Multivariate Modelling of Time Series Count Data: An Autoregressive Conditional Poisson Model *

Andréas Heinen and Erick Rengifo

Center for Operations Research and Econometrics Catholic University of Louvain Voie du Roman Pays, 34 1348 Louvain-la-Neuve Belgium heinen@core.ucl.ac.be rengifo@core.ucl.ac.be

March 2003

JEL Classification codes: C32, C35, G10. Keywords: Count data, time series, copula, market microstructure.

Abstract

This paper introduces a new multivariate model for time series count data. The Multivariate Autoregressive Conditional Poisson model (MACP) makes it possible to deal with issues of discreteness, overdispersion (variance greater than the mean) and both auto- and cross-correlation. We model counts as Poisson or double Poisson and assume that conditionally on past observations the means follow a Vector Autoregression. We use a copula to introduce contemporaneous correlation between the series. An important advantage of this model is that it can accommodate both positive and negative correlation among variables. As a feasible alternative to multivariate duration models, the model is applied to the submission of market orders and quote revisions on IBM on the New York Stock Exchange. We show that a single factor cannot explain the dynamics of the market process, which confirms that time deformation, taken as meaning that all market events should accelerate or slow down proportionately, does not hold. We advocate the use of the Multivariate Autoregressive Conditional Poisson model for the study of multivariate point processes in finance, when the number of variables considered simultaneously exceeds two and looking at durations becomes too difficult.

^{*}The authors would like to thank Luc Bauwens, Philippe Lambert and David Veredas for helpful discussions and suggestions. The usual disclaimers apply.

1 Introduction

This paper introduces new multivariate models for time series count data. The Multivariate Autoregressive Conditional Poisson model (MACP) makes it possible to deal with issues of discreteness, overdispersion (variance greater than the mean) and both cross- and serial correlation. We take a fully parametric approach and specify a marginal distribution for the counts where, conditionally on past observations, the means follow a Vector Autoregression (VAR). This enables to attain improved inference on coefficients of exogenous regressors relative to the static Poisson regression, which is the main concern of the existing literature, while modelling the serial correlation in a flexible way. In addition to the standard model based on the Poisson distribution, we introduce a model based on the double Poisson distribution of Efron (1986), which has an additional parameter to better fit the dispersion. In addition to several conditionally uncorrelated models, we estimate a model with a copula. which introduces contemporaneous correlation between the series. We use the multivariate normal copula, which is very flexible, since it makes it possible to accommodate both positive and negative correlation, something that is impossible in most existing multivariate count distributions. 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).

Many interesting empirical questions can be addressed by modeling a time series of count data. 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 tickby-tick data. 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 and correlation. Neglecting either of these 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). All of these problems are addressed simultaneously by the use of the multivariate autoregressive conditional Poisson model (MACP). In the simplest model the counts have a Poisson distribution and their mean, conditional on past observations, is autoregressive. Whereas, conditionally on past observations the model is equidispersed (the variance is equal to the mean), it is unconditionally overdispersed. We take a fully parametric approach and choose to model the conditional distribution explicitly and make specific assumptions about the nature of the autocorrelation and cross-correlation in the series.

Autoregressive Conditional Duration (ACD) models, introduced by Engle and Russell (1998) have been used widely to test market microstructure theories with tick-by-tick data in a univariate framework. However, extensions to more than one series have proven to be very difficult. The difficulty comes from the very nature of the data, which are by definition not aligned in time. The times at which an event of any type happens are random and therefore the data is not aligned in time, unlike usual multivariate data. Engle and Lunde (1999) suggest a model for the bivariate case, but the specification is not symmetric in the two processes. In this paper we suggest working with counts instead of durations, especially when there are more than two series. Any duration series can easily be made into a series

of counts by choosing an appropriate interval and counting the number of events that occur every period. The loss of information from considering counts is largely compensated for by the possibility of modeling interactions between several series. Our approach circumvents the problem that events in the different series are not 'aligned' in time (a typical problem found in multivariate duration data), by choosing an appropriate time interval for counts, which depends on the applications at hand. In order to capture the dynamic interactions between the series we model the conditional mean as a VAR-type structure, where the past of all series affects the current counts. We focus our attention to the VARMA(1,1) case. We estimate several alternative specifications, motivated largely by considerations of parsimony. When the number of series gets large, estimating fully unrestricted models is impossible. We therefore suggest to have a factor structure, whereby the conditional mean of every series depends on one lag of itself, one lag of the count and r factors of the cross-section of lagged counts. Another specification that is nested in this model is a pure common factor, in which the dynamics is common to all series, up to a multiplicative constant.

In terms of the distributional assumptions we consider several possibilities. Given that there is no standard way of modelling a multivariate count distribution, we start with counts, which are independently Poisson or double Poisson distributed, conditionally on past observations. This means that there is no contemporaneous correlation and that all the correlation goes through the mean. In order to introduce contemporaneous correlation we consider the multivariate Poisson (MVP) distribution, which has one parameter, which captures the covariance between all pairs of series. This is an improvement over the independence assumption, but it is not flexible enough. In order to improve over this 'one fits all' characteristic of the MVP, we use a multivariate normal copula to model the dependence between the series. This gives us more parameters to fit the dependence structure of the data. We apply a two-stage estimation procedure developed in Patton (2002), which consists in estimating first the marginal models and then the copula, taking the parameters of the marginal models as given. This considerably eases estimation of the model. The main advantages of the model we propose are that it is flexible, parsimonious and easy to estimate using maximum likelihood. The results are easy to interpret and standard hypothesis tests are available.

As a feasible alternative to multivariate duration models, the MACP is applied to the study of the submission of market orders and quote revisions on IBM at the New York Stock Exchange (NYSE). Similar data has been analyzed in Hasbrouck (1999) using spectral analysis, but in his case no econometric framework has been proposed to fit a model to the data, something that we do in this paper. Using data on IBM, we test and reject the hypothesis that the dynamics of buy and sell orders and quote revisions can be explained by a common factor. The paper is organised as follows. Section 2 reviews some of the econometric literature on time series and multivariate count data. Section 3 describes the main features of the models, and focuses on the conditional mean and the marginal distribution. Section 4 introduces copulas and shows how they can be used in the present context. Section 5 describes the data. Section 6 presents results based on the Poisson distribution, whereas section 7 presents and discusses estimates of the various mean structures in the framework of the double Poisson model. Section 8 concludes.

2 A review of multivariate time series count data models

Whereas there exists a large literature on time series models of count data, in which many very different approaches have been proposed, and a fairly large literature on multivariate count distributions, there are only relatively few papers dealing with multivariate time series of counts. Some approaches have been proposed to deal with longitudinal data, which have a time series dimension, but usually the time series aspect is considered as a nuisance, that pollutes estimates and needs to be controlled for, rather than as an interesting phenomenon in its own right. We will briefly review some of the methods that have been suggested in the various strands of the literature.

Winkelmann (2000), Chapter 5 reviews correlated count data, which includes multivariate data, panel and time series models. Most existing multivariate models for counts are based on some kind of common factor (often a latent variable), about which assumptions are made, that are most often not tested for, and which underlies the dynamics of all variables. As our empirical work demonstrates, a common factor is not always able to account for the dynamics of several series of counts. Most solutions that have been proposed in this literature suffer from either of the following limitations: the models only work in the bivariate case, the correlation is restricted to be positive, the correlation between all pairs is equal. For a more complete review, see Cameron and Trivedi (1998) chapter 8 on multivariate data and chapter 9 on longitudinal data.

One possibility to generate a multivariate Poisson distribution is to use a generalisation of the so-called 'trivariate reduction method', which consists in having a common additive Poisson term in all series. This makes for a very simple system of multivariate Poisson. For a K-dimensional vector of counts N_t , the variance-covariance matrix takes up the following form: $V(N_t) = L_t + \nu \iota \iota'$, where $L_t = diag(\mu_{i,t})$ is a diagonal matrix with the mean of the individual series and ν is the common covariance term. The correlation between any two series then takes the following form: $Corr(N_{i,t}, N_{j,t}) = \frac{\nu}{\sqrt{(\mu_{i,t}+\nu)(\mu_{j,t}+\nu)}}$. The limitations are that the correlation is the same between all pairs of variables and it is restricted to be positive. This can be generalised to the negative binomial case and the variance between any two pairs will be $Cov(N_{i,t}, N_{j,t}) = \nu(1 + \sigma)$, where σ is the dispersion coefficient of the negative binomial. The problem is that the limitations are the same as for the multivariate Poisson as far as the correlation structure is concerned. These models can be estimated by maximum likelihood.

