Dynamics of Trade-by-Trade Price Movements: Decomposition and Models

TINA HVIID RYDBERG
BNP Paribas, London

NEIL SHEPHARD
Nuffield College, University of Oxford

ABSTRACT

In this article 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 market microstructure using time series of transaction data. We use maximum likelihood estimation and testing methods to assess the fit of the model to one year of IBM stock price data taken from the New York Stock Exchange.

KEYWORDS: activity, decomposition, directions, GLARMA, size, transactions data.

Let $p(u)$ denote the price at time u of the most recent transaction of an asset. Here u is a continuous clock, but prices are only updated when a trade actually occurs. By construction the price can be written as

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

Tina Rydberg thanks the Danish National Research Council for its financial support through a postdoctoral fellowship. Neil Shephard thanks the ESRC for financial support through grant R00023839. Both authors are grateful for support from the Centre for Analytical Finance, Aarhus, Denmark. The authors are grateful for comments from participants at the conference on "Econometrics and Financial Time Series" at the Isaac Newton Institute, Cambridge University, October 12–16, 1998, where the details of our decomposition were first presented. We thank Richard Spady, Rob Engle, Frank Gerhard, Clive Bowsher, and Jeff Russell for various helpful conversations on these issues. The computations were carried out using software written in Ox [Doornik (2001)]. The research reported here was carried out before Tina Rydberg joined BNP Paribas. The views expressed here do not necessarily correspond to those of BNP Paribas. Address correspondence to Neil Shephard, Nuffield College, New Road, OX1 1NF, UK, or e-mail: neil.shephard@nuf.ox.ac.uk.

Journal of Financial Econometrics, Vol. 1, No. 1, pp. 2–25
DOI: 10.1093/jjfinec/nbg002

where $N(u)$ is the number of trades recorded in the interval from time 0 to time u and Z_t is the price movement associated with the *t–th* trade. In practice, many of the Z_t are exactly zero. Rogers and Zane (1998) and Rydberg and Shephard (2000) use this framework to study the evolution of transaction prices. They model $N(u)$ as a counting process, with new arrivals being generated by a Cox process, that is, a Poisson process with a random intensity. They are unspecific about the Z_t process beyond the use of simple descriptive Markov chains. Some of the econometric issues that arise with unequally spaced financial data are discussed at length in Engle and Russell (1998, 2002) and Engle (2000).

In this article we model the joint distribution of price movements Z_t, focusing on the econometric problem that the price movements are restricted to take on integer multiples of a smallest nonzero price change, that is, a tick. The tick size depends on the institutional setting. When normed, this means price movements can be thought of as being integers. To start with we will model the Z_t as being dependent only on itself. The problem of modeling using larger filtrations is dealt with later in the article. Then let $Z_t \in I$ be an integer process and $\mathcal{F}_t = \sigma(Z_s : s \leq t)$ be its natural filtration. We are primarily interested in the joint distribution of the movements which are given by

$$\Pr(Z_1, \ldots, Z_n | \mathcal{F}_0) = \prod_{t=1}^{n} \Pr(Z_t | \mathcal{F}_{t-1}), \tag{2}$$

using a prediction decomposition. The focus will be on specifying $\Pr(Z_t|\mathcal{F}_{t-1})$.

The main contribution of this article is to suggest decomposing Z_t into three components termed "activity," "direction" and "size." The variables measure, respectively, (i) if the price moved, (ii) which direction it moved, and (iii) how far it moved. Our claim is that this structure allows a relatively simple econometric analysis of sequential price movements.

A body of literature already exists on the modeling of trade-by-trade price dynamics. Russell and Engle (1998) suggest using a conditional multinomial model for specifying $\Pr(Z_t|\mathcal{F}_{t-1})$, in a sense generalizing previous work by Hausman, Lo, and MacKinlay (1992) on probit models for transaction data. Hasbrouck (1999) builds a class of dynamic latent variable models for efficient prices and traders' cost which uses an economically motivated truncation to force prices to live on a lattice. The general issue of discreteness of asset prices is discussed at length in Campbell, Lo, and MacKinlay (1997: 98–144).

