Modeling Time Series Count Data: An Autoregressive Conditional Poisson Model

Andréas Heinen

Department of Economics
University of California, San Diego
9500 Gilman Drive
La Jolla CA 92037
aheinen@weber.ucsd.edu

November 2000

Abstract

This paper introduces new models for time series count data. The Autoregressive Conditional Poisson model (ACP) makes it possible to deal with issues of discreteness, overdispersion (variance greater than the mean) and serial correlation. A fully parametric approach is taken and a marginal distribution for the counts is specified, where conditional on past observations the mean is autoregressive. This enables to attain improved inference on coefficients of exogenous regressors relative to static Poisson regression, which is the main concern of the existing literature, while modeling the serial correlation in a flexible way. A variety of models, based on the double Poisson distribution of Efron (1986) is introduced, which in a first step introduce an additional dispersion parameter and in a second step make this dispersion parameter time-varying. All models are estimated using maximum likelihood which makes the usual tests available. In this framework autocorrelation can be tested with a straightforward likelihood ratio test, whose simplicity is in sharp contrast with test procedures in the latent variable time series count model of Zeger (1988). The models are applied to the time series of monthly polio cases in the U.S between 1970 and 1983 as well as to the daily number of price change durations of .75$ on the IBM stock. A .75$ price-change duration is defined as the time it takes the stock price to move by at least .75$. The variable of interest is the daily number of such durations, which is a measure of intradaily volatility, since the more volatile the stock price is within a day, the larger the counts will be. The ACP models provide good density forecasts of this measure of volatility.

1 Introduction

Many interesting empirical questions can be addressed by modeling a time series of count data. Examples can be found in a great variety of contexts. In the area of accident prevention, Johansson (1996) used time series counts to assess the effect of lowered speed limits on the number of road casualties. In epidemiology, time series counts arise naturally in the study of the incidence of a disease. A prominent example is the time series of monthly cases of polio in the U.S. which has been studied extensively by Zeger (1988) and Brännäs and Johansson (1994). In the area of finance, besides the applications mentioned in Cameron and Trivedi (1996), counts arise in market microstructure as soon as one starts looking at tick-by-tick data. The price process for a stock can be viewed as a sum of discrete price changes. The daily number of these price changes constitutes a time series of counts whose properties are of interest.

Most of these applications involve relatively rare events which makes the use of the normal distribution questionable. Thus, modeling this type of series requires one to deal explicitly with the discreteness of the data as well as its time series properties. Neglecting either of these two characteristics would lead to potentially serious misspecification. A typical issue with time series data is autocorrelation and a common feature of count data is overdispersion (the variance is larger than the mean). Both of these problems are addressed simultaneously by using an autoregressive conditional Poisson model (ACP). In the simplest model counts have a Poisson distribution and their mean, conditional on past observations, is autoregressive. Whereas, conditional on past observations, the model is equidispersed (the variance is equal to the mean), it is unconditionally overdispersed. A fully parametric approach and choose to model the conditional distribution explicitly and make specific assumptions about the nature of the autocorrelation in the series. A similar modeling strategy has been explored independently by Rydberg and Shephard (1998), Rydberg and Shephard (1999b), and Rydberg and Shephard (1999a). Two generalisations of their framework are introduced in this paper. The first consists of replacing the Poisson by the double Poisson distribution of Efron (1986), which allows for either under- or overdispersion in the marginal distribution. In this context, two common variance functions will be explored. Finally, an extended version of the model is proposed, which allows for separate models of mean and of variance. It is shown that this model performs well in a financial application and that it delivers very good density forecasts, which are tested by using the techniques proposed by Diebold and Tay (1997).

The main advantages of this model are that it is flexible, parsimomious and easy to estimate using maximum likelihood. Results are easy to interpret and standard hypothesis test are available. In addition, given that the autocorrelation and the density are modeled explicitly, the model is well suited for both point and density forecasts, which can be of interest in many applications. Finally, due to its similarity with the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982), the present framework can be extended to most of the models in the ARCH class; for a review see Bollerslev, Engle, and Nelson (1994).

The paper is organised as follows. A great number of models of time series count data have been proposed, which will be reviewed in section 2. In section 3 the basic autoregressive Poisson model is introduced and some of its properties are discussed. Section 4 introduces

two versions of the Double Autoregressive Conditional Poisson (DACP) model, along with their properties. In section 5 the model is generalised to allow for time-varying variance, applications to the daily number of price changes on IBM and to the number of new polio cases are presented is section 6; section 7 concludes.

2 A review of models for count data in time series

Many different approaches have been proposed to model time series count data. Good reviews can be found both in Cameron and Trivedi (1979), Chapter 7 and in MacDonald and Zucchini (1997), Chapter 1. Markov chains are one way of dealing with count data in time series. The method consists of defining transition probabilities between all the possible values that the dependent variable can take and determining, in the same way as in usual time series analysis, the appropriate order for the series. This method is only reasonable, though, when there are very few possible values that the observations can take. A prominent area of application for Markov chains is binary data. As soon as the number of values that the dependent variable takes gets too large, these models lose tractability.

Discrete Autoregressive Moving Average (DARMA) models are models for time series count data with properties similar to those of ARMA processes found in traditional time series analysis. They are probabilistic mixtures of discrete i.i.d. random variables with suitably chosen marginal distribution. One of the problems associated with these models seems to be the difficulty of estimating them. An application to the study of daily precipitation can be found in Chang and Delleur (1984). The daily level of precipitation is transformed into a discrete variable based on its magnitude. The method of moments is used to estimate the parameters of the model by fitting the theoretical autocorrelation function of each model to their sample counterpart. Estimation of the model seems quite cumbersome and the model is only applied to a time series which can take at most three values.

McKenzie (1985) surveys various models based on "binomial thinning". In those models, the dependent variable y_t is assumed to be equal to the sum of an error term with some prespecified distribution and the result of y_{t-1} draws from a Bernoulli which takes value 1 with some probability ρ and 0 otherwise. This guarantees that the dependent variable takes only integer values. The parameter ρ in that model is analogous to the coefficient on the lagged value in an AR(1) model. This model, called INAR(1), has the same autocorrelations as the AR(1) model of traditional time series analysis, which makes it its discrete counterpart. This family of models has been generalised to include integer valued ARMA processes a well as to incorporate exogenous regressors. The problem with this type of models is the difficulty in estimating them. Many models have been proposed and the emphasis was put more on their stochastic properties than on how to estimate them.

Hidden Markov chains, advocated by MacDonald and Zucchini (1997) are an extension of the basic Markov chains models, in which various regimes characterising the possible values of the mean are identified. It is then assumed that the transition from one to another of these regimes is governed by a Markov chain. One of the problems of this approach is that it there is no accepted way of determining the appropriate order for the Markov chain. Whereas in some cases there is a natural interpretation for what might constitute a suitable regime, in most applications, and in particular in the applications considered in this paper,

this is not the case. Another problem is that the number of parameters to be estimated can get big, especially when the number of regimes is large. Finally, the results are, in most cases, not very easy to interpret.

Harvey and Fernandes (1989) use state-space models with conjugate prior distributions. Counts are modeled as a Poisson distribution whose mean itself is drawn from a gamma distribution. The Gamma distribution depends on two parameters a and b which are treated as latent variables and whose law of motion is $a_{t|t-1} = \omega a_{t-1}$ and $b_{t|t-1} = \omega b_{t-1}$. As a result, the mean of the Poisson distribution is taken from a gamma with constant mean but increasing variance. Estimation is done by maximum likelihood and the Kalman filter is used to update the latent variables.

In the static case overdispersion is usually viewed as the result of unobserved heterogeneity. One way of dealing with this problem is to keep an equidispersed distribution, but introduce new regressors in the mean equation which are believed to capture this heterogeneity. Zeger and Qaqish (1988) apply ths intuition to the time series case and add lagged values of the dependent variable to the set of regressors in a static Poisson regression. They adopt a generalised linear model formulation in which conditional mean and variance are modeled instead of marginal moments. For count data, the authors propose using the Poisson distribution with the *log*, its associated link function. The mean is set equal to a linear combination of exogenous regressors and the time series dependence is accounted for by a weighted sum of past deviations of the dependent variable from the linear predictor: $\log(\mu_t) = x_t^{'}\beta + \sum_{i=1}^{q} \theta_i \left[\log(y_{t-i}^{*}) - x_{t-i}^{'}\beta\right]$ where $y_t^{*} = \max(y_{t-i}, c)$. These models have not been used very much in applications. One of the weaknesses of this specification is that had hoc assumptions are needed to handle zeros.

Zeger (1988) extends the generalised linear models and introduces a latent multiplicative autoregressive term ϵ_t with unit expectation , variance σ^2 and autocorrelation $\rho_{\epsilon(\tau)}$ which is responsible for introducing both autoregression and overdispersion into the model. The dependent variable is is assumed to be a function of exogenous regressors x_t with conditional mean $\mu_t = \exp(x_t\beta)$, where β is the coefficient of interest. Conditionally on the latent variable, the model is equidispersed ($E[y_t|\epsilon_t] = V[y_t|\epsilon_t] = \mu_t\epsilon_t$), but the marginal variance depends both on the marginal mean and its square: $E[y_t] = \mu_t\epsilon_t$ and $V[y_t|\epsilon_t] = \mu_t + \sigma^2\mu_t^2$. In order to estimate this model with maximum likelihood, one would need to specify a density for $y_t|\epsilon_t$ and for ϵ_t. In most cases, no closed-form solution would be available. Instead, a quasilikelihood method is adopted which only requires knowledge of mean, variance and covariances of y_t. Given estimates of the parameters of the latent variable, obtained by the method of moments, the variance-covariance matrix V of the dependent variable is formed: $V = A + \sigma^2 A R_\epsilon A$, where $A = diag(\mu_i)$ and R is the autocorrelation matrix of the latent variable $R_\epsilon^{j,k} = \rho_\epsilon(|j-k|)$. The quasilikelihood approach then consists in using the inverse of the variance matrix V as a weight in the first order conditions. Since inversion of V is quite cumbersome when the time series is long, an approximation is proposed. This method can be viewed as a count data analog of the Cochrane-Orcutt method for normally distributed time series in which all the serial correlation is assumed to come from the error term. The method has been applied to sudden infant death syndrom by Campbell (1994). While conceptually this method is quite close to what is proposed in this paper in this paper, there are nonetheless important differences in that it is fundamentally a static

model, whereas the model in this paper is an explicitly dynamic one. The interest is not limited to getting correct inference about the parameters on the exogenous variables but also lies in adequately capturing the dynamics of the system. In order to achieve this, a more parametric approach is taken, which, among other things, allows forecasting.

3 The ACP Model

The first model proposed in this paper has counts follow a Poisson distribution with an autoregressive mean. The Poisson distribution is the natural starting point for counts. One characteristic of the Poisson distribution is that the mean is equal to the variance. This property is referred to as equidispersion. Most count data however exhibit overdispersion. Modeling the mean as an autoregressive process generates overdispersion in even the simple Poisson case.

In the simplest model, the counts are generated by a Poisson distribution

$$N_t|\mathcal{F}_{t-1} = P(\mu_t)\ , \tag{3.1}$$

with an autoregressive conditional intensity as in the ACD model of Engle and Russell (1998) or the conditional variance in the GARCH (Generalised Autoregressive Conditional Heteroskedasticity) model of Bollerslev (1986):

$$E[N_t|\mathcal{F}_{t-1}] = \mu_t = \omega + \sum_{j=1}^{p} \alpha_j N_{t-j} + \sum_{j=1}^{q} \beta_j \mu_{t-j}\ . \tag{3.2}$$

The following properties of the unconditional moments of the ACP can be estabished.