Marshall and Olkin (1990) propose a bivariate Poisson model based on mixing. The distribution of the data is the results of a mixture of a Poisson, whose mean depends on regressors and a Gamma-distributed multiplicative heterogeneity term with parameter $\frac{1}{\nu}$. This is similar to the Poisson random effects model suggested by Hausman, Hall, and Griliches (1984). As in the models described above, the fact that there is only one heterogeneity term restricts the variance-covariance matrix, which takes the following form: $V(N_t) = L_t + L_t \nu \iota t' L_t$. It can be seen that in that model the covariances are a product of the Gamma parameter and the product of the means, which means that the parameter of the mixing distribution is responsible both for the dispersion and the correlation, which is a limitation of this model. Several estimation methods have been proposed to deal with this model.

More general models can be obtained if the assumption of a common term is replaced

by the assumption of correlated terms. Aitchison and Ho (1989) consider a mixture of a Poisson model with a log-normally distributed error term. This allows for a fully general variance-covariance matrix and is more flexible than the models that have been proposed previously. Munkin and Trivedi (1999) propose to estimate this model using a simulated maximum likelihood approach. The problem is that the estimation of resulting models is quite cumbersome and requires numerically intensive methods.

Many different approaches have been proposed to model time series count data. Good reviews can be found both in Cameron and Trivedi (1998), Chapter 7 and in MacDonald and Zucchini (1997), Chapter 1. Chang, Kavvas, and Delleur (1984) apply Discrete Autoregressive Moving Average (DARMA) models, which are probabilistic mixtures of discrete i.i.d. random variables with suitably chosen marginal distributions. 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. 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. Zeger (1988) extends the generalised linear models and introduces a latent multiplicative autoregressive term with unit expectation, which is responsible for introducing both autoregression and overdispersion into the model. Harvey and Fernandes (1989) use state-space models in the univariate framework with conjugate prior distributions. Counts are modelled 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. Jorgensen, Lundbye-Christensen, Song, and Sun (1999) introduce a multivariate extension of state-space modelling for time series. The problem in this model is that there is a non-stationary latent factor common to all series, whereas we will see in our application that the dynamic of the variables we are dealing with cannot be accounted for by a single factor.

3 A VAR in counts

3.1 The conditional mean

In order to extend the Autoregressive Conditional Poisson model to a $(K \times 1)$ vector of counts N_t , we build a VAR-type system for the conditional mean. In a first step, we assume that conditionally on the past, the different series are uncorrelated. This means that there is no contemporaneous correlation and that all the dependence between the series is assumed to be captured by the conditional mean. We first propose to model counts as having a Poisson distribution with autoregressive means. 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. However most count data exhibit overdispersion. By modelling the mean as an autoregressive process, we generate overdispersion in even the simple Poisson case.

$$N_{i,t}|\mathcal{F}_{t-1} \sim P(\mu_{i,t}), \ \forall i = 1, \dots, K,$$

$$(3.1)$$

where \mathcal{F}_{t-1} is the information set generated by all K series up to and including time t-1, $N_t = (N_{1,t}, N_{2,t}, \ldots, N_{K,t})'$ and $\mu_t = (\mu_{1,t}, \mu_{2,t}, \ldots, \mu_{K,t})'$. The conditional means μ_t are assumed to follow a VARMA-type process:

$$E[N_t|\mathcal{F}_{t-1}] = \mu_t = \omega + \sum_{j=1}^p A_j N_{t-j} + \sum_{j=1}^q B_j \mu_{t-j}$$
(3.2)

which can always be written in VAR form:

$$\mu_t = c + G(L)N_t , \qquad (3.3)$$

where $G(L) = \sum_{i=1}^{\infty} G_i L^i = (I - \sum_{j=1}^{q} B_j L^j)^{-1} \sum_{j=1}^{p} A_j L^j$ and L is the lag operator. This model can be defined for arbitrary orders, but in practice, as is the case for the GARCH model, the most common is the (1,1) specification. For reasons of simplicity, in most of the ensuing discussion, we will focus on the most common (1,1) case and for notational simplicity, we will denote $A = \sum_{j=1}^{p} A_j$ and $B = \sum_{j=1}^{q} B_j$ and drop the index whenever there is no ambiguity.

There are several special cases of the general model that are interesting from the point of view of econometric modeling of economic or financial count data. The simplest case is the bivariate model, in which both the AR and the MA matrices A and B are left unrestricted. In systems with large K, which could be found, for instance when analysing a large group of stocks like the constituents of an index, the full approach would not be feasible, as the number of parameters would get too large. If we assume that A and B are of full rank, the number of parameters that has to be estimated in this model would be $2K^2 + K$. In situations where this is not an option, we propose to impose some additional structure on the process of the conditional mean. We distinguish the following cases of interest:

1. A or B are of full rank

No zero restrictions are imposed and all coefficients are estimated. This only makes sense in small systems.

2. A priori restrictions on A or B

It might be the case that financial theory gives the econometrician a clue as to which parameters should be zero. In that case these restrictions can be simply imposed and the model estimated in that form.

3. A or B are of reduced rank

 $A = \gamma \delta'$ where γ and δ are (K, r) matrices. This specification is useful for large systems, where estimation of the full system does not make sense or where the number of parameters is just too big for estimation. In that case a reasonable solution is to assume a factor structure, whereby lagged counts affect the conditional mean only through a series of r linear combinations of the original series. Examples of this include testing for the presence of common factors in market trading, as measured by number of transactions for a cross-section of stocks.

4. A or B are diagonal

If no cross-effects are assumed, then A or B can be assumed to be diagonal. If both A and B are assumed to be diagonal, then of course the system simplifies to a series of univariate models. The diagonality assumption imposed on B could mainly be a useful device to reduce the number of parameters in large systems.

5. Reduced rank and own effect

In this formulation it is assumed, that for every series the conditional mean depends on one lag of itself, one lag of the count and r factors of the cross-section of lagged counts. $A = diag(\alpha_i) + \gamma \delta'$ where γ and δ are (K, r) matrices. This is suited for large systems, where imposing a reduced rank is necessary for practical reasons, but there is reason to believe that every series' own past has explanatory power beyond the factor structure. In particular, the conditional mean can be specified as:

$$\mu_t = \omega + (diag(\alpha_i) + \gamma \delta')N_{t-1} + diag(\beta_i)\mu_{t-1} .$$
(3.4)

6. Common factor

In some cases one might want to assume that the dynamics of all the series under consideration is common, and that one factor explains the dynamics of the whole system. This can be obtained as a special case of our specification under the following set of assumptions: $A = \alpha \gamma \delta'$, $B = \beta I$, $\omega = c \gamma$, where α , β are scalars, $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_K)'$ and $\delta = (1, \delta_2, \ldots, \delta_K)'$, where we impose the normalisation $\delta_1 = 1$ in order to identify the model. This means that if we denote $f_t = \delta' N_t$, we have an autoregressive process for the factor:

$$\mu_t^0 = c + \alpha f_{t-1} + \beta \mu_{t-1}^0 ,$$

and $\mu_t = \gamma \mu_t^0$.

3.2 The conditional distribution

So far we have assumed that a model based on the Poisson distribution could accurately describe the data. 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 imaginable that the amount of overdispersion in the data is less than the overdispersion resulting from the autocorrelation, in which case an underdispersed marginal distribution might be appropriate. In order to account for these possibilities we consider the double Poisson distribution 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 ϕ 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:

$$f(y,\mu,\phi) = c(\mu,\phi) \left(\phi^{\frac{1}{2}}e^{-\phi\mu}\right) \left(\frac{e^{-y}y^{y}}{y!}\right) \left(\frac{e\mu}{y}\right)^{\phi y}$$
(3.5)

 $f(y, \mu, \phi)$ 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 $c(\mu, \phi)$, 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:

$$\frac{1}{c(\mu,\phi)} = 1 + \frac{1-\phi}{12\mu\phi} \left(1 + \frac{1}{\mu\phi}\right)$$
(3.6)

Furthermore 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 ϕ is smaller or larger than 1. In the case of the Double Poisson (DP hereafter), the distributional assumption 3.1 is replaced by the following:

$$N_{i,t}|\mathcal{F}_{t-1} \sim DP(\mu_{i,t},\phi_i), \ \forall i = 1,\dots,K.$$

$$(3.7)$$

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 $\frac{\mu}{\phi}$. Efron (1986) shows that this approximation is highly accurate, and we will use it in our more general specifications. With the double Poisson, the conditional variance is equal to:

$$V[N_t|\mathcal{F}_{t-1}] = \sigma_t^2 = \frac{\mu_t}{\phi}$$
(3.8)

The coefficient ϕ of the conditional mean will be a parameter of interest, as values different from 1 will represent departures from the Poisson distribution.