Other work which relates to this topic includes articles by Ghysels and Jasiak (1998), Meddahi, Renault, and Werker (1998), and Manganelli (2001), who combine GARCH models for price returns with ACD-style models for the times between trades. In this context this work has the disadvantage that the price process does not live on the required lattice observed in the data. The article by Darolles, Gourieroux, and Le Fol (2000), which we first saw after the circulation of the first draft of this article, is much closer to the framework of Rogers and Zane (1998) and Rydberg and Shephard (2000). Darolles, Gourieroux, and Le Fol (2000) use the structure of Equation (1), but assume the Z_t process is Markov living on the

points −1, 0, 1. This modeling assumption is combined with a reduction in the dataset by using quote data to allow them 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.

The article is organized as follows. Section 1 introduces the decomposition of the price movements. Section 2 presents the empirical results for the component models for the activity, direction, and size of price movements—together these models give an overall model of price movements. Section 3 places our suggestion in the context of the literature and suggests various extensions of the modeling approach. Section 4 concludes.

1 DECOMPOSITION OF PRICE MOVEMENTS

Potentially the distribution of $Z_t|\mathcal{F}_{t-1}$ 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

$$Z_t = A_t D_t S_t. \tag{3}$$

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:

$$D_t = -1,1 \quad \text{and} \quad S_t = 1,2,\ldots. \tag{4}$$

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 *inactive*. If $A_t = 1$ then there are *active* price movements. The nonzero 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 upward, otherwise it moves downward. Finally, S_t controls the *size* of price movements. This suggests the decomposition of price movements into

$$\Pr(Z_t = 0|\mathcal{F}_{t-1}) = \Pr(A_t = 0|\mathcal{F}_{t-1}), \tag{5}$$

while for $z_t \neq 0$,

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

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

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

- $\Pr(D_t|\mathcal{F}_{t-1}, A_t = 1)$—another binary process on $\{-1, 1\}$ modeling 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 modeling 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 appears difficult, we will see that our estimated specifications will have very simple interpretable structures which do 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: $\mathcal{F}_t^A = \sigma(A_s : s \le t)$, $\mathcal{F}_t^D = \sigma(D_s : s \le t)$, and $\mathcal{F}_t^S = \sigma(S_s : s \le t)$. Of course, $\mathcal{F}_t = \mathcal{F}_t^{A,D,S}$.

Finally, before we detail the modeling 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.

2 PRELIMINARY MODELS FOR THE COMPONENTS

2.1 The Data

To start our empirical modeling 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 article are for IBM stock recorded electronically at the New York Stock Exchange (NYSE) 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 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 4:00 P.M., 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 after 4:00 P.M..

For this article we constructed a single series by concatenating each of the above series (whose overnight effects were removed by deleting the action of removing the first 15 minutes of the day) for individual days. The size of the total dataset when all exchanges in the United States are considered is 413,906, this is too much data to initially handle and therefore we have limited our analysis to the trades performed at the NYSE (trades coded with an N). We have also deleted all trades that have an error code. This leaves us with a total of 173,146 observations to model. Of these, 33,184 are nonzero (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. Throughout we ignore the availability of quote data.

2.2 The Activity of Prices

2.2.1 Autologistic model Our initial parametric model for $\Pr(A_t|\mathcal{F}_{t-1})$ will be an autologistic model based on $\mathcal{F}_t$. Recall that for an autologistic we write

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

and

$$\theta_t^A = x_t'\beta + g_t, \quad \text{where } g_t = \sum_{j=1}^{p} \beta_j A_{t-j}, \tag{6}$$

with x_t being potential combinations and subsets of $\mathcal{F}_{t-1}$. This model structure was introduced by Cox (1958) [see Cox and Snell (1989) for an exposition] and has some significant advantages.[1] The log-likelihood for the autologistic is concave and so numerical optimization 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, e.g., 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 (A_t, D_t, and S_t) to enter the model. In practice, in order to reduce multicolinearity, it makes sense to transform the size variable S_t into

$$L_t = S_t - A_t,$$

which is zero unless S_t is bigger than one. We call L_t a large-move variable. After the model is fitted we will look at a portmanteau test to see its ability to capture the main features of the data. This is based on the residuals

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

which should be uncorrelated with zero mean and unit conditional (and unconditional) variance. The $\{u_t\}$ are then used inside a Box–Pierce statistic as a measure of residual dependence.

