Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2008) 2931–2944
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Multivariate reduced rank regression in non-Gaussian contexts, using copulas

Andréas Heinen[a], Erick Rengifo[b,*]

[a] *Departamento de Estadística, Universidad Carlos III de Madrid, Spain*
[b] *Department of Economics, Fordham University, 441 East Fordham Road, Dealy Hall, Room E513, Bronx, NY 10458, USA*

Available online 11 September 2007

Abstract

A new procedure is proposed that performs reduced rank regression (RRR) in non-Gaussian contexts based on multivariate dispersion models. Reduced-rank multivariate dispersion models (RR-MDM) generalize RRR to a very large class of distributions, which include continuous distributions like the normal, Gamma, inverse Gaussian, and discrete distributions like the Poisson, the binomial and the negative binomial. A multivariate distribution is created with the help of the Gaussian copula and estimation is performed using maximum likelihood. It is shown how this method can be amended to deal with the case of discrete data. A Monte Carlo simulation shows that the new estimator is more efficient than the traditional Gaussian RRR. In the framework of MDM's a procedure analogous to canonical correlations is introduced, which takes into account the distribution of the data. Finally, the method is applied to the number of trades of five US department stores on the New York Stock Exchange during the year 1999 and determine the existence of a common factor which represents sector specific news. This analysis is helpful in microstructure analysis to identify leaders from the point of view of dissemination of sectorial information.

Keywords: Multivariate dispersion model; Multivariate statistical analysis; Canonical correlations; Principal component analysis

1. Introduction

Reduced rank regression (RRR) is an important tool in multivariate statistical analysis. It provides interpretable results based on a low dimensional view of the data, allowing for parsimonious models. However, in all the classical references on RRR, like in Anderson (1984b) and in many other techniques of multivariate analysis, there is an implicit assumption of normality. For a survey including amongst others, discriminant analysis, factor analysis, canonical correlations and principal components, see Härdle and Simar (2003). Recently, there has been work on relaxing normality for many of these techniques. For instance in linear discriminant analysis, Zhu and Hastie (2003) analyze the case, in which data are classified into categories based on general types of distributions by using non-parametric techniques. In reduced rank models, a prominent example of this is Yee and Hastie (2003), who extend reduced-rank ideas to vector generalized linear models (VGLM), and base their development on the example of the reduced-rank multinomial logit model of Anderson (1984a), also called the stereotype model. Yee and Hastie (2003) also show the relation of RR-VGLM with

* Corresponding author. Tel.: +1 718 817 4061; fax: +1 718 817 3518.
E-mail address: rengifomina@fordham.edu (E. Rengifo).

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.08.012

many other classes of models, that have been proposed. Amongst them is canonical correspondence analysis of Ter Braak (1986), whose aim is to model how a group of exogenous variables influences a table of counts. Other applications of the RRR outside the normal errors framework include a paper by Fiocco et al. (2005) who proposed the use of RRR models to describe competing risk models with fewer parameters.

Our contribution is twofold. First, we propose a new procedure of RRR for potentially non-Gaussian data based on the multivariate dispersion model (MDM) of Song (2000). A MDM is a multivariate distribution obtained from taking univariate dispersion models and joining them with a multivariate Gaussian copula. Dispersion models, introduced by Jorgensen (1987), are a very general class of distributions which include, amongst others, continuous distribution like the normal and the Gamma, and discrete distributions like the Poisson and the binomial. We establish that under small dispersion asymptotics of Jorgensen (1987), our procedure converges to the RR-VGLM of Yee and Hastie (2003). We show in a Monte Carlo simulation that estimates of our model are more efficient than the traditional estimates of RRR.

Second, we introduce a procedure analogous to canonical correlations (CC) and principal component analysis (PCA), but which takes into account departures from normality in the distribution of the data. We show how this can be viewed under a tail area approximation as a maximization that looks like RRR, but with the appropriate deviance residual instead of the Gaussian residual.

We present an empirical market microstructure application used to determine the sectorial influence that certain assets exert, in terms of how much information they convey. This application is related to the theoretical and empirical literature about asymmetric information (the presence of informed and uninformed traders). Informed agents trade on the basis of private information and non-informed traders are looking for sources of information that can be used to determine the possibility of information arriving to the market. One alternative for them is to monitor the trading activity and abnormal behavior of those stocks that convey the most sectorial information. In order to implement this idea we use the trading activity of five of the most representative US department stores traded on the New York Stock Exchange (NYSE) during the year 1999 and we run a reduced-rank negative binomial PCA to analyze the existence of a common factor underlying the behavior of the stocks in this particular sector. This allows us to determine which assets contain more sectorial information and can lead the behavior of the other ones. This application contributes to the existing literature in that it presents a method to obtain relevant information from an exploratory analysis about the interactions and interdependence among stocks in a given sector.

The paper is organized as follows. Section 2 briefly explains Gaussian RRR, develops non-Gaussian RRR, introduces a CC for non-Gaussian data and explains how the procedures can be adapted for discrete variables. Section 3 presents some simulation results. Section 4 presents the empirical application and Section 5 concludes and presents some topics for future research.

2. Non-Gaussian RRR

In the first subsection we briefly present the Gaussian RRR, and we explain how we extend it to non-Gaussian contexts using multivariate dispersion models of Song (2000). In the next subsection we introduce the MDM-CC, a procedure analogous to CC, but which takes into account departures from normality in the distribution of the data. Finally we show how our procedures can be amended to deal with discrete variables.