The Double Poisson generalises the Poisson in the sense of allowing more flexible dispersion patterns. Another dimension in which we would like to generalise the present framework is in terms of correlation. The Multivariate Poisson (MVP) does this. This distribution considers that each variable is the sum of a Poisson term, whose mean depends on regressors and a Poisson term with mean ν , which is common to all variables and is responsible for the correlation amongst them. The distribution has the following form:

$$f(N_{1,t} = n_{1,t}, \dots, N_{K,t} = n_{K,t}, \mu_t) = \exp\left(-\sum_{i=1}^K \mu_{i,t} - \nu\right) \sum_{j=0}^{\min(N_{1,t},\dots,N_{K,t})} \frac{\nu^j}{j!} \prod_{i=1}^K \frac{\mu_{i,t}^{(N_{i,t}-j)}}{(N_{i,t}-j)!}$$

Although this distribution is more general than the Poisson, it has a few serious drawbacks. Firstly there is one parameter ν , which is equal to the covariance between any pairs of series. This makes ν a 'one fits all' covariance coefficient. This is obviously a serious limitation. In addition, given that it is a common term, which is responsible for the correlation, this model can only accommodate positive correlation between series. This distribution can be generalised easily to the multivariate negative binomial (MVNB), which combines the features of the MVP with the more flexible dispersion of the negative binomial. The limitations of the MVNB are therefore the same as the ones of the MVP.

The following properties about the unconditional moments of the MDACP can be established. As the MACP is a special case of the MDACP, we only present the more general results.

Proposition 3.1 (Unconditional mean of the MDACP(p,q)). The MACP is stationary if and only if the eigenvalues of (I - A - B) lie within the unit circle. In that case, the unconditional mean of the MDACP(p,q) is

$$E[N_t] = \mu = (I - A - B)^{-1}\omega$$
(3.9)

This proposition shows that, as long as the roots of the sum of the autoregressive coefficient matrices are within the unit circle, the model is stationary and the expression for its mean is identical to the one of a VARMA process.

Proposition 3.2 (Unconditional variance of the MDACP(1,1) Model). The unconditional variance of the MDACP(1,1) model, when the conditional mean is given by 3.2, is equal to:

$$vec(V[N_t]) = \left(I_{K^2} + \left(\left(I_{K^2} - (A+B) \otimes (A+B)'\right)^{-1} \cdot (A \otimes A'\right)\right) \cdot vec(\Omega), \quad (3.10)$$

where $\Omega = E\left(Var[N_t|\mathcal{F}_{t-1}]\right).$

Proof of Proposition 3.2. Proof in appendix

 Ω will take on different values for the models we have considered so far. For the ACP, the conditional distribution is equidispersed and therefore $\Omega = E[Var(N_t|\mathcal{F}_{t-1})] = diag(\mu)$. In the case of the DACP, the variance is equal to the ratio of the mean to the dispersion parameter: $\Omega = E[Var(N_t|\mathcal{F}_{t-1})] = diag(\frac{\mu_i}{\phi_i})$. In both cases, the covariances are zero and therefore the variance-covariance matrix is diagonal. For the MVP, the variance is composed of two parts: $\Omega = E[Var(N_t|\mathcal{F}_{t-1})] = E[diag(\mu_{i,t}) + \nu\iota\iota'] = diag(\mu_i) + \nu\iota\iota'$, where ι is a column vector of ones.

Proposition 3.3 (Autocovariance of the MDACP(1,1) Model). The autocovariance of the MDACP(1,1) model, when the conditional mean is given by 3.2, is equal to:

$$vec(Cov[N_t, N_{t-s}]) = \left[I \otimes A^{-1} \left((A+B)^s - B(A+B)\right)\right] \cdot \left(I_{K^2} + \left(\left(I_{K^2} - (A+B) \otimes (A+B)'\right)^{-1} \cdot (A \otimes A'\right)\right) \right) \cdot vec(\Omega)$$
(3.11)

where $\Omega = E(Var[N_t | \mathcal{F}_{t-1}]).$

Proof of Proposition 3.3. Proof in appendix

9

4 A general multivariate model using copulas

The models we mentioned so far are either uncorrelated contemporaneously, or, in the case of the MVP, introduced in the previous section, there is only one parameter to fit all the cross-correlations. This 'one-fits-all' characteristic of the covariance parameter of MVP and the absence of cross-correlation in the other models is largely unsatisfactory. In order to generate richer patterns of cross-correlation, we resort to copulas. Copulas are a very general way of introducing dependence among several series with known marginals. Copula theory goes back to the work of Sklar (1959), who showed that a joint distribution can be decomposed into its K marginal distributions and a copula, that describes the dependence between the variables. This theorem provides an easy way to form valid multivariate distributions from known marginals that need not be necessarily of the same distribution, i.e. it is possible to use normal, student or any other marginals, combine them with a copula and get a suitable joint distribution, which reflects the kind of dependence present in the series. In recent years, there have been many applications of copulas in finance, in particular Patton (2001) uses the theory of conditional copulas in order to get flexible multivariate density models, that allow for time-varying conditional densities of each variable, and for time-varying conditional dependence among all variables. His application is to the Japanese yen - U.S dollar and euro - U.S. dollar exchange rates. A more detailed account of copulas can be found in Joe (1997) and in Nelsen (1999).

4.1 Multivariate copulas

Let $H(y_1, \ldots, y_K)$ be a continuous K-variate cumulative distribution function with univariate margins $F_i(y_i)$, $i = 1, \ldots, K$, where $F_i(y_i) = H(\infty, \ldots, y_i, \ldots, \infty)$. According to Sklar (1959), there exists a function C, called copula, mapping $[0, 1]^K$ into [0, 1], such that:

$$H(y_1, \dots, y_K) = C(F_1(y_1), \dots, F_K(y_K)), \qquad (4.1)$$

And then, the joint density function is given by the product of the marginals and the copula density:

$$\frac{\partial H(y_1, \dots, y_K)}{\partial y_1 \dots \partial y_K} = \prod_{i=1}^K f_i(y_i) \frac{\partial C(F_1(y_1), \dots, F_K(y_K))}{\partial F_1(y_1) \dots \partial F_K(y_K)} .$$
(4.2)

With this we can define the copula of a multivariate distribution with Uniform [0, 1] margins as:

$$C(z_1, \dots, z_K) = H(F_1^{-1}(z_1), \dots, F_K^{-1}(z_K)) , \qquad (4.3)$$

where $z_i = F_i(y_i)$, for i = 1, ..., K.

It can be shown that any copula must be bounded by what is known as the Fréchet bounds:

$$W(F_1(y_1), \dots, F_K(y_K)) \le C(F_1(y_1), \dots, F_K(y_K)) \le M(F_1(y_1), \dots, F_K(y_K)), \quad (4.4)$$

where

$$W(F_1(y_1), \dots, F_K(y_K)) = \max\left(\sum_{i=1}^K F_i(y_i) - 1, 0\right) , \qquad (4.5)$$

$$M(F_1(y_1), \dots, F_K(y_K)) = \min(F_1(y_1), \dots, F_K(y_K)), \qquad (4.6)$$

which correspond respectively to the complete negative and positive dependence cases. As we can see with the use of the copulas we are able to map the univariate marginal distributions of K random variables, each supported in the [0, 1] interval, to their K-variate distribution, supported on $[0, 1]^K$, something that holds, no matter what the dependence among the variables is (including if there is none).

There exist many families of copulas, which have different properties in terms of the kind of dependence that they introduce between the variables they are applied to. Besides the well-known normal copula, which introduces linear and symmetric dependence, there exist many more exotic copulas. For instance the Clayton copula introduces more dependence in the left than in the right tail of the marginal distributions. This type of asymmetric dependence has been used by Patton (2002) to model joint returns for assets, which are more strongly correlated during downward than upward movements.