---

[1] A standard latent variable interpretation of these models is written as

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

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

Table 1 Autologistic model for activity

Variable	Coefficient	Std. Err.	Variable	Coefficient	Std. 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
L_{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
Constant	1.958	0.012	A_{t-20}	0.053	0.016
Q	$T\sum_{j=1}^{Q} r_j^2$			Log-likelihood = −82,313	
20	23.20	(31.41)			
100	698.4	(124.3)			
1500	5489	(1591)			

Many variables were tested. Reported model is what was left. Std. Err. denotes the standard deviation of the parameter estimates computed using likelihood theory assuming the model is correct. The r_j denotes the series correlation coefficient at lag j for the standardized residuals u_t. The figures in brackets correspond to 95 percentage points on the χ^2_Q distribution.

The results of the model fitting and diagnostic checking are shown in Table 1. At lag 2, L_t 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. To put it differently, this means that large movements are associated with subsequent high volatility. The direction variables are negative, which suggests that past decreases in prices tend to increase the chance that there will be a future movement in the market. This seems close to the leverage effect which is emphasized 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 and 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, Table 1 shows that the coefficients in front of the activity variables decay down—starting at 0.6 at lag 1 and falling to 0.2 at lag 3. 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 thousands of lags in the activity variable. In order to model this dependence parsimoniously we have to move away from autologistic models and into constructions which allow moving average-type behavior. This can be carried out by introducing GLARMA-type models.

2.2.2 GLARMA binary model We could generalize the autologistic structure of Equation (2) by allowing

$$g_t = \sum_{j=1}^{p} \gamma_j g_{t-j} + \sum_{j=1}^{q} \delta_j A_{t-j}, \tag{7}$$

but this is typically numerically unstable and thus 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 favor here puts

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

where

$$v_t = \frac{\{A_{t-1} - p(\theta_{t-1}^A)\}}{\sqrt{p(\theta_{t-1}^A)\{1 - p(\theta_{t-1}^A)\}}}. \tag{8}$$

Of importance is that $\{v_t\}$ 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 follows as $\{v_t\}$ has a zero mean and unit conditional and unconditional variance. This holds, whatever the process that generates the regressors. 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 invertibility 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. Further, following the initial draft of this article, Streett (2000) has shown that the above model has a unique stationary distribution.

In our numerical work we enforce covariance stationarity and invertibility constraints on the GLARMA representation of $\{g_t\}$. This is carried out by parameterizing the model in terms of the partial autocorrelations

$$\{\rho_j, j = 1, \ldots, p\}$$

and the inverse partial autocorrelations

$$\{\bar{\rho}_j, j = 1, \ldots, q\} \tag{9}$$

[e.g., Barndorff-Nielsen and Schou (1973) and Jones (1987)]. In this article we will report fitted empirical models using the ρ_j and $\bar{\rho_j}$ parameterization, as these have a simpler interpretation than the original γ_j and δ_j. In particular, a necessary and

Table 2 Estimation for activity using a GLARMA model parameterized in terms of the partial autocorrelations ρ_j and the inverse partial autocorrelations $\bar{\rho_j}$

Variable	Coefficient	Std. 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)	
L_{t-1}	−0.198	(.085)	
L_{t-2}	0.231	(.095)	
Q	$T\sum_{j=1}^{Q} r_j^2$		Log-likelihood = −81,927
20	21.88	(31.41)	
100	92.9	(124.3)	
1500	1447	(1591)	

Parameter estimates of the fitted GLARMA model, using a maximum likelihood criteria and the error term given by Equation (3). The figures in brackets are the standard errors of the regressors computed using the GLARMA model. Only the results obtained after model simplification are reported. Model order is selected using AIC. The r_j denotes the series correlation coefficient at lag j for the standardized residuals v_t. The figures in brackets correspond to 95 percentage points on the χ^2_Q distribution.

sufficient condition for covariance stationarity is that for all j, $|\rho_j| < 1$, while invertibility needs $|\bar{\rho_j}| < 1$ for all j. Throughout, the likelihood is maximized using analytic first derivatives and the BHHH algorithm.[2]