2.1. The Gaussian case

We are interested in the effect on an $(N \times K_1)$ matrix Y of an $(N \times K_2)$ matrix X of explanatory variables. Let $Y_i = (Y_{i,1}, Y_{i,2}, \ldots, Y_{i,K_1})'$ denotes the vector of the ith observation of the K_1 variables in Y, $\mu_i = (\mu_{i,1}, \mu_{i,2}, \ldots, \mu_{i,K_1})'$ denotes the corresponding mean vector and similarly $X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,K_2})'$. We assume that $Y_i \sim \mathcal{N}(\mu_i, \Omega)$ and that μ_i is linear in X. If the model is of full rank, we have:

$$\mu_i = \mu_i^{(\mathrm{f})} \equiv \omega + C^{(\mathrm{f})} X_i, \tag{2.1}$$

where $C^{(\mathrm{f})}$ has full rank, therefore $rank(C^{(\mathrm{f})}) = t = \min(K_1, K_2)$. In some cases the full model has too many parameters to estimate and for reasons of parsimony it is preferable to estimate a lower rank model. Alternatively, one might have some theoretical reasons for imposing a factor structure, as is often the case in financial models. In those cases the new assumption is that $rank(C^{(r)}) = r < t$. We can then write $C^{(r)}$ as the product of two matrices $A^{(r)}$ and $B^{(r)}$, of rank r

and the RRR model is

$$\mu_i = \mu_i^{(r)} \equiv \omega^{(r)} + A^{(r)} B^{(r)} X_i, \tag{2.2}$$

where $A^{(r)}$ is a $(K_1 \times r)$ matrix of regression coefficients and $B^{(r)}$ a $(r \times K_2)$ matrix of factor loadings.

Estimation of the parameters is done using least-squares or maximum likelihood (MLE), resulting in the following maximization:

$$\hat{\theta}^{(r)} = \underset{\theta^{(r)}, \Omega^{(r)}}{\arg\max}\, \mathscr{L}^G(Y, \mu^{(r)}, \Omega^{(r)}), \tag{2.3}$$

where μ^r is conformable with Y, $\theta^{(r)} = (\omega, vec(A^{(r)}), vec(B^{(r)}))$ and $L^G(Y, \mu^{(r)}, \Omega^{(r)})$ denotes the Gaussian log-likelihood function of the observations Y. As such, this is an underidentified system. In order to identify the model we impose some normalization. There are several possibilities and we choose to impose $B^{(r)} = [I_r, \tilde{B}^{(r)}]$, where I_r is the identity matrix of order r.

2.2. The general case

In the non-Gaussian case a similar procedure can be defined by replacing the multivariate Gaussian distribution by an MDM of Song (2000), obtained by applying a multivariate Gaussian copula to univariate dispersion models. We assume that conditionally on the explanatory variables X, $Y_{i,j}$ (the ith observation of the jth component) is distributed according to a univariate dispersion model distribution DM_j with mean $\mu_{i,j}$ and dispersion parameter γ_j:

$$Y_{i,j}|X_i \sim DM_j(\mu_{i,j}, \gamma_j). \tag{2.4}$$

These models are characterized by their density, which can be written as:

$$f(y; \mu, \gamma) = a(y; \gamma) \cdot \exp\left(-\frac{1}{2\gamma} d(y; \mu)\right), \tag{2.5}$$

where $a(y; \gamma)$ is positive, $d(y; \mu)$ is the unit deviance and γ is the dispersion parameter. For more details we refer to Jorgensen (1997). This is a very general class of distribution, which contains as special cases continuous distributions like the normal, the gamma, the inverse Gaussian, and discrete distributions like the Poisson, binomial, negative binomial and compound Poisson.

In order to get the multivariate version of the univariate *DM*, we briefly introduce the main ideas underlying the copulas. Copulas provide 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. A more detailed account of copulas can be found in Joe (1997) and in Nelsen (1999). Let $F(y_1, \ldots, y_K)$ 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:

$$F(y_1, \ldots, y_K) = C(F_1(y_1), \ldots, F_K(y_K)). \tag{2.6}$$

This function is unique only if the marginal distributions are continuous. The discrete case is explicitly studied in Section 2.4. The joint density function is given by the product of the marginals and the copula density:

$$\frac{\partial F(y_1, \ldots, y_K)}{\partial y_1 \cdots \partial y_K} = \prod_{i=1}^{K} f_i(y_i) \frac{\partial C(F_1(y_1), \ldots, F_K(y_K))}{\partial F_1(y_1) \cdots \partial F_K(y_K)}. \tag{2.7}$$

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

$$C(z_1, \ldots, z_K) = F(F_1^{-1}(z_1), \ldots, F_K^{-1}(z_K)), \tag{2.8}$$

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

Having defined dispersion models and copulas, we can write the density of a vector of variables Y which follows an MDM as the product of its marginals DM_j times the Gaussian copula, i.e. we write the density of a matrix Y which follows an MDM as the product of its marginals DM_j times the Gaussian copula:

$$f(Y_i;\mu_i,\gamma,\Omega)=c(q_i,\Omega)\prod_{j=1}^{K_1}f_j(Y_{i,j};\mu_{i,j},\gamma_j), \tag{2.9}$$

where $c(q,\Omega)=|\Omega|^{-1/2}\exp(\frac{1}{2}(q'(I_{K_1}-\Omega^{-1})q)$ is the multivariate Gaussian copula, where I_{K_1} is the identity matrix of order K_1, $\mu_{i,j}$ is the mean of Y_j at observation i, which is the jth component of μ_i defined in Eq. (2.2), $\gamma=(\gamma_1,\ldots,\gamma_{K_1})$ is the vector of dispersion parameters, $f_j(Y_{i,j};\mu_{i,j},\gamma_j)$ is the p.d.f. corresponding to the marginal model DM_j, and Ω is a correlation matrix. By definition $c(q,I)=I_{K_1}$. The vector

$$q_i=(\Phi^{-1}(z_{i,1}),\ldots,\Phi^{-1}(z_{i,K_1}))', \tag{2.10}$$

where $\Phi(.)$ is the distribution function of a standard normal random variable, collects the quantiles of the $z_{i,j}$, which are the probability integral transforms (PITs) of the data under the marginal densities:

$$z_{i,j}=F_j(Y_{i,j};\mu_{i,j},\gamma_j), \tag{2.11}$$

where $F_j(Y_{i,j};\mu_{i,j},\gamma_j)$ is the c.d.f. corresponding to the marginal model DM_j. One caveat applies at this point: the method relies on the fact that the marginal distributions are correctly specified. In empirical work this assumption should be tested. We suggest to apply the tests proposed by Diebold et al. (1998) in the context of the evaluation of density forecasts. The basic idea underlying these tests is to make sure that the $z_{i,j}$'s are uniform [0, 1] and i.i.d., ignoring estimation error. The uniformity assumption can also be tested with a Kolgomorov–Smirnov test but again being aware that this test will not be asymptotically distribution free when the parameters are estimated. There is however recent literature extending these tests that takes into account estimation error like Bai (2003) who uses the martingalization argument. For a recent survey see Corradi and Swanson (2006).

In the case of the MDM, we replace the linear mean equation (2.2) by

$$\mu_i^{(r)}=g^{-1}(\omega^{(r)}+A^{(r)}B^{(r)}X_i), \tag{2.12}$$

where $g(\mu_i)=(g_1(\mu_{i,1}),\ldots,g_{K_1}(\mu_{i,K_1}))$ and $g^{-1}(.)$ is the inverse of the link function, as defined in the literature of generalized linear models (GLM). The *link function* $g=g(\mu)$ relates the linear predictor $x'\beta$ to the mean μ. For example, the Poisson model with mean $\mu=\exp(x'\beta)$ corresponds to the log link function $g=\ln\mu$. For variable j, the mean is:

$$\mu_{i,j}^{(r)}=g_j^{-1}(\omega_j^{(r)}+A_j^{(r)}B^{(r)}X_i), \tag{2.13}$$

where $A_j^{(r)}$ is the jth row of $(K_1\times r)$ matrix $A^{(r)}$.

The estimation is done by MLE using the MDM distribution instead of the multivariate normal to build the likelihood function. Denote by $\hat{\theta}^{(r)}$ the estimator of the RRR of rank r:

$$\hat{\theta}^{(r)}=\underset{\theta^{(r)},\Omega^{(r)}}{\arg\max}\,\mathscr{L}(\theta^{(r)},\Omega^{(r)}), \tag{2.14}$$

where

$$\mathscr{L}(\theta^{(r)},\Omega^{(r)})=\mathscr{L}(\theta^{(r)})+\log(c(q,\Omega^{(r)})) \tag{2.15}$$

and

$$\mathscr{L}(\theta^{(r)})=\sum_{i=1}^{N}\sum_{j=1}^{K_1}\log(f_j(Y_{i,j};g_j^{-1}(\omega_j^{(r)}+A_j^{(r)}B^{(r)}X_i),\gamma_j)) \tag{2.16}$$

is the part of the likelihood that comes from the marginal models. Maximizing $\mathscr{L}(\theta^{(r)})$ alone corresponds to the joint estimation of common parameters of the mean, under the assumption of uncorrelated marginal distributions.

The minimization procedure is implemented by numerical optimization of the likelihood using standard fminunc routine of Matlab which, given that we do not provide the gradient, reduces to using the multidimensional unconstrained nonlinear minimization (Nelder–Mead simplex (direct search) method). By definition, $\mathscr{L}(\theta^{(r)}, I) = \mathscr{L}(\theta^{(r)})$, since $c(q, I) = 1$.

It is clear that if the marginals are Gaussian, the procedure outlined above reduces to the classic RRR, considered by Izenman (1980), by virtue of the fact that using a multivariate Gaussian copula along with Gaussian marginals is equivalent to using a multivariate Gaussian distribution.

In order to determine the optimal rank for the RRR one can follow Fiocco et al. (2005) and decide based on the Akaike information criterion (AIC) and on the Bayesian information criterion (BIC). Moreover, and mainly for a rank two model, a biplot can be used (see Fiocco et al., 2005).

Song (2000) shows, in the context of panel data models for distributions in the exponential family, that the generalized estimating equation (GEE) approach of Zeger and Liang (1986) provides estimates that approximate the MDM estimates of the same model, under the conditions of small dispersion asymptotics of Jorgensen (1997). A similar relationship exists between RRR using MDM and the RR-VGLM procedure proposed by Yee and Hastie (2003). This means that the benefits of our procedure relative to RR-VGLM in terms of efficiency should be particularly important in data with large dispersion. This can be seen in the following development for multivariate exponential dispersion models (MED), which parallels Song (2000). The exponential dispersion model is a dispersion model with mean μ_i, dispersion γ_i and $d(y_i, \mu_i) = a(\mu_i) + b(y_i) + c(\mu_i)y_i$, for given functions a, b and c.

Assume $Y_i \sim MED_{K_1}(\mu_i, \gamma)$ and define $var(Y_{i,j}) = \gamma_j v(\mu_{i,j})$ and $\mu_i^{(r)} = g^{-1}(\omega^{(r)} + A^{(r)}B^{(r)}X_i)$ is the linear reduced rank predictor. Then, under small dispersion asymptotics and Theorem 2 of Song (2000), the log-likelihood (2.15) can be approximated by:

$$-\frac{N}{2}\log|\Omega^{(r)}| + \sum_{i=1}^{N}\sum_{j=1}^{K_1}\log(a(Y_{i,j};\gamma_j)) - \frac{1}{2\gamma_j}\sum_{i=1}^{N}(Y_i - \mu_i^{(r)})' V_i^{-1/2}\Omega^{(r)-1}V_i^{-1/2}(Y_i - \mu_i^{(r)}), \tag{2.17}$$