Most of the literature on copulas is concerned with the bivariate case. We are, however trying to specify a general type of multivariate count model, not limited to the bivariate case. Whereas there are many alternative formulations in the bivariate case, the number of possibilities for multi-parameter multivariate copulas is rather limited. We choose to work with the most intuitive one, which is arguably the Gaussian copula, obtained by the inversion method (based on Sklar (1959)). This is a K-dimensional copula such that:

$$C(z_1, \dots, z_K; \Sigma) = \Phi^K(\Phi^{-1}(z_1), \dots, \Phi^{-1}(z_K); \Sigma) , \qquad (4.7)$$

and its density is given by,

$$c(z_1, \dots, z_K; \Sigma) = |\Sigma|^{-1/2} \exp\left(\frac{1}{2}(q'(I_K - \Sigma^{-1})q)\right),$$
 (4.8)

where Φ^K is the K-dimensional standard normal multivariate distribution function, Φ^{-1} is the inverse of the standard univariate normal distribution function and $q = (q_1, \ldots, q_K)'$ with normal scores $q_i = \Phi^{-1}(z_i)$, $i = 1, \ldots, K$. It is shown in Joe (1997) that the Gaussian copula attains the lower Fréchet bound, independence or the upper Fréchet bound, when the correlation parameter is equal to -1, 0 or 1, respectively. Furthermore, it can be seen that if Y_1, \ldots, Y_K are mutually independent, the matrix Σ is equal to the identity matrix I_K and the copula density is then equal to 1.

4.2 Copulas and discrete distributions

In the present paper we are using discrete marginals, either Poisson or Double Poisson, whose support is the set of integers, instead of continuous ones, which are defined for real values. In that case, a multivariate probability function h is obtained by taking the Radon-Nikodym derivative of $H(y_1, \ldots, y_K)$ in (4.1) with respect to the counting measure,

$$h(y_1, \dots, y_K) = P(Y_1 = y_1, \dots, Y_K = y_K) = \sum_{j_1=1}^2 \cdots \sum_{j_K=1}^2 (-1)^{j_1 + \dots + j_K} C(z_{1j_1}, \dots, z_{Kj_K}; \Sigma) ,$$
(4.9)

where $z_{i1} = F_i(y_i)$ and $z_{i2} = F_i(z_i-)$, i.e. the left hand limit of F_i at y_i which is equal to $F_i(y_i-1)$ when the distribution is Poisson, where the support of F_i is the set of all integers. If the marginal distributions functions are all continuous then C is unique. However when the marginal distributions are discrete, this is no longer the case and the copula is only uniquely identified on $\bigotimes_{i=1}^{K} Range(F_i)$, a K-dimensional set, which is the Cartesian product of the range of all marginals. Moreover, a crucial assumption, which underlies the use of copulas, is that the marginal models are well specified and that the probability integral transformation (PIT) of the variables under their marginal distributions is that the Probability Integral Transformation Theorem (PITT) of Fisher (1932) does not apply, and the uniformity assumption does not hold, regardless of the quality of the specification of the marginal model. The PITT states that if Y is a continuous variable, with cumulative distribution F, then

$$Z = F(Y)$$

is uniformly distributed on [0, 1].

Denuit and Lambert (2002) use a continuousation argument to overcome these difficulties and apply copulas with discrete marginals. The main idea of continuousation is to create a new random variable Y^* by adding to a discrete variable Y a continuous variable U valued in [0, 1], independent of Y, with a strictly increasing cdf, sharing no parameter with the distribution of Y, such as the Uniform [0, 1] for instance:

$$Y^* = Y + (U - 1)$$
.

As the authors point out, continuousation does not alter the concordance between pairs of random variables. Intuitively, two random variables Y_1 and Y_2 are concordant, if large values of Y_1 are associated with large values of Y_2 . Concordance is an important concept, since it underlies many measures of association between random variables, such as Kendall's tau for instance. It is easy to see that continuousation does not affect concordance, since $Y_1^* > Y_2^* \iff Y_1 > Y_2$.

Using continuousation, Denuit and Lambert (2002) state a discrete analog of the PITT. If Y is a discrete random variable with domain χ , in **N**, such that $f_y = P(Y = y), y \in \chi$, continuoused by U, then

$$Z^* = F^*(Y^*) = F^*(Y + (U - 1)) = F([Y^*]) + f_{[Y^*]+1}U = F(Y - 1) + f_yU$$

is uniformly distributed on [0, 1], and [Y] denotes the integer part of Y.

In this paper, we will use the continuoused version of the probability integral transformation in order to test the correct specification of the marginal models. If the marginal models are well-specified, then Z^* , the PIT of the series under the estimated distribution and after continuousation, is uniformly distributed. We will also use Z^* as an argument in the copula, since, provided that the marginal model is well specified, this will ensure that the conditions for use of a copula are met.

One remark needs to be made concerning the use of continuousation in the present context. In a sense the lack of identifiability of the copula outside of the range of the cumulative distribution of the marginal model is less acute in the time-varying distribution case, as the number of points at which the copula is observed increases, relative to the static case. In order to illustrate this point, let's consider the case of Bernoulli variables, which are in a sense, the 'most discrete' possible random variables. The problem we describe is the same with the Poisson or the Double Poisson distribution. We consider the Bernoulli variables Y_i , for $i = 1, \ldots, K$, whose cumulative density functions F_i can only take 3 possible values:

$$Z_{i,t} = F_i(Y_{i,t}) = \begin{cases} 0 & \text{if } y_{i,t} \le 0\\ p_i & \text{if } 0 < y_{i,t} < 1\\ 1 & \text{otherwise} \end{cases}$$

The copula is then only identified on the set $S = \bigotimes_{i=1}^{K} \{0, p_i, 1\}$. Therefore it is impossible to distinguish two copulas which have the same values on S, but are different on $[0, 1]^n \cap \overline{S}$. In the case where the distributions are time-varying, we have:

$$F_{i,t}(Y_{i,t}) = \begin{cases} 0 & \text{if } y_{i,t} \le 0\\ p_{i,t} & \text{if } 0 < y_{i,t} < 1\\ 1 & \text{otherwise} \end{cases}$$

The copula is now identified on the set $\bigcup_{t=1}^{T} \bigotimes_{i=1}^{K} \{0, p_{i,t}, 1\}$, which is obviously much larger a set than S. Nonetheless, it remains true in the time-varying case, that the non-corrected Z-statistic is not uniformly distributed, which, alone, justifies the use of continuousation.

4.3 Copulas and Time Series

In the present paper we use a two stage maximum likelihood estimator suggested by Patton (2002). Patton (2002) develops a theory of conditional copulas for the estimation of copulas when the marginal models are time-varying. The estimation is based on maximum likelihood, where the likelihood is broken down into a part relating to the marginal models and a part for the copula:

$$L_{X_1X_2}(\theta) = L_{X_1}(\theta_1, \theta_0) + L_{X_2}(\theta_2, \theta_0) + L_C(\theta_1, \theta_2, \theta_0, \theta_c)$$

His method is applicable when the parameters of the joint distribution can be partitioned into elements relating exclusively to the marginal distributions and elements relating to the copula, which corresponds to the case where there is no θ_0 in the previous equation. In most models we are interested in, it is impossible to separate the marginal models (they have common parameters, as in the factor models, or one series depends on the conditional mean of the other series).

Several estimators are possible. The full information estimator, which consists in maximising the joint likelihood, taking all interactions into consideration is by definition the most efficient, even though, depending on the applications its implementation might be problematic. A two-step procedure, where the marginal models are estimated first, and their quantiles are used in a second stage, to estimate the parameters of the copula is shown to perform very well on currency data. In the first stage, one computes

$$\hat{\varphi} = \underset{\varphi \in \Phi}{\arg \max} L_{X_1}(\theta_1, \theta_0) + L_{X_2}(\theta_2, \theta_0)$$
(4.10)

where $\hat{\varphi} = [\hat{\theta_0}', \hat{\theta_1}', \hat{\theta_2}']'$ and in the second stage,

$$\hat{\theta}_c = \underset{\theta_c \in \Theta_C}{\arg \max} L_C(\theta_1, \theta_2, \theta_0, \theta_c)$$
(4.11)

where $\hat{\theta} = [\hat{\varphi}', \hat{\theta_c}']'$.

Patton (2002) shows the asymptotic normality for the two-stage estimator based on the work of Newey and McFadden (1994). The two-stage estimates of the copula parameters are less asymptotically efficient than the full information estimates. To take into consideration this fact, Patton adjusts the two-stage estimator taking a single iteration of the Newton-Raphson algorithm starting from the two-stage parameter estimate. With this modification the updated estimates attain the minimum asymptotic variance bound.

4.4 The Model

Having dealt with the problems due to the discreteness and the time-varying nature of the marginal density in earlier sections, we proceed with the estimation of the model. The joint density of the counts in the Double Poisson case with the Gaussian copula is:

$$h(N_{1,t},\ldots,N_{K,t},\theta,\Sigma) = \prod_{i=1}^{K} DP(N_{i,t},\mu_{i,t},\phi_i) \cdot c(q_t;\Sigma) ,$$

 $DP(N_{i,t}, \mu_{i,t}, \phi_i)$ denotes the Double Poisson density as a function of the observation $N_{i,t}$, the conditional mean $\mu_{i,t}$ and the dispersion parameter ϕ_i . c denotes the copula density of a multivariate normal and $\theta = (\omega, vec(A), vec(B))$.