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) large moves have not changed very much. Table 2 shows that both L_{t-1} and L_{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 parameterization 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 nonstationary. More realistically, we need a more intricate model of activity which takes into account

[2] A plain BHHH method was used mapping the real variables being maximized into partial autocorrelations and inverse partial autocorrelations using the transform $x/(1 + |x|)$. No interventions or numerical problems were encountered. The prediction decomposition of the GLARMA likelihood function is initialized by setting $g_0, \ldots, g_{-p+1}, v_0, \ldots, v_{-q+1}$ to zero.

intraday, intraweek, and month effects on the series. Work on this topic is reported in the next section.

2.3 The Direction of Price Changes

An important feature of our decomposition is that we are now able to focus on a model of the direction of the price changes, given that the price has changed: $\Pr(D_t|\mathcal{F}_{t-1}, A_t = 1)$.

Again we will use an autologistic model, but this time the outcome variable will live on the support $\{-1, 1\}$, rather than $\{0, 1\}$. After testing out insignificant explanatory variables we end up with directions and large direction as the only information of significance, where large direction is given by

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

Furthermore, let

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

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 this measurement of time to be extremely significant statistically.

Then D_{T_t} will be a vector of the last k price changes different from zero. We refer to the ith element of this vector as $D_{T_t,i}$, which is the ith 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 zero.

This gives us an autologistic model for

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

and so

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

By going from the general model to the specific we find that direction has only a 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.

Table 3 Estimation results for the direction of active trade using an autologistic model

Variable	Estimate	Std. Err.	Variable	Estimate	Std. Err.
D_{t-1}	−2.192	.043	LD_{t-1}	0.629	.180
D_{t-2}	−0.672	.033	LD_{t-2}	−0.506	.160
D_{t-4}	0.296	.030	LD_{t-3}	−0.837	.200
D_{t-5}	0.395	.033	LD_{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	Constant	−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-likelihood = −17,947
20	39.25	(31.41)	
100	101.4	(124.3)	
1500	1510	(1591)	

The figures in the column marked Std. Err. are the standard errors on the regressors computed using the autologistic model. The r_j denotes the series correlation coefficient at lag j for the standardized residuals u_t. The figures in brackets correspond to 95 percentage points on the χ^2_Q distribution.

The lagged values of D_t and LD_t seem reasonably straightforward. If the price moved on the last trade then there is a large chance that this movement will be reversed 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 0.5.

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. This means they work in activity time not in trading time. They suggest that the sign of the last active trade has a sustained 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 nonreversal.

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 tends 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 i.i.d. Bernoulli—an implication of the model suggested by Rogers and Zane (1998)] the log-likelihood is −22,966 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 $\{D_{T_t,i}, i = 1, 2, \ldots, 10\}$, then we have a log-likelihood of −20,416 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 $\{D_{t-i}, i = 1, 2, \ldots, 20\}$, which is modeling directions using data ordered in transaction time rather than activity time. This has a quite high log-likelihood of −18,419, 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 for 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—for a review of empirical work on this topic, see Campbell, Lo, and MacKinlay (1997: 99–107). This means people buying shares from market makers have to pay higher prices for them than those selling them to market makers [an elegant model of bid and ask dynamics is given by Hasbrouck (1999)]. 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 direction sequences. This means that changes in the traded price are almost certainly reversed.

2.4 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 article, 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 large movements $S_t - 1$. Recall the negative binomial (NegBin) is a generalization of the Poisson, allowing over-dispersion [see, e.g., Johnson, Kotz, and Kemp (1992: 204–205)]. The model will have

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

Table 4 Estimation for excess price movements using a NegBin-based GLARMA model parameterized in terms of the partial autocorrelations ρ_j and the inverse partial autocorrelations $\bar{\rho_j}$

Variable	Coefficient	Std. Err.	Variable	Coefficient	Std. Err.
Constant	−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)
$D_{(T_t,1)} - 5$	−0.121	(.067)	$D_{(T_t,1)} - 6$	−0.206	(.064)
α	0.0713				
Q	$T\sum_{j=1}^{Q} r_j^2$			Log-likelihood = −1478	
20	31.88	(31.41)	$E(v_t) = -.001$		
100	85.39	(124.3)	$\mathrm{Var}(v_t) = 1.207$		
1500	1743	(1591)			