where $V_i = diag(v(\mu_{i,1}), \ldots, v(\mu_{i,K_1}))$, and the order of approximation is equal to $o(\gamma_{\max})$, where $\gamma_{\max} = \max(\gamma_1, \ldots, \gamma_{K1})$. The first order condition of this equation with respect to $\theta^{(r)}$, the vector containing the reduced-rank coefficients, is given as

$$\sum_{i=1}^{N}\left(\frac{\partial\mu_i'}{\partial\theta^{(r)}}\right)\Sigma^{(r)-1}(Y_i - \mu_i^{(r)}) = 0, \tag{2.18}$$

where $\Sigma^{(r)} = \gamma V_i^{1/2}\Omega^{(r)}V_i^{1/2}$. Finally, this can be written as

$$\sum_{i=1}^{N}\left(\frac{\partial\eta_i'}{\partial\mu_i}\right)W^{(r)}(Y_i - \mu_i^{(r)})X_j = 0, \tag{2.19}$$

where η is the linear predictor and $W^{(r)-1} = (\frac{\partial\eta_i'}{\partial\mu_i})'\Sigma^{(r)}(\frac{\partial\eta_i'}{\partial\mu_i})$ is the variance of the adjusted dependent variable, as defined in the GLM literature. Eq. (2.19) is the first order condition of Yee and Hastie (2003).

2.3. CC and PCA

CC between two sets of variables Y and X, is a very widely used technique of multivariate analysis. It can be thought of as finding successive pairs of linear combinations of X and Y, which are most correlated, in order to summarize the dependence between the two sets of variables. PCA is a similar technique with only one set of variables, where the aim is to find linear combinations that best represent the variation in the original data set.

Even though there is no reference to any specific distribution in these techniques, and they are typically used in practice for all sorts of data, there exist strong links between CC (or PCA) and RRR under the assumption of Gaussian errors. In particular Tso (1981) shows that MLE of a reduced-rank model under normal errors is equivalent to CC. Izenman (1980) shows, in the Gaussian case, how principal components and CC can be obtained as special cases

of RRR under different assumptions on the variance–covariance matrix. Izenman (1980) works in the context of the following estimation procedure:

$$\hat{\theta}^{(r)} = \underset{\theta^{(r)} \in \Theta^{(r)}}{\arg\min} \sum_{i=1}^{N} [(Y_i - \omega^{(r)} - A^{(r)} B^{(r)} X_i)' \Omega^{-1} (Y_i - \omega^{(r)} - A^{(r)} B^{(r)} X_i)] \tag{2.20}$$

and shows that when $\Omega = \Sigma_{YY} \equiv Var(Y)$, this is equivalent to CC, and when $X = Y$ and $\Omega = I_{K_1}$, the procedure is equivalent to PCA of Y. Moreover the procedure in (2.20) is equivalent to MLE under normality and with a given variance–covariance matrix Ω. We use this last equivalence as the basis for a new procedure which we call MDM-CC. It is analogous to CC but takes into account the distribution of the data. We have shown in Section 2 how we can take into account the distribution of the data in an RRR using the MDM. In that case, the traditional variance–covariance matrix is replaced as a measure of dependence by a copula variance–covariance matrix, which is the variance of the normal score q associated with the original data. We propose to estimate MDM distributions instead of the Gaussian with a given copula variance–covariance matrix equal to the unconditional copula variance–covariance of the data. This essentially mimics the way in which CC are obtained from Gaussian RRR, but in the case of the reduced-rank multivariate dispersion model (RR-MDM). By doing this, we are able to capture more general dependence structures than the variance–covariance and we are also able to take into account the right distribution of the data.

Denote by $\hat{\theta}^{(r)}_{\Omega}$ the estimator of the RRR when the copula variance–covariance of the data is assumed to be Ω:

$$\hat{\theta}^{(r)}_{\Omega} = \underset{\theta^{(r)} \in \Theta^{(r)}}{\arg\min}\, \mathscr{L}(\theta^{(r)}, \Omega). \tag{2.21}$$

By analogy to the Gaussian RRR, we want Ω to be Ω_{YY}, the unconditional copula covariance matrix of the dependent variable. We want to consider the unconditional variance–covariance of the dependent variables (i.e. without considering the impact of the explanatory variables X), thus we use the unconditional distribution of Y, which consists in taking a distribution with a constant mean for every variable in Y. We denote the c.d.f. of the unconditional distribution of Y_j by F^0_j. As we are using the multivariate Gaussian copula to model the dependence, we have to map the covariance matrix of Y into the corresponding copula covariance. In order to do this, we note that the input into the copula is the normal quantile $q^0_{i,j}$ of the PIT $z^0_{i,j} = F^0_j(Y_{i,j}, \mu^0_{i,j})$ of the raw data. Several possibilities arise at this stage. Firstly, we can have a distribution with no other parameter than the mean. Examples of this are the exponential in the continuous case or the Poisson in the discrete case. As mentioned before, if we believe that the data follow such a distribution, the assumption should be tested, for instance with the density forecast evaluation techniques of Diebold et al. (1998) or newer techniques like Bai (2003) (see also Corradi and Swanson, 2006), and if it is found to be satisfactory we can proceed. However, if the data are assumed to follow a distribution $F^0_j(., \gamma_j)$, which depends on some unknown parameter γ_j, then this parameter has to be estimated first. An estimate $\hat{\Omega}_{YY}$ of the unconditional Gaussian copula variance–covariance matrix of the dependent variables Y can thus be obtained as

$$\hat{\Omega}_{YY} = Var[q^0] = \frac{1}{N} \sum_{i=1}^{N} q^0_i q^{0'}_i. \tag{2.22}$$

If the marginals are correctly specified, the q^0_is will be standard normals and $\hat{\Omega}_{YY}$ will be a correlation matrix. In order to impose this restriction which might not hold exactly in finite samples, we can impose it and use instead $Corr[q^0]$. We denote the PIT of the data as

$$q^0_i = (q^0_{i,1}, \ldots, q^0_{i,K_1}), \tag{2.23}$$

$$q^0_{i,j} = \Phi^{-1}(z^0_{i,j}), \tag{2.24}$$

$$z^0_{i,j} = F^0_j(Y_{i,j}), \tag{2.25}$$

where F^0_j is the c.d.f. of the jth variable under a distribution characterized only by its mean and dispersion, which corresponds to the assumption that the variables Y depend on a constant only (without variables X):

$$Y_{i,j} \sim DM_j(\mu^0_{i,j}, \gamma_j), \tag{2.26}$$

where $\mu_{i,j}^0$ is the constant mean of the variable Y_j at observation i, and γ_j the corresponding dispersion parameter. Of course, as noted in Section 2, when we consider this procedure in the Gaussian case, we get back the original CC.

