BIN models for trade-by-trade data. Modelling the number of trades in a fixed interval of time.

Tina Hviid Rydberg
Paribas, London & Nuffield College, Oxford OX1 1NF, UK.
tina.rydberg@paribas.com

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

November 30, 1999

Abstract

In this paper we propose a simple time series model of the number of transactions made in intervals of length Δ seconds. We call this model the BIN model. The properties of the BIN model are evaluated while we explore connections between this model and Cox processes — that is Poisson processes with random intensities. We apply the modelling framework to data on trades in IBM shares.

Some keywords Aggregation, BIN model, Compound processes, Cox process, Trade-by-trade data.

All of the papers on trade-by-trade dynamics written by Rydberg/Shephard can be downloaded at

www.nuff.ox.ac.uk/users/rydberg/publications.htm

UNFINISHED PAPER.

1 INTRODUCTION

1.1 Background

In a recent paper, stimulated by the advent of trade-by-trade datasets recorded on commonly traded stocks, futures and derivatives, Rydberg and Shephard (2000) proposed to model an asset price $p(u)$ at time u using a compound Poisson process

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

where $\{N(u)\}_{u\geq 0}$ 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 (2000) specified $N(u)$ to be a counting process[1], modelled as a Cox process — that is a Poisson process with a random

[1] There are many equivalent definitions of a counting process. The one which is most helpful in our context states that if $\{N(u)\}_{u\geq 0}$ is a process with state space $\mathbf{Z}\cup\{+\infty\}$ and non–decreasing right continuous paths, then $\{N(u)\}_{u\geq 0}$ is a counting process. Since the paths are non–decreasing and right continuous we automatically get that $\{N(u)\}_{u\geq 0}$ is càdlàg (continu à droit – limite à gauche).

intensity. In general, the dynamics of the Cox and price movements processes can be jointly adapted to a wide class of filtrations involving just their own past or more extensive information sets (e.g. volume of transaction and an indicator of whether the initiator of the transaction is a buyer or a seller). 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.

From an economic viewpoint we are typically interested in comparing the rate of return on holding the asset with that obtainable by other risky investments or riskless interest rate bearing accounts. In order to do this we have to compute the return over a fixed length of time, which we write as $\Delta > 0$. Then these returns will be based around the difference

$$\begin{aligned} p_n &= p\left\{(n+1)\,\Delta-\right\} - p(n\Delta) \\ &= \sum_{t=1}^{N\{(n+1)\Delta-\}} Z_t - \sum_{t=1}^{N(n\Delta-)} Z_t \\ &= \sum_{t=N(n\Delta-)+1}^{N\{(n+1)\Delta-\}} Z_t. \end{aligned}$$

This shows that the number of trades in the interval $[n\Delta, (n+1)\,\Delta)$ plays a crucial role. To reflect this we will work with

$$N_n = N\left\{(n+1)\,\Delta-\right\} - N(n\Delta), \tag{2}$$

the number of trades in that time interval. This binning operation is illustrated in Figure 1, which displays the value of the $N(u)$ process against u. Binning partitions time into sections and we count the number of trades in that interval. We have to take care in dealing with the open and closed aspects of the interval, but this is straightforward to do. The important point to remember here is that if $N_n = 0$ then $p_n = 0$ immediately, while for $N_n > 0$ the prices can change. Hence N_n plays a vital role in determining the activity in the changes in the price level. For small values of Δ there will be a negligible loss in information in doing this compared to studying the complete record of the $\{N(u)\}$ process. Indeed if Δ equals one second, for the NYSE data there is no loss at all.

Example 1 *The very simplest example of this is where we suppose that the* $\{N_n\}$ *and* $\{\mathrm{Z_t}\}$ *processes are stochastically independent and covariance stationary. Further assume that the* $\{\mathrm{Z_t}\}$ *are independent and identically distributed. Then, writing* $F_{n\Delta-}$ *as the information about the* $\{N_n\}$ *sequence available infinitesimally before time* $n\Delta$*, assuming the moments exist*

$$\begin{aligned} Var(p_n|\mathcal{F}_{n\Delta-}) &= E\left\{Var(p_n|N_n)|\mathcal{F}_{n\Delta-}\right\} + Var\left\{E(p_n|N_n)|\mathcal{F}_{n\Delta-}\right\} \\ &= Var(Z_t)E(N_n|\mathcal{F}_{n\Delta-}) + E(Z_t)^2 Var\left(N_n|\mathcal{F}_{n\Delta-}\right) \end{aligned}$$

Hence predicting the variance of the price over the next period of length Δ *requires us to model the mean and variance of the future number of trades. Of course in practice* $E(Z_t)$ *will be tiny and so what matters in the above setup is really only* $E(N_n|F_{n\Delta-})$*. If we additionally set* $E(Z_t) = 0$ *then*

$$\begin{aligned} \mathrm{Cov}(p_n^2, p_{n+s}^2) &= \mathrm{E}\left\{\mathrm{Cov}\left(p_{n+s}^2, p_n^2|N_{n+s}, N_n\right)\right\} + \mathrm{Cov}\left\{\mathrm{Var}\left(p_{n+s}|N_{n+s}\right), \mathrm{Var}(p_n|N_n)\right\} \\ &= Var(Z_t)^2 \mathrm{Cov}\left(N_n, N_{n+s}\right). \end{aligned}$$

Hence we can obtain volatility clustering, with the autocorrelation of squared price changes being proportional to that of counts.

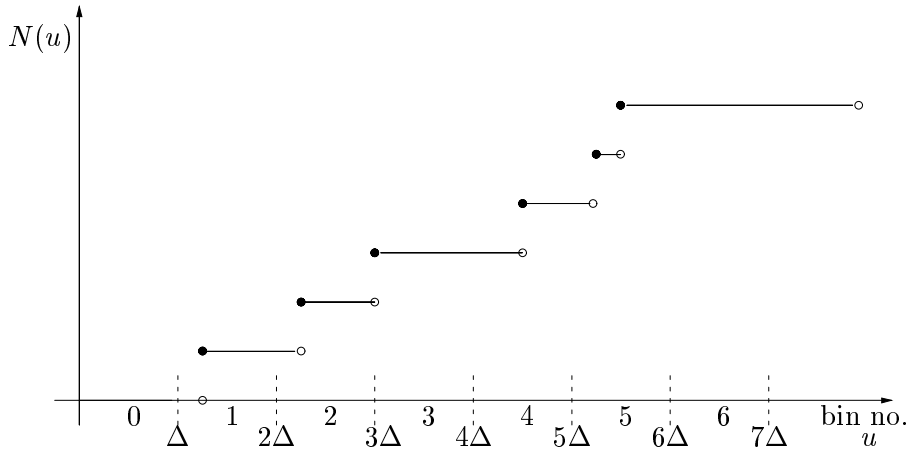

Figure 1: *Example of a counting process and bins of size* Δ. *Notice that with the definition of* N_n *given in (2) we have* $N_0 = 0, N_1 = 1, N_2 = 1, N_3 = 1, N_4 = 1, N_5 = 2, N_6 = 0$. *A "*$\circ$*" denotes a point where the value is not obtained, where as "*$\bullet$*" signifies that the value is obtained. This is important as the* $N(u)$ *process has discrete jumps.*

Example 2 *We suppose that the* $\{N_n\}$ *and* $\{\mathrm{Z_t}\}$ *processes are stochastically independent and covariance stationary processes. Also assume that* $\{\mathrm{Z_t}\}$ *has a zero mean and we write* $\rho(s) = Cor(Z_t, Z_{t+s})$. *Then*[2]

$$Var(p_n|\mathcal{F}_{n\Delta-}) = Var(Z_t)\left\{E\left(N_n|\mathcal{F}_{n\Delta-}\right) + 2\sum_{k=1}^{\infty}\Pr(N_n = k|\mathcal{F}_{n\Delta-})\sum_{j=1}^{k}(k-j)\,\rho(j)\right\}.$$

If Δ *is very small then* $\Pr(N_n = 1|\mathcal{F}_{n\Delta-}) \simeq E\left(N_n|\mathcal{F}_{n\Delta-}\right)$ *and*

$$Var(p_n|\mathcal{F}_{n\Delta-}) \simeq Var(Z_t)E\left(N_n|\mathcal{F}_{n\Delta-}\right),$$

which is the same results we achieved when the $\{\mathrm{Z_t}\}$ *were uncorrelated. A more specific result is obtained if we assume the* $\{Z_t\}$ *have a first order moving average representation (*Rydberg and Shephard (2000) for evidence of the empirical support for this*) for then*

$$Var(p_n|\mathcal{F}_{n\Delta-}) = Var(Z_t)\left[E\left(N_n|\mathcal{F}_{n\Delta-}\right)\{1+2\rho(1)\} - 2\rho(1)\Pr(N_n > 0|\mathcal{F}_{n\Delta-})\right].$$

For large values of Δ *this is going to behave like*

$$Var(p_n|\mathcal{F}_{n\Delta-}) \simeq Var(Z_t)E\left(N_n|\mathcal{F}_{n\Delta-}\right)\{1+2\rho(1)\}.$$

In applied work $\rho(1) \simeq -0.24$ *for IBM stocks on the NYSE in 1995 (*Rydberg and Shephard (2000)*). Hence we would not expect* $Var(p_n|\mathcal{F}_{n\Delta-})/E\left(N_n|\mathcal{F}_{n\Delta-}\right)$ *to be a constant as* Δ *changes, although it will have an asymptote of* $Var(Z_t)\{1+2\rho(1)\}$ *for large values of* Δ.

These two examples are, of course, only illustrative for they impose empirically unattractive constraints on the model. In particular, detailed empirical modelling would require us to move

[2] This uses the result that

$$Var(p_n|N_n) = Var(Z_t)\left\{N_n + 2\sum_{j=1}^{N_n}(N_n - j)\,\rho(j)\right\}.$$

away from the assumption that the $\{N_n\}$ and $\{\mathrm{Z_t}\}$ processes are stochastically independent (see, for example, Russell and Engle (1998) and Rydberg and Shephard (1998a)). However, with careful thought it may be possible to do this without loss of mathematical tractability.

The examples demonstrate, we believe, that if our goal is to model price changes in calander time then we should focus on building models of $\{N_n\}$.

1.2 BIN models

In order to model the sequence $\{N_n\}$ we suggest using BIN models. A direct link between BIN models and Cox processes will be provided in Section 3, but for now the BIN model specifies the one-step ahead forecast distribution of the $\{N_n\}$ series using a counting distribution. There is a large existing literature on time series models of counts (see, for example, the reviews by MacDonald and Zucchini (1997), Grunwald, Hyndman, Tedesco, and Tweedie (1997) and Cameron and Trivedi (1998, Ch. 7)), however the particular form of model we choose is inspired by the econometric literature on ARCH models, see Bollerslev, Engle, and Nelson (1994) for a review. In particular we will write $N_n|\mathcal{F}_{n\Delta-} \sim Po(\lambda_n)$, allowing λ_n to depend upon $\mathcal{F}_{n\Delta-}$, the information available infinitesimally before time $n\Delta$, in a particular way. Here $Po(\alpha)$, denotes a Poisson distribution with mean α. This time series setup is a special case of models put forward by Zeger and Qaqish (1988) and Shephard (1994).

In our treatment we will assume λ_n is a *linear* function of past data — the linearity is the vital part of this development. The simplest example[3] of this is where, for $\alpha, \gamma > 0$,

$$N_n|\mathcal{F}_{n\Delta-} \sim Po(\lambda_n), \qquad \lambda_n = \alpha + \gamma N_{n-1}.$$

This very simple model has been suggested by Grunwald, Hyndman, Tedesco, and Tweedie (1997, p. 11, example 2) in the context of checking regularity conditions required to prove geometric ergodicity of generic first order autoregressions. Further, this model is observationally identical to a compound model

$$N_n = \sum_{j=1}^{N_{n-1}} W_{n,j} + Z_n, \quad \text{where} \quad W_{n,j} \overset{iid}{\sim} Po(\gamma), \quad Z_n \overset{iid}{\sim} Po(\alpha)$$

with the $\{W_{n,j}\}$ and $\{Z_n\}$ processes being independent[4].

The BIN model is precisely the analog of the squares of an ARCH(1)[5] model (due to Engle (1982)). It is easy to see that unlike the squares of an ARCH, $\{N_n\}$ is always covariance stationary and geometrically ergodic if $\gamma < 1$. No other condition is needed. In particular

$$E(N_n) = \frac{\alpha}{1-\gamma}, \qquad Cor(N_n, N_{n-s}) = \gamma^{|s|} \qquad \text{and} \qquad Var(N_n) = \frac{E(N_n)}{1-\gamma^2}.$$

[3] The most straightforward discrete time alternative to this model structure would be to write down a state space model with

$$N_n|\lambda_n \sim Po(\lambda_n),$$

where $\log \lambda_n$ is a Gaussian ARMA model — see Gamerman (1992). Methods for carrying out likelihood inference for this class of model (and a wider set of models) were developed in Shephard and Pitt (1997) and simplified by Durbin and Koopman (1997).