Std. Err. denotes the standard deviation. The figures in brackets are the standard errors on the regressors computed using the GLARMA model. Model order selected using AIC. The r_j denotes the series correlation coefficient at lag j for the standardized residuals u_t. The figures in brackets correspond to 95 percentage points on the χ^2_Q distribution.

implying

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

Notice that μ_t will be typically very small and so ${\mu_t}^2$ 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

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

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

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

as our basic parametric model.[3] The NegBin distribution was selected because it is simple and familiar.

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

[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 methodologic viewpoint this raises no new issues.

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. First, big price movements tend to be preceded by nonsymmetric decreases in price (directions being negative). This is the familiar dynamic leverage effect [see Nelson (1991)], and so its presence is not surprising.

The second form of nonsymmetry is a contemporaneous one and is simply dependent on the significance of D_t (not its lags). The negative sign associated with it suggests big decreases are more common than big increases. 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 upward over time due to the predominance of small positive movements (over small negative movements), but tends to fall back sometimes with quite large decreases.

The above nonsymmetry 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 large movements (movements of more than one tick) downward are, on average, around twice as large as large movements upward.

Judging from Table 4, the NegBin-based GLARMA model fails slightly, as the variance of the residuals is slightly larger 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 condition on a wider information set and so 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 the GLARMA(2,1) model. This has a log-likelihood of -1502 and Box–Pierce of 16.6. Thus we can see that it is the time-series modeling aspect of this particular model which dominates the fitting of this series.

3 COMMENTS

3.1 Predictive Distributions

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

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

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

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

For s very large, the multistep ahead forecast distribution of price movements will approach the unconditional distribution of our fitted model.[4] Although this feature is of little economic meaning, it can be a useful diagnostic check on the fitted model.

3.1.3 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 that 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, this can be carried out for very large values of N (in the simulations discussed below $N = 10{,}000$) even if s is large.

Table 5 shows the center of the two-step ahead forecast distribution of price moves based on different histories. The histories are made as simple as possible. We assume that (i) trading has not been moving the prices for some time, that is, $\mathcal{F}^A = (1, 1, 0, \ldots, 0)$; and (ii) the last two trades before we forecast both moved the prices, that is, $\mathcal{F}^D = (?, ?, -1, 1, -1)$, where we will replace ?, ? by moves of one tick in various directions. We then tabulate the forecast distribution over all possible one-tick price changes in the last two periods. Column 1 shows the impact of two down movements having just been observed and this is seen to give an increased probability for moving one tick up. Column 4 has the opposite observation, namely two up movements. Here we have the opposite result that the probability of moving down is increased. The middle two columns correspond to "bid-ask bouncing." This decreases the probability of moving away from the current price level. From all the columns it can be seen that the predominant behavior is mean reversion of one tick size and that the last directions are impor-

[4] We should note that this conclusion is based on an assumption that the Z_t process is stationary, which we have not proven.

Table 5 The effect on the predictive distribution of future price movements of conditioning on particular past events

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

An example of this is where the last two trades each resulted in a one tick move downward. This is denoted by $\{-1, -1\}$. More generally it shows the simulated conditional probabilities for having moved x ticks, after 2 or 10 trades, given the history. $\mathcal{F}^A = (1, 1, 0, \ldots, 0)$ and $\mathcal{F}^D = (?, ?, -1, 1, -1)$. ?, ? is given in the top row of the table. The estimated probabilities are based on $N = 10{,}000$ simulations.

tant 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 tick down (up) with probability 0.990 (0.991) and at least two ticks down (up) with probability 0.707 (0.756).

In Table 5 we also give the 10-step ahead forecast. In this case we see that after a movement of two one-ticks down (up), the price after 10 trades will be at least two ticks down (up) with probability 0.779 (0.809).

3.2 Previous Work

3.2.1 Conditional multinomial models In a recent article, Russell and Engle (1998) suggested modeling price movements using a conditional multinomial distribution. Their article can be viewed as a time-series extension of a 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 if it is -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 on some features of the filtration of Y_t, $\mathcal{F}_t^Y = \sigma(Y_s, s \leq t)$.