Proposition 3.1 (Unconditional mean of the ACDP(p,q) or the ACP(p,q)). *The unconditional mean of the DACP(p,q) or the ACP(p,q) is*

$$E[N_t] = \mu = \frac{\omega}{1 - \sum_{j=0}^{\max(p,q)} (\alpha_j + \beta_j)}\ .$$

This proposition shows that, as long as the sum of the autoregressive coefficients is less than 1, the model is stationary and the expression for its mean is identical to the mean of an ARMA process.

Proposition 3.2 (Unconditional variance of the ACP(1,1) Model). *The unconditional variance of the ACP(1,1) model, when the conditional mean is given by*

$$E[N_t|\mathcal{F}_{t-1}] = \mu_t = \omega_1 + \alpha_1 N_{t-1} + \beta_1 \mu_{t-1}\ ,$$

is equal to

$$V[N_t] = \sigma^2 = \frac{\mu(1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2)}{1 - (\alpha_1 + \beta_1)^2} \geq \mu\ .$$

Proof of Proposition 3.2. Proof in appendix □

Proposition 3.2 shows that the ACP exhibits overdispersion, even though it uses an equidispersed marginal distribution. The model is overdispersed, as long as $\alpha_1 \neq 0$ and the amount of overdispersion is an increasing function of α_1 and also, to a lesser extent, of β_1. The following proposition establishes an expression for the autocorrelation function of the ACP.

Proposition 3.3 (Autocorrelation of the ACP(1,1) Model). *The unconditional autocorrelation of the ACP(1,1) model is given by*

$$Corr[N_t, N_{t-s}] = (\alpha_1 + \beta_1)^{s-1} \frac{\alpha_1 \left(1 - \beta_1(\alpha_1 + \beta_1)\right)}{1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2} .$$

Proof of Proposition 3.3. Proof in appendix □

It can be shown that the result in Proposition 3.3 holds also for the GDACP developed in the next section and for both the exponential and the Weibull versions of the ACD model of Engle and Russell (1998).

For the ACP the likelihood is

$$l_t(\theta) = N_t \ln(\mu_t) - \mu_t - \ln(y_t!) ,$$

where θ is a vector containing all the parameters of the autoregressive conditional intensity and the conditional intensity μ_t is autoregressive as in 3.2. The score and Hessian take the very simple form:

$$\frac{\partial l_t}{\partial \theta} = \left(\frac{N_t - \mu_t}{\mu_t} \right) \frac{\partial \mu_t}{\partial \theta} ,$$

$$\frac{\partial^2 l_t}{\partial \theta^2} = -\frac{N_t}{\mu_t^2} \frac{\partial^2 \mu_t}{\partial \theta^2} ,$$

where

$$\frac{\partial \mu_t}{\partial \theta} = z_t^{'} + \sum_{t=1}^{p} \beta_t \frac{\partial \mu_{t-1}}{\partial \theta} , \tag{3.3}$$

and

$$z_t = [1, N_{t-1}, N_{t-2}, \ldots , N_{t-p}, \mu_{t-1}, \mu_{t-2}, \ldots , \beta_q \mu_{t-q}] . \tag{3.4}$$

The score can easily be seen to be equal to zero in expectation. It is of interest in this model to test whether there is significant autocorrelation. In most time series applications a massive rejection can be expected. This can be done very simply with a likelihood ratio

test (LR). The statistic will be equal to twice the difference between the unrestricted and the restricted likelihoods, which follows the usual χ^2 distribution with two degrees of freedom. Both the restricted and the unrestricted models are easy to estimate. The simplicity of this test contrast sharply with the difficulties associated with tests and estimation of the autocorrelation of a latent variable in the model of Zeger (1988) which have been recently addressed in Davis and Wang (2000).

4 The DACP Model

This section introduces two models based on the double Poisson distribution, which differ only by their variance function. The DACP1 has its variance proportional to the the mean, whereas the DACP2 has a quadratic variance function.

In some cases one might want to break the link between overdispersion and serial correlation. It is quite probable that the overdispersion in the data is not attributable solely to the autocorrelation, but also to other factors, for instance unobserved heterogeneity. It is also possible that the amount of overdispersion in the data is less than the overdipsersion resulting from the autocorrelation, in which case an underdispersed marginal distribution might be appropriate. In order to account for these possibilities the Poisson is replaced by the double Poisson, introduced by Efron (1986) in the regression context, which is a natural extension of the Poisson model and allows one to break the equality between conditional mean and variance. This density is obtained as an exponential combination with parameter γ of the Poisson density of the observation y with mean μ and of the Poisson with mean equal to the observation y, which can be thought of as the likelihood function taken at its maximum value.

The density of the Double Poisson is:

$$f(y,\mu,\gamma) = \left(\gamma^{\frac{1}{2}} e^{-\gamma\mu}\right)\left(\frac{e^{-y}y^y}{y!}\right)\left(\frac{e\mu}{y}\right)^{\gamma y} . \tag{4.1}$$

$f(y,\mu,\gamma)$ is not strictly speaking a density, since the probabilities don't add up to 1, but Efron (1986) shows that the value of the multiplicative constant $c(\mu,\gamma)$which makes it into a real density is very close to 1 and varies little across values of the dependent variable. He also suggests an approximation for this constant:

$$\frac{1}{c(\mu,\gamma)} = 1 + \frac{1-\gamma}{12\mu\gamma}\left(1+\frac{1}{\mu\gamma}\right) .$$

As a consequence, he suggests maximising the approximate likelihood (leaving out the highly nonlinear multiplicative constant) in order to find the parameters and using the correction factor when making probability statements using the density.

The advantages of using this distribution are that it can be both under- and overdispersed, depending on whether γ is smaller or larger than 1. This will prove particularly useful in a later section of this paper, when two separate processes are used for the variance and mean, because this density ensures that there is no possibility that the conditional

mean become larger than the conditional variance, which would not be allowable with strictly overdispersed distributions. In the case of the Double Poisson (DP hereafter), the distributional assumption 3.1 is replaced by the following:

$$N_t|\mathcal{F}_{t-1} = DP(\mu_t, \gamma) \ . \tag{4.2}$$

It is shown in Efron (1986) (Fact 2) that the mean of the Double Poisson is μ and that the variance is approximately equal to $\frac{\mu}{\gamma}$. Efron (1986) shows that this approximation is highly accurate, and it will be used in the more general specifications.

For the simplest model, called the DACP1, the variance is a multiple of the mean:

$$V[N_t|\mathcal{F}_{t-1}] = \sigma_t^2 = \frac{\mu_t}{\gamma} \ . \tag{4.3}$$

The coefficient γ of the conditional mean will be a parameter of interest, as values different from 1 will represent departures from the Poisson distribution. The following proposition gives an expression for the variance of the DACP1.

Proposition 4.1 (Unconditional variance of the DACP1(1,1) Model). *The unconditional variance of the DACP1(1,1) model, when the conditional mean is given by*

$$E[N_t|\mathcal{F}_{t-1}] = \mu_t = \omega_1 + \alpha_1 N_{t-1} + \beta_1 \mu_{t-1} \ ,$$

is equal to

$$V[N_t] = \sigma^2 = \frac{1}{\gamma} \frac{\mu(1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2)}{1 - (\alpha_1 + \beta_1)^2} \geq \mu \ .$$

Proof of Proposition 4.1. Proof in appendix □

Proposition 4.1 shows that, like the ACP, the DACP1 model exhibits overdispersion. Results for the ACP(1,1) are obtained by setting $\gamma = 1$. The amount of overdispersion is an increasing function of α_1 and also, to a lesser extent, of β_1. When α_1 is zero, the overdispersion of the model comes exclusively from the Double Poisson distribution and is purely a function of the parameter γ, a measure of the departure from the Poisson distribution. The overdispersion can be seen to be a product of two terms which can be interpreted as the overdispersion due to the autocorrelation and the overdispersion of the marginal distribution, which is due to other factors:

$$\frac{\sigma^2}{\mu} = \frac{1}{\gamma} \frac{1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2}{1 - (\alpha_1 + \beta_1)^2} \ .$$

Another popular way of parameterising the dispersion is to allow for a quadratic relation between variance and mean, and this, along with the distributional assumption 4.2 defines the DACP2:

$$V[N_t|\mathcal{F}_{t-1}] = \sigma_t^2 = \mu_t + \delta\mu_t^2 \ . \tag{4.4}$$

This parameterisation can be used along with the expression for the variance of the double Poisson, replacing γ in 4.1 with $\frac{\mu_t}{\sigma_t} = \frac{1}{1+\delta\mu_t}$. One potential problem with this specification, however, is that even though it can accommodate some underdispersion when $\delta < 0$, it is then not possible to exclude negative values of the variance when $\mu_t < -\frac{1}{\delta}$. For this reason, when underdispersion is observed, the DACP1 should be preferred.
In this case the unconditional variance takes a somewhat different form, which is the object of the following proposition.

Proposition 4.2 (Unconditional variance of the DACP2(1,1) Model). *The unconditional variance of the DACP2(1,1) model, when the conditional mean is given by*

$$E[N_t|\mathcal{F}_{t-1}] = \mu_t = \omega_1 + \alpha_1 N_{t-1} + \beta_1 \mu_{t-1} \ ,$$

is equal to

$$V[N_t] = \sigma^2 = \frac{\mu(1-(\alpha_1+\beta_1)^2+\alpha_1^2)(1+\delta\mu)}{1-\delta\alpha_1^2-(\alpha_1+\beta_1)^2} \geq \mu \ .$$

Proof of Proposition 4.2. Proof in appendix □

This shows that the DACP2 is overdispersed in general (when $\delta > 0$, which is the case considered in this paper) and that overdispersion is an increasing function of the parameters of the mean equation and of the dispersion parameter. The two effects cannot be separated here as they could in the case of the DACP1.
As mentioned earlier on, the autocorrelation of both versions of the DACP is identical to the one of the ACP, which means that the dispersion properties of the marginal density do not affect the time series properties of the model.

For the DACP1 the likelihood is:

$$l_t(\theta,\gamma) = -\frac{1}{2}\ln(\gamma) - \frac{\mu_t}{\gamma} + N_t(\ln(N_t)-1) - \ln(N_t!) + \frac{N_t}{\gamma}\left(1+\ln\left(\frac{\mu_t}{N_t}\right)\right) \ ,$$

where θ is a vector containing all the parameters of the autoregressive conditional intensity and the conditional intensity μ_t is autoregressive:

$$E[N_t|\mathcal{F}_{t-1}] = \mu_t = \omega + \sum_{j=1}^{p}\alpha_j N_{t-j} + \sum_{j=1}^{q}\beta_j\mu_{t-j} \ .$$

Differentiating with respect to θ and γ,

$$\frac{\partial l_t}{\partial \theta} = \gamma \quad \frac{N_t - \mu_t}{\mu_t} \quad \frac{\partial \mu_t}{\partial \theta}$$
$$\frac{\partial l_t}{\partial \gamma} = \frac{1}{2}\frac{1}{\gamma} + (N_t - \mu_t) + N_t \ln\left(\frac{\mu_t}{N_t}\right) \ .$$

The first condition with respect to θ is just the ACP condition multiplied by the dispersion parameter γ. Taking the expectation, one gets:

$$E\left[\frac{\partial l_t}{\partial \theta}\right] = 0 \ .$$

The DACP1 is therefore a quasi maximum likelihood estimator: it is consistent as long as the mean is correctly specified, even if the true density is not double Poisson. The Hessain is given by:

$$\frac{\partial^2 l_t}{\partial \theta \partial \theta'} = -\frac{\gamma \quad N_t}{\mu_t^2} \quad \frac{\partial^2 \mu_t}{\partial \theta^2}$$
$$\frac{\partial^2 l_t}{\partial \theta \partial \gamma} = \frac{N_t - \mu_t}{\mu_t}\frac{\partial \mu_t}{\partial \theta}$$
$$\frac{\partial^2 l_t}{\partial \gamma^2} = -\frac{1}{2} \quad \frac{1}{\gamma^2} \ ,$$

where $\frac{\partial \mu_t}{\partial \theta}$ and z_t are as in 3.3 and 3.4.
Taking the expectation of the cross-derivative, one gets:

$$E\left[\frac{\partial^2 l_t}{\partial \theta \partial \gamma}\right] = 0 \ .$$

The cross derivative has an expectation of zero, so the expected Hessian is a block-diagonal matrix, which means that it is efficient to estimate γ independently from θ and that the variance of the estimators of the mean and dispersion are just the inverse of the diagonal elements of the Hessian.