Song (2000) shows that under a tail area approximation, the MDM approximates the multivariate dispersion density of Jorgensen and Lauritzen (1998). This density is based on the deviance residual $r(Y, \mu) = d^{1/2}(Y, \mu)$, where $d(Y, \mu)$ is the deviance of GLM models (see McCullagh and Nelder, 1976). More specifically, this density looks very much like the multivariate normal in which the Pearson residual has been replaced by the deviance residual. The loglikelihood can be written as

$$-\frac{1}{2}\log|\Omega| + \sum_{i=1}^{N}\sum_{j=1}^{K_1}\log(a(Y_{i,j};\gamma_j)) - \frac{1}{2}\sum_{i=1}^{N} r(Y_i,\mu_i)'\Sigma^{-1}r(Y_i,\mu_i), \tag{2.27}$$

where $\Sigma = diag(\gamma_j)\,\Omega\,diag(\gamma_j)$. Under that approximation, the MDM-CC can be seen to be the analogue of a traditional CC, but with the appropriate deviance residual instead of the Gaussian one. For instance in the case of the gamma distribution, the deviance residual takes the form $r = Y - \mu/\mu$, as opposed to the Gaussian one, which is simply $r = (Y - \mu)$. One remark needs to be made about MDM-CC: unlike CC, MDM-CC is not symmetric in the variables X and Y, and one therefore needs to choose which set of variables is a priori thought of as determining the other. This is certainly a limitation of the procedure, but it is inevitable if the distribution is taken into account. Another comment needs to be made about PCA. The above development has been made in the case of CC, but all the results are valid for PCA. In that case, considering $Y = X$ and $\Omega = I$ will yield a MDM-PCA, which takes into account the distribution of Y_j.

2.4. The discrete case

So far we have implicitly considered the case of a continuous distribution. However, our method should be amended to deal with discrete distributions. Our presentation in this section closely follows Heinen and Rengifo (2007) who use continued extension for multivariate time series of counts.

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 distribution is distributed uniformly on the [0, 1] interval. The problem with discrete 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 (2005) use a continued extension argument to overcome these difficulties and apply copulas with discrete marginals. The main idea of continued extension 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 c.d.f., 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, continued extension 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 continued extension does not affect concordance, since $Y_1^* > Y_2^* \Leftrightarrow Y_1 > Y_2$.

Using the continued extension, Denuit and Lambert (2005) state a discrete analog of the PITT. If Y is a discrete random variable with domain χ, in $\mathbf{N}$, such that $f_y = P(Y = y)$, $y \in \chi$, continued by U, then

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

is uniformly distributed on [0, 1], and $[Y]$ denotes the integer part of Y. In the case of discrete data, we use Z^* as an argument in the copula, instead of Z, since, provided that the marginal model is well specified, this will ensure that the

conditions for use of a copula are met. In terms of the method we propose, this amounts to replacing the expression above for $z_{i,j}$ and $z^0_{i,j}$ by their continued extension versions:

$$z_{i,j} = F^*(Y^*_{i,j}) = F(Y_{i,j} - 1) + f(Y_{i,j}) * U_{i,j}$$

and

$$z^0_{i,j} = F^{*,0}(Y^*_{i,j}) = F^0(Y_{i,j} - 1) + f^0(Y_{i,j}) * U_{i,j},$$

where $Y^*_{i,j}$ are the continued extension version of the original data $Y_{i,j}$:

$$Y^*_{i,j} = Y_{i,j} + (U_{i,j} - 1),$$

F^*, F and f are, respectively, the continued extension version c.d.f., the c.d.f. and p.d.f. of Y, and $F^{*,0}$, F^0 and f^0 are the same for the unconditional distribution. Finally the $U_{i,j}$ are independent uniform random variables on [0, 1].

In order to avoid the noise introduced by the uniform random number, we average the estimates over a certain number of runs, as proposed in Machado and Santos Silva (2005), in the context of quantile regression for counts. Our estimator is therefore the average over M uniform draws of the continued extension estimates:

$$\hat{\theta}_{\text{cont}} = \frac{1}{M} \sum_{l=1}^{M} \hat{\theta}^{(l)}_{\text{cont}}, \tag{2.28}$$

where $\hat{\theta}^{(l)}_{\text{cont}}$ is the estimate of the parameters obtained with the lth random uniform draw $U^{(l)}$.