The $q_{i,t}$, gathered in the vector q_t are the normal quantiles of the $z_{i,t}$:

$$q_t = (\Phi^{-1}(z_{1,t}), \dots, \Phi^{-1}(z_{K,t}))',$$

where the $z_{i,t}$ are the PIT of the continuous ed count data, under the marginal densities: $z_{i,t} = F^*(N_{i,t}^*) = F(N_{i,t} - 1) + f(N_{i,t}) * U_{i,t}$,

The $N_{i,t}^*$ are the continoused version of the original count data $N_{i,t}$:

$$N_{i,t}^* = N_{i,t} + (U_{i,t} - 1)$$
.

Finally the $U_{i,t}$ are uniform random variable, on [0, 1]. Taking logs, one gets:

$$log(h_t) = \sum_{i=1}^{K} log(DP(N_{i,t}, \mu_{i,t}, \phi_i)) + log(c(q_t; \Sigma))]$$

We consider two estimators as in Patton (2002), a full information estimator and a two-stage estimator. Given that we use the multivariate normal copula, the second step of the two-stage procedure does not require any optimisation, as the MLE of the variance-covariance matrix of a multivariate normal with a zero mean, is simply the sample counterpart:

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} q_t q_t$$

It is important to realise that correct specification of the density in the marginal models is crucial to the specification of the copula, as any mistake would have as a consequence the fact that the uniformity assumption is violated which would invalidate the use of copulas. We evaluate models on the basis of their log-likelihood, but also on the basis of their Pearson residuals, which are defined as: $\epsilon_t = \frac{N_t - \mu_t}{\sigma_t}$. If a model is well specified, the Pearson residuals will have variance one and no significant autocorrelation left. Another tool for checking the specification is the $z_{i,t}$'s. If the model is well specified, the $z_{i,t}$'s should be uniformly distributed and serially uncorrelated. We will check this for all the models we estimate.

5 Data description

We are working on the IBM stock and we consider the number of the various events, that constitute the market process, which happen in 5-minute intervals. The data we use was taken from the Trades and Quotes (TAQ) dataset, produced by the New York Stock Exchange (NYSE). This data set contains every trade and quote posted on the NYSE, the American Stock Exchange and the NASDAQ National Market System for all securities listed on NYSE. We first remove any trades that occurred with non-standard correction or G127 codes (both of these are fields in the trades data base on the TAQ CD's), such as trades that were cancelled, trades that were recorded out of time sequence, and trades that were called for delivery of the stock at some later date. Any trades that were recorded to have occurred before 9:40am or after 4pm (the official close of trading) were removed. The reason for starting at 9:40 instead of 9:30am, the official opening time, is that we wanted to make sure that none of the opening transactions were accidentally included in the sample, or that there would not be artificially low numbers of events at the start of the day, due to the fact that part of the first interval was taking place before the opening transaction. This could have biased estimates of intradaily seasonality. An unfortunate feature of the TAQ dataset is that the initiating side of the transaction is not known. We therefore have to infer the sign of the transactions. Transactions were classified as buys or sells according to a procedure proposed and tested by Lee and Ready (1991), that is now common in the empirical market microstructure literature. This procedure identifies the

standing quote at a given trade, calculates the mid-quote, and compares the price at which the trade occurred with the mid-quote. If the trade price was higher than the mid-quote, the trade is considered "buyer-initiated". If the trade price was lower than the mid-quote, the trade is considered "seller-initiated". If the trade price was exactly the mid-quote, then it is considered "indeterminate". We use the tick test to classify the trades for which comparison with the prevailing quote is inconclusive. The principle of this test is that if a transaction takes place at a higher price than the previous one, it must have been a buyerinitiated transaction. Due to problems in the trade and quote recording process, Lee and Ready (1991) suggest that using quotes that are at least five seconds old as the standing quote for a trade is preferable to using quotes immediately prior to a trade.

The events we analyze are the number of buyer- and seller-initiated trades, the number of quote updates, that change either the bid or the ask (or both) and the number of quote revisions with only changes in the depth and no price changes. The data used was from January 1st 1998 to March 31st 1998. This means that the sample covers 61 trading days, that represent 4636 observations, as there are 76 5-minute intervals every day between 9:40 AM and 4 PM. The descriptive statistics are given in Table 1. The means of the series are relatively small, which makes the use of a continuous distribution like the normal problematic. As can be seen, the data exhibits significant overdispersion (the variance is greater than the mean), which could be due alternatively to autocorrelation or to overdispersion in the marginal distribution. The presence of overdispersion is confirmed by looking at the histogram of the data in Figure 1, which shows that, whereas the probability mass is fairly concentrated around the mean, there exist large outliers. There is significant autocorrelation in each series, as can be seen from the Ljung-Box Q-statistic shown here at order 20. The order of 20 was chosen arbitrarily, the results are the same at any lags between 1 and 200 at least. Table 2 presents the contemporaneous correlation matrix between the four series we analyze. It is noteworthy to see that there is a negative correlation between buyer- and seller-initiated trades, which is something that most existing multivariate count models can not deal with. Figure 2 shows the auto- and cross-correlations of the vector of market-events, up to 380 5-minute intervals, which corresponds to 5 trading days. A very striking pattern of seasonality appears in the first series, to a lesser extent in the fourth one. Clearly looking only at contemporaneous correlation does not reveal the full picture, there is a very significant and systematic link across time between the various market events. The correlations move from positive to negative in a systematic way, which seems to be due to the presence of diurnal seasonality of the U-shape type, which is commonly found in time series based on high-frequency data.

6 A first model

In the present section we estimate a first model, designed to deal with the stylized facts documented in the previous section, specifically the overdispersion and the cross-correlation present in the vector of market events. For reasons of parsimony we estimate a mean structure based on a common factor, a series-specific lagged term in the moving average part and a diagonal autoregressive part. The most basic distributional assumption is the Poisson, and therefore our first model is the Multivariate Autoregressive Conditional Poisson model (MACP). Our goal in this section is to see how well this relatively simple, parsimonious

	Buys	sells	quotes	depth
14	10.10	<u> </u>	quotes	4.10
Mean	10.12	5.94	6.08	4.10
Std.Dev.	6.21	4.54	3.61	3.46
Dispersion	3.81	3.47	2.14	2.88
Maximum	43	37	22	26
Minimum	0	0	0	0
Q(20)	6,268.3	$2,\!416.8$	$4,\!400.7$	7,444.3

Table 1: Descriptive statistics

Descriptive statistics for the number of buyer-initiated and seller-initiated transactions, as well as the number of quote changes and depth changes. The number of observations is 4636. Q(20) is the Ljung-Box Q-statistic of order 20 on the series. The dispersion refers to the ratio of the variance to the mean.

Table 2: Correlation Matrix of the raw market events data

	Buys	sells	quotes	depth
Buys	1.00	-0.07	0.48	0.14
Sells	-0.07	1.00	0.33	0.11
Quotes	0.48	0.33	1.00	0.13
Depth	0.14	0.11	0.13	1.00

model, based on an equidispersed distribution can perform.

As can be seen from table 3, in the Multivariate Autoregressive Conditional Poisson Model (MACP), the factor corresponds basically to the buyer- and seller-initiated transactions (the weights δ on the quote and depth revisions are negative, but not significant). The factor impacts buys, sells and quote revisions with a price change positively and significantly. The own past effect is very significant for all variables, as are the estimates of the parameter on the lagged conditional mean. The process exhibits fairly strong autocorrelation, as can be seen from the AR coefficients. We checked for the stationarity of the model by looking at the eigenvalues of A + B, which are all within the unit circle, with a value of the largest one at 0.95. Next we look at the Pearson residuals from the model, which are defined as $\epsilon_t = \frac{N_t - \mu_t}{\sigma_t}$. If a model is well specified, the Pearson residuals will have variance one and neither significant auto- nor cross-correlation left. It is quite obvious that the variance of the Pearson residuals is larger than 1, more so for the trades than for the quotes, but very significantly so for all variables. This represents a failure of the Poisson distribution. As can be seen from figure 5, which shows the auto- and cross-correlation of the standardised residuals, there is no more correlation left. This means that our model is able to whiten the residuals and that it is therefore able to account for the dynamics of the data in a very satisfactory way. This is especially striking when we compare the result with figure 2. However, there is still seasonality present in the graphs. Another feature of the data we have not taken into account is contemporaneous correlation. The model we have estimated is based on uncorrelated Poisson distributions, conditionally on the past. The only channel for contemporaneous correlation is through the conditional mean. As can be seen form the graphs, there is significant contemporaneous correlation left, and this can only be captured by a model with correlated marginal distributions.