For the DACP2 the likelihood is more complicated. It can be obtained from the likelihood of the DACP1 by the following transformation:

$$L_t(\theta, \delta) = l_t\left(\theta, \gamma\left(\mu_t(\theta), \delta\right)\right) \ .$$

$$L_t(\theta, \delta) = -\tfrac{1}{2}\ln\left(1 + \delta\mu_t\right) - \tfrac{\mu_t}{1+\delta\mu_t} + N_t\left(\ln\left(N_t\right) - 1\right) - \ln\left(N_t!\right)$$
$$+\frac{N_t}{1 + \delta\mu_t}\left(1 + \ln\left(\frac{\mu_t}{N_t}\right)\right) \ ,$$

where θ is a vector containing all the parameters of the autoregressive conditional intensity and the conditional intensity μ_t is autoregressive:

$$E[N_t|\mathcal{F}_{t-1}] = \mu_t = \omega + \sum_{j=1}^{p} \alpha_j N_{t-j} + \sum_{j=1}^{q} \beta_j \mu_{t-j}$$

Differentiating with respect to θ and γ,

$$\frac{\partial L_t}{\partial \theta} = \frac{1}{1+\delta\mu_t}\left(-\frac{\delta}{2} + \frac{N_t}{\mu_t} - \frac{1}{1+\delta\mu_t}\left(1+\delta N_t\left(1+\ln\left(\frac{\mu_t}{N_t}\right)\right)\right)\right)\frac{\partial \mu_t}{\partial \theta}$$
$$\frac{\partial L_t}{\partial \delta} = \frac{\mu_t}{1+\delta\mu_t}\left(-\frac{1}{2} - \frac{1}{1+\delta\mu_t}\left(\mu_t + N_t\left(1+\ln\left(\frac{\mu_t}{N_t}\right)\right)\right)\right) .$$

It can be seen, by setting δ equal to zero in the score with respect to the parameters of the mean, that one gets back the first order condition of the Poisson model:

$$\frac{\partial L_t}{\partial \theta} = \frac{N_t - \mu_t}{\mu_t}\frac{\partial \mu_t}{\partial \delta} .$$

$$\frac{\partial^2 L_t}{\partial \theta \partial \theta'} = \frac{-1}{(1+\delta\mu_t)^2}\left(\frac{1}{2} + \delta\frac{N_t}{\mu_t} + \frac{2\delta}{1+\delta\mu_t}\left(1+\delta N_t\left(1+\ln\left(\frac{\mu_t}{N_t}\right)\right)\right)\right)\frac{\partial^2 \mu_t}{\partial \theta^2}$$
$$\frac{\partial^2 L_t}{\partial \theta \partial \gamma} = \frac{-1}{(1+\delta\mu_t)^2}\left(\frac{1}{2} + N_t + N_t\frac{1-\delta^2\mu_t^2}{(1+\delta\mu_t)^2}\left(1+\ln\left(\frac{\mu_t}{N_t}\right)\right)\right)\frac{\partial \mu_t}{\partial \theta}$$
$$\frac{\partial^2 L_t}{\partial \gamma^2} = \left(\frac{\mu_t}{1+\delta\mu_t}\right)^2\left(\frac{1}{2} + \frac{2}{1+\delta\mu_t}\left(\mu_t + N_t\left(1+\ln\left(\frac{\mu_t}{N_t}\right)\right)\right)\right)$$

Again $\frac{\partial \mu_t}{\partial \theta}$ and z_t are as in 3.3 and 3.4.

In these models excess overdispersion can be tested with a simple Wald test for $\gamma = 1$ in the DACP1 and for $\delta = 0$ in the DACP2. Autocorrelation can be tested with a likelihood ratio test of a model in which the autoregressive parameters of the conditional mean are restricted to be zero against a more general alternative. One potential problem may lie in the fact that the double Poisson models are estimated by approximate maximum likelihood. The test is really an approximate likelihood ratio where the approximation error $log\left(\frac{c(\mu_u,\gamma_u)}{c(\mu_r,\gamma_r)}\right)$ is assumed to be close to zero.

In this section the equidispersion assumption of the Poisson model was relaxed and replaced it by two alternative specifications of the mean variance relation. In the first one, the conditional variance was allowed to be a multiple of the conditional mean, whereas in the DACP2 model, the variance was a quadratic function of the mean. In both of these formulations, however, the mean and variance are bound to covary in a rather restricted way. In the DACP1, by construction, the mean and variance have a correlation of 1. In the DACP2 this isn't the case, but mean and variance have to move together. Whenever the mean increases, the variance increases as well. This might be too restrictive in certain

situations, for example with a rare disease where one does not really know whether a disease is spreading out or whether the large number of cases was just an isolated occurrence. When the number of cases is small it tends to be followed by small counts and both mean and variance are then small. On the contrary, a larger count can be followed either by a large count (the disease is spreading out) or by a small count, in which case the mean does not change much, but the variance increases significantly as a result of this uncertainty. In order to deal with series of that sort, more flexible models have to be considered.

5 Extension to time-varying variance

This section introduces two different types of generalisations of the double Poisson models. The first, which will be called Generalised DACP (GDACP), adds a GARCH variance function to the DACP. The second type of extension consists of modeling the dispersion parameters of the DACP1 and DACP2 as autoregressive variables and the resulting models are called the Generalised DACP1 (GDACP1) and Generalised DACP2 (GDACP2).

The advantage of the Double Poisson distribution is that it allows for separate models of the mean and of the dispersion. In the regression applications of Efron (1986), the mean is made to depend on one set of regressors and the dispersion depends on a different set, via a logistic link function. This is a natural parameterisation for the cases most often analysed in the biostatistics literature, where certain variables are thought to affect the dispersion of the dependent variable. For example, in the toxoplasmosis data analysed by Efron (1986), which involves a double logistic, the dependent variable is the percentage of people affected by the disease in every city, the annual rainfall in every city is the explanatory variable and the sample size in each city determines the amount of overdispersion. However, in the time series case which is the focus of this paper, there are in general no such variables. This is obviously true for univariate time series but it is also the case for many other applications. In the univariate time series context dispersion will be modeled as a function of the past of the series. Two alternative specifications will be proposed, one in terms of the variance and the other in terms of the dispersion parameter.

The model which is written in terms of the variance will be examined first. This is a natural way to think about this problem since in the modeling of economic time series, based implicitly on thinking about the normal distribution, the focus has been more on models of the variance than on models of the dispersion. Moreover in terms of writing an autoregressive model of dispersion, it is not obvious what the lagged variable should be. As a consequence, it will also be difficult to calculate the theoretical unconditional moments of the model. In the case of variance however, one can adopt GARCH type variance functions and choose conditional variance as the lagged variable. With a GARCH process for the conditional variance, the dispersion parameter of the double Poisson can be expressed as a function of both the conditional mean and variance. The mean specification will be the same as in 3.2, but in addition, the conditional variance will be modeled as:

$$h_t \equiv E\left[\sigma_t^2 | \mathcal{F}_{t-1}\right] = \omega_2 + \alpha_2 \left(N_{t-1} - \mu_{t-1}\right)^2 + \beta_2 h_{t-1} \ . \tag{5.1}$$

From here on coefficients appearing in the conditional mean are indexed by 1 and coefficients of the conditional variance process are indexed by 2. For this specification, the dispersion

coefficient γ of the simple double Poisson is expressed in terms of the conditional mean and variance of the process:

$$\gamma_t = \frac{\mu_t}{h_t} \,. \tag{5.2}$$

The following proposition gives an expression for the unconditional variance of the GDACP when both the conditional mean and variance have one autoregressive and one moving average parameter.

Proposition 5.1 (Unconditional variance of the GDACP(1,1,1,1) Model). *The variance of the generalised model is given by:*

$$V[N_t] = \sigma^2 = \frac{\omega_2}{(1-\alpha_2-\beta_2)} \frac{(1-(\alpha_1+\beta_1)^2+\alpha_1^2)}{(1-(\alpha_1+\beta_1)^2)} \,.$$

Proof of Proposition 5.1. Proof in appendix □

The variance is a product of the unconditional mean of the variance process and a term depending on the mean equation parameters. Some insight can be gained by dividing both sides of this equation by the unconditional mean of the counts and making use of 3.2:

$$\frac{\sigma^2}{\mu} = \frac{\omega_2}{(1-\alpha_2-\beta_2)} \frac{\omega_1(1-(\alpha_1+\beta_1)^2+\alpha_1^2)}{(1+\alpha_1+\beta_1)} \,.$$

It is now apparent that the overdispersion of the GDACP is a product of the long term mean of the variance process and of a term which is decreasing in α_1 and β_1. This somewhat counterintuitive result is due to the fact that the variance process gives rise to most of the overdispersion and that the autoregressive parameters of the mean, while increasing the unconditional variance, actually increase the mean even more for a given ω_1, which in aggregate decreases the overdispersion.

It is of interest to see how the GDACP relates to the models presented in the previous sections. The GDACP model nests the plain double Poisson (obtained when $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 0$) with $\gamma = \frac{\omega_2}{\omega_1}$ in the case of the DACP1 and $\delta = \frac{\omega_2 - \omega_1}{\omega_1^2}$ in the case of the DACP2) and the plain Poisson (obtained when $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 0$ and $\omega_2 = \omega_1$), but it does not nest non-trivial ACP or DACP models (where $\alpha_1 \neq 0$ and $\beta_1 \neq 0$). A consequence of this is that it is not possible to test the GDACP against the previous models with a test of nested hypotheses. One would have to resort to tests of non-nested hypotheses, which in the present case would be quite cumbersome, due to the fact that the conditional mean and variance are unobserved autoregressive processes.

The likelihood function of the GDACP is given below:

$$l_t(\theta_1, \theta_2)) = \frac{1}{2} \ln\left(\frac{\mu_t}{h_t}\right) - \frac{\mu_t^2}{h_t} + N_t \left(\ln\left(N_t\right) - 1\right) - \ln\left(N_t!\right) + \frac{N_t \mu_t}{h_t} \left(1 + \ln\left(\frac{\mu_t}{N_t}\right)\right) \,.$$

where μ_t and h_t are as defined above and where $\theta_1 = [\omega_1, \alpha_1, \beta_1]$ and $\theta_2 = [\omega_2, \alpha_2, \beta_2]$.