3. Simulation

In order to evaluate the performance of our estimator we conduct several simulation studies. In all cases we generate 500 replications of a 10-dimensional data set consisting of a (500×10) matrix of normally distributed explanatory variables X with correlation ranging from 0.12 to 0.59, and a 10-dimensional matrix Y, generated from the distribution $DM(\mu, \gamma)$, which is allowed to depend on a dispersion parameter γ, according to:

$$Y_{i,j}|X_i \sim DM(\mu_{i,j}, \gamma_j). \tag{3.1}$$

The dispersion models used are the normal, the gamma (with shape parameter equal to 2) and the Poisson distributions. The size of our simulations is the result of compromise between meaningfulness of the tests and computing time. We used the Matlab built-in random number generator commands normrnd, poissrnd and gamrnd for the normal, Poisson and gamma, respectively. The conditional mean is specified as:

$$\mu_i = g^{-1}(\omega^{(r)} + C^{(r)} X_i), \tag{3.2}$$

where $C^{(r)} = A^{(r)} B^{(r)}$ is given and assumed to be a rank-2 matrix, and $g^{-1}(.)$ is the inverse of the canonical link function, which is the exponential function in the case of the Poisson distribution and the identity function in the case of the Gaussian. For the gamma case, we also use the exponential function, which avoids the possibility of having a negative mean. Depending on the particular dispersion model, we use the appropriate deviance residual as explained in Section 2.3. Note that the density depends on a dispersion parameter in the case of the Gaussian and the gamma distributions, but not of the Poisson. In the various simulations, we compare the estimators to the traditional Gaussian-RRR which, according to Tso (1981), reduces to simply canonical correlations analysis as soon as the errors are normally distributed. We apply this procedure on the raw data Y (RRR) and on the data to which we apply the link function $g^{-1}(Y)$ (RRR-link). We assume that applying the correct link function to the data will improve the performance of RRR.

We present and evaluate the performance of the models based on three statistical tests: the bias, the efficiency and the mean squared error (MSE) of our parameter estimates of the given C matrix. In order to have a clear idea of the comparison of the models we present the results in terms of the Euclidean norm $\|C\| = trace(C'C)^{1/2}$. In terms of these statistics, the best model would be the one with smallest bias, variance (measured by the efficiency) and mean

Table 1
Bias, efficiency and mean squared error of MDM-RRGR

Model	*Bias*	*Eff*	*MSE*	RRR	RRR-link
RRR	0.17	0.88	0.88		
RRR-link	0.14	0.69	0.69	62	
MDM-RRGR	0.13	0.59	0.60	76	69

The model estimates are based on Gamma marginals and multivariate Gaussian copula (MDM-RRGR). The competing models are the Gaussian-RRR estimated on the raw data (RRR) and on the data to which we apply the link function $g^{-1}(Y)$ (RRR-link). The table presents the bias, efficiency and MSE of the parameter estimates according to Eqs. (3.3), (3.4) and (3.5). The last two columns contain the number of times (in percentage) that the model *MDM-RRGR* is better than the competing models.

squared errors. Accordingly, we define:

$$Bias = \left\| \frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} \hat{C}^i - C \right\|, \tag{3.3}$$

$$Efficiency = \|V\|, \tag{3.4}$$

$$MSE = \|W\|, \tag{3.5}$$

where $V = (v_{i,j}) = (\frac{1}{N_{\text{sim}}} \sum_{k=1}^{N_{\text{sim}}} (\hat{C}_{i,j}^{(k)} - \bar{C}_{i,j}^{(k)})^2)$, $W = (w_{i,j}) = (\frac{1}{N_{\text{sim}}} \sum_{k=1}^{N_{\text{sim}}} (\hat{C}_{i,j}^{(k)} - C_{i,j}^{(k)})^2)$, and $\bar{C}_{i,j}^{(k)}$ is the mean of the (i, j)th component of the estimated C matrix.

Finally and in order to be sure that the results are not driven by some outliers, we present the number of times (in percentage) that our estimates are better than the competing ones.

In terms of computational time, it takes more or less 3 days to do the reduced rank gamma regression (MDM-RRGR) and about 14 days to do the reduced rank Poisson regression (MDM-RRPR). This happens because in the last case we perform simulations in three different cases: the simple MDM-RRPR, the MDM-RRPR continued with a simple uniform draw and the MDM-RRPR continued averaging over 20 uniform draws of the continued estimates. The average time per simulation in the RRGR is 2 min and 52 s, and the average time per simulation in the reduced rank Poisson regression is 2 min 8 s. Using the normalization defined above, we have not had numerical problems in the optimization procedure. All the simulations where done using a Pentium IV 3.2 GHz.

3.1. Reduced rank gamma regression (MDM-RRGR)

In the first simulation, we generate data according to a gamma distribution, with shape parameter γ equal to (2, 2, 2). We estimate parameters of RRR, RR-link and RRGR. Table 1 presents (in terms of the Euclidean norm) the bias, efficiency and MSE of the parameter estimates of our *MDM-RRGR* compared with the *RRR* and the *RRR-link*. We can appreciate that our procedure is the most efficient and that also has the smallest bias. Moreover, we present in the last two columns the number of times (in percentage) that our procedure is better than the other ones. With this we can see that over 70% of the time, our estimates are closer to the given values of the C matrix.

3.2. Reduced rank poisson regression (MDM-RRPR)

As discussed in Section 2.4, MDM models include many discrete distributions. For this kind of distributions we propose to apply the continued extension in order to satisfy the conditions for use of the copula. First we estimate the simple MDM-RRPR model, which uses the Poisson distribution, but does not consider the continued extension (RRPR). Then, we estimate a RRPR model with a simple continued extension, i.e. we use a single uniform random variable [0, 1] ($M = 1$ in Eq. (2.28), RRPR-cont). Finally we average in each run over $M = 20$ draws of the uniform (RRPR-cont-20). There is a trade-off between more efficiency and simulation time, and we chose $M = 20$ based on a graph of the added benefit of an additional simulation. Arguably, efficiency gains would bit somewhat higher if we moved to 100 uniform draws. In a sense our results here can be seen as being lower bounds on the gains that could be obtained from a higher number of uniform draws. Table 2 presents the bias, efficiency and MSE of the parameter