[4] This setup is close to those presented by McKenzie (1985) and McKenzie (1988) who work with $W_{n,j} \overset{iid}{\sim} Bern(\gamma)$ in order to ensure that the marginal distribution of the counts is always Poisson — this is not an aim our modelling framework has. Estimation of this alternative process, which is harder than the BIN model, has recently been studied by Freeland and McCabe (1998).

[5] In the ARCH(1) case the squares are

$$p_n^2|\mathcal{F}_{n-1} \sim \chi_1^2\sigma_n^2, \quad \sigma_n^2 = \alpha + \gamma p_{n-1}^2.$$

For this process $\{p_n^2\}$ is covariance stationary if $\gamma^2 < 1/3$.

This model easily generalises and can be directly linked into Cox processes in a natural way. These points will be spelt out in the main sections of our paper.

One of the major advanges of using counts as the basis of modelling is that it is relatively easy to extend to the multivariate case. This maybe necessary if we are modelling the joint evolution of a vector of prices or we are modelling the number of trades and quote revisions in a single stock. Let us write $\{N(u)\}_{u\geq 0}$ as a $K \times 1$ vector which has as its $j-th$ element the number of trades recorded up until time u in the $j-th$ stock. Then define the vector differencing operator

$$N_n = N\{(n+1)\Delta-\} - N(n\Delta).$$

The simplest multivariate model would put

$$E(N_n|\mathcal{F}_{n\Delta-}) = \lambda_n, \qquad \text{where} \quad \lambda_n = \alpha + \gamma N_{n-1}.$$

Here all of the elements of the vector α and the matrix γ have to be non-negative.

1.3 Related work

Papers which have previously just looked at the $\{N(u)\}$ process include Engle and Russell (1998) and, subsequently, Meddahi, Renault, and Werker (1998), and Ghysels, Jasiak, and Gourieroux (1998). These papers model the time between trades. Let τ_t be the time of the $t-th$ trade. Then it is given by

$$\tau_t = \min_u \{N(u) = t\}, \qquad t = 1, 2, ..., N(S),$$

recalling that $N(S)$ is the number of trades in the period of length S we are studying. Then the length of time between trades[6] is

$$L_t = \tau_t - \tau_{t-1}, \qquad t = 1, 2, ...$$

For our data a small number of these times are exactly zero.

In the econometric literature an influential model of the durations is the autoregressive conditional duration (ACD) model of Engle and Russell (1998). This puts

$$L_t = \varepsilon_t \psi_t, \qquad \varepsilon_t > 0, \qquad E(\varepsilon_t) = 1$$

and the ε_t's are independent identical distributed (*i.i.d.*), with

$$\psi_t = \alpha + \sum_{j=1}^{p} \gamma_j L_{t-j} + \sum_{j=1}^{q} \beta_j \psi_{t-j}.$$

Here $\psi_t = E(L_t|\mathcal{F}_{t-1})$, the conditional expected waiting time. The model has many similarities with earlier work by Wold (1948) and Cox (1972). In practice Engle and Russell (1998) have used an exponential or Weibull distribution on the $\{\varepsilon_t\}$. Straightforward alternative structures would be to parameterise the $\log \psi_t$ instead of the ψ_t.

A key feature of this model is that, conditional on $\psi_0, ..., \psi_{q+1}, L_0, ..., L_{p+1}$, the likelihood can be computed via a prediction decomposition. Further the number of terms that need to be evaluated is only $N(S)$, rather than the number of seconds. Hence this model is much faster to fit than the BIN model above.

[6] From a statistical viewpoint we can think of $\{L_t\}$ as a time series of duration times. There is an enormous literature on the analysis of durations, although most of it does not have a time series interpretation. We refer to Lancaster (1990), Synder and Miller (1991) and Cox and Oakes (1984) for general discussions of this literature.

The ACD model implies the following structure for the price level of the stock. It has the evolution according to the process

$$p\left(\sum_{t=1}^{n} L_t\right) = p(0) + \sum_{t=1}^{n} Z_t,$$

which tells us the price after n irregularly space times of trades (i.e. it tells us the price at a random point in time in the future). In continuous time this has a less elegant expression as

$$p(u) = p(0) + \sum_{t=1}^{N(u)} Z_t, \qquad \text{where} \qquad N(u) = \arg\max_n \left\{ \sum_{t=1}^{n} L_t \leq u \right\}.$$

This implies

$$\begin{aligned} N_n &= N\left\{(n+1)\,\Delta-\right\} - N(n\Delta) \\ &= \arg\max_r \left\{ \sum_{t=1}^{r} L_t < (n+1)\,\Delta \right\} - \arg\max_r \left\{ \sum_{t=1}^{r} L_t \leq n\Delta \right\}. \end{aligned}$$

The complexity of the relationship between the $\{L_t\}$ and $\{N_n\}$ implies studying the behaviour of returns in calendar time $\{p_n\}$ implied by ACD style models is difficult even if we make very simple assumptions on the $\{Z_t\}$ process. This is not the case with BIN based models.

There are other papers which relate to our approach. Following Rydberg and Shephard (2000), Rydberg and Shephard (1998a) have studied the dynamics of the $\{Z_t\}$ process within the context of our compound process framework, while not discussing the specification of the counting process $\{N(u)\}$. A multivariate generalisation of the model is given by Rydberg and Shephard (1998b). An alternative attempt to model the dynamics of the $\{Z_t\}$ has also been previously proposed by Hausman, Lo, and MacKinlay (1992) and Russell and Engle (1998).

Our models are also related to stochastic volatility (SV) or time deformation models, see e.g. Clark (1973), Hull and White (1988), Stein and Stein (1991), Ghysels, Harvey, and Renault (1996) and Barndorff-Nielsen and Shephard (1999). In SV models Brownian motion is deformed, while in the compound process the Cox process is a deformed Poisson process. Barndorff-Nielsen and Shephard (1999) have studied the connection between the modelling framework given in (1) and a SV model in a thickly traded market. Those results have been elaborated by Frey and Runggaldier (1998).

1.4 Structure of the paper

The paper is organized as follows. In Section 2 we define BIN models and derive various properties of the models. In Section 3 we derive the properties of $\{N_n\}$ when the counting process $\{N(u)\}$ is modelled as a Cox process. We show that for certain types of intensity processes for the Cox process, the properties of BIN models are similar to that of the Cox process. This allows us to coherently move from continuous time to discrete time. In this section we also discuss the time aggregation of BIN models. In Section 4 we look at an empirical implementation of our model structure on IBM data, while section 5 concludes. The paper has a number of theorems. We have collected the proofs of these theorems in an Appendix, which is given after the conclusions.

2 A BIN MODEL

2.1 Model structure

2.1.1 First order model

Our class of discrete time processes, which we label BIN models, are based on writing down a structure with $N_n|\mathcal{F}_{n\Delta-} \sim Po(\lambda_n)$, allowing λ_n to depend upon $\mathcal{F}_{n\Delta-}$, the information available

at time $n\Delta-$, in a particular way. We assume λ_n is a *linear* function of past data. The simplest example of this is where, for $\alpha, \gamma > 0$,

$$N_n|\mathcal{F}_{n\Delta-} \sim Po(\lambda_n), \qquad \lambda_n = \alpha + \gamma N_{n-1}. \tag{3}$$

We call this a BIN model of order one (BIN(1)). This is precisely the analog of the squares of an ARCH(1) model (due to Engle (1982)).

This model is of autoregressive type for

$$\begin{aligned} N_n &= \lambda_n + u_n \\ &= \alpha + \gamma N_{n-1} + u_n, \end{aligned} \tag{4}$$

where $u_n = N_n - \lambda_n$ is such that $E(u_n|\mathcal{F}_{n\Delta-}) = 0$[7]. Many of the interesting features of the BIN model follow from this structure.

2.1.2 Stochastic properties of BIN(1) models

The main stochastic features of the first order BIN model are remarkably simple. They are given by the following theorem.

Theorem 2.1. If $\{N_n\}$ is generated by (3) for $n \in Z$ with $\alpha, \gamma > 0$, then if $\gamma < 1$ we have that the second order[8] properties of the model are

$$\kappa_1 = \frac{\alpha}{1-\gamma}, \qquad Cor(N_n, N_{n-s}) = \gamma^{|s|} \qquad \text{and} \qquad \kappa_2 = \frac{\kappa_1}{1-\gamma^2},$$

where κ_r denotes the $r-th$ cumulant of $\{N_n\}$. Further, if we write $k(\theta \ddagger N_n) = \log\left[E\left\{\exp\left(\theta N_n\right)\right\}\right]$ then

$$k(\theta \ddagger N_n) = \alpha\left(e^{\theta} - 1\right) + k\left\{\gamma\left(e^{\theta} - 1\right) \ddagger N_n\right\}.$$

From this we have

$$\kappa_3 = \frac{3\kappa_2}{1-\gamma^3} \quad \text{and} \quad \kappa_4 = \frac{\left(1 + 5\gamma^3\right)\kappa_3 + 2\gamma^2\kappa_2}{1-\gamma^4}.$$

Finally all of the moments of $\{N_n\}$ are bounded if $\gamma < 1$.

□

Proof of Theorem 2.1: See the Appendix.

Theorem 2.2. If $\{N_n\}$ is generated by (3) for $n \in Z$ with $\alpha, \gamma > 0$, then $\{N_n\}$ has an ergodic distribution if $\gamma < 1$. Further the convergence to the ergodic distribution is geometrically fast when measured using a total variation norm.

□

Proof of Theorem 2.2: Given by Grunwald, Hyndman, Tedesco, and Tweedie (1997, Case II of Proposition 3).

□

Remark 1 Theorem 2.1 is a very simple result for it says that only $\gamma < 1$ is required for covariance stationarity. This is a much simpler result than the parallel results for the squares of an ARCH(1) and for an ACD(1) model for both of these processes impose additional constraints on the dynamics in order for the variance to exist.

Remark 2 The autocorrelation function of $\{N_n\}$ has to be non-negative. This feature is also present in ARCH and ACD models.

Remark 3 Theorems 2.1 and 2.2 show that the condition $\gamma < 1$ implies both covariance and strict stationarity.

[7] In fact u_n is a Martingale since λ_n is the compensator of N_n.

[8] Second order properties, in this context, are the mean, variance and autocovariance function of the process.

2.1.3 Inference for BIN(1) models

A major advantage of the BIN style models is that we can use a conditional likelihood function to estimate the parameters. Recall the prediction decomposition writes

$$\begin{aligned}\log f(N_2,...,N_T|N_1) &= \sum_{n=2}^{T} \log f(N_n|N_{n-1}) \\ &= \sum_{n=2}^{T} -\lambda_n + N_n \log \lambda_n - \log N_n!.\end{aligned}$$

The score vector has the form, for a parameter vector ψ,

$$\frac{\partial \log f(N_2,...,N_T|N_1)}{\partial \psi} = \sum_{n=2}^{T} \frac{\partial \lambda_n}{\partial \psi}\left(\frac{N_n}{\lambda_n}-1\right), \quad \text{where} \quad \frac{\partial \lambda_n}{\partial \psi} = \begin{pmatrix} 1 \\ N_{n-1} \end{pmatrix}.$$

Importantly $E\left(N_n|\mathcal{F}_{n\Delta-}\right) = \lambda_n$ and so $\left\{\frac{\partial \lambda_n}{\partial \psi}\left(\frac{N_n}{\lambda_n}-1\right)\right\}_{n>0}$ is a bivariate Martingale difference sequence. Likewise the observed information is

$$\begin{aligned}-\frac{\partial^2 \log f(N_2,...,N_T|N_1)}{\partial \psi \partial \psi'} &= -\sum_{n=2}^{T} \frac{\partial^2 \lambda_n}{\partial \psi \partial \psi'}\left(\frac{N_n}{\lambda_n}-1\right) + \sum_{n=2}^{T} \frac{\partial \lambda_n}{\partial \psi}\frac{\partial \lambda_n}{\partial \psi'}\frac{N_n}{\lambda_n^2} \\ &= \sum_{n=2}^{T} \frac{\partial \lambda_n}{\partial \psi}\frac{\partial \lambda_n}{\partial \psi'}\frac{N_n}{\lambda_n^2},\end{aligned}$$

due to the linearity of λ_n. Hence, unlike ARCH(1) models, BIN(1) models have globally concave likelihood functions. It is sometimes convenient to approximate the observed information by

$$\mathcal{I}_T = \sum_{n=2}^{T} \frac{\partial \lambda_n}{\partial \psi}\frac{\partial \lambda_n}{\partial \psi'}\frac{1}{\lambda_n},$$

which is a term by term conditional expected information measure.