The score of the model is given by

$$\frac{\partial l_t}{\partial \mu_t} = \left(\frac{1}{2\mu_t} + 2\frac{N_t - \mu_t}{h_t} + \frac{N_t}{h_t} \ln\left(\frac{\mu_t}{N_t}\right) \right) \frac{\partial \mu_t}{\partial \theta_1}$$
$$\frac{\partial l_t}{\partial h_t} = \left(\frac{1}{2h_t} - \frac{\mu_t}{h_t}\frac{N_t - \mu_t}{h_t} - \frac{N_t \mu_t}{h_t^2} \ln\left(\frac{\mu_t}{N_t}\right) \right) \frac{\partial h_t}{\partial \theta_2} .$$

The Hessian is

$$\frac{\partial l_t^2}{\partial \theta_1 \partial \theta_1'} = \left(-\frac{1}{2\mu_t^2} - \frac{2}{h_1} + \frac{N_t}{\mu_t h_t} \right) \frac{\partial^2 \mu_t}{\partial \theta_1 \theta_1'}$$
$$\frac{\partial l_t^2}{\partial \theta_1 \partial \theta_2'} = \left(\frac{2\mu_t}{h_t^2} - \frac{N_t}{h_t^2}\left(2 + \ln\left(\frac{\mu_t}{N_t}\right)\right) \right) \frac{\partial \mu_t}{\partial \theta_1}\frac{\partial h_t}{\theta_2'}$$
$$\frac{\partial l_t^2}{\partial \theta_2 \partial \theta_2'} = \left(\frac{1}{2h_t^2}\left(\frac{2\mu_t}{h_t^2}\left(\frac{N_t - \mu_t}{h_t} + \frac{N_t}{h_t}\ln\left(\frac{\mu_t}{N_t}\right)\right)\right)\right) \frac{\partial^2 h_t}{\partial \theta_2 \partial \theta_2'} .$$

The GDACP model is convenient because it has many features which make it familiar for time series econometricians, however it does not nest simpler models like the DACP. This makes testing the general model against the more restrictive ones difficult. To remedy this, one can model the dispersion parameter of the double Poisson directly. Unlike the variance, when modeling time-varying dispersion, there is no natural candidate lagged dependent variable. As a consequence a logistic link function will be used as in Efron (1986):

$$\gamma_t = \frac{M}{1 + \exp(-\lambda_t)} , \tag{5.3}$$

along with an autoregressive process for λ:

$$\lambda_t = \omega_2 + \alpha_2 \frac{N_{t-1} - \mu_{t-1}}{\sqrt{\mu_t \gamma_t}} + \beta_2 \lambda_{t-1} .$$

M is the maximum possible overdispersion, which, as in Efron (1986), will not be estimated but set by trial and error. In all the applications it will be set to 10. Unlike in the previous model, the unconditional moments are not easy to compute explicitly. In the present formulation, however, the regressor is a martingale difference sequence which makes λ_t a stationary ARMA process. It can be seen that this dispersion model nests the DACP1 (when $\alpha_2 = \beta_2 = 0$ and the ACP (when in addition, $\omega_2 = log(M-1)$), but not the DACP2. It is possible however to build a model which nests the DACP2 by setting δ defined in 4.4 to be a function of λ as in 5.3. These models will be called GDACP1 and GDACP2 respectively by reference to the DACP1 and DACP2 models that they generalise. Table 1 contains a summary of the model specifications and the chart on page 29 shows how the various models relate to each other. Testing the time-varying overdispersion in the GDACP1 and GDACP2 can be done with a likelihood test against the null hypothesis of DACP1 and DACP2 respectively. Gurmu and Trivedi (1993) consider artificial regression

tests for dispersion models which could easily be generalised to time-varying overdispersion. These tests would then be very similar to GARCH tests with the exception that they are based on the deviation instead of the residuals.

An advantage of the double Poisson over any overdispersed densities alluded to earlier on becomes more obvious now. The double Poisson distribution can be either over- or underdispersed, which means that the conditional variance can be either larger or smaller than the conditional mean. With a strictly overdispersed distribution, there would be difficulies each time the mean gets larger than the variance in the GDACP or when the dispersion gets smaller than 1 for the GDACP1 or GDACP2, which could happen in the numerical optimisation of the likelihood or when making out-of-sample predictions. There is no constraint on the parameters of the model which could ensure that this does not happen. With the double Poisson, however, there is no need to worry about this possibility.

The models can easily be generalised to include explanatory variables. It is possible to multiply the conditional mean, variance or overdispersion by a function of exogenous regressors in exactly the same way as in Zeger (1988). In the case of the Poisson, this can be done by using the natural link function, which is the exponential. Instead of using directly μ_t and h_t, one could use $\mu_t \exp X\delta_1$ and $h_t \exp X\delta_2$ as the new conditional mean and variance. Another possibility is to include regressors directly in the conditional mean or variance equation. Obviously the model proposed here can also be extended to most of the variations of the GARCH family, which is a very rich class of models.

6 Data analysis

In order to demonstrate how the model works, it is applied to two datasets in completely different areas. The first is a medical example and the second is an application to stock market data.

6.1 Incidence of poliomyelitis in the U.S.

The first application of the model is to the monthly number of cases of poliomyelitis in the United States between 1970 and 1983, which was analysed as an example in Zeger (1988). Zeger estimated three models, two of which assume that repeated observations are independent and follow a negative binomial distribution in one case or have a constant coefficient of variation in the other case. The model proposed by Zeger considers that there is a latent process which generates overdispersion and autocorrelation. The time series of polio cases seems to have become a benchmark for time series count models. It has been analysed by Brännäs and Johansson (1994), who study properties of different estimators for the parameters of the latent variable in the Zeger model which are responsible for overdispersion and autoregression. Davis and Wang (2000) propose a test for the presence of a latent variable and a correction for the standard errors of the coefficients on exogenous regressors based on asymptotic theory. They apply their method to the polio series to show that their theoretical standard errors are close to simulated ones. Jorgensen, Lundbye-Christensen, Song, and Sun (1999) propose a nonstationary state space model and apply it to the same dataset. Fahrmeir and Tutz (1994) reports results from the estimation of a log-linear Poisson model which includes lagged dependent variables.

Upon examining the series of polio cases (see Figure 2), it is not at all obvious whether this dataset provides evidence of declining incidence of polio in the U.S. which is one of the questions of interest. The histogram reveals that, with the exception of an outlier at 14 which could well be a recording error, the counts are very small with more than 60% of zeros. The series is clearly overdispersed with a mean of 1.33 and a variance of 3.50. Figure 3 shows the autocorrelogram of the number of polio cases after an outlier has been taken away. There is significant autocorrelation in levels and squares, and there is a clear pattern in the sign of the autocorrelation, which is first positive, then negative and positive again at very high lags.

Results from models without exogenous regressors are reported in table 2. The outlier of 14 has been taken out, as it had a strong impact on some of the coefficients. Models are evaluated on the basis of their log-likelihood, but also on the basis of their Pearson residuals, which are defined as: $\epsilon_t = \frac{1}{\sqrt{T-K}} \frac{N_t - \mu_t}{\sigma_t}$, where a correction for the number of degrees of freedom was used. If a model is well specified, the Pearson residuals will have variance one and no significant autocorrelation left. Autocorrelation is tested with a likelihood ratio test (LR) which gives a test statistic of 57.4 much larger than the $\chi^2_{[2]}$ 5% value of 5.99. The simplest model is the ACP which, while providing a good fit and capturing some autocorrelation, leaves an overdispersed residual with a variance of 1.70. The DACP1 and DACP2 correct for this, reducing this variance to 1.05 and .96 respectively. Both a Wald test, which in this case is the same as a simple t-test of the null that $\gamma = 0$ in the DACP1 or $\delta = 0$ in the DACP2, and a likelihood ratio (LR) test reject the ACP model. As was to be expected, the double Poisson makes it possible to fit the autocorrelation and the overdispersion at the same time.

Most parameters are significant at the 5% level. The parameters of the conditional mean for the three models are very similar, which indicates that the ACP was adequately capturing the time series aspect of the data while missing some of the dispersion. They imply a relatively high degree of persistence with the sum of the autoregressive parameters around .75 for most models. Amongst the more general models the GDACP2 performs best in terms of the dispersion, since it leaves a standard residual with a variance of 1.00, but the GDACP1 seems to provide a slightly better fit. The GDACP1 is preferred to the DACP1 that it nests as shown by a LR test statistic of 8, but it is not possible to reject the DACP2 against the more general GDACP2 (the LR test statistic is .8). For an overview of the models, see table 1 and the chart on page 29. The model which is formulated with a GARCH variance function does not perform as well as the dispersion models.

Figure 4 shows the autocorrelations of the Pearson residuals for the models considered so far. There is a great overall reduction in the level of autocorrelation below the Bartlett confidence interval which lies at .154, except for very few values. The Ljung-Box statistic for autocorrelation which does not allow to reject the null hypothesis of zero autocorrelations for any lags in the original series is reversed for all lags for the standardised residuals of the models. The pattern observed in the autocorrelogram of the series has been replaced by a pattern at a higher frequency. It can be expected that this pattern will disappear with the inclusion of seasonality dummies. Figure 5 reveals that all models have very small autocorrelations in squares and have lost the low frequency pattern that is present in the original series.

In order to compare the results to the results of Zeger (1988), models using the same set

of regressors were estimated. These include trigonometric seasonality variables at yearly and half yearly frequencies as well as a time trend, since interest lies in whether or not the present dataset provides evidence of declining incidence of polio in the U.S. The results, shown in table 3 are qualitatively similar to Zeger's results. In all the models the coefficient on the time trend is negative and insignificant. This coefficient is systematically smaller in absolute value across all the specifications than what Zeger reports. The coefficients of the seasonality variables are more or less the same as in Zeger (1988). The results seem to be even closer to what Fahrmeir and Tutz (1994) report in Table 6.2 than to Zeger's results. Magnitudes vary a little but the signs are always the same. In all models, the seasonality variables are significant as a group. Again autocorrelation is tested and the null of no serial correlation is rejected with a test statistic of 24 very much in excess of the 5.99 value under the null. The overall picture for the models including exogenous regressors is the same as for the ones which do not. LR tests reject the Poisson model in favour of the double Poisson. The DACP2 is the best amongst the simpler models both in terms of the likelihood and in the modeling of dispersion, since it leaves an almost perfectly equidispersed residual. The GDACP1 fits better, but does slightly worse than the DACP2 in terms of the residuals. A LR test rejects the DACP1 in favour of the GDACP1, but it is not possible to reject the DACP2 against the alternative hypothesis of GDACP2 at the 5% level.

The autocorrelations of the standard errors are shown in figure 6 and 7 and they are quite similar to the previous ones. This is somewhat of a surprise, as it could be expected that what seems to be a systematic seasonal pattern in the autocorrelations would disappear after inclusion of seasonality variables. Maybe this apparent pattern is not very important, since it is clearly below the Bartlett significance level and insignificant also according to the Ljung-Box statistic. Alternatively it could suggest that the seasonality has not been taken into account properly, but a little experimenting with alternative parameterisations suggests that this is not the case. This example shows that the ACP family of models is able to capture autocorrelation and dispersion successfully and whiten the residuals, while getting similar results as Zeger (1988) in terms of coefficients on exogenous regressors and their standard errors.

6.2 Daily number of price changes

The second application of the model that is to the daily number of price change durations of .75$ on the IBM stock on the New York Stock Exchange. A .75$ price-change duration is defined as the time it takes the stock price to move by at least .75$. The variable of interest is the daily number of such durations, which is a measure of intradaily volatility, since the more volatile the stock price is within a day, the more often it will cross a given threshold and the larger the counts will be. Midquotes from the Trades and Quotes (TAQ) dataset will be used to compute the number of times a day that the price moves by at least .75$, using the "five second rule" of Lee and Ready (1991) to compute midquotes prevailing at the time of the trade. For robustness it is required that the following midquote not revert the price change.

Let us denote by S_t the midquote price of the asset and by τ_n the times at which the threshold is crossed:

$$\tau_{n+1} = \inf_t \{t > \tau_n : |S_t - S_{\tau_n}| \geq d\} \tag{6.1}$$

The durations are then defined as $\Delta t_n = \tau_{n+1} - \tau_n$ and the object of interest is the daily number of such durations, which is a measure of intradaily volatility. Unlike the volatility that can be extracted from daily returns, for example with GARCH, the counts are a daily measure based on the price history within a day and this will contain information that volatility measures based on daily data are missing. For such a series interest lies amongst other things in forecasting, as volatility is an essential indicator of market behaviour as well as an input in many asset pricing problems.

