	68.34
D_{t-1}	-0.084	0.013	-6.693	$\log(\text{dur})_{t-2}$	-0.044	0.005	-8.82
A_{t-2}	0.294	0.016	18.598	$\log(\text{dur})_{t-3}$	-0.039	0.005	-7.92
E_{t-2}	0.432	0.094	4.58	$\log(\text{dur})_{t-4}$	-0.034	0.005	-6.86
A_{t-3}	0.256	0.016	16.12	$\log(\text{dur})_{t-5}$	-0.035	0.005	-7.00
A_{t-4}	0.179	0.016	11.21	$\log(\text{dur})_{t-6}$	-0.032	0.005	-6.38
A_{t-5}	0.173	0.016	10.86	$\log(\text{dur})_{t-7}$	-0.017	0.005	-3.47
A_{t-6}	0.114	0.016	7.08	$\log(\text{dur})_{t-8}$	-0.026	0.005	-5.23
A_{t-7}	0.113	0.016	6.99	$\log(\text{dur})_{t-9}$	-0.019	0.005	-3.80
A_{t-8}	0.088	0.016	5.42	$\log(\text{dur})_{t-10}$	-0.012	0.005	-2.46
A_{t-9}	0.080	0.016	4.92	$\log(\text{dur})_{t-12}$	-0.018	0.005	-3.53
A_{t-10}	0.043	0.016	2.62	$\log(\text{dur})_{t-13}$	-0.019	0.005	-3.80
A_{t-11}	0.066	0.016	4.11	$\log(\text{dur})_{t-14}$	-0.013	0.005	-2.57
A_{t-12}	0.077	0.016	4.74	$\log(\text{dur})_{t-15}$	-0.012	0.005	-2.43
A_{t-13}	0.041	0.016	2.50	$\log(\text{dur})_{t-16}$	-0.022	0.005	-4.37
A_{t-14}	0.074	0.016	4.54	$\log(\text{dur})_{t-17}$	-0.015	0.005	-3.06
A_{t-15}	0.052	0.016	3.16	$\log(\text{dur})_{t-18}$	-0.014	0.005	-2.88
A_{t-16}	0.039	0.016	2.39	$\log(\text{dur})_{t-20}$	-0.017	0.005	-3.47
A_{t-17}	0.033	0.016	2.02	$\log10(\text{vol})_t$	0.133	0.010	13.02
A_{t-18}	0.042	0.016	2.58	$\log10(\text{vol})_{t-1}$	-0.377	0.010	-36.20
A_{t-19}	0.051	0.016	3.17	$\log10(\text{vol})_{t-2}$	-0.034	0.011	-3.23
A_{t-20}	0.058	0.016	3.58	$\log10(\text{vol})_{t-3}$	0.022	0.011	2.07
Const.	1.246	0.081	15.38	$\log10(\text{vol})_{t-4}$	0.031	0.010	2.95

Table 11: *Estimation for the activity including contemporaneous and lagged the* $\log$*(duration) and the* $\log_{10}$*(volume). Log Likelihood = Log Likelihood = -79513.864. Improvement in likelihood is 3,349.095 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.*

5.4 Previous works

5.4.1 Conditional multinomial models

In a recent highly stimulating paper Russell and Engle (1998) have suggested modelling price movements using a conditional multinomial distribution. 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 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.

Variable	estimate	Std. Err.	t–stat.	Variable	estimate	Std. Err.	t–stat.
D_{t-1}	-2.192	0.046	-47.79	$\log10(\text{vol})_{t-1}$	-0.078	0.020	-3.96
D_{t-2}	-0.629	0.036	-16.30	$\log10(\text{vol})_{t-2}$	0.055	0.019	2.86
D_{t-3}	-0.036	0.034	2.01				
D_{t-4}	0.171	0.033	9.90				
D_{t-5}	0.267	0.033	12.14				
D_{t-6}	0.232	0.032	10.21				
D_{t-7}	0.131	0.032	6.44				
D_{t-8}	0.097	0.030	5.19				
ED_{t-1}	0.889	0.187	3.74				
ED_{t-2}	-0.422	0.156	-3.13				
ED_{t-3}	-0.788	0.194	-4.38				
ED_{t-5}	-0.691	0.190	-3.76				
M_{t-1}	-0.271	0.020	-13.35				
M_{t-2}	0.179	0.016	11.41				
M_{t-3}	-0.099	0.015	-6.68				
M_{t-4}	-0.032	0.014	-2.17				
M_{t-5}	-0.035	0.014	-2.51				
Const.	-0.03	0.078	-0.38				