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)]:

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

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

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

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 generalized 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 Equation (3). In particular, if they define $v_t = (v_{1t}, v_{2t}, v_{3t}, v_{4t})'$ as

$$v_{it} = \frac{I(Y_t = i) - p_{it}}{\sqrt{p_{it}(1 - p_{it})}}, \tag{10}$$

and

$$\theta_t^* = \alpha^* + g_t, \tag{11}$$

then the system would be modeled as

$$g_t = \sum_{j=1}^{p}\gamma_j g_{t-j} + \sigma v_t + \sigma\sum_{j=1}^{q}\delta_j v_{t-j}, \tag{12}$$

where α^* is a vector, while $\{\gamma_j\}, \sigma$ 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 purposes of 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 we use is very different. Our main advantages are parsimony, easy interpretability via the decomposition, and options for extensions.

3.2.2 Hasbrouck's truncation model Hasbrouck (1999) introduced a dynamic model for the evolution of bid and ask prices 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 articles include Bollerslev and Melvin (1994) and Harris (1994).

The Hasbrouck (1999) 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.

3.3 Conditioning on Signs

In recent work carried out independently from our own, Granger (1998) emphasized the potential importance of modeling separately the direction (sign) and the size of stochastic processes. Typically he models these two variables independently, while we emphasize the sequential nature of our decomposition, which is empirically vital for our problem and more general. Abstracting from that detail, we see that our analysis can be viewed within his framework when the activity, A_t, is conditioned to be one.

3.4 Explanatory Variables

3.4.1 Deterministic seasonality It may be 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, e.g., Rydberg and Shephard (2000)]. It is a straightforward task to include this information within our model, allowing seasonality to influence any or all of the submodels for activity, direction, and size. No new issues are raised by this, and we will give empirical results on this in a later subsection.

3.4.2 Exogenous variables Our modeling 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

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

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 of attention, as it allows lagged information to improve the predictions of future price movements given the history of the Y_t process.

Table 6 Estimation for activity using a binary GLARMA model parameterized in terms of the partial autocorrelations ρ_j and the inverse partial autocorrelations $\bar{\rho_j}$

Variable	Estimate	Std. Err.	Variable	Estimate	Std. Err.
Constant	−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)
L_{t-1}	−0.219	(.085)	$\log(\text{dur})_{t-5}$	−0.017	(.005)
L_{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-likelihood = −81,839	
20	19.06	(31.41)			
100	87.2	(124.3)			
1500	1437	(1591)			

Estimation included lagged log(duration + 1) and log(volume) (which tests out). Std. Err. denotes the standard deviation.

This 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: chap. 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 standardized 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. Notice that in Tables 6 and 7 (as well as other tables given below), many of these fixed explanatory variables do not appear, as they were insignificant.

The empirical model for directions, reported in Table 7, is interesting because it completely tests out the effect of duration on the prediction of direction, 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 expect them to be influenced by other activity series. The *range* variable is a trend variable, which we interpret as a monthly seasonal variable rather than a typical trend, as we only have a year of

Table 7 Estimation for the direction including lagged durations (which test out) and volume

Exp. Var.	Estimate	Std. Err.	Exp. Var.	Estimate	Std. Err.
D_{t-1}	−2.192	.043	LD_{t-1}	0.620	.180
D_{t-2}	−0.671	.033	LD_{t-2}	−0.506	.160
D_{t-4}	0.298	.030	LD_{t-3}	−0.851	.200
D_{t-5}	0.395	.033	LD_{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	Constant	−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-likelihood = −17,938
20	47.78	(31.41)	
100	110.4	(124.3)	
1500	1513	(1591)	

Improvement in the log-likelihood for introducing these variables is 9. The figures in the column marked Std. Err. are the standard errors on the regressors computed using the autologistic model. The r_j denotes the series correlation coefficient at lag j for the standardized residuals u_t. The figures in brackets correspond to 95 percentage points on the χ^2_Q distribution.

data. We used this *range* variable rather than a full set of monthly seasonal variables 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 we report the estimated activity and direction processes using contemporaneous 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 is quite large. Activity is affected positively 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 is high the same thing happens.