In order to go beyond the limitations of the MACP, we estimate a model with a Multivariate Poisson distribution. This distribution is characterised by a single parameter, which is meant to capture the correlation between all variables. The mean structure of this model is nearly the same as the one of the MACP, and the covariance coefficient is positive and highly significant (see Table 3 MVP panel). The likelihood improves slightly, but a likelihood-ratio test would clearly reject the uncorrelated model. While the MVP represents a small improvement over the MACP, there remain the problems that the covariance parameter is restricted to be positive, whereas we can see from figure 5 that for instance the correlation between the errors in buys and sells is around -0.30, which is actually much larger in absolute value than the correlation of -0.07 found in the raw data. The main limitation of course is that a single parameter cannot explain the complex structure with some positive and some negative correlations of varying magnitudes that prevail among the series.

The next model we estimate solves this problem by using copulas as a way of introducing dependence amongst the various series. We keep the Poisson assumption and use continuousation as in Denuit and Lambert (2002). As discussed earlier we estimate the model in two ways. Firstly we apply a two-stage procedure proposed in Patton (2002), which, given that we use a multivariate normal copula, involves only computing the sample counterpart of the variance-covariance of the normal quantiles of the continuoused probability integral transformation (PIT) of the data. Secondly, we estimate the MACP-copula model in one step. The results (which are not shown) of both procedures are extremely close, both in terms of likelihood and parameter values, which is consistent with the finding of Patton (2002) for the case of a bivariate model of exchange rates. This justifies using only two-step procedures, which is a great advantage, most especially in terms of computational effort for larger models. Table 4 shows the correlation matrix implied by the parameters of the copula. A comparison with the contemporaneous correlation in the standardised residuals from the MACP shows that we capture the patterns very well, both in terms of magnitude and sign. This represents a very significant edge over existing models. Most multivariate models (even in the static case) are not able to capture negative correlations among series, but as our application shows, this is an important feature of our data.

An assumption underlying the use of copulas is the uniformity of Z, the PIT of the observations under the marginal distributions. 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, we show quantiles of Z plotted against quantiles of the uniform distribution. The closer the plot is to a 45-degree line, the closer the distribution is to a uniform. The quantile plot of the MACP is shown in figure 3. 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 0.7 and 1. This holds for all series and is a sign of the failure of the Poisson distribution to accommodate the tails of the distribution. This is due to the fact that the Poisson is equidispersed and suggests using an overdispersed distribution, like the double Poisson. Figure 4 shows the autocorrelation of the Z variables, which are mostly under the significance level. The correlations for higher powers of Z are very similar and are not shown here. Another problem with this first model, that can be appreciated from the correlations of the residuals in figure 5, is the presence of seasonality, which is very systematic, even though it is below the significance level, making explicit modelling of this phenomenon necessary.

7 The Multivariate Double Poisson model

We now turn to estimation of a model which solves the problems mentioned above. In order to better fit the dispersion we use the double Poisson distribution and we model seasonality using a series of half hourly dummy variables. We estimate three different mean structures, which have very different implications in terms of finance. Firstly we estimate a model which has the same mean structure as the Poisson model described above, then we present a model with only one factor and finally we present an unrestricted model.

7.1 Factor structure and own effect

The first model has the same mean structure as the MACP model we presented in the previous section, and contains in addition 13 diurnal seasonality variables. The results are shown in table 5 (right panel). The parameters of the mean process are essentially the same as the ones in the MACP, which is not surprising, given that estimates of the MACP are quasi-maximum likelihood estimates (QMLE), which deliver consistent parameters, even in the case of a misspecified distribution. The eigenvalues of A + B are still high but smaller than 1, which means that the model is stationary. A likelihood ratio test shows that the seasonality variables (the estimates are not shown) are jointly significant, with a test statistic of 192.1, which implies a strong rejection of the model without seasonality. The coefficients on the seasonality exhibit the well-documented U-shape, which means that there is more activity at the beginning and end of the trading day and less at lunch time. The dispersion parameter ϕ of the double Poisson is also very significant and is very different from 1, which corresponds to the Poisson case. This means that the Poisson distribution is strongly rejected and that we now have a much better model for the conditional distribution. This can also be seen form the variance of the Pearson residuals, which are now much closer to 1. Furthermore, we observe that the likelihood improves with respect to the MACP model. A likelihood ratio test of the MDACP model with seasonality against the MACP leads to an overwhelming rejection of the Poisson model without seasonality, with a test statistic of 5486, which is much larger than the critical value of the Chi-square distribution with 17 degrees of freedom. Visual inspection of the Q-Q plots of the Z statistic of this model in figure 7 reveals that indeed the distribution is much better specified, since the Q-Q plots nearly coincide with the 45-degree line. This means that with the use of the double Poisson we satisfy the uniformity assumption, which is the theoretical basis for using copulas. This gives us increased confidence in the estimates of the copula parameters. The autocorrelations of the Z statistic, shown in figure 8 are essentially not significant, which indicates that the dynamics of the series is well accounted for. A look at the correlations of the Pearson residuals of the series, depicted in figure 6 shows that there is no more seasonal pattern left and the correlations are well below significance. Only the contemporaneous correlations are left in the plots. In order to model this we estimate, as before, a multivariate normal copula. As this model is somewhat more involved in terms of the number of parameters, we only use the two-step procedure of Patton (2002), given its very good performance relative to the one-step procedure in the simpler MACP case. Table 7 shows the correlation matrix implied by the parameters of the copula. These values are similar to the copula estimates form the MACP, which suggests, that improving the conditional distribution and introducing seasonality hardly impacts the correlation structure in the data, which we take as a sign of the robustness of the result. A comparison with the contemporaneous correlation in the standardised residuals from the MDACP, shows that we capture the patterns very well, both in terms of magnitude and sign.

7.2 Is a single factor enough to model trades and quotes jointly?

The second model is the most parsimonious. The moving average term is restricted to a factor structure. This means that the past of the counts only affects any given series through a fixed linear combination of the variables. This corresponds to the case where $A = \alpha \gamma \delta'$ and $B = \beta I$. The assumption of a common factor in the dynamics of all series corresponds to the idea that there are times when all components of the market process speed up or slow down proportionately, for instance when new information arrives and times between events shrink. According to Hasbrouck (1999), time deformation in the context of a vector of market events can be characterised by the existence of such a common factor. The idea of time deformation, which goes back to Clark (1973), is that one can differentiate between real time and operational time. In empirical implementations, it was often considered that the price process is a geometric Brownian motion in operational time, but in real time, we observe clustering of volatility, due to the irregular arrival of information, often proxied by volume. If this is the case, all elements of the market process should exhibit strong co-movements.

The results for this model are shown in table 6. All series enter the factor positively and very significantly, even though the coefficients are greater for buys and sells than for quote and depth. The impact of the common factor is also very significant for all four series and it is greatest for the buys and smallest for the depth changes. The common memory parameter β is high and very significant. Visual inspection of the Q-Q plots of the Z statistic of this model in figure 9 reveals that, whereas the results are much better than for the Poisson case, they are not quite as good as the double Poisson with a factor and an own effect. The Q-Q plots are very close to the 45-degree line, but there are small but systematic discrepancies. In figure 10, the autocorrelations of the Z statistic show a clear failure of the model. There is significant and persistent autocorrelation in all four series. This means that the factor structure is unable to capture the dynamic relation between the variables. This can also be seen from figure 11, where there are some systematic patterns left in the residuals, most strikingly so in the auto- and cross-correlations of depth changes, but also in the first-order autocorrelation of all series, which are significantly different from zero. This confirms the findings in Hasbrouck (1999) that a single factor is not enough to account for the dynamics of the vector of market events. Thus, time deformation in its simplest form does not seem to be supported by the data.

7.3 A full model

The third model we estimate is the least restrictive. The moving average matrix is left unrestricted and we estimate all 16 parameters of the A matrix. This allows us to get a better idea of the interactions between the various components of the market process. The estimates are shown in table 5 (left panel). As can be seen in the table, the transactions vary together in a block, as do the quote related variables. All own effects are significant and positive, which means that the series are persistent. It is interesting to note that depth changes with no change in the bid or the ask have a negative impact on subsequent buys and sells, whereas changes in the bid or ask have a significant and positive impact on subsequent trades. As suggested by Engle and Lunde (1999), changes in either the bid or the ask are caused by the specialist changing perception about the value of the stock and suspecting the presence of informed traders. Quote changes which only change the posted depth, but neither the bid nor the ask are potentially related to movements in the limit order book. We therefore take a great number of depth changes to be a sign of a slow market with no informed traders. The effect of trades on subsequent quote and depth changes is not significant. The likelihood increases only slightly with respect to the reference model, even though a likelihood ratio test would favour the more general model. The dispersion of the Pearson residuals is still very close to one, which means that the changes in the mean structure hardly affect the quality of the specification.