As can be seen from the plotted series in Figure 9, the counts have episodes of high and low mean as well as variance, which suggest that autoregressive modeling should be appropriate. The histogram reveals that the data range from 0 to 30. The data is overdispersed with a mean of 5.98 and a variance of 20.4. Also, Figure 8 shows that the autocorrelations of the series in level and squares are clearly significant up to a relatively large number of lags (the Bartlett's 95% confidence interval under the null of iid is .10). Significant autocorrelation is also present in the third and fourth powers, although to a lesser degree.

A series of models is estimated, which range from the simple ACP to the most general GDACP and the results are reported in table 4. Most parameters are significant with t-statistics well above 2. All the models imply that the series is quite persistent with $\alpha_1 + \beta_1$ in the order of .82 to .94, which . The GDACP implies that the variance is also very persistent with a value of .912 for the sum of the autoregressive parameters. Again the simple Poisson model is rejected in favour of the DACP1 and DACP2. The DACP2 has standardised errors with a variance very near 1. The model in variance does not perform well, either in likelihood or in terms of its standard errors. The DACP1 is rejected in favour of the GDACP1, but the DACP2 cannot be rejected at the 5 % level. The preferred model is then DACP2 which has the best likelihood of all the models and which leaves almost an equidispersed error term. Another way to check the specification is to look at the autocorrelation of the Pearson residuals, which should be white noise if the time series dependence has been well accounted for by the model. In Figure 10 shows the autocorrelogram of the standardised errors of the various models. This reveals that the standard errors have no more autocorrelation left and they have lost any of the systematic patterns that were present in the original series. Figure 11 shows that the same is true for the squared standard errors. The autocorrelogram of the residuals to the third and fourth power show very low autocorrelation, which indicates that the models capture the serial correlation in the four first moments well.

The models are also evaluated with respect to the quality of their out-of-sample forecasts. The models are estimated on a starting sample of 202 observations, then a one-step-ahead forecast is calulated and the model is reestimated for every period with all the available information. First the quality of the point forecasts from each one of the models is evaluated. The Root Mean Squared Errors (RMSE) of the various models are quite close to each other. The models which do best are the DACP2 and GDACP2, but the ACP and he DACP1 do almost as well. In terms of the Standardised Sum of Errors the DACP2 performs best with a variance of 1.29, the closest to one from all models. A decomposition of the forecasting error shows that the bias is very small in most models which means

that the forecasts are right on average. The variance proportion of the forecast for most models is around .25% which means that the variance of the forecast and the variance of the original are quite close to each other. The remaining part of the forecasting error is unsystematic forecasting error. It is a good sign for the forecasts that most of the error is of the unsystematic type. Using that measure, the DACP2 is again seen to be the best with 77% of the forecasting error being of the unsystematic type.

Another way to test the accuracy of the models is to evaluate how good they are at forecasting the density of the counts out-of-sample. For this purpose the method developed by Diebold and Tay (1997) is used, which consists in computing the cumulative probability of the observed values under the forecast distribution. If the density from the model is accurate, these values will be uniformly distributed and will have no significant autocorrelation left neither in level nor when raised to integer powers. In order to assess how close the distribution of the Z variable is to a uniform, quantile plots of Z against quantiles of the uniform distribution are shown. The closer the plot is to a 45%-line, the closer the distribution is to a uniform. The quantile plots of the various models are shown in figure 12. The Z statistic for most models is quite close to the 45%-line. The GDACP1 is the model which performs best. The Poisson model gives too little weight to large observations as is reflected in the fact that the curve is clearly below the 45%-line between .6 and 1. This is present to a certain degree in all of the plots, but less so for more sophisticated models. The ACP and GDACP1 give too little weight to zeros whereas all other models attribute too much probability mass to them. All the models seem to have difficulty with the right tail: they cannot quite accommodate as many large values as are present in the data. Diebold and Tay (1997) propose to graphically inspect the correlogram along with the usual Bartlett confidence intervals. For the present case, this means that all correlations smaller in absolute value than .141 can be considered not to be significant. The autocorrelations for the models are displayed in figures 13 and 14. In general the models perform very well with only very few significant correlations left. It is difficult to discriminate between models base on these autocorrelograms. It seems though that the GDACP1 has large negative autocorrelations for small lags, both in levels and in squares. In terms of the Ljung-Box, which has acceptable properties according to the Monte Carlo study in Brännäs and Johansson (1994), the GDACP and GDACP2 have the highest p-values for levels and squares, followed by the DACP1. The GDACP which does quite poorly in-sample seems to capture the time-series adequately out-of-sample.

7 Conclusion

This paper introduces new models for time series count data. These models have proved very flexible and easy to estimate. They make it possible to correct standard errors and improve inference on parameters of exogenous regressors compared to Static Poisson regression like other methods, particularly the one proposed by Zeger (1988), which has attracted a lot of attention recently, while modeling the time series dependence in a more flexible way. It is shown that these models perform well and can explain both autocorrelation and dispersion in the data. The biggest advantage of this framework is that it is possible to apply straightforward likelihood based tests for autocorrelation or overdispersion. Finally

this method also makes it possible to perform point and density forecasts, which successfully pass a series of forecast evaluation tests. An interesting question is how this model can be generalised in a multivariate framework and this is the object of ongoing research.

8 Appendix

Proof of Proposition 3.1. Same as the proof of lemma 1 in Engle and Russell (1998). □

Proof of Proposition 4.1. Upon substitution of the mean equation in the autoregressive intensity, one obtains:

$$\mu_t - \mu = \alpha_1(N_{t-1} - \mu) + \beta_1(\mu_{t-1} - \mu) \ ,$$

$$\mu_t - \mu = \alpha_1(N_{t-1} - \mu_{t-1}) + (\alpha_1 + \beta_1)(\mu_{t-1} - \mu) \ .$$

Squaring and taking the expectation gives:

$$E[(\mu_t - \mu)^2] = \alpha_1^2 E\left[(N_{t-1} - \mu_{t-1})^2\right] + (\alpha_1 + \beta_1)^2 E\left[(\mu_{t-1} - \mu)^2\right] \ .$$

Using the law of iterated expectations and substituting the conditional variance σ_{t-1} for its expression, one gets:

$$E\left[(\mu_t - \mu)^2\right] = \alpha_1^2 E[\frac{\mu_{t-1}}{\gamma}] + (\alpha_1 + \beta_1)^2 E\left[(\mu_{t-1} - \mu)^2\right] \ . \tag{8.1}$$

Collecting terms, one gets:

$$V[\mu_t] = E[(\mu_t - \mu)^2] = \frac{\alpha_1^2 \mu}{\gamma(1 - (\alpha_1 + \beta_1)^2)} \ . \tag{8.2}$$

Now, applying the following property on conditional variance

$$V[y] = E_x\left[V_{y|x}(y|x)\right] + V_x\left[E_{y|x}(y|x)\right] \ , \tag{8.3}$$

to the counts, one obtains:

$$E\left[(N_t - \mu)^2\right] = E\left[(N_t - \mu_t)^2\right] + E\left[(\mu_t - \mu)^2\right] \ . \tag{8.4}$$

Again using the law of iterated expectations, substituting the conditional variance σ_t for its expression, then making use of the previous result, and after finally collecting terms, one gets the announced result. □

Proof of Proposition 4.2. This proof is similar to the proof of Proposition 4.1. When substituting the conditional variance σ_t^2 for its expression, instead of 8.1, one gets:

$$E\left[(\mu_t - \mu)^2\right] = \alpha_1^2 E[\mu_{t-1} + \delta\mu_{t-1}^2] + (\alpha_1 + \beta_1)^2 E\left[(\mu_{t-1} - \mu)^2\right] \ .$$

Similarly, instead of 8.2, one gets:

$$V[\mu_t] = E[(\mu_t - \mu)^2] = \frac{\alpha_1^2 \mu(1 + \delta\mu)}{1 - \alpha_1^2\delta - (\alpha_1 + \beta_1)^2} \ .$$

Now, applying 8.3 to the counts, one obtains 8.4. The result then follows after applying the same steps as in the previous proof. □

Proof of Proposition 3.3. As a consequence of the martingale property, deviations between the time t value of the dependent variable and the conditional mean are independent from the information set at time t. Therefore:

$$E[(N_t - \mu_t)(\mu_{t-s} - \mu)] = 0 \qquad \forall\, s \geq 0 \ .$$

By distributing $N_t - \mu_t$, one gets:

$$Cov[N_t, \mu_{t-s}] = Cov[\mu_t, \mu_{t-s}] \qquad \forall\ s \geq 0 \ . \tag{8.5}$$

By the same "non-anticipation" condition used above, it must be true that:

$$E[(N_t - \mu_t)(N_{t-s} - \mu)] = 0 \qquad \forall\ s \geq 0 \ .$$

Again, distributing $N_t - \mu_t$, one gets:

$$Cov[N_t, N_{t-s}] = Cov[\mu_t, N_{t-s}] \qquad \forall\ s \geq 0 \ . \tag{8.6}$$

Now,

$$\begin{aligned} Cov[\mu_t, \mu_{t-s}] &= \alpha_1 Cov[N_t, \mu_{t-s+1}] + \beta_1 Cov[\mu_t, \mu_{t-s}] \\ &= (\alpha_1 + \beta_1) Cov[\mu_t, \mu_{t-s}] \\ &= (\alpha_1 + \beta_1)^s V[\mu_t] \ . \end{aligned}$$

The first line was obtained by replacing μ_t by its expression, the second line by making use of 8.5, the last line follows from iterating line two.

$$Cov[\mu_t, \mu_{t-s+1}] = \alpha_1 Cov[\mu_t, N_{t-s}] + \beta_1 Cov[\mu_t, \mu_{t-s}] \ .$$

Rearranging and making use of 8.6, one gets:

$$
\begin{aligned}
Cov[N_t, N_{t-s}] &= \frac{1}{\alpha_1} Cov[\mu_t, \mu_{t-s+1}] - \frac{1}{\beta} Cov[\mu_t, \mu_{t-s}] \\
&= \frac{1}{\alpha_1} \left(1 - \beta(\alpha + \beta)\right) (\alpha_1 + \beta_1)^s V[\mu_t] \; .
\end{aligned}
$$

Finally,

$$
Corr[N_t, N_{t-s}] = \frac{1}{\alpha_1} \left(1 - \beta_1(\alpha_1 + \beta_1)\right) (\alpha_1 + \beta_1)^s \frac{V[\mu_t]}{V[N_t]} \; .
$$

Replacing $V[\mu_t]$ and $V[N_t]$ by their respective values, the result follows. □

Proof of Proposition 5.1. This proposition can be proved by proceeding in a similar fashion as in the Proof of 3.2, but making use of the expression for the conditional variance h_t. As a result of the first step, instead of 8.2, one obtains:

$$
V[\mu_t] = E[(\mu_t - \mu)^2] = \frac{\alpha_1^2 E[h_{t-1}]}{1 - (\alpha_1 + \beta_1)^2} \; . \tag{8.7}
$$

It can be seen from 5.1 that:

$$
E[h_t] = \frac{\omega_2}{1 - \alpha_2 - \beta_2} \; . \tag{8.8}
$$

Applying property 8.3, making use of 8.8 and after simplification, one gets the announced result. □

References

Bollerslev, Tim, 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 52, 5–59.

——— , Robert F. Engle, and Daniel B. Nelson, 1994, Arch models, in Robert F. Engle, and Daniel L. McFadden, ed.: *Handbook of Econometrics, Volume 4* (Elsevier Science: Amsterdam, North-Holland).