Table 8 Estimation for activity using a binary GLARMA model parameterized in terms of the partial autocorrelations ρ_j and the inverse partial autocorrelations $\bar{\rho_j}$

Variable	Estimate	Std. Err.	Variable	Estimate	Std. Err.
Constant	−1.460	0.045	Range	0.157	0.028
ρ_1	0.9998		LD_{t-1}	−0.215	0.086
ρ_2	−0.624		LD_{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-likelihood $= -79{,}022$	
20	39.07	(31.41)			
100	115.0	(124.3)			
1500	1490	(1591)			

The fitted model includes contemporaneous and lagged log(duration +1) and log(volume). Std. Err. denotes the standard deviation. Again, the fitted model reports only effects which are not tested out.

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

The estimated NegBin-based GLARMA model for the large variable $S_t - 1$ is reported in Table 10 using lagged durations and volume. The AIC measure selected a GLARMA(2,1) structure. Of interest is that 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 greater impact, with longer lags than that given in the tables having positive but insignificant effects on the large variable. Of importance is that the *range* and *quadr* variables are taken as estimating the seasonal component of the process. This is significant, which is unsurprising given that 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 explanatory variables.

Overall lagged explanatory variables improve the likelihood function by around 24, which is modest given we have included 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

Table 9 Estimation for the direction including lagged and contemporaneous durations and volume

Variable	Estimate	Std. Err.	Variable	Estimate	Std. Err.
D_{t-1}	−2.173	.043	LD_{t-1}	0.638	.181
D_{t-2}	−0.663	.033	LD_{t-2}	−0.490	.161
D_{t-4}	0.298	.030	LD_{t-3}	−0.830	.199
D_{t-5}	0.395	.033	LD_{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	Constant	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-likelihood = −17,837
20	47.76	(31.41)	
100	112.3	(124.3)	
1500	1540	(1591)	

Improvement in the log-likelihood for introducing the contemporaneous variables is 101. The figures in the columns marked Std. Err. are the standard errors on the regressors computed using the autologistic model. The r_j denotes the series correlation coefficient at lag j for the standardized residuals u_t. The figures in brackets correspond to 95 percentage points on the χ^2_Q distribution.

overdispersion, suggesting some improvement could 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 effect on the large movement 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 large movement variable. Of interest is that 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 toward the end of the year. However, this effect is not very significant.

Table 10 Estimation for excess price movements (S_{t-1}) using a NegBin GLARMA model parameterized in terms of the partial autocorrelations ρ_j and the inverse partial autocorrelations $\bar{\rho_j}$

Variable	Coefficient	Std. Err.	Variable	Coefficient	Std. Err.
Constant	−5.546	(.152)	σ	0.175	
Tuesday	0.683	(.184)	Wednesday	0.489	(.222)
Range	0.391	(.124)	Quadratic	−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)			
$D_{(T_t,1)} - 6$	−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-likelihood = −1,454	
20	27.05	(31.41)	$E(v_t) = -.002$		
100	119.2	(124.3)	$\text{Var}(v_t) = 1.166$		
1500	1393	(1591)			

Estimation includes lagged log(duration + 1) and the log(volume). Std. Err. denotes the standard deviation. The figures in brackets are the standard errors on the regressors computed using the GLARMA model. Model order selected using AIC. The r_j denotes the series correlation coefficient at lag j for the the standardized residuals v_t. The figures in brackets correspond to 95 percentage points on the χ^2_Q distribution.

Table 11 Estimation for excess price movements (S_{t-1}) using a NegBin GLARMA model parameterized in terms of the partial autocorrelations ρ_j and the inverse partial autocorrelations $\bar{\rho_j}$

Variable	Coefficient	Std. Err.	Variable	Coefficient	Std. Err.
Constant	−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)
$\sum_{j=1}^{6} D_{(T_t,1)-j}$	−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)	$\sum_{j=3}^{5} log(vol)_{t-j}$	0.062	(.023)
Q	$T\sum_{j=1}^{Q} r_j^2$			Log-likelihood = −1,355	
20	19.21	(31.41)	$E(v_t) = -.001$		
100	111.6	(124.3)	$\text{Var}(v_t) = 1.171$		
1500	1587	(1591)			

Estimation includes contemporaneous and lagged log(duration + 1) and log(volume). Std. Err. denotes the standard deviation. The figures in brackets are the standard errors on the regressors computed using the GLARMA model. Model order selected using AIC. The r_j denotes the series correlation coefficient at lag j for the the standardised residuals u_t. The figures in brackets correspond to 95 percentage points on the χ^2_Q distribution.