Table 12: *Estimation for the direction including lagged* log*(duration+1) and the log10(volume). Log Likelihood = -18182.17. Improvement in likelihood 9.163 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.*

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 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 (2). 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

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

Variable	estimate	Std. Err.	t–stat.	Variable	estimate	Std. Err.	t–stat.
D_{t-1}	-2.181	0.046	-47.49	$\log(\text{dur})_t$	-0.105	0.010	-10.27
D_{t-2}	-0.593	0.037	-16.22	$\text{log10}(\text{vol})_t$	-0.026	0.009	-2.93
D_{t-3}	0.069	0.034	2.00	$\text{log10}(\text{vol})_{t-1}$	-0.066	0.020	-3.33
D_{t-4}	0.331	0.033	9.89	$\text{log10}(\text{vol})_{t-2}$	0.058	0.020	2.96
D_{t-5}	0.398	0.033	12.05				
D_{t-6}	0.327	0.032	10.11				
D_{t-7}	0.199	0.032	6.29				
D_{t-8}	0.149	0.030	4.99				
ED_{t-1}	0.713	0.187	3.82				
ED_{t-2}	-0.478	0.156	-3.06				
ED_{t-3}	-0.827	0.193	-4.28				
ED_{t-5}	-0.706	0.192	-3.67				
M_{t-1}	-0.278	0.020	-13.63				
M_{t-2}	0.179	0.016	11.366				
M_{t-3}	-0.101	0.015	-6.755				
M_{t-4}	-0.031	0.014	-2.188				
M_{t-5}	-0.035	0.014	-2.528				
Const.	-0.562	0.100	-5.62				

Table 13: *Estimation for the direction including contemporaneous and lagged the* $\log$*(duration) and the* $\log_{10}$*(volume). Log Likelihood = -18125.342. Improvement in likelihood 65.991 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.*

then model this system as

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

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

An interesting additional feature of the Russell and Engle (1998) specification is their equation (32) which imposes symmetry on the up/down dynamics of the process. This is carried out a priori and would seem motivated by parsimony, as well as an informal discussion in Section 3 of the paper. Our empirical results on size suggests this constraint of symmetry is overwhelmingly rejected by the data.

Overall we can see that our analysis is quite closely related to that of Russell and Engle (1998). Our goals are the same, although the technology that we use is very different. Our main advantages are: parsimony, interpretation and options for extensions. It is not clear if we have any disadvantages, although we have to wait to see the effectiveness of our moving average terms on our modelling fitting to be sure of this statement.

5.4.2 Hasbrouck's truncation model

Hasbrouck (1996) introduced a dynamic model for the evolution of the bid and ask price of quotation data. Let us write μ_t denote the theoretical efficient price in the market and

α_t, β_t represent the ask and bid costs respectively. The Hasbrouck argued for a structure where the bid price is $Floor(\mu_t - \alpha_t)$ and the ask price is $Ceiling(\mu_t - \beta_t)$. 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.

5.5 Market betas

A traditional way of modelling individual stock prices is to regress stock returns on the return of an index. This can be carried out here by looking at the return on the index since the last trade of the individual stock we are modelling. Of course, individual stocks may lead or lag the index and so this regression will have to be carried out using dynamic structures.

5.6 Software

All the models we have discussed in this paper have been fitted in STATA 5, StataCorp (1997), although most commonly used statistical software has routines which allow the fitting of generalized linear models (special cases of which are binary and logarithmic regressions). Example macros for our empirical results are available from the authors.

5.7 Non-parametric autoregressions

Our work can be extended to allow the construction of autogressions of the forecast distribution for a general stationary time series. Let us be interested in a sequence $\{Y_t\}$, with corresponding filtration , $\mathcal{F}_t = \sigma(Y_s, s \leq t)$. We will allow $\Pr(Y_t|\mathcal{F}_{t-1})$ to possess atoms. We will focus on using directional information in our analysis, although doing so would be natural in many problems.