We typically find the maximum likelihood estimator by iteration. A convenient way of carrying this out is to compute

$$\widetilde{\psi}^{i+1} = \widetilde{\psi}^{i} + \left(\sum_{n=2}^{T} \frac{\partial \lambda_n}{\partial \psi}\frac{\partial \lambda_n}{\partial \psi'}\frac{1}{\lambda_n}\right)^{-1} \sum_{n=2}^{T} \frac{\partial \lambda_n}{\partial \psi}\left(\frac{N_n}{\lambda_n}-1\right), \tag{5}$$

for the change in the parameters at each iteration is simply a regression with

$$X_n = \frac{\partial \lambda_n}{\partial \psi}\frac{1}{\sqrt{\lambda_n}}, \quad Y_n = \sqrt{\lambda_n}\left(\frac{N_n}{\lambda_n}-1\right), \quad n = 2,...,T.$$

As the likelihood function is globally concave this algorithm should perform reliably. Further we can test for serial dependence using likelihood ratio, score or Wald principles. All are straightforward in this context.

A major theoretical advantage of BIN models is that inference is, in a sense, robust to misspecification in the Poisson assumption. This is a result of the score equation continuing to be a sum of martingale differences so long as $\lambda_n = E(N_n|\mathcal{F}_{n\Delta-})$, that is we correctly model the one-step ahead conditional mean. So long as $\gamma < 1$ the maximum likelihood estimator will be consistent, although a sandwich estimator would have to be used to correctly model its asymptotic distribution.

2.1.4 General order model

Naturally all of the above results extend to the BIN(p) model where

$$\lambda_n = \alpha + \sum_{s=1}^{p} \gamma_s N_{n-s},$$

with the constraints that $\alpha > 0$ and $\{\gamma_s > 0\}$. In particular if $\sum_{s=1}^{p} \gamma_s < 1$ then the $\{N_n\}$ are covariance stationary with second order properties

$$\kappa_1 = \frac{\alpha}{1 - \sum_{s=1}^{p} \gamma_s}, \quad \kappa_2 = \frac{\kappa_1}{1 - \sum_{s=1}^{p} \gamma_s \rho_s},$$

where $\rho_s = Cor(N_n, N_{n-s})$. In turn the $\{\rho_s\}$ can be found by solving the usual Yule-Walker equations

$$\rho_s = \gamma_1 \rho_{s-1} + ... + \gamma_s \rho_0, \quad \text{for } s = 1, 2, ..., p.$$

It is important to note that the feature that the $\{\gamma_s > 0\}$ means that the autocorrelation function has to be non-negative. In particular we can write

$$\rho_s = \sum_{j=1}^{p} b_j \varphi_j^{|s|}, \quad \text{where } \{\varphi_j\} \text{ are the roots to } \quad \sum_{j=0}^{p} z^{p-j} \gamma_j = 0,$$

where $\gamma_0 = 1$ and $\sum_{j=1}^{p} b_j = 1$. As all the $\{\gamma_s\}$ are positive, the roots $\{\varphi_j\}$ and weights $\{b_j\}$ must be real and positive.

The conditional likelihood function of this model is again easy to work with. In particular

$$\begin{aligned} \log f(N_{p+1}, ..., N_T | N_1, ..., N_p) &= \sum_{n=p+1}^{T} \log f(N_n | N_{n-1}, ..., N_{n-p}) \\ &= \sum_{n=p+1}^{T} -\lambda_n + N_n \log \lambda_n - \log N_n!, \end{aligned}$$

is concave in the parameters of the model, while the score and observed information is straightforward to compute iteratively.

2.2 Moving average models

An alternative generalisation of the BIN(1) model is to set

$$N_n | \mathcal{F}_{n\Delta -} \sim Po(\lambda_n), \qquad \lambda_n = \alpha + \gamma N_{n-1} + \delta \lambda_{n-1}, \tag{6}$$

which we will label a BIN(1,1) model. Sufficient conditions for λ_n to be almost-surely non-negative is that $\alpha, \gamma, \delta \geq 0$.

This model is inspired by the GARCH model due to Bollerslev (1986) and Taylor (1986). This model is of autoregressive moving average (ARMA) type for

$$\begin{aligned} N_n &= \lambda_n + u_n & (7) \\ &= \alpha + \gamma N_{n-1} + \delta \lambda_{n-1} + u_n \\ &= \alpha + \gamma N_{n-1} + \delta \left(N_{n-1} - u_{n-1} \right) + u_n & (8) \\ &= \alpha + (\gamma + \delta) N_{n-1} + u_n - \delta u_{n-1}, & (9) \end{aligned}$$

where $u_n = N_n - \lambda_n$ is such that $E(u_n | \mathcal{F}_{n\Delta -}) = 0$. Theorem 2.3 collects the main features of this model.

Theorem 2.3 If $\{N_n\}$ is generated by (6) for $n \in Z$ with $\alpha, \gamma, \delta > 0$, then if $\gamma + \delta < 1$, we have that

$$\kappa_1 = \frac{\alpha}{1-(\delta+\gamma)}, \qquad \kappa_2 = \kappa_1 \frac{1-\delta^2-2\gamma\delta}{1-(\delta+\gamma)^2} = \kappa_1 + \kappa_1 \frac{\gamma^2}{1-(\delta+\gamma)^2},$$

and

$$\rho_1 = \frac{\gamma\{1-\delta(\gamma+\delta)\}}{1+\delta^2-2\delta(\gamma+\delta)}, \quad \rho_s = \rho_1(\gamma+\delta)^{s-1}, \quad s = 2, 3, ...$$

Further, the joint cumulant generating function of N_n and λ_n is, for an index $\theta = (\theta_1, \theta_2)'$,

$$k(\theta \ddagger N_n, \lambda_n) = \alpha\left(\theta_2 + e^{\theta_1} - 1\right) + k\left\{\gamma\left(\theta_2 + e^{\theta_1} - 1\right), \delta\left(\theta_2 + e^{\theta_1} - 1\right) \ddagger N_n, \lambda_n\right\}.$$

□

Proof of Theorem 2.3: See the Appendix.

Remark 4 The constraints require the ARMA representation to have an autoregressive parameter which is positive and the moving average being negative. In addition the autoregressive parameter has to be at least as big as the absolute value of the moving average parameter.

Remark 5 The autocorrelation function is non-negative at all lags.

Remark 6 The properties of $\{\lambda_n\}$ are sometimes helpful. In particular $E(\lambda_n) = \kappa_1$ and

$$\begin{aligned} Var(\lambda_n) &= \kappa_2 - \kappa_1 \\ &= \kappa_1 \frac{\gamma^2}{1-(\delta+\gamma)^2} \\ &= \frac{\alpha\gamma^2}{\{1-(\delta+\gamma)\}\left\{1-(\delta+\gamma)^2\right\}}. \end{aligned}$$

The conditional likelihood function for the BIN(1,1) model is not immediately available to us, instead we work with

$$\begin{aligned} \log f(N_1, N_2, ..., N_T|\lambda_1) &= \sum_{n=1}^{T} \log f(N_n|\mathcal{F}_{n\Delta-}) \\ &= \sum_{n=1}^{T} -\lambda_n + N_n \log\lambda_n - \log N_n!, \end{aligned}$$

which depends upon the unknown (random) λ_1 as well as the unknown parameters α, γ, δ.

In the GARCH literature a similar dependence on the unknown initial conditional variance is present in the conditional likelihood function. A number of approaches have been used: (i) treat λ_1 as an unknown parameter to be estimated along with the other parameters, (ii) set λ_1 to equal κ_1, (iii) set λ_1 equal to the variance of a group of initial datapoints. All of these are asymptotically negligible and can be justified in a variety of ways.

The score function of this model is more complicated than that of its BIN(1) cousin. This is because

$$\frac{\partial \log f(N_1, N_2, ..., N_T|\lambda_1)}{\partial\psi} = \sum_{n=1}^{T} \frac{\partial\lambda_n}{\partial\psi}\left(\frac{N_n}{\lambda_n} - 1\right), \quad \text{where} \quad \frac{\partial\lambda_n}{\partial\psi} = \begin{pmatrix} 1 \\ N_{n-1} \\ \lambda_{n-1} \end{pmatrix} + \delta\frac{\partial\lambda_{n-1}}{\partial\psi}.$$

The complication is that the third element of $\partial\lambda_n/\partial\psi$ has an extra term which means we have to compute these terms recursively starting at $\partial\lambda_1/\partial\psi$. This seemingly marginal change in the score has some important implications. In particular

$$\frac{\partial^2 \log f(N_1, N_2, ..., N_T|\lambda_1)}{\partial\psi\partial\psi'} = \sum_{n=1}^{T} \frac{\partial^2\lambda_n}{\partial\psi\partial\psi'}\left(\frac{N_n}{\lambda_n} - 1\right) - \sum_{n=1}^{T} \frac{\partial\lambda_n}{\partial\psi}\frac{\partial\lambda_n}{\partial\psi'}\frac{N_n}{\lambda_n^2}.$$

We can compute

$$\frac{\partial^2\lambda_n}{\partial\psi\partial\psi'} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{\partial\lambda_{n-1}}{\partial\psi'} & & \end{pmatrix} + \begin{pmatrix} 0 & 0 & \\ 0 & 0 & \frac{\partial\lambda_{n-1}}{\partial\psi} \\ 0 & 0 & \end{pmatrix} + \delta\frac{\partial^2\lambda_{n-1}}{\partial\psi\partial\psi'} \quad \text{starting at} \quad \frac{\partial^2\lambda_1}{\partial\psi\partial\psi'}.$$

Suppose that λ_1 is parameter free, then $\partial^2\lambda_1/\partial\psi\partial\psi'$ is zero and the upper left hand block of four elements of $\partial^2\lambda_n/\partial\psi\partial\psi'$ are exactly zero. Hence if we know δ the log-likelihood function is concave in α and γ. However, in general this is not the case and the log-likelihood function is only asymptotically concave with respect to all three parameters.

We typically approximate the observed information to produce

$$\mathcal{I}_T = \sum_{n=2}^{T} \frac{\partial\lambda_n}{\partial\psi}\frac{\partial\lambda_n}{\partial\psi'}\frac{1}{\lambda_n},$$

which for large T is likely to behave well. This allows us to again produce a regression interpretation of the iterative updating of approximations to the maximum likelihood estimator via (5). In this case we are not guaranteed to converge to a unique maximum of the likelihood function, although for large samples this should be satisfactory. Alternatively we could use a more sophisticated quasi-Newton optimisation routine which can perform helpful line searches.

2.3 Generalised version

2.3.1 General case

This model structure generalises straightforwardly to a BIN model of order p, q (BIN(p, q)) where

$$N_n|\mathcal{F}_{n\Delta-} \sim Po(\lambda_n), \qquad \lambda_n = \alpha + \sum_{j=1}^{p}\gamma_j N_{n-j} + \sum_{j=1}^{q}\delta_j\lambda_{n-j}.$$

Notice that we are following standard ARMA notation here and using p to denote the number of autoregressive terms in the model and q to denote the number of moving average ones.

Although imposing the constraints $\alpha > 0$, $\{\gamma_j > 0\}$ and $\{\delta_j > 0\}$ seems natural in this context, these constraints are not necessary in order to impose that, with probability one, λ_n is a non-negative sequence when $q \geq 1$. The following theorem gives the result for the BIN(p,1) model. More general results were developed, within the context of GARCH models but applicable to all positive processes, by Nelson and Cao (1992).

Theorem 2.4 In the BIN(p,1) model the $\{\lambda_n\}$ sequence will be non-negative with probability one iff

$$\alpha \geq 0, \quad \delta_1 \geq 0, \quad \sum_{j=0}^{k}\gamma_{j+1}\delta_1^{k-j} \geq 0 \text{ for } k = 0, ..., p-1.$$

Proof of Theorem 2.4: This result follows immediately from the results on non-negative processes studied in the context of GARCH models by Nelson and Cao (1992).

As before this model has an ARMA representation,

$$\begin{aligned} N_n &= \lambda_n + u_n \\ &= \alpha + \sum_{j=1}^{p} \gamma_j N_{n-j} + \sum_{j=1}^{q} \delta_j \left(N_{n-j} - u_{n-j}\right) + u_n \\ &= \alpha + \sum_{j=1}^{\max(p,q)} \phi_j N_{n-j} + u_n - \sum_{j=1}^{q} \delta_j u_{n-j}, \end{aligned} \tag{10}$$

where $u_n = N_n - \lambda_n$ is again a MD sequence and $\phi_j = \gamma_j + \delta_j$. This model is covariance stationary when $\sum_{j=1}^{\max(p,q)} \phi_j < 1$. Under this condition

$$\kappa_1 = \frac{\alpha}{1 - \sum_{j=1}^{\max(p,q)} \phi_j}, \quad Cov(N_n, \lambda_n) = Var(\lambda_n) = \kappa_2 - \kappa_1,$$

and

$$\begin{aligned} Var(u_n) &= Var(N_n) + Var(\lambda_n) - 2Cov(N_n, \lambda_n) \\ &= \kappa_1. \end{aligned}$$

Then κ_2 and the autocorrelation function of $\{N_n\}$ can be computed using results on variances and autocorrelations of stationary linear ARMA$\{\max(p,q), q\}$ process (e.g. Brockwell and Davis (1987, Ch. 3)). In particular various numerical routines are available which can compute these terms for any value of p and q.