Brännäs, Kurt, and Per Johansson, 1994, Time series count data regression, *Communications in Statistics: Theory and Methods* 23, 2907–2925.

Cameron, A. Colin, and Pravin K. Trivedi, 1979, *Regression Analysis of Count Data* (Cambridge University Press: Cambridge).

Cameron, Colin A., and Pravin K. Trivedi, 1996, Count data models for financial data, in Maddala G.S., and C.R. Rao, ed.: *Handbook of Statistics, Volume 14, Statistical Methods in Finance* (Elsevier Science: Amsterdam, North-Holland).

Campbell, M.J., 1994, Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature, *Journal of the Royal Statistical Society A* 157, 191–208.

Chang, Tiao J., M.L. Kavvas, and J.W. Delleur, 1984, Daily precipitation modeling by discrete autoregressive moving average processes, *Water Resources Research* 20, 565–580.

Davis, Richard A., William Dunsmuir, and Yin Wang, 2000, On autocorrelation in a poaisson regression model, *Biometrika* 87, 491–505.

Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay, 1997, Evaluating density forecasts, U Penn Working Paper.

Efron, Bradley, 1986, Double exponential families and their use in generalized linear regression, *Journal of the American Statistical Association* 81, 709–721.

Engle, Robert, 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* 50, 987–1008.

Engle, Robert F., and Jeffrey Russell, 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* 66, 1127,1162.

Fahrmeir, Ludwig, and Gerhard Tutz, 1994, *Multivariate Statistical Modeling Based on Generalized Linear Models* (Springer-Verlag: New York).

Gurmu, Shiferaw, and Pravin K. Trivedi, 1993, Variable augmentation specification tests in the exponential family, *Econometric Theory* 9, 94–113.

Harvey, A.C., and C. Fernandes, 1989, Time series models for count or qualitative observations, *Journal of Business and Economic Statistics* 7, 407–417.

Johansson, Per, 1996, Speed limitation and motorway casualties: a time series count data regression approach, *Accident Analysis and Prevention* 28, 73–87.

Jorgensen, Bent, Soren Lundbye-Christensen, Peter Xue-Kun Song, and Li Sun, 1999, A state space model for multivariate longitudinal count data, *Biometrika* 96, 169–181.

Lee, Charles M., and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 66, 733–746.

MacDonald, Iain L., and Walter Zucchini, 1997, *Hidden Markov and Other Models for Discrete-valued Time Series* (Chapman and Hall: London).

McKenzie, Ed., 1985, Some simple models for discrete variate time series, *Water Resources Bulletin* 21, 645–650.

Rydberg, Tina H., and Neil Shephard, 1998, A modeling framework for the prices and times of trades on the nyse, To appear in Nonlinear and nonstationary signal processing edited by W.J. Fitzgerald, R.L. Smith, A.T. Walden and P.C. Young. Cambridge University Press, 2000.

———, 1999a, Dynamics of trade-by-trade movements decomposition and models, Working paper, Nuffield College, Oxford.

———, 1999b, Modelling trade-by-trade price movements of multiple assets using multivariate compound poisson processes, Working paper, Nuffield College, Oxford.

Zeger, Scott L., 1988, A regression model for time series of counts, *Biometrika* 75, 621–629.

———, and Bahjat Qaqish, 1988, Markov regression models for time series: A quasi-likelihood approach, *Biometrics* 44, 1019–1031.

Model	Mean μ_t	Variance σ_t^2	Dispersion
ACP	$\omega_1 + \alpha_1 N_{t-1} + \beta_1 \mu_{t-1}$	μ_t	1
DACP1	$\omega_1 + \alpha_1 N_{t-1} + \beta_1 \mu_{t-1}$	$\frac{\mu_t}{\gamma}$	$\frac{1}{\gamma}$
DACP2	$\omega_1 + \alpha_1 N_{t-1} + \beta_1 \mu_{t-1}$	$\mu_t + \delta\ \mu_t^2$	$1 + \delta\mu_t$
GDACP	$\omega_1 + \alpha_1 N_{t-1} + \beta_1 \mu_{t-1}$	$\omega_2 + \alpha_2 N_{t-1} + \beta_2 (N_t - \mu_t)^2$	$\frac{\mu_t}{\sigma_t^2}$
GDACP1	$\omega_1 + \alpha_1 N_{t-1} + \beta_1 \mu_{t-1}$	$\frac{\mu_t}{\gamma_t}$	$\gamma_t = \frac{M}{1+\exp(-\lambda_t)}$ $\lambda_t = \omega_2 + \alpha_2 \frac{N_{t-1}-\mu_{t-1}}{\sqrt{\mu_{t-1}\gamma_{t-1}}} + \beta_2 \lambda_{t-1}$
GDACP2	$\omega_1 + \alpha_1 N_{t-1} + \beta_1 \mu_{t-1}$	$\mu_t + \delta_t\ \mu_t^2$	$\delta_t = \frac{M}{1+\exp(-\lambda_t)}$ $\lambda_t = \omega_2 + \alpha_2 \frac{N_{t-1}-\mu_{t-1}}{\sqrt{\mu_{t-1}+\delta_{t-1}\mu_{t-1}^2}} + \beta_2 \lambda_{t-1}$

Table 1: Summary of the model specifications

Coefficients	ACP	DACP1	DACP2	GDACP	GDACP1	GDACP2
ω_1	.29	.28	.56	.12	.37	.34
	(2.46)	(1.49)	(2.00)	(.93)	(1.72)	(1.52)
α_1	.23	.23	.36	.08	.28	.28
	(4.83)	(2.84)	(3.47)	(1.33)	(2.49)	(2.33)
β_1	.55	.56	.21	.82	.44	.46
	(5.02)	(3.13)	(.92)	(5.31)	(2.09)	(2.12)
ω_2				.20	-2.89	-3.45
				(1.35)	(-2.73)	(-1.22)
α_2				.06	-.21	.26
				(1.51)	(-2.66)	(1.25)
β_2				.84	-.09	-.12
				(10.5)	(-.24)	(-.13)
dispersion		.62	.53			
		(7.40)	(2.36)			
Log-L	-261.8	-250.2	-247.8	-254.1	-246.2	-248.2
Variance of s.e	1.70	1.05	.96	1.11	1.02	1.00

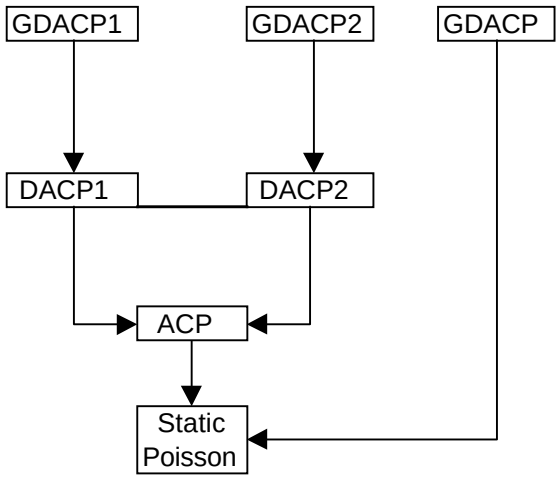
Table 2: Results from the polio dataset: no exogenous variables (T-statistics are given in parenthesis).

Coefficients	ACP	DACP1	DACP2	GDACP	GDACP1	GDACP2	Zeger
ω_1	.15	.37	.14	.15	.19	.15	
	(1.64)	(1.25)	(1.01)	(1.29)	(1.18)	(1.07)	
α_1	.16	.23	.17	.12	.17	.18	
	(2.92)	(2.18)	(1.75)	(1.78)	(1.77)	(1.62)	
β_1	.73	.52	.72	.74	.70	.71	
	(7.65)	(2.14)	(4.81)	(5.01)	(4.37)	(4.68)	
ω_2				1.45	-2.72	-3.81	
				(1.99)	(-2.18)	(-1.20)	
α_2				.30	-.21	.33	
				(-2.01)	(-2.48)	(1.30)	
β_2				.51	-.07	-.13	
				(2.90)	(-.14)	(-.14)	
trend	-.0013	-.0021	-.0010	-.0003	-.0019	-.0014	-.0044
	(-1.05)	(-1.22)	(-.50)	(-.13)	(-.57)	(-.96)	(-1.62)
cos12	-.26	-.24	-.29	-.02	-.18	-.25	-.11
	(-3.14)	(-1.89)	(-2.39)	(-.12)	(-1.51)	(-2.02)	(-.69)
sin12	-.37	-.33	-.36	-.20	-.40	-.40	-.48
	(-3.60)	(-2.19)	(-2.65)	(-1.29)	(-2.55)	(-2.82)	(-2.82)
cos6	.18	.14	.21	.02	.21	.22	.20
	(1.93)	(1.03)	(1.45)	(.14)	(1.55)	(1.52)	(1.43)
sin6	-.40	-.41	-.38	-.20	-.40	-.40	-.41
	(-4.08)	(2.91)	(-2.68)	(-1.27)	(-1.96)	(-2.37)	(-2.93)
dispersion		.69	.38				
		(7.16)	(2.07)				
Log-L	-246.5	-239.6	-238.2	-250.5	-235.9	-236.4	
Variance of s.e	1.51	.98	1.01	1.11	1.04	1.18	

Table 3: Results from the polio dataset: exogenous variables (T-statistics are given in parenthesis).

Coefficients	ACP	DACP1	DACP2	GDACP	GDACP1	GDACP2
ω_1	.79 (6.35)	.73 (3.16)	.69 (3.13)	.66 (3.43)	.82 (3.50)	.66 (3.12)
α_1	.42 (17.5)	.41 (8.75)	.41 (8.55)	.25 (5.61)	.39 (6.82)	.39 (7.04)
β_1	.45 (13.2)	.53 (13.22)	.48 (7.34)	.57 (7.62)	.47 (6.37)	.50 (7.35)
ω_2				1.07 (2.64)	-3.76 (-3.89)	1.71 (1.11)
α_2				.21 (4.16)	-.09 (-2.42)	-.04 (-.60)
β_2				.70 (11.2)	-.32 (-.94)	.59 (1.61)
dispersion		.52 (13.9)	.15 (6.11)			
Log-L	-1062.7	-1009.3	-1001.2	-1023.6	-1004.1	-998.8
Variance of s.e	1.99	1.04	1.02	1.10	1.04	1.02
Forecast						
RMSE	4.25	4.25	4.24	4.43	4.29	4.24
SSE	2.55	1.46	1.29	1.51	3.12	1.40
Bias %	.03	.03	.02	.08	.04	.03
Variance %	.22	.22	.21	.35	.27	.22
Covariance %	.75	.75	.77	.57	.69	.75

Table 4: Number of price changes larger than $.75 on the IBM stock (T-statistics are given in parenthesis.).