Suppose we impose a discrete lattice on this structure at points

$$\{c_j; j = -d, -d+1, ..., 0, ..., d-1, d\}.$$

Then it maybe sensible to work sequentially in the following manner:

$$\Pr(Y_t|\mathcal{F}_{t-1} > c_d) = \Pr(Y_t > c_d|\mathcal{F}_{t-1}, Y_t > c_{d-1}) \Pr(Y_t > c_{d-1}|\mathcal{F}_{t-1}),$$

for we can now model $\Pr(Y_t > c_d|\mathcal{F}_{t-1}, Y_t > c_{d-1})$ as a binary GLARMA model. We can see this argument chains, requiring us to sequentially model

$$\Pr(Y_t > c_j|\mathcal{F}_{t-1}, Y_t > c_{i-j}), \qquad j = d, d-1, ..., -d+1.$$

We finish off this model by working with

$$\Pr(Y_t > c_{-d}|\mathcal{F}_{t-1}).$$

This structure has some direct advantages. In some sense this is a non-parametric approach for we do not write down a direct model for all of $Y_t|\mathcal{F}_{t-1}$. The conditional probability structures means that data at lower quantiles helps us to change the probability statements we make in higher quantiles.

6 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.

As we write this paper we have not fitted moving average style models and so we cannot, as yet, fit the long run dynamic behaviour of the activity variable. However, we are currently writing software to do this and so this paper will be updated when this has been carried out.

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.

7 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.

A Appendix — data

The data set used in this paper consists of trade data recorded electronically at the New York Stock Exchange. Specifically we limit our study to the trades in IBM stock in 1995. We 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 have worked only on changes in tick sizes and we have concatenated the days. In order to clean out over night effects we let the first tick size recorded in the day be the difference between to first price after 9.45 and the last price just before that time. The length of the total data set when all exchanges in the US is considered 413,906, this is too much data set to 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 174,045 observations.

B Appendix — econometric background

B.1 Autologistics

In this Appendix we will remind readers about autologistics and the properties of the logarithmic distributions as they will be used extensively in the paper as models for activity, direction and size. For clarity of exposition we will write Y_t as the dependent random variable and y_t as its observed value. Sometimes we will be concerned with the role of a set of explanatory variables which we will write as X_t. This particular notation is only used in this Section of the paper.

A standard regression model for binary data is the logistic regression (a straightforward alternative is the probit regression), which puts

$$\Pr(Y_t = 1|X_t) = p(\theta_t), \qquad p(\theta_t) = \exp(\theta_t)/\left\{1 + \exp(\theta_t)\right\}$$

and $\theta_t = X_t\beta$. The classic reference is Cox and Snell (1989)[8]. Maximum likelihood estimation is particularly easy for this model as the log-likelihood is concave and numerical optimisation is known to be both reliable and very fast even when there are a large number of explanatory variables (e.g. McCullagh and Nelder (1989)).

In the time series case the autologistic model allows lagged information to enter linearly into θ_t. The canonical example is the first order Markov autologistic

$$\Pr(Y_t = 1|Y_{t-1} = y_{t-1}) = p(\theta_t),$$

where $\theta_t = \mu + \beta y_{t-1}$. This was introduced in Cox (1958). Conditional (on initial values) maximum likelihood estimation is straightforward using a prediction decomposition.

From our modelling perspective it straightforward to note that although $Y_t = 0, 1$, precisely the same framework can be used for any model where there are only two possible outcomes. In our case this will be useful as we will sometime wish to model $Y_t = \pm 1$.

B.2 Autoregressions on positive integers

In later sections we will need to model the size of actual price movements, which are defined to be strictly positive integers. For the NYSE the vast majority of these price movements will be one, while a few will be two with very occasional movements outside these two points — with falling probabilities as the sizes get larger. We will use this structure to build our models, however in other markets rather different shapes will be needed. In actively traded exchange rates Hasbrouck (1996) notes clustering of sizes on particular points of the space. We ignore this type of structure in our model, although in theory this is straightforward to take into account.

[8] 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.