Table 2
Bias, efficiency and mean squared error of MDM-RRPR

Model	*Bias*	*Eff*	*MSE*	RRR	RRR-link	RRPR	RRPR-cont
RRR	7.36	5.27	5.33				
RRR-link	0.92	0.87	0.88	100			
RRPR	3.73	4.53	4.65	100	9		
RRPR-cont	0.27	0.74	0.74	100	86	100	
RRPR-cont-20	0.26	0.68	0.68	100	94	100	85

The model estimate is based on Poisson marginals and multivariate Gaussian copula. This table presents the results from the procedure without considering the discreteness of the marginal distributions (RRPR) and the application of a simple continued extension argument procedure (RRPR-cont). Finally, it presents the results of a Monte Carlo simulation on the continued estimates (RRPR-cont-20). The last estimator is therefore the average over $M = 20$ uniform draws of the continued estimates according to Eq. (2.28). The competing models are the Gaussian-RRR estimated on the raw data (RRR) and on the data to which we apply the link function $g^{-1}(Y)$ (RRR-link). The table presents the bias, efficiency and MSE of the parameter estimates according to Eqs. (3.3), (3.4) and (3.5). The last two columns contain the number of times (in percentage) that the model RRPR-cont and RRPR-cont-20 are better than the competing models, including the *RRPR*.

estimates of the *RRPR*, the *RRPR-cont*, the *RRPR-cont*-20 and compares them with Gaussian RRR (RRR) and Gaussian RRR on transformed data (RRR-link). First of all the RRPR is better than RRR, but not than RRR-link. The second and very striking result is how much better the estimates are with only one uniform draw. This procedure is more efficient than all the other procedures by a very large margin. Finally averaging procedure over $M = 20$ draws instead of doing a simple continued extension has only a small impact on the quality of our estimates, as it improves results by 4–8%. In order to be sure that the result is not driven by some outliers, we check that a similar picture holds in terms of the number of times that the model outperforms the benchmarks. We verify that the RRPR-cont outperforms the RRR, RRR-link and the RRPR without continued extension, respectively, 100%, 86% and 100% of the time, and the RRPR-cont-20 outperforms RRR, RRR-link, RRPR without continued extension and RRPR-cont, respectively, 100%, 94%, 100% and 85% of the time, which establishes that averaging over continued extension estimates does help.

4. Application: sector- and stock-specific news

Much of the microstructure literature is based on the existence of asymmetric information and consequently of two types of traders: the uninformed who trade for liquidity reasons and informed traders who possess private information. This private information can be market base (macroeconomic), sector- or stock-specific information. Moreover, through the trading process this information is disseminated to the public and therefore trading conveys information. According to Admati and Pfleiderer (1988) and Easley and O'Hara (1992) frequent trading implies that news is arriving to the market. Thus a higher number of trades in a given time interval is a signal for the arrival of news.

Based on this, the trading activity of one stock does not only convey information about that specific stock, but can also contain information about the whole sector that this asset belongs to. Accordingly, in recent years the focus of empirical microstructure has shifted from the study of an individual asset to the analysis of the cross-sectional interactions amongst stocks. For example, Hasbrouck and Seppi (2001) document the existence of commonalities in order flow that are responsible for about two-thirds of the commonalities in returns, using principal components analysis and CC on the stocks of the Dow Jones industrial average. Heinen and Rengifo (2007), using the Multivariate autoregressive conditional double poisson model (MDACP) showed that the dynamic interactions among assets are important to determine which assets contain more sectorial information than can affect the trading behavior of the remaining assets within the sector.

We analyze the same data as Heinen and Rengifo (2007), but in this paper we are only interested in getting some insights from an exploratory analysis about the existence of a common factor underlying the behavior of the assets, i.e. a factor that captures the sector specific news. In this sense, this analysis is helpful for the purpose of identifying leaders from the point of view of dissemination of sectorial information.

4.1. Data description

We work with the five most important US department stores traded on the NYSE during the year 1999: May Department Stores (MAY), Federated Department Stores (FD), J.C. Penney Company, Inc. (JCP), Dillar's Inc. (DDS)

Table 3
Descriptive statistics

	DDS	FD	JCP	MAY	SKS
No. trades	55,399	100,928	108,392	90,881	59,725
Market Cap	1647	8945	3538	11,226	1612
Mean	2.93	5.34	5.73	4.81	3.16
Median	2.00	5.00	5.00	4.00	3.00
Std. Dev.	2.57	3.56	3.89	3.04	2.84
Dispersion	2.25	2.38	2.64	1.92	2.55
Skewness	1.74	1.17	1.39	0.91	2.27
Kurtosis	10.24	5.54	6.70	4.25	13.76
Maximum	37	35	38	22	32
Minimum	0	0	0	0	0

Descriptive statistics for the number of trades. The number of observations is 18,900. The market capitalization is in millions of US$. The dispersion refers to the ratio of the variance to the mean.

and Saks Inc. (SKS). We work with the number of trades in 5-min intervals. The data we use was taken from the Trades and Quotes (TAQ) data set, produced by the 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 canceled, 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 9:45 AM or after 4 PM (the official close of trading) were removed. The reason for starting at 9:45 instead of 9:30 AM, 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.

The data used were from January 2nd 1999 to December 30th 1999. This means that the sample covers 252 trading days, that represent 18,900 observations, as there are 75 5-min intervals every day between 9:45 AM and 4 PM. The descriptive statistics are given in Table 3. 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 exhibit 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 Fig. 1, which shows that, whereas the probability mass is fairly concentrated around the mean, there exist large outliers. Table 4 presents the contemporaneous correlation matrix among the five series we analyze, obtained using the Gaussian copula on the data with negative binomial (NB2) marginals. One property of the negative binomial distribution is that it captures the overdispersion present in certain discrete data.

4.2. Estimation results

We estimate an MDM-PCA of rank 1 on the number of transactions of the five stocks described above. We use the negative binomial as marginal distribution. It is the member of the family of dispersion models that could best fit the number of transactions per 5-min intervals, which are overdispersed and have small means making the use of the normal distribution inappropriate.