2.3.2 Special case

The special case of $q = 1$ is noteworthy when $p \geq 1$. In this case

$$\begin{aligned} N_n &= \alpha + \sum_{j=1}^{p} \left(\gamma_j + \delta_j\right) N_{n-j} + u_n - \delta u_{n-1} \\ &= \alpha + \sum_{j=1}^{p} \phi_j N_{n-j} + u_n - \delta u_{n-1}. \end{aligned} \tag{11}$$

As a result the Yule-Walker equations are, for $s > 1$,

$$Cov(N_n, N_{n-s}) = \sum_{j=1}^{p} \phi_j Cov(N_{n-j}, N_{n-s}).$$

Hence the damping down behaviour is determined by the autoregressive roots of the model, while the moving average just influences the initial conditions.

2.3.3 Component based parametrisations

The use of high frequency data means we will typically use BIN models on very long time series and be interested in forecasting future counts many time periods ahead. In the context of GARCH models Engle and Lee (1999) have suggested the use of a component parameterisation in order to focus on multistep forecasting. We will show that their arguments carry over to BIN models.

A basic BIN(1,1) model can be written with

$$\begin{aligned}\lambda_n &= (1-\gamma-\delta)\mu + \gamma N_{n-1} + \delta\lambda_{n-1}, \quad \text{where} \quad \mu = \alpha/(1-\gamma-\delta).\\ &= \mu + \gamma\left(N_{n-1}-\mu\right) + \delta\left(\lambda_{n-1}-\mu\right).\end{aligned}$$

This can be extended by allowing

$$\begin{aligned}\lambda_n &= \mu_n + \gamma\left(N_{n-1}-\mu_{n-1}\right) + \delta\left(\lambda_{n-1}-\mu_{n-1}\right)\\ \mu_n &= \mu^* + \rho\mu_{n-1} + \phi\left(N_{n-1}-\lambda_{n-1}\right).\end{aligned}$$

Here, if $\rho > \gamma+\delta$, we can think of μ_n as the longer term component of the conditional mean. We will call this model the component BIN model and require that all of its parameters to be non-negative. Further, we will need that $\rho < 1$ and $\gamma+\delta < 1$ for covariance stationarity.

Remark 7 The component BIN model has five parameters, the same number as its corresponding "reduced form" BIN(2,2) model which puts

$$\begin{aligned}\lambda_n &= (1-\gamma-\delta)\mu^* + (\delta+\rho-\phi)\lambda_{n-1} + \{\phi(\gamma+\delta)-\rho\delta\}\lambda_{n-2}\\ &\quad + (\phi+\gamma)N_{n-1} - \{\rho\gamma+\phi(\gamma+\delta)\}N_{n-2}.\end{aligned}$$

Note that the constant in front of the N_{n-2} has to be negative, while the sign of the reduced form parameters for λ_{n-1} and λ_{n-2} are uncertain.

Remark 8 A potential problem with this type of model is the possibility that λ_n may become negative. In the Appendix to their paper Engle and Lee (1999) studied, using generic results of Nelson and Cao (1992), exactly this problem in the context of component GARCH models, noting that their results "are true for any dependent variable in the same dynamic structure as long as the dependent variables are non-negative". The precise sufficient conditions they showed were that

$$\begin{aligned}1 &\geq \rho > \gamma+\delta > 0\\ \delta &> \phi > 0\\ \mu^* &> 0, \gamma > 0.\end{aligned}$$

The first and third of these constraints seem natural and desirable in any case. The second is easy to impose numerically.

2.4 Other issues

2.4.1 Negative dependence

All of the models we have so far introduced imply the non-negative serial correlation amongst the $\{N_n\}$ series. Empirically it maybe helpful to broaden this framework. One possible approach is to write, for example

$$N_n|\mathcal{F}_{n\Delta-} \sim Po\left\{\lambda_n \exp\left(\mu_n\right)\right\}, \qquad \lambda_n = \alpha + \gamma N_{n-1},$$

where μ_n is a linear function of past pieces of information. The simplest case is where

$$\mu_n = \zeta N_{n-1},$$

requiring that ζ be negative. We call this effect "cooling", as it reduces the conditional mean of the process. Again $\{N_n\}$ is a Markov chain. If $\Pr(N_n > 1)$ is small we have that

$$\exp\left(\mu_n\right) \simeq \begin{cases} 1 & \text{if} \quad N_{n-1}=0 \\ \exp(\zeta) & \text{if} \quad N_{n-1}=1 \end{cases}.$$

implying

$$\lambda_n \exp(\mu_n) \simeq \begin{cases} \alpha, & \text{if } N_{n-1}=0 \\ (\alpha+\gamma)\exp(\zeta), & \text{if } N_{n-1}=1 \end{cases}.$$

Hence the model can induce negative dependence as long as $\zeta < \log(\alpha+\gamma)$. If $\zeta > 0$ then this model can become nonstationary and so some care should be taken.

This model generalises to allow

$$\mu_n = \sum_{j=1}^{r} \zeta_j N_{n-j}, \quad \text{and} \quad \lambda_n = \alpha + \sum_{j=1}^{p} \gamma_j N_{n-j} + \sum_{j=1}^{q} \delta_j \lambda_{n-j},$$

which we write as a generalised BIN, GBIN(p,q,r), model. It is important to note that μ_n is only influenced by the last r values of $\{N_n\}$ and so has quite short memory compared to λ_n. Hence if we use this model on empirical data we would expect the role of μ_n will be to deal with the short term dynamics of the process, while λ_n can model the longer term effects.

2.4.2 Weak BIN models

So far we have used a Poisson assumption for the distribution of $N_n|\mathcal{F}_{n\Delta-}$. When we later study the effect of changing Δ we will not be able to maintain the Poisson assumption for different values of Δ. The same issues appear in the literature on the temporal aggregation of GARCH processes (see Drost and Nijman (1993)). In order to have the appropriate vocabulary to deal with this issue we here define three different types of BIN models, each with

$$\lambda_n = \alpha + \sum_{j=1}^{p} \gamma_j N_{n-j} + \sum_{j=1}^{q} \delta_j \lambda_{n-j}.$$

The three definitions of the BIN models are:

- Strong BIN models have $N_n|\mathcal{F}_{n\Delta-} \sim Po(\lambda_n)$.
- Semi-strong BIN models have $\mathrm{E}(N_n|\mathcal{F}_{n\Delta-}) = \lambda_n$. This model is more flexible than the Strong BIN model for it could allow, for example, the conditional variance of the process to depend on the filtration not only through λ_n.
- Weak BIN models have λ_n being the best linear predictor of N_n in terms of lagged versions of $\{N_n\}$. This means that the Wold representation of the covariance stationary process $\{N_n\}$ can be exactly represented in terms of a parameterised BIN model. Further, so long as the moments exist, weak BIN models have the same second order properties as strong BIN models.

We can think of the these three different specifications as representing, respectively, (i) parametric models of the counts, (ii) a Martingale difference decompositions for $N_n - \lambda_n$, (iii) a covariance representations of the counts. In terms of inference these three modelling frameworks allow: likelihood inference, quasi-likelihood inference and method of moment based inference, respectively.

2.4.3 Diagnostic checking of count models

One use of BIN models is that we can use them as alternative hypotheses in formal testing procedures. Think of the following setup. Let the current model of the conditional mean be $E(N_n|\mathcal{F}_{n\Delta-}) = \lambda_n$, while we wish to test for unmodelled serial dependence. An alternative model could be that

$$E(N_n|\mathcal{F}_{n\Delta-}) = \lambda_n + \beta N_{n-1},$$

while we have a null as $\beta = 0$. The score for β is

$$\sum_{n=1}^{T} N_{n-1} \left(\frac{N_n}{\lambda_n + \beta N_{n-1}} - 1 \right),$$

while the second derivative of the log-likelihood is

$$\sum_{n=1}^{T} N_{n-1}^2 \frac{N_n}{\lambda_n + \beta N_{n-1}}.$$

Hence a standard score statistic for the null hypothesis is

$$\begin{aligned} S &= T \left\{ \frac{1}{T} \sum_{n=1}^{T} \frac{1}{\lambda_n} \left(N_{n-1} N_n - N_{n-1} \lambda_n \right) \right\}^2 / \left(\frac{1}{T} \sum_{n=1}^{T} N_{n-1}^2 \frac{N_n}{\lambda_n} \right) \\ &\xrightarrow[H_0]{D} \chi_1^2. \end{aligned}$$

If α is very small then $N_{n-1} N_n = 0$ unless both of the counts are one.

The important special case of where $\lambda_n = \alpha = E(N_n) = Var(N_n)$, a constant, yields a simplification

$$\begin{aligned} S &= T \left\{ \frac{1}{T} \sum_{n=1}^{T} \frac{1}{\alpha} \left(N_{n-1} N_n - \alpha N_{n-1} \right) \right\}^2 / \left(\frac{1}{T} \sum_{n=1}^{T} N_{n-1}^2 \frac{N_n}{\alpha} \right) \\ &\simeq T \left\{ \frac{1}{T} \sum_{n=1}^{T} \left(N_{n-1} N_n - \alpha^2 \right) \right\}^2 / \left\{ \alpha^3 \left(\alpha + 1 \right) \right\} \\ &\simeq T r_1^2 / \left\{ \alpha \left(\alpha + 1 \right) \right\}, \end{aligned}$$

where r_1 is the first order serial correlation coefficient. This test is very different than that for the Gaussian first order autoregression, which ignores the $\alpha(\alpha + 1)$ term. It suggests if α is very small then even small serial correlation coefficients would present grave evidence against the null hypothesis of no serial dependence.

3 BIN AND COX PROCESSES

3.1 Background

BIN models specify the distribution of $\{N_n | \mathcal{F}_{n\Delta-}\}$ and so are inherently a discrete time model fixed to some sampling interval Δ. If we wish to study how the properties of the number of trades in intervals of length Δ change with Δ then it is useful to think of $N(u)$ as a continuous time process which is discretised to produce the $\{N_n\}$ via

$$N_n = N\left\{ (n+1)\Delta - \right\} - N(n\Delta).$$

In this section we follow the suggestion of Rydberg and Shephard (2000) of using Cox processes to model $N(u)$. From this model we derive the implied second order properties of the $\{N_n\}$. We will show that for a fixed value of Δ some configurations of Cox processes lead to second order properties for $\{N_n\}$ which are exactly the same as those implied by the BIN model. This allows us to think of the BIN model as an approximation to the Cox process and to predict how the weak BIN model will behave if applied to the same data record with a variety of values of Δ.

In order to start this study we will need some definitions. We first define a stochastic process $\Lambda(u)$ such that:

- $\Lambda(0) = 0$.
- $\Lambda(u) < \infty$ for all $u < \infty$.
- $\Lambda(u)$ has non–decreasing realizations.

This process is said to be a random measure. With this definition we are able to recall the formal definition of a Cox process.

Definition 3 (e.g. Grandell (1997)). Let $\left\{\widetilde{N}(u)\right\}$ be a standard Poisson process and further let Λ, a random measure, and $\widetilde{N}$ be independent of each other. Then the point process $N(u) = \widetilde{N}\{\Lambda(u)\}$ is called a Cox process.

A elegant discussion of Cox processes from the viewpoint of subordination is given in Cox and Miller (1965, p. 154). The subordination interpretation of continuous time stochastic processes is familiar in economics in the case of Brownian motion, for a subordinated Brownian motion is a stochastic volatility model (e.g. Clark (1973) and Barndorff-Nielsen and Shephard (1999)). The Cox process is the point process analogue of this.

3.2 Integrated intensity

The random measure Λ is linked to the intensity λ via the integral equation[9]

$$\Lambda(u) = \int_0^u \lambda(s)ds,$$

where $\{\lambda(u)\}$ is a positive stochastic process. As a result, in the context of Cox processes we will call $\{\Lambda(u)\}$ the *integrated intensity* process, although in the probability literature it is usually called the intensity measure.

Remark 9 Immediately

$$N_n|\lambda_n \sim Po(\lambda_n), \qquad \text{where} \qquad \lambda_n = \Lambda\{(n+1)\Delta-\} - \Lambda(n\Delta). \tag{12}$$

It is possible to work out the autocorrelation pattern of $\{N_n\}$ simply under the condition that $\{\lambda(u)\}$ is covariance stationary. This work follows some related ideas on stochastic volatility due to Barndorff-Nielsen and Shephard (1999). These general results will be helpful in allowing us to formally relate BIN models to Cox processes. First we need some notation.