8 Conclusion

We have proposed new models for multivariate time series of count data. These models have proved very flexible and easy to estimate. We discuss how to adapt copulas to the case of time series of counts and show that the Multivariate Autoregressive Conditional Poisson model (MACP) can accommodate many features of multivariate count data, such as discreteness, overdispersion (variance greater than the mean) and both auto- and cross-correlation. Hypothesis testing in this context is straightforward, because all the usual likelihood-based tests can be applied. An important advantage of this model is that it can accommodate both positive and negative correlation among variables, which most multivariate count models cannot do, and this is shown to be important in our financial application. As a feasible alternative to multivariate duration models, the model is applied to the submission of market orders and quote revisions on IBM on the New York Stock Exchange. More precisely we distinguish the number of buying and selling market orders, as well as the number of quote revisions with and without price revisions. We show that a single factor cannot explain the dynamics of the market process, which confirms that time deformation, taken as meaning that all market events should accelerate or slow down proportionately, does not hold. We advocate the use of the Multivariate Autoregressive Conditional Poisson model for the study of multivariate point processes in finance, when the number of variables considered simultaneously exceeds two and looking at durations becomes too difficult. Plans for further research include evaluating the forecasting ability of these models, both in terms of point and density forecasts and applying these models to more detailed tick-by-tick data sets.

9 Appendix

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

Proof of Proposition 3.2. Upon substitution of the mean equation in the autoregressive intensity, one obtains:

$$\mu_t - \mu = A(N_{t-1} - \mu) + B(\mu_{t-1} - \mu)$$
(9.1)

$$\mu_t - \mu = A(N_{t-1} - \mu_{t-1}) + (A + B)(\mu_{t-1} - \mu)$$
(9.2)

Squaring and taking expectations gives:

$$V[\mu_t] = AE\left[(N_{t-1} - \mu_{t-1})(N_{t-1} - \mu_{t-1})'\right]A + (A+B)V[\mu_{t-1}](A+B)'$$
(9.3)

Using the law of iterated expectations and denoting $\Omega = V[N_t | \mathcal{F}_{t-1}]$, one gets:

$$V[\mu_t] = A\Omega A + (A+B)V[\mu_{t-1}](A+B)'$$
(9.4)

Vectorialising and collecting terms, one gets:

$$vec(V[\mu_t]) = \left(I_{K^2} - (A+B) \otimes (A+B)'\right)^{-1} \cdot (A \otimes A') \cdot vec(\Omega)$$
(9.5)

Now, applying the following property on conditional variance

$$V[y] = E_x \left[V_{y|x}(y|x) \right] + V_x \left[E_{y|x}(y|x) \right]$$
(9.6)

to the counts and vectorialising, one obtains:

$$vec(V[N_t]) = vec(\Omega) + vec(V[\mu_t])$$
(9.7)

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.

$$vec(V[N_t]) = \left(I_{K^2} + \left(\left(I_{K^2} - (A+B) \otimes (A+B)'\right)^{-1} \cdot (A \otimes A'\right)\right) \cdot vec(\Omega)$$
(9.8)

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:

$$E[(N_t - \mu_t)(\mu_{t-s} - \mu)'] = 0 \qquad \forall s \ge 0$$
(9.9)

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

$$Cov[N_t, \mu_{t-s}] = Cov[\mu_t, \mu_{t-s}] \qquad \forall \ s \ge 0$$
(9.10)

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

$$E[(N_t - \mu_t)(N_{t-s} - \mu)'] = 0 \qquad \forall \ s \ge 0$$
(9.11)

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

$$Cov[N_t, N_{t-s}] = Cov[\mu_t, N_{t-s}] \qquad \forall \ s \ge 0$$
(9.12)

Now,

$$Cov[\mu_t, \mu_{t-s}] = ACov[N_t, \mu_{t-s+1}] + BCov[\mu_t, \mu_{t-s}]$$

= (A + B)Cov[\mu_t, \mu_{t-s}]
= (A + B)^s V[\mu_t] (9.13)

The first line was obtained by replacing μ_t by its expression, the second line by making use of 9.10, the last line follows from iterating line two.

$$Cov[\mu_t, \mu_{t-s+1}] = ACov[\mu_t, N_{t-s}] + BCov[\mu_t, \mu_{t-s}]$$
(9.14)

Rearranging and making use of 9.12, one gets:

$$ACov[N_t, N_{t-s}] = Cov[\mu_t, \mu_{t-s+1}] - BCov[\mu_t, \mu_{t-s}]$$

= ((A + B)^s - B(A + B)) V[\mu_t] (9.15)

Under the condition that A is invertible, which is not an innocuous assumption, as it excludes the pure factor model, we get after vectorialising:

$$vec(Cov[N_t, N_{t-s}]) = \left[I \otimes \left(A^{-1}(A+B)^s - A^{-1}B(A+B)\right)\right] vec(V[\mu_t])$$
(9.16)

After substituting in 9.8, we get:

$$vec(Cov[N_t, N_{t-s}]) = \left[I \otimes A^{-1} \left((A+B)^s - B(A+B)\right)\right] \cdot \left(I_{K^2} + \left(\left(I_{K^2} - (A+B) \otimes (A+B)'\right)^{-1} \cdot (A \otimes A'\right)\right) \right) \cdot vec(\Omega)$$
(9.17)

_	_	

Table 3: Maximum Likelihood Estimates of the MACP models.

The table presents the Maximum Likelihood Estimates of the Multivariate Autoregresive Conditional Poisson model (MACP) and Multivariate Poisson Model (MVP) on counts based on IBM at intervals of 5 minutes for the period January 1998 to the end of March 1998. The t-statistics are presented in parenthesis. We impose the normalisation $\delta_1 = 0.25$ in order to identify the model. $\epsilon_t = \frac{N_t - \mu_t}{\sigma_t}$ are the Pearson residuals from the model. The mean equation is:

θ	MACP				MVP			
	Buys	sells	quotes	depth	Buys	sells	quotes	depth
ω_i	0.426	0.396	0.461	0.282	0.329	0.290	0.342	0.197
	(11.84)	(13.03)	(10.42)	(12.81)	(9.47)	(10.33)	(8.47)	(9.60)
α_{11}	0.128				0.128			
	(10.08)				(10.77)			
α_{22}		0.129				0.128		
		(15.57)				(16.12)		
α_{33}			0.159				0.158	
			(15.55)				(16.19)	
α_{44}			· · · ·	0.211			. ,	0.209
				(39.25)				(39.45)
γ	0.376	0.227	0.241	0.004	0.367	0.220	0.234	0.001
	(8.02)	(20.76)	(13.63)	(0.16)	(8.33)	(20.64)	(13.72)	(0.17)
δ	0.250	0.244	-0.001	-0.010	0.250	0.253	-0.005	-0.008
		(10.14)	(-0.06)	(-1.41)		(10.44)	(-0.32)	(-1.01)
β	0.684	0.654	0.609	0.721	0.687	0.661	0.616	0.725
,	(97.64)	(68.40)	(40.73)	(96.99)	(99.32)	(71.27)	(41.98)	(99.10)
ν					0.274			
					(15.98)			
LogL	-52,856				-52,729			
Eigenval	0.95	0.93	0.79	0.77				
$Var(\epsilon_t)$	2.68	2.81	1.66	1.98				

 $\mu_t = \omega + (diag(\alpha_i) + \gamma \delta')N_{t-1} + diag(\beta_i)\mu_{t-1}$

Table 4: Correlation Matrix of the Q estimated by the MACP model.

The table presents the correlation matrix of Q, the normal quantiles based on the probability integral transformation, Z, of the continuoused count data under the marginal densities using the MACP models in the two step procedure; i.e.:

 $z_{i,t} = F^*(N_{i,t}^*) = F(N_{i,t} - 1) + f(N_{i,t})U_{i,t}$

	COPULA - MACP4						
	Buys	sells	quotes	depth			
Buys	1.00	-0.32	0.33	0.12			
Sells	-0.32	1.00	0.22	0.16			
Quotes	0.33	0.22	1.00	0.09			
Depth	0.12	0.16	0.09	1.00			



Figure 1: Histograms of the counts

Table 5: Maximum Likelihood Estimates of the MDACP models.