GDACP1
GDACP2
GDACP
DACP1
DACP2
ACP
Static
Poisson

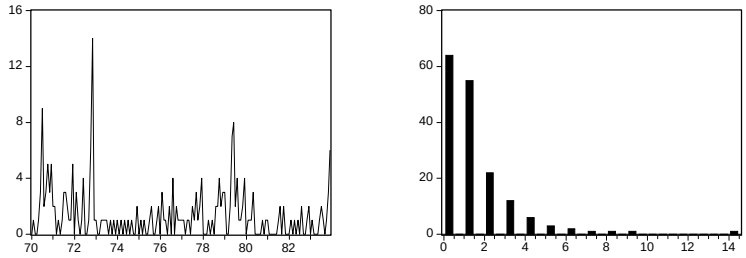


Figure 2: Plot and histogram of the monthly number of polio cases

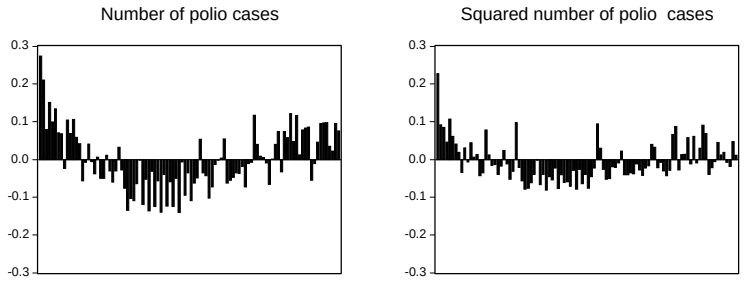


Figure 3: Autocorrelations of the number of polio cases

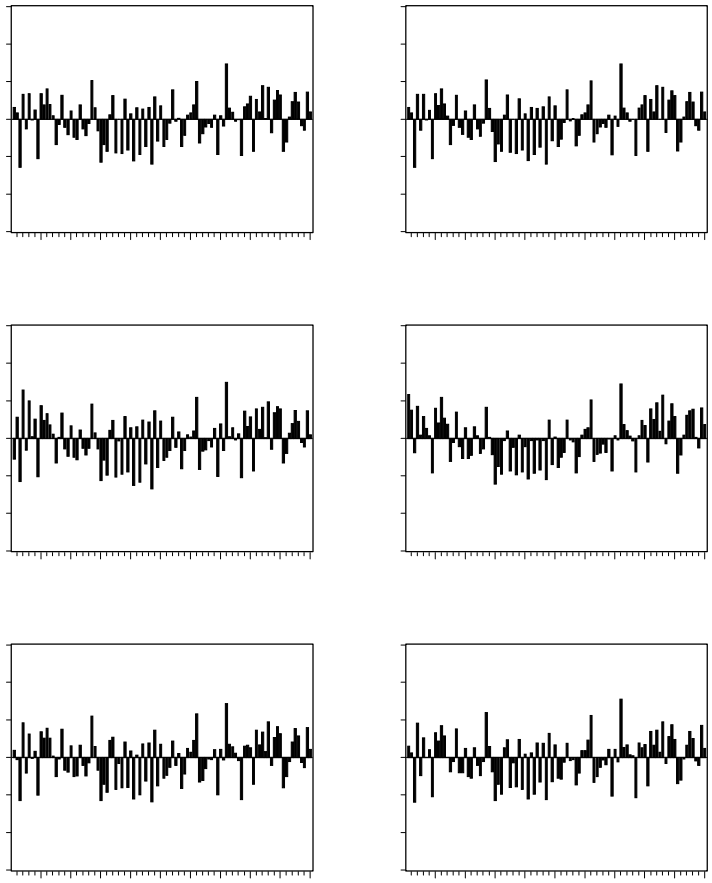


Figure 4: Autocorrelations of the Pearson residuals: no regressors

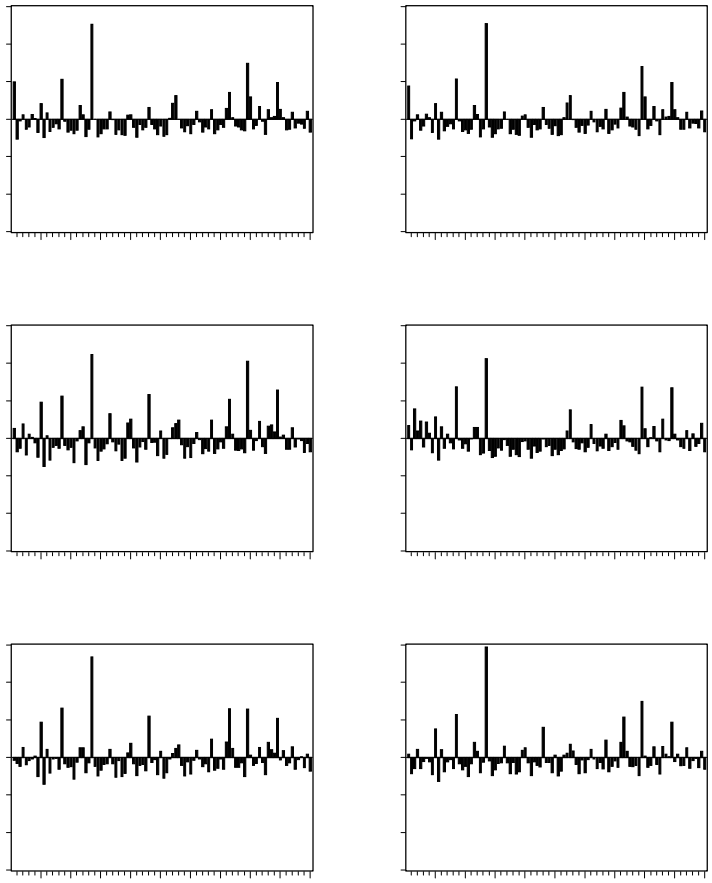


Figure 5: Autocorrelations of the squared Pearson residuals: no regressors

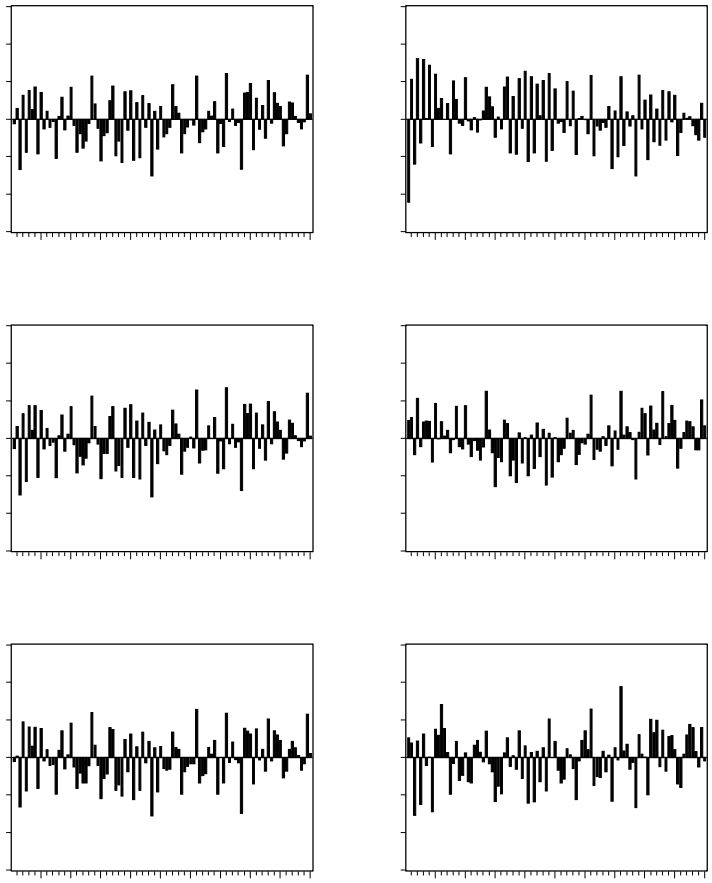


Figure 6: Autocorrelations of the Pearson residuals with time trend and seasonality

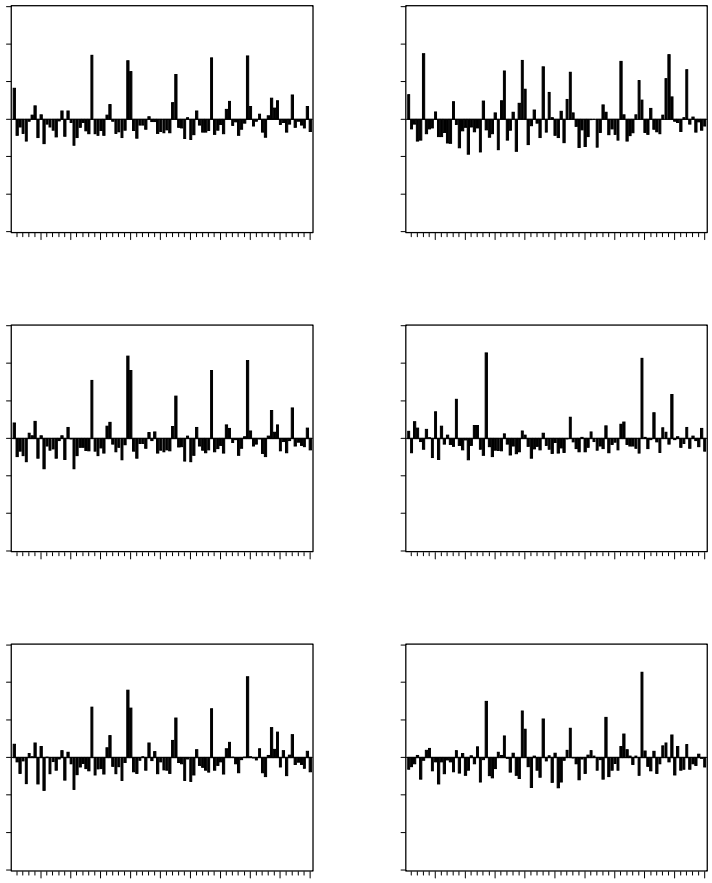


Figure 7: Autocorrelations of the squared Pearson residuals with time trend and seasonality

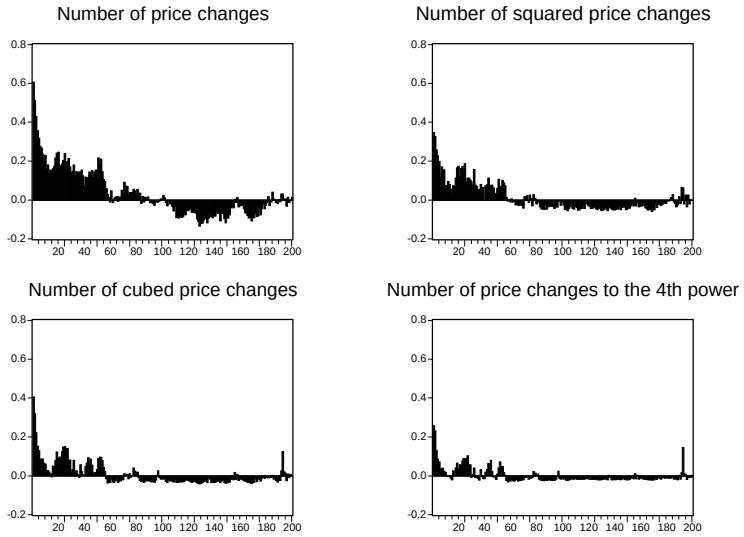


Figure 8: Autocorrelations of the daily number of price changes larger than $ .75 on IBM

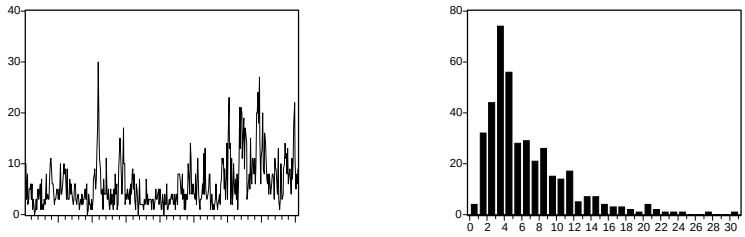


Figure 9: Plot and histogram of the daily number of price changes larger than $ .75 on IBM