Definition 4 Let ξ, ω^2 and r denote, respectively, the mean, the variance and the autocorrelation function of the process $\lambda(u)$. Further define

$$r^{**}(s) = \int_0^s \int_0^v r(u)dudv \quad \text{and} \quad \diamondsuit r^{**}(s) = r^{**}(s+\Delta) - 2r^{**}(s) + r^{**}(s-\Delta).$$

The next Lemma will look at the variance of the integrated intensity and the autocovariance function of the $\{\lambda_n\}$ process.

Lemma 3.2.1 (Barndorff-Nielsen and Shephard (1999)). Assuming that $\lambda(u)$ is square integrable and stationary, then

$$Var\{\Lambda(u)\} = 2\omega^2 r^{**}(u) \quad \text{and} \quad Cov\{\lambda_n, \lambda_{n+s}\} = \omega^2 \diamondsuit r^{**}(\Delta s).$$

□

[9] This type of relation is familar in econometrics for integrated volatility is the integral of the instantaneous volatility in a stochastic volatility model. See, for example, Barndorff-Nielsen and Shephard (1999).

3.3 Properties of counts

The Corollary given below will use the results in Lemma 3.2.1 to give us the variance of the number of counts in intervals of length of Δ. Further, it gives the fundamental autocovariance function, that is the second order properties, of the $\{N_n\}$ process. The result is completely general.

Corollary 3.3.1 Assuming $\{\lambda(u)\}$ is covariance stationary and square integrable, then

$$E\{N_n\} = E(\lambda_n) = \Delta\xi,$$

$$Var\{N_n\} = Var\{\lambda_n\} + E(\lambda_n) = 2\omega^2 r^{**}(\Delta) + \Delta\xi,$$

and

$$Cov\{N_n, N_{n+s}\} = Cov\{\lambda_n, \lambda_{n+s}\} = \omega^2 \diamondsuit r^{**}(\Delta s). \tag{13}$$

□

In turn the autocorrelation function is given in the following corollary.

Corollary 3.3.2 Assuming $\{\lambda(u)\}$ is covariance stationary and square integrable, then

$$Cor\{N_n, N_{n+s}\} = q \diamondsuit r^{**}(\Delta s), \qquad \text{where} \qquad q = \frac{\omega^2}{2\omega^2 r^{**}(\Delta) + \Delta\xi}.$$

□

Corollaries 3.3.1 and 3.3.2 give a complete discussion of the second order properties of the $\{N_n\}$ process for any Cox process. We now turn our attention to special cases which illustrate the power of this analysis.

Example 5 Suppose that

$$r(s) = \exp(-\beta^* |s|)$$

for some $\beta^* > 0$. Then, for $s > 0$,

$$\diamondsuit r^{**}(\Delta s) = \beta^{*-2}(1 - e^{-\beta^*\Delta})^2 e^{-\beta^*\Delta(s-1)}$$

which falls exponentially with s. Hence

$$Cor\{N_n, N_{n+s}\} = \frac{\omega^2}{2\omega^2 r^{**}(\Delta) + \Delta\xi} \diamondsuit r^{**}(\Delta s) = ce^{-\beta^*\Delta(s-1)}.$$

where

$$c = \frac{\omega^2(1 - e^{-\beta^*\Delta})^2}{2\omega^2\left(\beta^*\Delta + e^{-\beta^*\Delta} - 1\right) + \beta^{*2}\Delta\xi}.$$

This is an interesting result for the correlogram for $\{N_n\}$ is precisely the correlogram of a weak BIN(1,1) model[10]. Then, when the moments exist,

$$E(N_n) = \frac{1}{1 - \gamma - \beta}, \qquad Var(N_n) = E(N_n)\left\{\frac{1 + \beta^2 - 2(\gamma + \beta)\beta}{1 - (\gamma + \beta)^2}\right\}$$

and

$$Cor(N_n, N_{n-s}) = (\gamma + \beta)^{s-1} \frac{\{1 - \beta(\gamma + \beta)\}\gamma}{1 + \beta^2 - 2\beta(\gamma + \beta)}, \qquad s > 0.$$

That is the model has the same second order properties as a BIN(1, 1), which is identical to that derived for the Cox process with $r(s) = \exp(-\beta^* |s|)$. In particular we can relate,

[10] This style of result is similar to work on stochastic volatility and GARCH models put forward by Meddahi and Renault (1996) and Barndorff-Nielsen and Shephard (1999).

$$\exp(-\beta^*\Delta) \qquad \text{to} \qquad (\gamma+\beta),$$

and

$$\Delta\xi \quad \text{to} \quad \frac{1}{1-\gamma-\beta}$$

as well as other features in the model. The continuous time model thus predicts, for example, how $\gamma+\beta$ will vary with different values of Δ.

Remark 10 The autocorrelation $r(s)=e^{-\beta^*|s|}$ results from at least two interesting processes. The simplest is the Ornstein-Uhlenbeck process

$$d\lambda(u) = -\beta^*\lambda(u)dt + dz(\beta^* u), \tag{14}$$

with $0<\beta^*<\infty$, where the process z is a homogeneous Lévy process with positive increments with $Var(\{z(1)\})$ existing. This class of processes is studied at length in Barndorff-Nielsen and Shephard (1999). The same autocorrelation function results from the 'constant elasticity of variance' process

$$d\lambda(u) = -\beta^*\left\{\lambda(u)-\xi\right\}dt + \gamma\left\{\lambda(u)\right\}^d dW(u), \quad d\geq 1/2, \tag{15}$$

where $W(u)$ is standard Brownian motion. This general structure, which is always covariance (and strictly) stationary if $0<\beta^*<\infty$, has been highlighted by Meddahi and Renault (1996).

Example 6 The exponential damp down generalises to

$$r(s) = \sum_{j=1}^{P} w_j \exp\left(-\beta_j^*|s|\right), \quad \text{such that} \quad \sum_{j=1}^{P} w_j = 1.$$

In particular if we write $r_j^{**}(s)=\int_0^s r_j^*(u)du$, then

$$\diamondsuit r^{**}(s) = \sum_{j=1}^{P} w_j \diamondsuit r_j^{**}(s) = \sum_{j=1}^{P} v_j \exp\left\{-\beta_j^*\Delta\left(s-1\right)\right\},$$

where

$$v_j = w_j\beta_j^{*-2}\left\{1-\exp\left(-\beta_j^*\Delta\right)\right\}^2.$$

Remark 11 This is the autocorrelation function of a weighted sum of independent OU processes and constant elasticity of variance models. This allows us to model processes with short and longer term memory components in the intensity. This process has the same autocorrelation function as a weak BIN(p,1) model.

Remark 12 An alternative structure is to write down a Gaussian OU process for the log of the intensity process

$$d\log\lambda(t) = -\beta^*\left\{\log\lambda(t)-\mu\right\}dt + \varsigma db(t), \quad \beta^*>0, \tag{16}$$

where $b(t)$ is a Brownian motion. This process has some advantages as it has a simple strong solution, while

$$r_{\log}(u) = cor\left\{\log\lambda(t+u),\log\lambda(t)\right\} = \exp\left(-\beta^*|u|\right), \quad \beta^*>0.$$

However, it is not obvious how to work with Λ in this framework without making discretisation errors.

3.4 Temporal aggregation

By specifying a Cox process for $\{N(u)\}$ we have been able to study the implied second order properties of $\{N_n\}$ at different values of Δ. For some choices of stochastic model for the intensity the second order properties of $\{N_n\}$ are always of the form which would have been generated by a weak BIN process. Hence we are able to study the effect of time aggregation, the lengthening of Δ, on the dynamics of the BIN model.

Instead of working via a continuous time Cox process, we could study the effect of time aggregation on the discrete time BIN models directly. This is perhaps less elegant and links less well with our continuous time model of the price given in (1), but it does have the advantage of being mathematically less demanding. In order to carry this out, let us define for an integer Δ, N_n as

$$N_n = N\{(n+1)\Delta -\} - N(n\Delta) = \sum_{j=0}^{\Delta-1} N^{(1)}_{n\Delta+j},$$

where

$$N^{(1)}_n = N\{(n+1) -\} - N(n),$$

the number of trades we observe during an interval of length one. The approach we follow is to assume $\left\{N^{(1)}_n\right\}$ is a weak BIN process and ask if we can derive the second order properties of N_n.

The effect on the second order properties of time aggregation is a classic time series problem, with a large associated literature. A complete treatment for any covariance stationary BIN model, in the frequency domain, is available by working with the spectrum of the ARMA representation of the process (see, for example, the textbook exposition in Harvey (1989, p. 321) of flow variables). In the special case of a BIN(1,1) model there are some very elegant results pointed out, in the context of a GARCH process, by Drost and Nijman (1993) which carry over to our setup. If we write the main parameters of the BIN(1,1) model as in (6), then N^{Δ}_n has a weak BIN(1,1) representation with autoregressive root of $(\gamma+\beta)^{\Delta}$. This precisely matches the previous prediction from the analysis of Cox processes.

4 EMPIRICAL WORK

THIS SECTION IS EXTREMELY PRELIMINARY.

4.1 Background

The intensity of trading on the NYSE varies considerably through time. In this subsection we study the basic features of the observed sequence of $\{N(u)\}$ for the IBM stock during the whole of 1995, where the trades took place on the floor of the NYSE. The top graph of Figure 2 shows an estimate of the average number of trades which occur at each second for each day of the week, written $\{D_{jn}, j = 1, ..., 5\}$. The estimate is generated using a natural cubic spline with a different bandwidth selected by generalised cross-validation for each day of the week (see, for example, Green and Silverman (1994)). We can see that for each day trading is brisk in the morning hours, slows down around lunch time and picks up again in the afternoon. In addition there are changes in these patterns between the days of the week. In particular Monday mornings and Friday afternoons are comparatively inactive, while the first 30 minutes of Friday mornings are the most active trading period of the week during 1995 for the IBM stock. Finally, we can see that the first ten minutes of each day are unlike most of the rest of the day — for the activity rate changes very dramatically during this time.

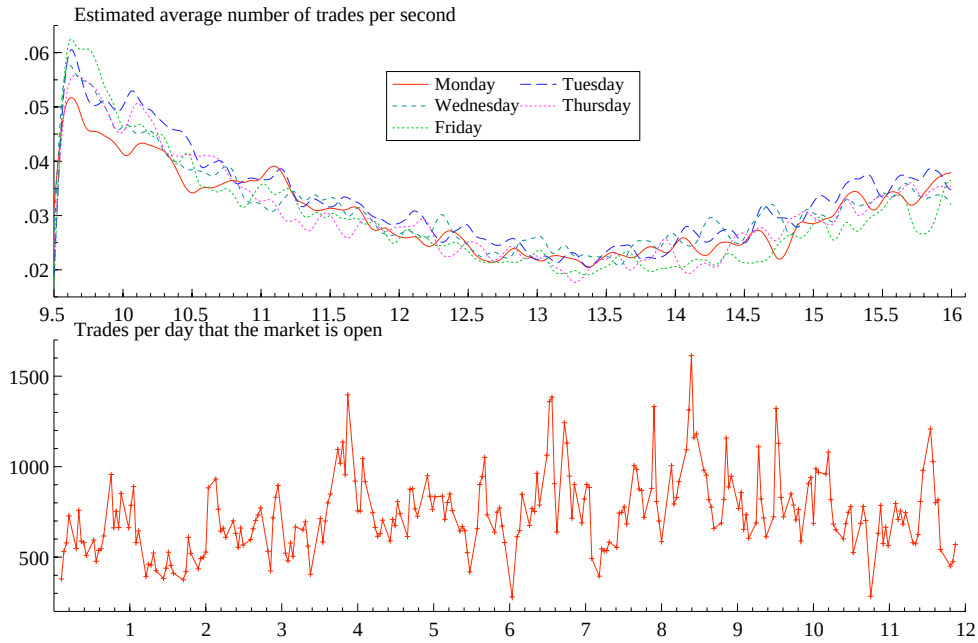

Figure 2: *Top graph is the intensity of the trading per second. Estimated daily curves using a spline with the smoothness penalty selected using generalised cross validation. Bottom graph is the number of trades per day for each day the market is open during the year.*

The bottom graph of Figure 2 shows the number of trades on each day that the NYSE was open. We can see very significant changes in the activity level during the year, with low levels at the beginning of the year and high levels in September and October. Some of the variation of this series arises due to a seasonal component. However, there is also important serial dependence in the series.

4.2 Dynamics of $\{N_{jn}\}$

When we construct our counts, we can do this for each second of each day. In order to have a notation for this we introduce an additional subscript, j to denote the $j-th$ active day of 1995. Throughout we will study the dynamics of the random field $\{N_{jn}\}$ with Δ set to one second, focusing on estimating very (perhaps overly) simple models which allow interpretation. Our first analysis is to look at the daily time series generated by looking at the difference between the $\{N_{jn}\}$ sequence in each day and the corresponding daily seasonal pattern given in Figure 2. For each day we computed the correlogram and plotted the average of these 251 correlograms in Figure 3. This picture shows a negative correlation at lag one, followed by very significant (although quite small) correlations at longer lags. These die down quite slowly, but are mostly irrelevant after 5000 seconds.