The table presents the Maximum Likelihood Estimates of the Multivariate Autoregresive Conditional Double Poisson (MDACP) models on counts based on data of IBM at intervals of 5 minutes for the period January 1998 to the end of March 1998. These models consider the seasonality presented in the data and solved it by the use of 15 minutes dummies. The t-statistics are presented in parenthesis. We impose the normalisation $\delta_1 = 0.25$ in order to identify the model. $\epsilon_t = \frac{N_t - \mu_t}{\sigma_t}$ are the Pearson residuals from the model. The equations for the full model and the model with a factor and an own effect are respectively: $\mu_t = \omega + AN_{t-1} + diag(\beta_i)\mu_{t-1}$

θ	MDACP full model			MDACP with factor and own effect				
	Buys	Sells	Quotes	Depth	Buys	Sells	Quotes	Depth
ω_i	0.372	0.467	0.431	0.316	0.398	0.381	0.445	0.302
	(2.13)	(2.22)	(2.25)	(2.30)	(2.20)	(2.21)	(2.30)	(2.34)
α_{1i}	0.170	0.058	0.001	0.013	0.106			
	(2.42)	(2.24)	(0.09)	(1.19)	(2.25)			
α_{2i}	0.048	0.151	-0.016	-0.020		0.110		
	(2.30)	(2.39)	(-1.11)	(-1.68)		(2.31)		
α_{3i}	0.044	0.038	0.132	0.011			0.139	
	(2.29)	(2.17)	(2.38)	(1.33)			(2.35)	
α_{4i}	-0.003	-0.016	0.016	0.198			. ,	0.196
	(-0.84)	(-1.95)	(1.57)	(2.42)				(2.47)
γ		. ,		. ,	0.271	0.167	0.177	-0.014
					(1.92)	(2.31)	(2.33)	(-1.28)
δ					0.250	0.223	-0.038	0.016
						(4.76)	(-0.77)	(0.84)
β	0.734	0.694	0.653	0.716	0.738	0.702	0.657	0.726
	(42.59)	(25.82)	(25.57)	(46.16)	(43.57)	(27.16)	(25.88)	(48.72)
ϕ	0.379	0.358	0.583	0.493	0.378	0.357	0.583	0.492
	(46.58)	(45.33)	(45.71)	(45.82)	(46.56)	(45.24)	(45.74)	(46.07)
LogL	-48,100				-48,113			
Eigenval	0.93	0.92	0.81	0.79	0.93	0.91	0.81	0.79
$Var(\epsilon_t)$	0.99	0.98	0.97	0.95	0.99	0.98	0.98	0.95

 $\mu_t = \omega + (diag(\alpha_i) + \gamma \delta')N_{t-1} + diag(\beta_i)\mu_{t-1}$

Table 6: Maximum Likelihood Estimates of the MDACP models.

The table presents the Maximum Likelihood Estimates of the Multivariate Autoregressive Conditional Double Poisson (MDACP) models on counts based on data of IBM at intervals of 5 minutes for the period January 1998 to the end of March 1998. These models consider the seasonality presented in the data and solved it by the use of 15 minutes dummies. The t-statistics are presented in parenthesis. We impose the normalisation $\delta_1 = 0.25$ in order to identify the model. $\epsilon_t = \frac{N_t - \mu_t}{\sigma_t}$ are the Pearson residuals from the model. The equation for the model MDACP with factor only is: $\mu_t^0 = c + \alpha f_{t-1} + \beta \mu_{t-1}^0$

and $\mu_t = \gamma \mu_t^0$

θ	MDACP	with fact	tor only ar	nd seasonality
	Buys	Sells	Quotes	Depth
ω_i	0.259	0.439	0.306	0.633
	(1.94)	(2.29)	(2.25)	(2.37)
α_{1i}				
α_{2i}				
α_{3i}				
α_{4i}				
γ	0.474	0.221	0.255	0.091
	(2.39)	(2.38)	(2.39)	(2.35)
δ	0.250	0.191	0.155	0.100
		(13.84)	(7.10)	(7.62)
β	0.702	0.702	0.702	0.702
	(53.58)	(53.58)	(53.58)	(53.58)
ϕ	0.370	0.338	0.572	0.374
	(46.01)	(44.46)	(45.22)	(44.23)
LogL	-48,972			
$Var(\epsilon_t)$	1.04	1.00	0.99	0.95

Table 7: Correlation Matrix of the Q estimated by the MDACP model.

The table presents the correlation matrix of Q, based on the probability integral transformation, Z, of the continuoused count data under the marginal densities estimated using the MDACP models by the two-step procedure

	COPULA – MDACP4S					
	Buys	sells	quotes	depth		
Buys	1.00	-0.33	0.30	0.11		
Sells	-0.33	1.00	0.20	0.13		
Quotes	0.30	0.20	1.00	0.08		
Depth	0.11	0.13	0.08	1.00		



Figure 2: Auto- and cross-correlogram of the data



Figure 3: Histogram of the Z statistics of the MACP model



Figure 4: Autocorrelogram of the Z statistics of the MACP model



Figure 5: Auto- and cross-correlogram of the errors from the MACP model



Figure 6: Auto- and cross-correlogram of the errors from the MDACP model with seasonality



Figure 7: Histogram of the Z statistics or the MDACP model with seasonality



Figure 8: Autocorrelogram of the Z statistics of the MDACP model with seasonality



Figure 9: Histogram of the Z statistics of the MDACP model with factors only and seasonality



Figure 10: Autocorrelogram of the Z statistics or the MDACP model with factors only and seasonality



Figure 11: Auto- and cross-correlogram of the errors from the MDACP model with factors only and seasonality



Figure 12: Histogram of the Z statistics of the full MDACP model with seasonality



Figure 13: Autocorrelogram of the Z statistics of the full MDACP model with seasonality



Figure 14: Auto- and cross-correlogram of the errors from the full MDACP model with seasonality

References

- Aitchison, J., and C. H. Ho, 1989, The multivariate poisson-log normal distribution, *Econo*metrica 76, 643–651.
- Cameron, A. Colin, and Pravin K. Trivedi, 1998, 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).
- 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.
- Clark, Peter K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41.
- Denuit, M., and P. Lambert, 2002, Constraints on concordance measures in bivariate discrete data, Discussion Paper 02-12, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
- Efron, Bradley, 1986, Double exponential families and their use in generalized linear regression, *Journal of the American Statistical Association* 81, 709–721.
- Engle, Robert F., and Asger Lunde, 1999, Trades and quotes: A bivariate process, Working paper, University of California, San Diego.
- Engle, Robert F., and Jeffrey Russell, 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* 66, 1127,1162.
- Fisher, R. A., 1932, Statistical Methods for Research Workers.
- Harvey, A.C., and C. Fernandes, 1989, Time series models for count or qualitative observations, Journal of Business and Economic Statistics 7, 407–417.
- Hasbrouck, Joel, 1999, Trading fast and slow: Security markets events in real time, mimeo.
- Hausman, Jerry A., Bronwyn H. Hall, and Zvi Griliches, 1984, Econometric models for count data with an application to the patents-r and d relationship, *Econometrica* 52, 909–938.
- Joe, Harry, 1997, *Multivariate Models and Dependence Concepts* (Chapman and Hall: London).
- 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).
- Marshall, A. W., and I. Olkin . , 1990, (IMS Lecture Notes Monograph Series).

- McKenzie, Ed., 1985, Some simple models for discrete variate time series, *Water Resources Bulletin* 21, 645–650.
- Munkin, Murat K., and Pravin K. Trivedi, 1999, Simulated maximum likelihood estimation of multivariate mixed-poisson regression models, with applications, *Econometrics Journal* 2, 29–48.
- Nelsen, Roger B, 1999, An introduction to Copulas (Springer: New York).
- Newey, Whitney K, and Daniel McFadden, 1994, Large sample estimation and hypothesis testing, in Robert F. Engle, and Daniel McFadden, ed.: *Handbook of Econometrics*., vol. 4 (North-Holland: Amsterdam).
- Patton, Andrew, 2001, Estimation of copula models for time series of possibly different lengths, University of California, San Diego, Discussion Paper 01-17.
- ——, 2002, Skewness, asymmetric dependence, and portfolios, mimeo.
- Sklar, A, 1959, Fonctions de répartitions à *n* dimensions et leurs marges, *Public Institute* of Statistics of the University of Paris 8, 229–231.
- Winkelmann, Rainer, 2000, Econometric Analysis of Count Data (Springer: Berlin).
- Zeger, Scott L., 1988, A regression model for time series of counts, *Biometrika* 75, 621–629.