B.3 Logarithmic

The logarithmic density (see, for example, Johnson, Kotz, and Kemp (1992, Ch. 7))

$$\Pr(Y_t = y_t) = \kappa \frac{p^{y_t}}{y_t}, \qquad p \in (0,1), \qquad \kappa = \frac{-1}{\log(1-p)}, \qquad y_t = 1, 2, 3, ...$$

is reasonably unfamiliar to most econometricians and so we will review a few of its basic results as we will use it extensively later to model the size of price changes[9]. The mean and variance are given by

$$\mathrm{E}Y_t = \frac{\kappa p}{1-p}, \quad \mathrm{Var}Y_t = \kappa p \frac{(1-\kappa p)}{(1-p)^2} = \frac{\kappa p}{(1-p)^2} - (\mathrm{E}Y_t)^2 .$$

Hence

$$CV = \frac{\mathrm{Var}Y_t}{(\mathrm{E}Y_t)^2} = \frac{1}{\kappa p} - 1 = \frac{\log(1-p)}{-p} - 1.$$

The shape of tails of the density are best seen on the log-scale for

$$\log\left\{\Pr(Y_t = y_t)\right\} = \log\left\{\frac{-1}{\log(1-p)}\right\} + y_t \log p - \log y_t,$$

which is of the same type as that of the generalized hyperbolic distribution, which has shown very capable of fitting log returns of financial data, see e.g. Rydberg (1997) and Rydberg (1999). In particular, when p is close to one, then $-\log y_t$ will dominate. Unfortunately, having p even close to one is very unlikely for our applications.

We are going to use the logarithmic distribution as a model for the size of change, i.e. as a model for the variable Y_t. However, we are going to introduce dependency in the parameter p by letting it be a logistic transform of past history, in other words, let,

$$p(\theta_t) = \exp(\theta_t)/\left\{1 + \exp(\theta_t)\right\}$$

where θ_t again depends linearly on some transformation of past history. A simple example of this is where $\theta_t = \mu + \beta y_{t-1}$[10]. By introducing past dependence in the parameter we mix different logarithmic distributions with different tail behaviour. To give an idea of the tail behaviour of the logarithmic density, see Figure 10 which shows four scenarios.

[9] This framework can be generalised to the (positive) power series family (see, for example, Johnson, Kotz, and Kemp (1992, Section 2.2)) where

$$\Pr(Y_t = y_t) \propto a_{y_t} \theta^{y_t}, \qquad y_t = 1, 2, ...$$

(The standard setup typically allows $y_t = 0$, but we exclude this here — but this is a trivial extension). Special cases of this are the Poisson and negative binomial distributions.

[10] This type of univariate autoregression does have some worrying features. If we were to maintain exactly this structure but replace the logarithmic distribution by its cousin the Poisson distribution, then this autoregression is strictly non-stationary if $\beta > 0$. Hence one might anticipate the same result carrying over to this situation. There seem two responses to this: — take a transform of lagged y_t (e.g. $\log y_t$) or not worry as in practice as this model will in practice have a more complicated structure which imposes stationarity in other features of the model.