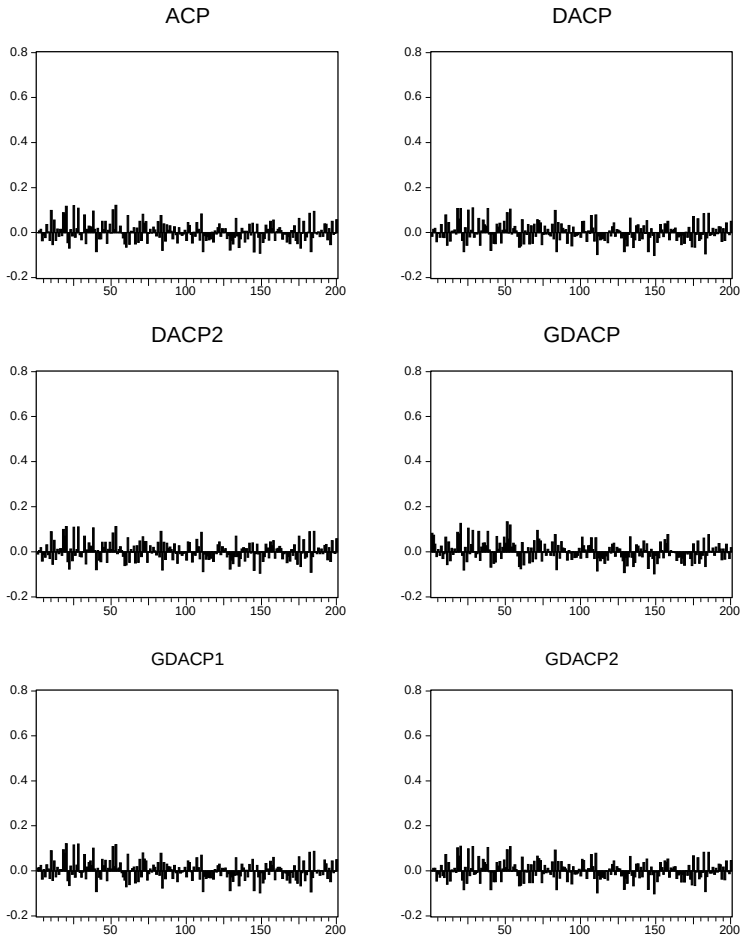


Figure 10: Autocorrelations of the Pearson residuals for the IBM price-change data

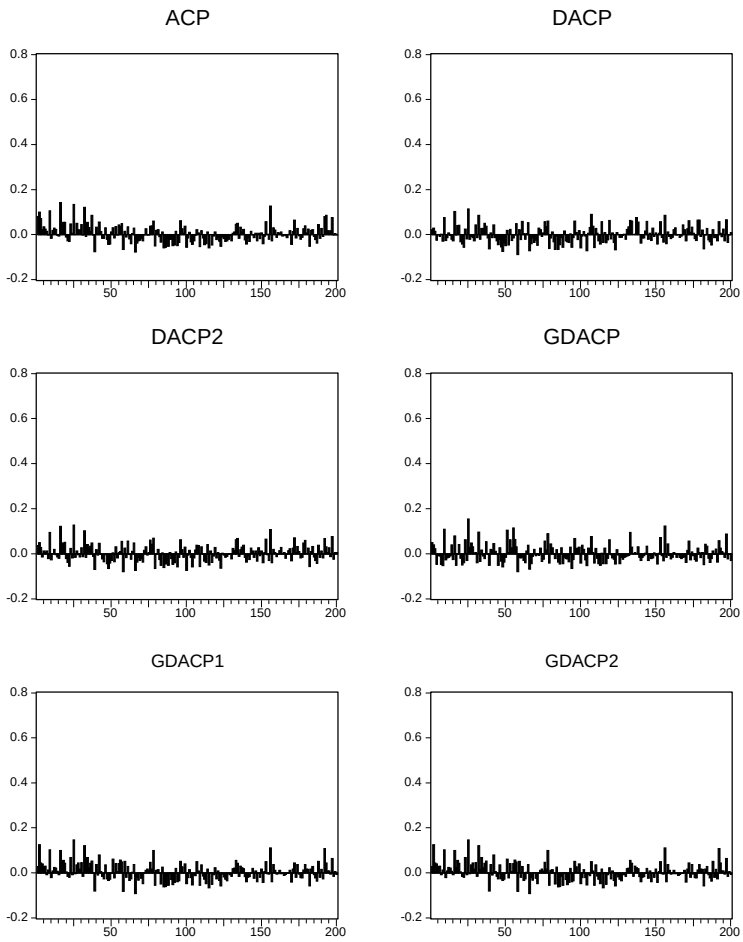


Figure 11: Autocorrelations of the squared Pearson residuals for the IBM price-change data

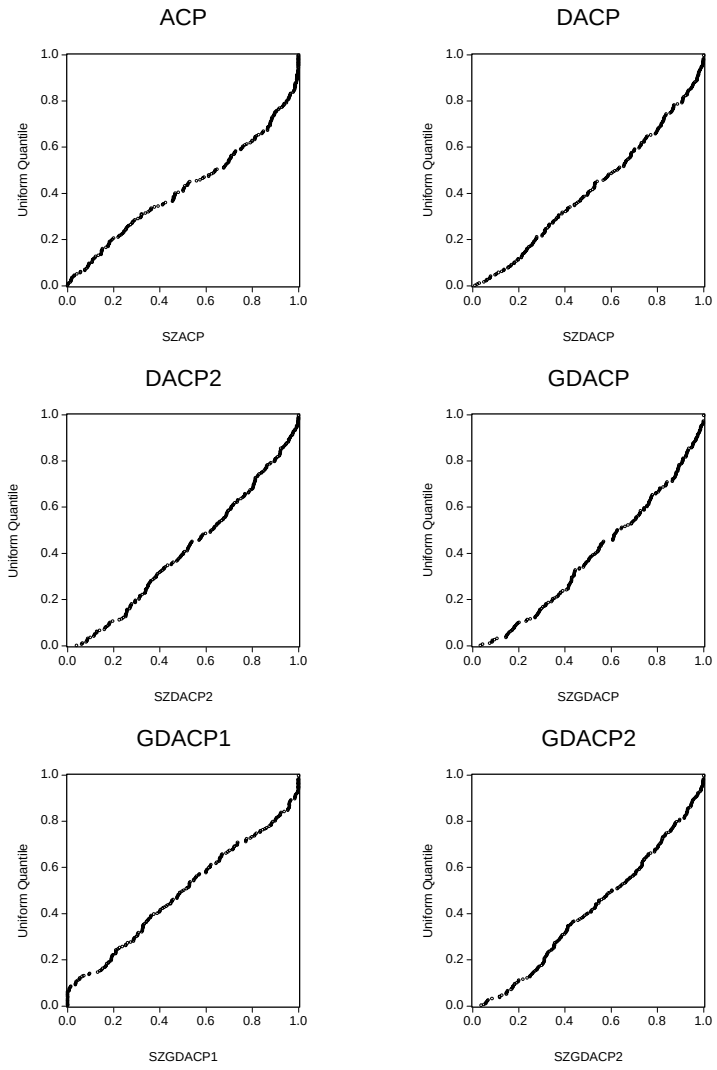


Figure 12: Plot of the quantiles of the distribution of Z against the uniform quantiles

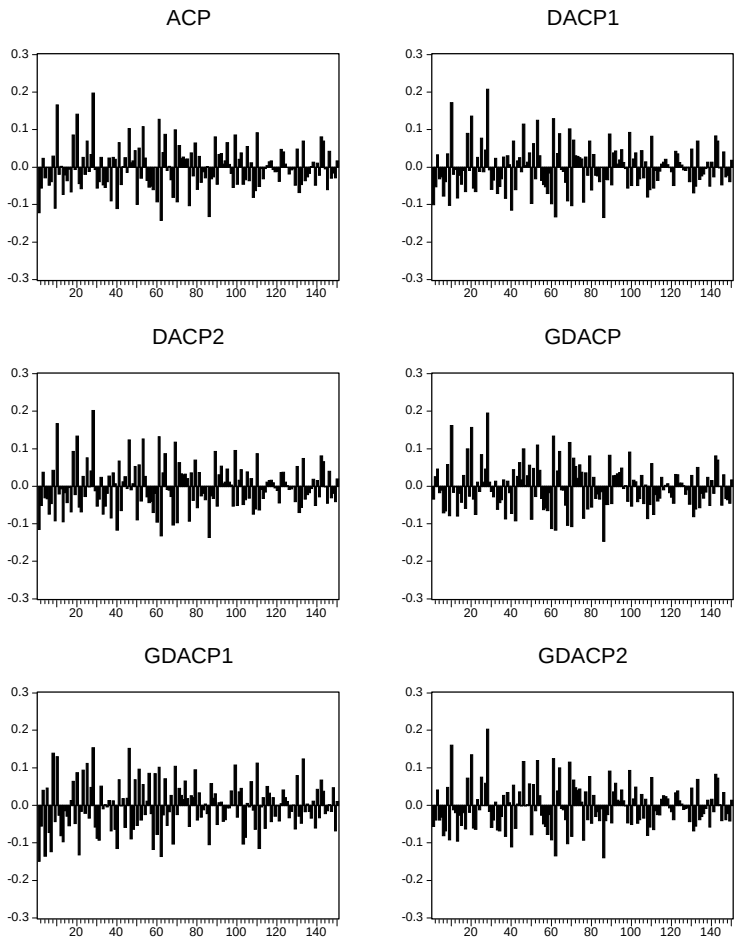


Figure 13: Autocorrelations of the Z statistics for the IBM price-change data

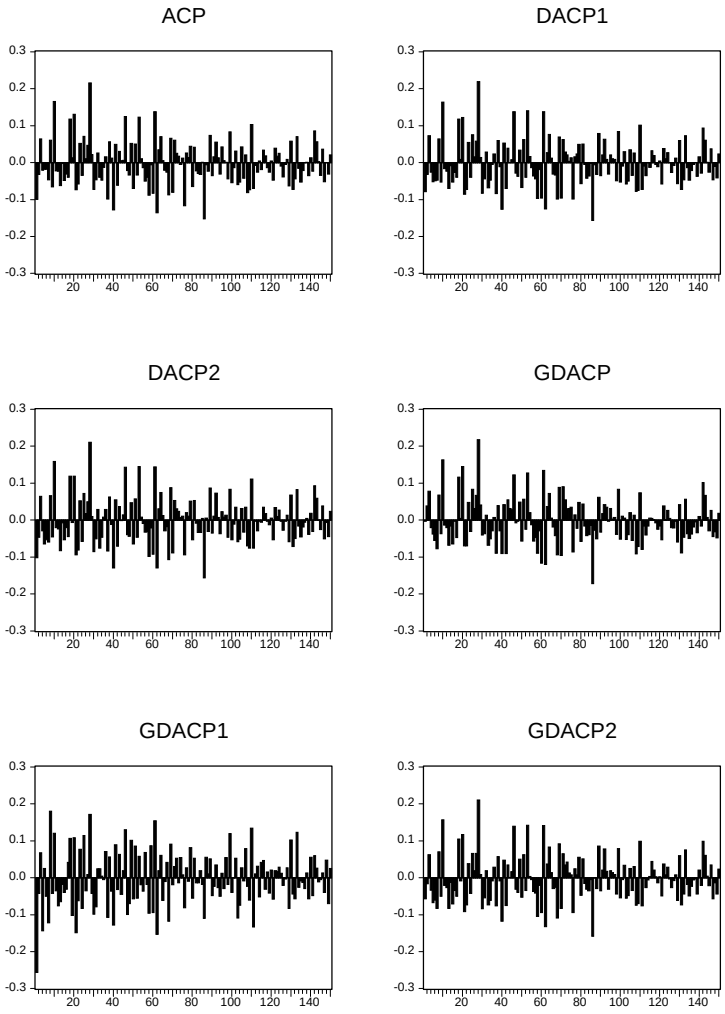


Figure 14: Autocorrelations of the squared Z statistics for the IBM price-change data