The negative correlation at small lags seem a consistent feature of this data when we take Δ to be a second. The positive correlations at other lags are more important to the overall dynamics of the counting process as they are sustained over a large number of lags.

4.3 Every day is different (EDID): 4 Tuesdays in 1995

Our empirical analysis will be initially based on modelling each day separately, starting off with

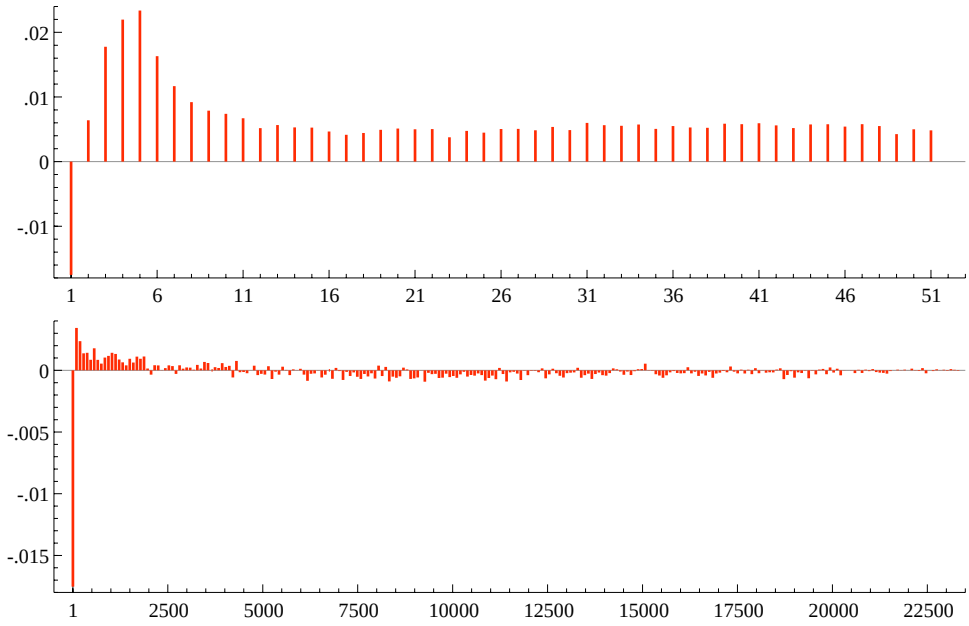

Figure 3: *Averaged correlogram for 251 active days. For each day we computed the correlogram for the day using the 23,400 second by second data. Top correlogram shows first 50 lags, bottom a thinned version of 23,400 lags.*

the first four Tuesdays in 1995. We call this the "every day is different" (EDID) model. We are able to estimate models of this type, as we have a large amount of data for each day, so long as the dynamics of the process is not very extensive nor long lived.

Table 1 looks at the dependence structure in the counts. We define M_{jn} to be $I(N_{jn} > 0)$, then the Table shows that for the second Tuesday there are 762 seconds in which there are trades. However, there were only five seconds when there were consecutive trades — which is massively under what we would expect from a heterogeneous Poisson process. This forces us towards a Cox or BIN process of some type. At longer lags of time the dependence becomes positive, rather than negative. The EDID model we initially analyse has the following form

$$N_{jn}|\mathcal{F}_{jn\Delta-} \sim Po\left\{\lambda_{jn} + \beta x_{jn}\right\},$$

where $\{x_{jn}\}$ denotes the seasonal pattern of the process estimated using the splines described above and

$$\lambda_{jn} = \gamma_j N_{jn-1} + \delta_j \lambda_{jn-1}, \quad \text{where} \quad \lambda_{j0} = \beta_j x_{j1} \frac{(1-\delta_j)}{(1-\delta_j-\gamma_j)}.$$

Inevitably the choice of this EDID model structure is arbitrary. We make the following points:

- The unconventional choice of the additive structure $\lambda_{jn} + \beta_j x_{jn}$, rather than the more obvious $\lambda_{jn} \exp(\beta_j \log x_{jn})$, is motivated by a desire to calculate analytically temporal aggregation features of this process. The lack of an intercept in the model is due to it testing out in some initial empirical experiments.

- We ignore cooling effects (modelling the first few lags which have negative dependence), although this will lead to empirical failures of the model, again in order to allow us to understand temporal aggregation.

	Tuesdays in 1995			
	First	Second	Third	Fourth
$\sum M_n$	377	762	480	662
$\sum M_n M_{n-1}$	9	5	11	8
$\sum M_n M_{n-2}$	12	14	27	14
$\sum M_n M_{n-3}$	19	26	37	22
$\sum M_n M_{n-4}$	14	50	22	27
$\sum M_n M_{n-5}$	6	54	16	36
$\sum M_n M_{n-6}$	8	48	15	42
E Ind	6	25	10	19
SE Ind	2	5	3	4

Table 1: $M_{jn} = I(N_{jn} > 0)$, *an indicator if there is a trade during the n-th time period of length* Δ. *"Independence" figure is the expected number of* $\sum M_{jn}M_{jn-s}$ *for all* s *when* $N(u)$ *is a time heterogenous Poisson process with expected number of events over the 23400 seconds of* $\sum M_{jn}$.

- The choice of initial condition, which is arbitrary as N_{jn} is non-stationary over n due to the time varying regression effect, is taken as the unconditional expectation of the approximating model which does have a stationary solution

$$N_{jn}|\mathcal{F}_{jn\Delta-} \sim Po\left\{\lambda_{jn} + \beta_j x_{j1}\right\}, \quad \lambda_{jn} = \gamma_j N_{jn-1} + \delta_j \lambda_{jn-1}.$$

Table 2 gives results for the estimation of this model, together with the corresponding result when we take out all of the dynamics in the process. The Table also gives some summary measures of fit through the use of the residuals

$$u_{jn} = N_{jn} - \lambda_{jn} - \beta_j x_{jn},$$

computed for each second of each day. If the models were correct then they should indicate white noise. To assess this we work with $\{r_{sj}, s = 1, 2, ...\}$, the serial correlation coefficients, of the residuals for the j-th day of the series $\{u_{jn}\}$. To summarise these coefficients we will use conventional Box-Pierce statistics

$$BP(q)_j = T\sum_{s=1}^{q} r_{sj}^2.$$

In the context of models with estimated dynamics we refer this statistic to a χ^2_{q-2} distribution, although this has yet to be theoretically justified. Finally, we expect misspecification in the model for the first few lags and so we sometimes look at the difference of the Box-Pierce statistics over various lags. In particular we favour looking at $BP(100)_j - BP(20)_j$, which only involves correlations at longer than 20 seconds. This should be roughly a χ^2_{80} if the model is well specified at those longer lags.

However, we have already noted that our model ignores cooling and so we would expect to fail conventional Box-Pierce type measures of fit. In order to reinforce this point the Table also gives the corresponding results for the cooled version of the model with 8 lags

$$N_{jn}|\mathcal{F}_{jn\Delta-} \sim Po\left\{\left(\lambda_{jn} + \beta_j x_{jn}\right)\exp\left(\mu_n\right)\right\}, \quad \mu_n = \sum_{j=1}^{8} \zeta_j N_{n-j}.$$

	Tuesdays in 1995							
	First		Second		Third		Fourth	
β_j	0.313	0.242	0.459	0.504	0.424	0.324	0.636	0.467
γ_j	0.00304	0	0.00458	0	0.00849	0	0.00406	0
δ_j	0.991	0	0.992	0	0.975	0	0.987	0
$log-like$	-1904.7	-1913.6	-3304.9	-3332.1	-2261.7	-2282.7	-2945.3	-2954.7
$BP(20)$	40.4	54.6	111	151	92.9	132	62.0	72.7
$BP(20,100)$	89.7	97.7	72.3	121	133	168	106	116
	FULLY COOLED MODEL — EIGHT LAGS							
γ_j	0.00177		0.00339		0.00413		0.00160	
δ_j	0.995		0.993		0.986		0.996	
$log-like$	-1893.7		-3258.8		-2240.0		-2923.6	
$BP(20)$	8.0		14.6		13.0		11.4	
$BP(20,100)$	93.0		69.7		131		104	

Table 2: *Fitting of the BIN model using a EDID structure. Above the triple line the model is estimated without any cooling — estimated BIN(1,1) model and a model with no dynamics (the dynamic parameters are set to zero in the Table). Below the line the same model is estimated but employing 8 lags of cooling. The eight parameters of the cooling structure plus the corresponding β_j are not reported here.*

With this feature the fitted model generally does an adequate job at fitting the data. However, the cooled model does not really change the longer term features of the model, which are the ones which predominately interest us.

Overall the introduction of the time series modelling of $\{N_{jn}\}$ has an important impact on the log-likelihood function. However, the improvement in fit is not overwhelming. Further, the modelling of the short-term dynamics via cooling is as important from a likelihood viewpoint.

4.4 Day-by-day dynamic (DBDD) model

The EDID model allows all of the parameters of the model to change every day and so it is impossible to use this model structure to understand any aspect of the evolution of trading activity over a number of days. In this subsection we constrain the EDID model to have common features over the day. We call this approach a day-by-day dynamic (DBDD) model.

We will assume that the dynamic parameters of the model, the $\{\gamma_j, \delta_j\}$ do not change over time. We will provide an empirical assessment of the validity of this assumption. This provides a link between the models for each day. For the moment we will allow the regression effect $\{\beta_j x_{jn}\}$ to vary over days, freely estimating the $\{\beta_j\}$ parameters. As such this DBDD model again provides no concrete link between the activity in one day and the next. We will also set

$$\lambda_{j0} = \beta_j x_{j1} \frac{(1-\delta)}{(1-\delta-\gamma)},$$

which varies because of the changing $\{\beta_j\}$ parameters.

The DBDD model structure has one parameter per day plus two dynamics parameters for $\{\delta, \gamma\}$. To see the effectiveness of this model structure we have fitted it to the first 100 active days of 1995. The results are given in Table 3 and Figure 4. The Table focuses mostly on the likelihood ratio statistics for each day, looking at assessing the change in the likelihood caused by the constraint of a common dynamic structure across days. To benchmark these results we

also give the likelihood ratios for the case where there are no dynamics structure assumed at all — for it is now possible that this model structure would provide a better fit than our common dynamics model. The Table also gives some diagnostics for the model. This is based on an Box-Pierce type statistic generated from the averaged correlogram for the residuals — that is for each day we compute the residual and their corresponding correlogram. These are averaged and we place them into the following Box-Pierce type statistic

$$\overline{BP}(q) = TN \sum_{s=1}^{q} r_{s.}^2, \quad \text{where} \quad r_{s.}^2 = \frac{1}{N} \sum_{j=1}^{N} r_{sj}.$$

This statistic should be approximately χ_q^2, although we see this fails rather dramatically at $q = 20$. The differenced version is more satisfactory, again because it is not so effected by the cooling effect we have not modelled.

Figure 4 gives a detailed description of the fitted model. An important feature of the figure is that we give, in the top left hand time series graph, an estimate of the average level of activity estimated by the model

$$\overline{\lambda}_j = \widehat{\beta}_j \overline{x}_j \frac{(1-\delta)}{(1-\delta-\gamma)},$$

where $\overline{x}_j$ is the average level of the series for that day of the week over the entire year. This picture also gives the time series of $\overline{N}_{.j}$ and it shows these two series are very close. Hence we can see that a good approximation to

$$\widehat{\beta}_j = \overline{N}_{j.} \frac{(1-\delta-\gamma)}{\overline{x}_j(1-\delta)}.$$

This observation will become crucial later.

Figure 4 also gives the cumulative correlograms for the residual for the first six days of the year for the DBDD model. This shows the scatter of this measure of memory in the residuals. The bottom left graph draws $\left\{r_{s.}^2\right\}$ for the DBDD model and the corresponding EDID model. It is not possible to see the difference between these two graphs. Both correlograms look reasonable outside the first few lags. The figure also gives, as dots, the corresponding correlogram for the non-dynamic model, which seems poor at all lags. The corresponding cumulative correlograms are given in the bottom right hand side of the figure. They show how poor the non-dynamic model is, while the DBDD model seems as satisfactory as the EDID model. However, both of these models seem to allow very small amounts of positive correlation at long lags. This is a potentially serious problem with the model.