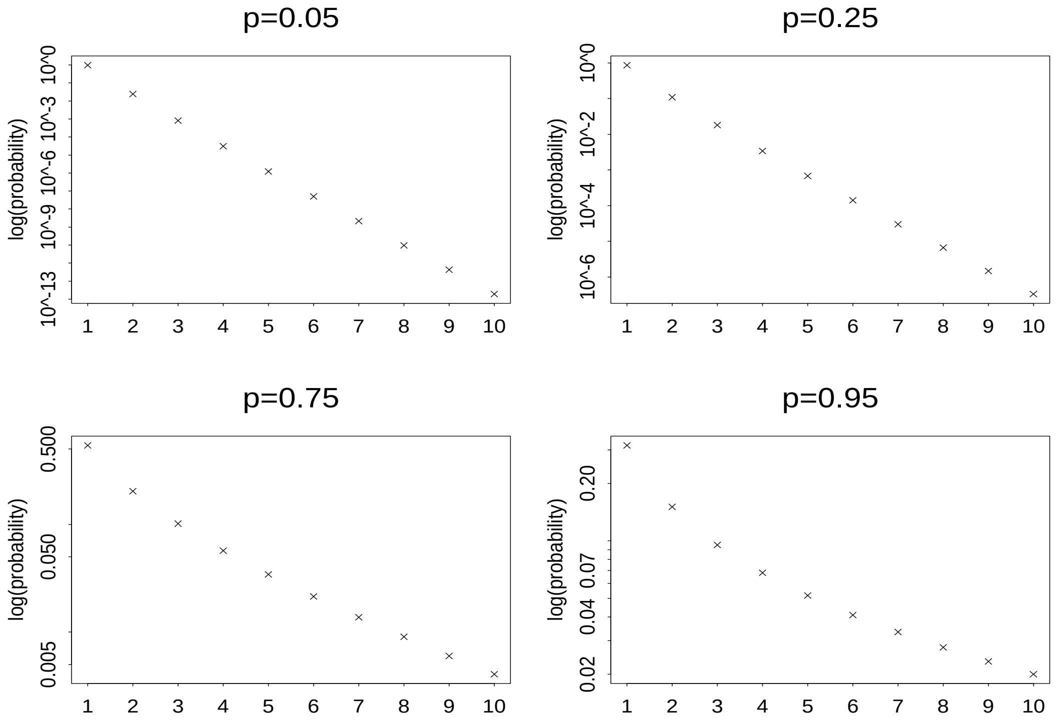

Figure 10: *Plot of the log probabilities for the logarithmic distribution for 4 different parameter values.*

B.4 Three straightforward alternatives

B.4.1 Geometric

The geometric density Johnson, Kotz, and Kemp (1992, Ch. 5))

$$\Pr(Y_t = y_t) = (1-p)p^{y_t - 1}, \qquad p \in (0,1), \qquad y_t = 1, 2, 3, ...$$

has the simple interpretation as the discrete density with a constant hazard rate.

$$\mathrm{E}Y_t = \frac{1}{1-p}, \quad \mathrm{Var}Y_t = \frac{2p}{(1-p)^2} - (\mathrm{E}Y_t)^2 = \frac{2p-1}{(1-p)^2}$$

It generally has thinner tales than the logarithmic. We can make it into an autoregressive model using the usual trick of setting $\theta_t = \mu + \beta y_{t-1}$.

B.4.2 Non-zero Poisson

The non-zero Poisson has

$$\Pr(Y_t = y_t) = \frac{1}{1-e^{-\lambda}} \frac{e^{-\lambda}\lambda^{y_t}}{y_t!}, \qquad \lambda > 0, \qquad y_t = 1, 2, 3, ...$$

which is the standard Poisson model with the zero point of support removed.

$$\mathrm{E}Y_t = \frac{\lambda}{1-e^{-\lambda}}, \qquad \mathrm{Var}Y_t = \frac{\lambda(1+\lambda)}{1-e^{-\lambda}} - (\mathrm{E}Y_t)^2 .$$

A standard way of allowing λ to vary over time is to write it as $\lambda(\theta_t) = \exp(\theta_t)$ where θ_t is a linear function of transforms of the relevant filtration. A simple example of this is where $\theta_t = \mu + \beta \log y_{t-1}$.

B.4.3 Two parameter mixture model

A more flexible model can be obtained by mixing a geometric distribution with a beta. In particular define

$$\begin{aligned} f(Y_t &= y_t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 (1-p)p^{y_t-1}p^{\alpha-1}(1-p)^{\beta-1}dp \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y_t-1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+y_t)}, \qquad \alpha,\beta>0, \qquad y_t=1,2,.... \end{aligned}$$

By selecting α, β we can separately determine the centrality and scale of this distribution. However, as we have mixed with a geometric, it must have its mode at one. This seems a natural assumption in our setup.

References

Azzalini, A. (1983). Maximum likelihood estimation of order M for stationary stochastic processes. *Biometrika 70*, 381–387.

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

Barndorff-Nielsen, O. E. and N. Shephard (1998). Incorporation of a leverage effect in a stochastic volatility model. Unpublished discussion paper: Nuffield College, Oxford.

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

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.

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.

Geweke, J. (1986). Modelling the persistence of conditional variances: a comment. *Econometric Reviews 5*, 57–61.

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.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev. Economic Studies 65*, 361–93.

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.

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. (1997). The normal inverse Gaussian Lévy process: simulation and approximation. *Communications in Statistics: Stochastic Models 13*, 887–910.

Rydberg, T. H. (1999). Generalized hyperbolic diffusions with applications towards finance. *Mathematical Finance 9*. Forthcoming.

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.

StataCorp (1997). *Stata Statistical Software: Release 5.* College Station, Texas: Stata Corporation.