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

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 extent of overdispersion in the fitted model is modest.

4 CONCLUSION

In this article we proposed a decomposition of the price movements of trade-by-trade datasets. The decomposition means we have to sequentially model price activity, direction of moves, and size of moves. Each modeling exercise is straightforward and interpretable. A number of extensions of the modeling 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) modeling of two or more asset prices simultaneously, and (ii) using trade data on the same stock but collected on different exchanges.

Received June 14, 2001; revised January 10, 2002; accepted May 9, 2002

REFERENCES

Barndorff-Nielsen, O. E., and G. Schou. (1973). "On the Reparameterization of Autoregressive Models by Partial Autocorrelations." *Journal of 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*, 355–372.

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

Cox, D. R. (1958). "The regression analysis of binary sequences (with discussion)." *Journal of the Royal Statistical Society, Series B 20*, 215–242.

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

Darolles, S., C. Gourieroux, and G. Le Fol. (2000). "Intra-day Transaction Price Dynamics." *Annales d'Economie et Statistique 60*, 207–238.

Doornik, J. A. (2001). *Ox: Object Oriented Matrix Programming, 3.0*. London: Timberlake Consultants Press.

Engle, R. F. (2000). "The Econometrics of Ultra-High Frequency Data." *Econometrica 68*, 1–22.

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.

Engle, R. F., and J. Russell. (2002). "High-Frequency and Transaction data." In Y. Ait-Sahalia and L. P. Hansen (eds.), *Handbook of Financial Econometrics*. Amsterdam: North-Holland.

Ghysels, E., and J. Jasiak. (1998). "GARCH for Irregularly Spaced Financial Data: the ACD-GARCH Model." *Studies in Nonlinear Dynamics and Econometrics 2*, 133–149.

Granger, C. W. J. (1998). "Comonotonicity." Lecture to the 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. (1999). "The Dynamics of Discrete Bid and Ask Quotes." *Journal of Finance 54*, 2109–2142.

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). *Dynamic Econometrics*. Oxford: Oxford University Press.

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

Jones, M. C. (1987). "Randomly Choosing Parameters for the Stationary and Invertibility Region of Autoregressive-Moving Average Models." *Applied Statistics 36*, 134–138.

Manganelli, S. (2001). "Duration, Volume and Volatility Impact of Trades." Unpublished paper, European Central Bank.

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

McFadden, D. L. (1984). "Qualitative Response Models." In Z. Griliches and M. Intriligator (eds.), *The Handbook of Econometrics*, vol. 2, Amsterdam: North-Holland.

Meddahi, N., E. Renault, and B. Werker. (1998). "Modeling 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.

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 the Workshop on Mathematical Finance, University of Bremen, Bremen Germany, February 1998.

Russell, J. R., and R. F. Engle. (1998). "Econometric Analysis of Discrete-Valued, Irregularly-Spaced Financial Transactions Data Using a New Autoregressive Conditional Multinomial Model." Unpublished paper, Graduate School of Business, University of Chicago. Presented at the Second International Conference on High Frequency Data in Finance, Zurich, Switzerland, April 1998.

Rydberg, T. H., and N. Shephard. (2000). "A Modeling Framework for the Prices and Times of Trades Made on the NYSE." In W. J. Fitzgerald, R. L. Smith, A. T. Walden, and P. C. Young (eds.), *Nonlinear and Nonstationary Signal Processing*, Cambridge: Isaac Newton Institute and Cambridge University Press.

Shephard, N. (1994). "Autoregressive Based Generalized Linear Models." Unpublished paper, Nuffield College, Oxford. Presented at the Econometric Society World Congress, Tokyo, August 1995.

Streett, S. (2000). "Some Observation Driven Models for Time Series." Unpublished PhD dissertation, Department of Statistics, Colorado State University.