	DBDD	EDID	No dynamics
γ	0.00578		
δ	0.987		
LR (compared to DBDD)		933	-4356
$BP(20)$	5262	5563	8426
$BP(20,100)$	268	190	3623

Table 3: *Various models for the first 100 active days of 1995, with no cooling. LR is twice the difference in the log-likelihood compared to the DBDD model. EDID has 198 more parameters than DBDD. The "no dynamics" model has 2 less parameters than DBDD. BP(20) should be less than 30, BP(100,20) less than 100.*

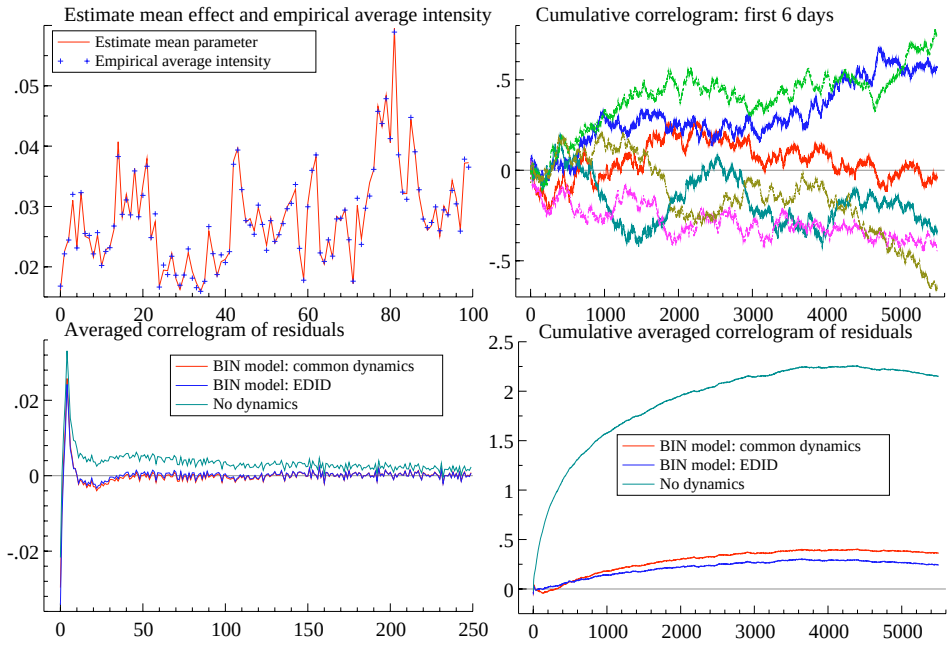

Figure 4: *Top left:* $\overline{\lambda}_j$ *is graphed against day* (j) *with the sample daily average of the* N_{jt}. *Top right: cumulative correlograms for the residuals from the fitted model for the first three days. Bottom left: averaged correlogram for residuals for fitted model, for EDID model (almost identical) and for the no dynamics case. Bottom right: corresponding cumulative correlograms.*

4.5 Modelling daily totals

We have seen the intimate connection between the best fitting parameters $\{\beta_j\}$ and the daily totals of trades $\left\{\overline{N}_{j.}\right\}$. It thus becomes important to be able to predict this quantity if we are to create a model which allows us to flow from one day to the next. An initial analysis of the daily totals is give in Figure ?. This shows the distinct seasonal pattern in these totals (estimated using a spline) and some serial dependence in the residuals, which is well approximated by a first order autoregression with a parameter of around 0.55.

Our basic model for

$$\overline{N}_{j.} \simeq \kappa_0 + \kappa_1 \overline{N}_{j-1.} + \kappa_2 \lambda_{Sj-1}.$$

5 MULTIVARIATE MODELS

5.1 Framework

An advantage of our approach to dealing with the intensity of trading activity is that the extension to the multivariate case is straightforward. Here we will briefly discuss this development, leaving a full treatment to be carried out elsewhere.

Let us write $\{N(u)\}_{u\geq 0}$ as a $K \times 1$ vector which has as its $j-th$ element the number of trades recorded up until time u in the $j-th$ stock. Then define the vector differencing operator

$$N_n = N\left\{(n+1)\,\Delta-\right\} - N(n\Delta).$$

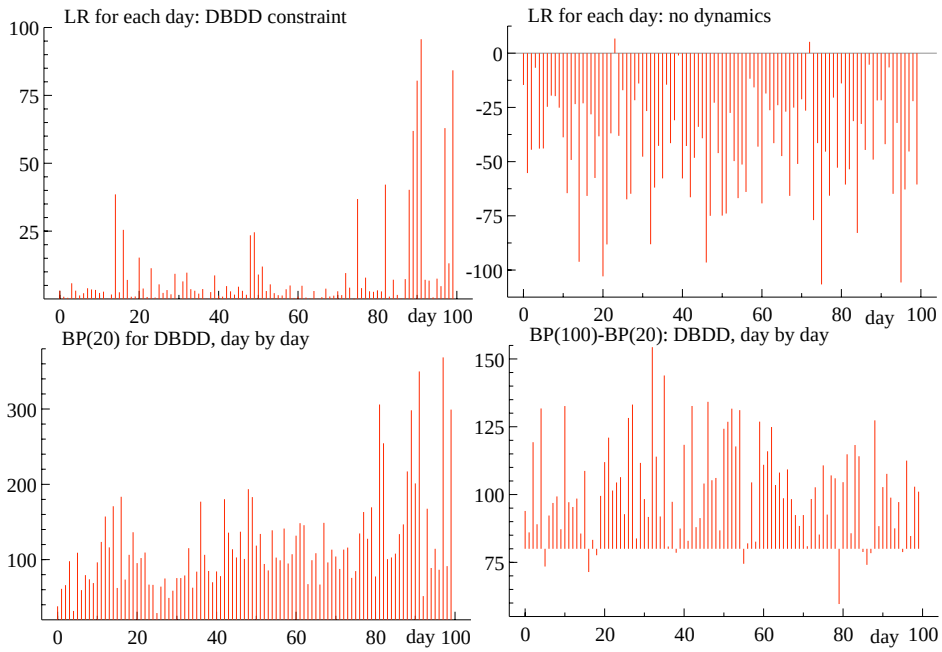

Figure 5: *Top left: LR to test, each day, the constraint that the dynamics are common. Values larger than 6 are worrying. Top right: LR to test, for each day, the constraint that the dynamics should be set to be zero. Values less than zero support the dynamic model. Bottom left:* $BP(20)_j$ *for each day using residuals from the common dynamic model. Values much larger than 30 are worrying.* $BP(100,20)_j$ *for each day using residuals from the common dynamic model. Values larger than 100 are worrying.*

It represents the number of trades occurring in a time interval of length Δ. A standard multivariate BIN(1) model would have the $\{N_{jn}|\mathcal{F}_{n\Delta-}\}$ being conditionally independent over j and

$$N_{jn}|\mathcal{F}_{n\Delta-} \sim Po(\lambda_{jn}),$$

where

$$\lambda_n = \alpha + \gamma N_{n-1},$$

and $\mathcal{F}_{n\Delta}$ is the information set available at time $n\Delta$ for predicting the vector of counts. Each element of the $K \times 1$ vector α and the $K \times K$ matrix γ needs to be non-negative. If γ is diagonal then the series will be independent. More generally γ can be non-symmetric. We will assume that the eigenvalues of γ, which can be complex, are less than one in absolute value. Of course we could generalise this structure by using a multivariate Poisson type model which allowed dependence between the $\{N_{jn}\}$ even when $\gamma = 0$ — see Johnson, Kotz, and Balakrishnan (1997, Ch. 37).

In the BIN model we have that

$$N_n = \alpha + \gamma N_{n-1} + u_n, \quad \text{where} \quad u_n = N_n - \lambda_n.$$

As a result, for $s > 0$,

$$E(N_n) = (I - \gamma)^{-1}\alpha, \quad Cov(N_n, N_{n-s}) = \gamma^s Cov(N_n)$$

and

$$Cov(N_n) = \gamma Cov(N_n)\gamma' + diag\{E(N_n)\}.$$

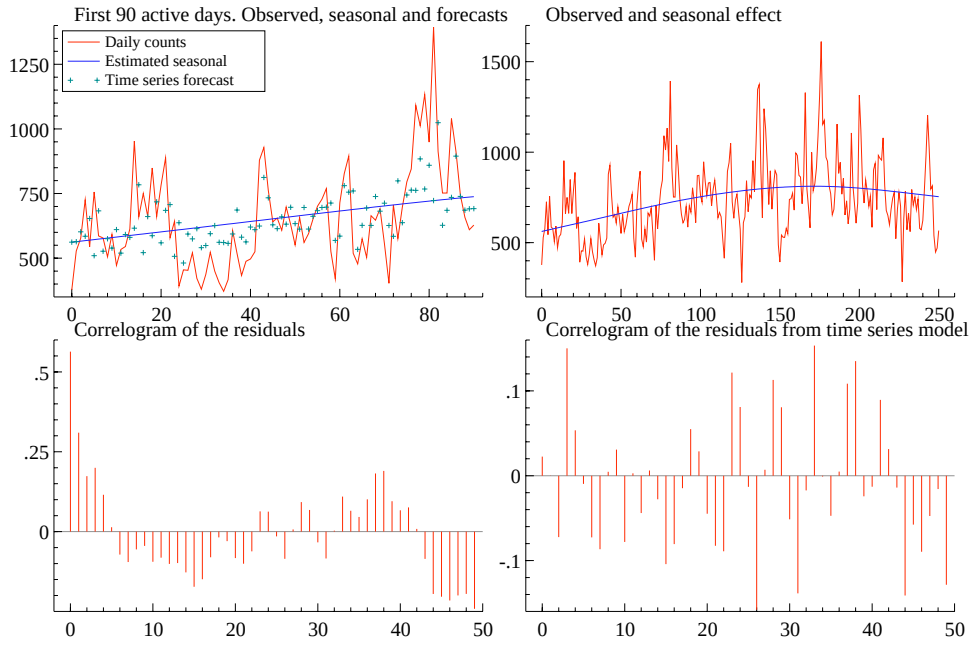

Figure 6:

As a result

$$vec\{Cov(N_n)\} = (I - \gamma \otimes \gamma)^{-1} vec\left[diag\left\{E(N_n)\right\}\right].$$

This model structure clearly generalises to where

$$\lambda_n = \alpha + \gamma N_{n-1} + \delta \lambda_{n-1}.$$

Again we will require that $K \times 1$ vector α and the $K \times K$ matrices γ and δ to be non-negative. This model has an ARMA(1,1) structure

$$\begin{aligned} N_n &= \alpha + \gamma N_{n-1} + \delta \lambda_{n-1} + u_n \\ &= \alpha + \gamma N_{n-1} + \delta \left(N_{n-1} - u_{n-1}\right) + u_n \\ &= \alpha + (\gamma + \delta) N_{n-1} + u_n - \delta u_{n-1}, \end{aligned}$$

which can be analysed using standard multivariate ARMA models with white noise error terms (see, for example, Reinsel (1993, Ch. 2)). The extension to deal with more complicated dynamics is immediate.

5.2 Cox processes

We can think about N_n as a discretisation of a multivariate Cox process. Let us define a $K \times 1$ vector stochastic process $\Lambda(u)$ such that for each element $\Lambda_j(u)$, $j = 1, ..., K$:

- $\Lambda_j(0) = 0$.
- $\Lambda_j(u) < \infty$ for all $u < \infty$.
- $\Lambda_j(u)$ has non–decreasing realizations.

Then we give a formal definition of a multivariate Cox process.

Definition 7 Let $\left\{\widetilde{N}(u)\right\}$ be a standard vector Poisson process and further let Λ, a random measure, and $\widetilde{N}$ be independent of each other. Then the point process $N(u) = \widetilde{N}\{\Lambda(u)\}$ is called a multivariate Cox process.

Importantly, for i not equal to j

$$N_j(u) \perp\!\!\!\perp N_i(u)|\Lambda(u)$$

implying any codependence between the marginal Cox processes can only be introduced via the vector integrated intensity process $\Lambda(u)$ which we setup via the integral equation

$$\Lambda(u) = \int_0^u \lambda(s)ds,$$

where the intensity process $\{\lambda(s)\}_{s\geq 0}$ is a positive stochastic process. Further

$$\Pr\left(N_n = 0|\lambda_n\right) = \prod_{j=1}^{K}\left(1-\lambda_{jn}\right) + o(\Delta), \qquad \text{where} \qquad \lambda_n = \Lambda\left\{(n+1)\,\Delta\right\} - \Lambda(n\Delta),$$

implying the probability of observing a trade in more than one asset in a very small period of time is close to zero.

How do we construct a co-dependent intensity process $\{\lambda(s)\}$? One way of carrying this out is to work with a log-based OU process

$$d\log\left\{\begin{array}{c}\lambda_1(t)\\ \lambda_2(t)\\ \\ \lambda_K(t)\end{array}\right\} = -\beta^*\left[\log\left\{\begin{array}{c}\lambda_1(t)\\ \lambda_2(t)\\ \\ \lambda_K(t)\end{array}\right\} - \mu\right]dt + \varsigma db(t), \quad \beta^* > 0,$$

where $b(t)$ is vector standard Brownian motion. An alternative is to work with a factor style model of intensity. This will be based on $P+K$ common latent non-negative intensities $\{\lambda_j(s)\}$, which will be assumed independent of one another, with

$$\lambda(s) = B\left\{\begin{array}{c}\lambda_1(t)\\ \lambda_2(t)\\ \\ \lambda_P(t)\end{array}\right\} + \left\{\begin{array}{c}\lambda_{P+1}(t)\\ \lambda_{P+2}(t)\\ \\ \lambda_{P+K}(t)\end{array}\right\}.$$

In this model each element of the B matrix will be constrained to be non-negative. This style of model is easier to work with in terms of integrated intensity than the log-based vector OU process.

6 ACKNOWLEDGMENTS

Tina Rydberg thanks the Danish National Research Council for their financial support through a post-doctoral fellowship. Both authors are grateful for support from The Centre for Analytical Finance, Aarhus, Denmark. We thank Richard Spady for various helpful comments on this line of research, while we have a big debt to Ole Barndorff-Nielsen for his work on SV models (with Neil Shephard) which allowed us to better understand the behaviour of $\{N_n\}$. We alone are responsible for any errors in the paper.

7 APPENDIX

Proof of Theorem 2.1 We can construct a Martingale difference (MD) error term $\{u_n\}$ by writing

$$\begin{aligned} N_n &= \lambda_n + (N_n - \lambda_n) \\ &= \alpha + \gamma N_{n-1} + u_n. \end{aligned}$$

It follows immediately that $\kappa_1 = \alpha/(1-\gamma)$. Further

$$\begin{aligned} \kappa_2 &= E\{Var(N_n|\mathcal{F}_{n\Delta-})\} + Var\{E(N_n|\mathcal{F}_{n\Delta-})\} \\ &= E(\lambda_n) + Var(\lambda_n) \\ &= E(N_n) + \gamma^2 Var(N_{n-1}), \end{aligned}$$

which gives the result for κ_2 immediately. The result on the autocorrelation follows as, for $s > 0$,

$$\begin{aligned} Cor(N_n, N_{n-s}) &= \gamma Cor(N_{n-1}, N_{n-s}) + Cor(u_n, N_{n-s}) \\ &= \gamma Cor(N_{n-1}, N_{n-s}) \\ &= \gamma^s. \end{aligned}$$

Now we turn to the cumulant generating function

$$\begin{aligned} m(\theta \ddagger N_n) &= E\{m(\theta \ddagger N_n|\mathcal{F}_{n\Delta-})\} \\ &= E\{\exp(\psi\lambda_n)\}, \quad \text{where} \quad \psi = \left(e^\theta - 1\right) \\ &= m(\psi \ddagger \lambda_n) \\ &= e^{\alpha\psi} m(\psi\gamma \ddagger N_{n-1}). \end{aligned}$$

Under stationarity then

$$\begin{aligned} \log m(\theta \ddagger N_n) &= k(\theta \ddagger N_n) \\ &= \alpha\psi + k(\psi\gamma \ddagger N_n) \\ &= \alpha\left(e^\theta - 1\right) + k\left\{\gamma\left(e^\theta - 1\right) \ddagger N_n\right\}. \end{aligned}$$

We write, for $r > 0$,

$$S_r(\theta) = \gamma^r e^{r\theta} \frac{\partial^r k\left\{\gamma\left(e^\theta - 1\right) \ddagger N_n\right\}}{\partial\left\{\gamma\left(e^\theta - 1\right)\right\}^r}.$$

A property of the $S_r(\theta)$ is that

$$\frac{\partial S_r(\theta)}{\partial\theta} = S_{r+1}(\theta) + rS_r(\theta), \quad r = 1, 2, ...$$

This allows us to compute

$$\frac{\partial k(\theta \ddagger N_n)}{\partial\theta} = \alpha e^\theta + S_1(\theta) \quad \text{and} \quad \frac{\partial^2 k(\theta \ddagger N_n)}{\partial\theta^2} = \frac{\partial k(\theta \ddagger N_n)}{\partial\theta} + S_2(\theta),$$

and recursively

$$\frac{\partial^r k(\theta \ddagger N_n)}{\partial\theta^r} = \frac{\partial k^{r-1}(\theta \ddagger N_n)}{\partial\theta^{r-1}} + \frac{\partial^{r-2} S_2(\theta)}{\partial\theta^{r-2}}, \quad r = 3, 4, ...$$

Finally, of course,

$$\kappa_r = \left.\frac{\partial^r k(\theta \ddagger N_n)}{\partial \theta^r}\right|_{\theta=0} = \left.\frac{\partial^r k\left\{\gamma\left(e^\theta - 1\right) \ddagger N_n\right\}}{\partial \left\{\gamma\left(e^\theta - 1\right)\right\}^r}\right|_{\theta=0}.$$

Then we note that

$$S_r(0) = \gamma^r e^{r\theta}\kappa_r.$$

As a result, for example,

$$\kappa_1 = \alpha + \gamma\kappa_1, \quad \kappa_2 = \kappa_1 + \gamma^2\kappa_2, \quad \kappa_3 = 3\kappa_2 + \gamma^3\kappa_3,$$

and

$$\begin{aligned}
\kappa_4 &= \kappa_3 + \frac{\partial^2 S_2(0)}{\partial\theta^2} = \kappa_3 + \frac{\partial S_3(0)}{\partial\theta} + 2\frac{\partial S_2(0)}{\partial\theta} \\
&= \kappa_3 + \{S_4(0) + 3S_3(0)\} + 2\{S_3(0) + 2S_2(0)\} \\
&= \left(1 + 5\gamma^3\right)\kappa_3 + 2\gamma^2\kappa_2 + \gamma^4\kappa_4.
\end{aligned}$$

Further, all higher order cumulants will be such that $(1-\gamma^r)\kappa_r$ is a function of cumulants of order less than r and so all the moments of the marginal distribution of $\{N_n\}$ exist if and only if $\gamma < 1$ by induction.
□

Proof of Theorem 2.3. Now

$$\begin{aligned}
N_n &= \lambda_n + (N_n - \lambda_n) \\
&= \alpha + (\gamma + \delta) N_{n-1} + u_n - \delta u_{n-1},
\end{aligned}$$

is a constrained ARMA(1,1) representation with a MD error term if $\gamma + \delta < 1$. It follows immediately that $\kappa_1 = \alpha/(1 - \gamma - \delta)$. Further as

$$Cov(N_n, \lambda_n) = E\left(\lambda_n^2\right) - \{E(\lambda_n)\}^2 = Var(\lambda_n),$$

so

$$\begin{aligned}
\kappa_2 &= E\{Var(N_n|\mathcal{F}_{n\Delta-})\} + Var\{E(N_n|\mathcal{F}_{n\Delta-})\} \\
&= E(\lambda_n) + Var(\lambda_n) \\
&= E(N_n) + \gamma^2 Var(N_{n-1}) + \delta^2 Var(\lambda_{n-1}) + 2\gamma\delta Cov(N_{n-1}, \lambda_{n-1}) \\
&= \kappa_1 + \gamma^2\kappa_2 + \left(\delta^2 + 2\gamma\delta\right) Var(\lambda_n) \\
&= \kappa_1 + \gamma^2\kappa_2 + \left(\delta^2 + 2\gamma\delta\right)(\kappa_2 - \kappa_1).
\end{aligned}$$

implying the required result as $Var(\lambda_n) = \kappa_2 - \kappa_1$. The autocorrelation function follows from standard ARMA theory. The moment generating function of the model is most easily expressed via the bivariate recursion

$$\begin{aligned}
m(\theta \ddagger N_n, \lambda_n) &= E\{\exp(\theta_1 N_n + \theta_2\lambda_n)\} \\
&= E[\exp\{(\psi + \theta_2)\lambda_n\}] \quad \text{where} \quad \psi = \left(e^{\theta_1} - 1\right) \\
&= e^{\alpha(\psi+\theta_2)} E[\exp\{(\gamma(\theta_2 + \psi) N_n + \delta(\theta_2 + \psi)\lambda_n)\}] \\
&= e^{\alpha(\psi+\theta_2)} m\{\gamma(\theta_2 + \psi), \delta(\theta_2 + \psi) \ddagger N_n, \lambda_n\},
\end{aligned}$$

yielding the cumulant generating function stated in the theorem.
□

References

Barndorff-Nielsen, O. E. and N. Shephard (1999). Non–Gaussian OU based models and some of their uses in financial economics. Unpublished discussion paper: Nuffield College, Oxford.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *J. Econometrics 51*, 307–327.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). ARCH models. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2959–3038. North-Holland.

Brockwell, P. J. and R. A. Davis (1987). *Time Series: Theory and Methods*. New York: Springer-Verlag.

Cameron, C. A. and P. K. Trivedi (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Clark, P. K. (1973). A subordinated stochastic process model with fixed variance for speculative prices. *Econometrica 41*, 135–156.

Cox, D. R. (1972). The statistical analysis of dependencies in point processes. In P. A. P. Lewis (Ed.), *Symposium on Point Processes*, pp. 55–66. New York: Wiley and Sons.

Cox, D. R. and H. D. Miller (1965). *The Theory of Stochastic Processes*. London: Chapman & Hall.

Cox, D. R. and D. Oakes (1984). *Analysis of Survival Data*. London: Chapman & Hall.

Drost, F. C. and T. E. Nijman (1993). Temporal aggregation of GARCH processes. *Econometrica 61*, 909–927.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika 84*, 669–84.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. F. and G. G. J. Lee (1999). A permanent and transitory component model of stock return volatility. Chapter 20, pp. 475–497.

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

Freeland, K. and B. McCabe (1998). Forecasting discrete valued low count time series. Working paper, Dept of Economics, Liverpool University.

Frey, R. and W. Runggaldier (1998). Nonlinear filtering techniques for estimation and risk management in partially observed stochastic volatility models. Unpublished paper: Mathematics Department, EZH.

Gamerman, D. (1992). A dynamic approach to the statistical analysis of point processes. *Biometrika 79*, 39–50.

Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.

Ghysels, E., J. Jasiak, and C. Gourieroux (1998). Stochastic volatility duration models. Unpublished paper: Pennsylvania State University. Presented at Second international conference on high frequency data in finance, Zurich, Switzerland, April.

Grandell, J. (1997). *Mixed Poisson Processes*, Volume 77 of *Monographs on Statistics and Applied Probability*. London: Chapman and Hall.

Green, P. and B. W. Silverman (1994). *Nonparameteric Regression and Generalized Linear Models: A Roughness Penalty Approach*. London: Chapman & Hall.

Grunwald, G. K., R. J. Hyndman, L. Tedesco, and R. L. Tweedie (1997). A unified view of linear AR(1) models. Unpublished paper: Dept. of Statistics, Colorado State University.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

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.

Hull, J. and A. White (1988). An analysis of the bias in option prices caused by stochastic volatility. *Advances in Futures and Options Pricing Research 3*, 29–61.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1997). *Discrete Multivariate Distributions*. New York: John Wiley.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.

MacDonald, I. L. and W. Zucchini (1997). *Hidden Markov and other models for discrete-values time series*. London: Chapman and Hall.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin 21*, 645–50.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability 20*, 822–35.

Meddahi, N. and E. Renault (1996). Aggregation and marginalization of GARCH and stochastic volatility models. Unpublished paper: CREST-INSEE.

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

Nelson, D. B. and C. Q. Cao (1992). Inequality constraints in the univariate GARCH model. *J. Business and Economic Statist. 10*, 229–35.

Reinsel, G. C. R. (1993). *Elements of Multivariate Time Series Analysis*. New York: Springer-Verlag.

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

Rydberg, T. H. and N. Shephard (1998a). Dynamics of trade-by-trade price movements: decomposition and models. Working paper, Nuffield College, Oxford. Presented at Workshop on Econometrics and Finance, Issac Newton Institute, Cambridge University, October 1998.

Rydberg, T. H. and N. Shephard (1998b). Modelling trade-by-trade price movements of multiple assets using multivariate compound Poisson processes. Working paper, Nuffield College, Oxford. Presented at Conference on Statistics and Finance, Hong Kong University, July 1999.

Rydberg, T. H. and N. Shephard (2000). A modelling 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. Forthcoming.

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

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–67.

Stein, E. M. and J. Stein (1991). Stock price distributions with stochastic volatility: an analytic approach. *Rev. Financial Studies 4*, 727–752.

Synder, D. L. and M. I. Miller (1991). *Random Point Processes in Time and Space* (2 ed.). New York: Springer-Verlag.

Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester: John Wiley.

Wold, H. (1948). On stationary point processes and Markov chains. *Skand. AktuarTidskr. 31*, 229–240.

Zeger, S. L. and B. Qaqish (1988). Markov regression models for time series, a quasi likelihood approach. *Biometrics 44*, 1019–1032.