In particular, in this market microstructure application, we are interested in the factor loadings that represent the weights of a linear combination that explains the comovements of the whole system. We consider these loadings as proxies of sector-specific news since they represent the marginal contribution of each particular stock to the overall variability of our five-variate system. Table 5 presents these factor loadings and compares them to the factor loadings found using the principal components analysis. According to our model, the ranking of sectorial influence is MAY, DDS, FD, JCP and SKS. The ranking according to the PCA is: JCP, FD, MAY, DDS and SKS. The first of these rankings is closely related to the one found by Heinen and Rengifo (2007). We see that this ranking is highly correlated with the market capitalization of the stocks (MAY, FD, JCP, DDS and SKS) with a Spearman rank correlation of 0.70. Looking at the PCA results we observe a completely different picture. In this case the ranking is exactly the same as

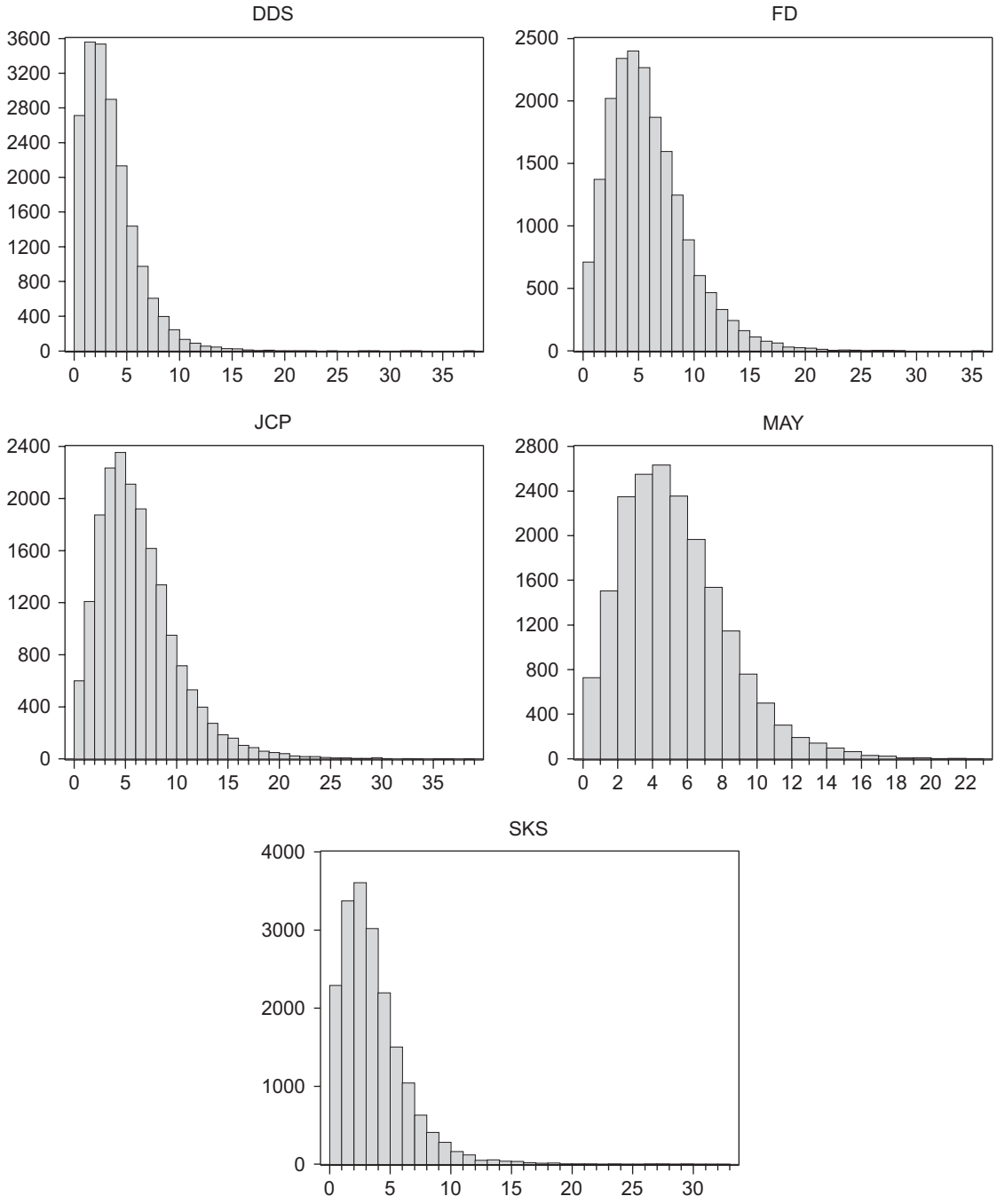

Fig. 1. Histogram of the data.

the ranking implied by the number of trades of the stocks in our sample (see Table 3). Clearly, taking into account the correct distribution of the data and implicitly, the correct link function, matters in the estimation: the Spearman rank correlation between both sets of results is 0.10.

Considering the specific sector that we study in this paper, and looking at our results and similar results already found in the literature, we can argue that the sectorial information content of individual stocks is closely related to their market capitalization. This result implies that in general traders tend to closely follow the activity of the stocks with the

Table 4
Correlation matrix of the trades data

	DDS	FD	JCP	MAY	SKS
DDS	1.00				
FD	0.27	1.00			
JCP	0.24	0.29	1.00		
MAY	0.25	0.30	0.31	1.00	
SKS	0.12	0.10	0.15	0.12	1.00

Table 5
Common factor results

	DDS	FD	JCP	MAY	SKS
Factor loadings MDM-PCA	1.000	0.738	0.570	1.405	0.387
Factor loadings PCA	1.000	2.128	2.749	1.570	0.630

Factor loadings of MDM-PCA of rank 1 and PCA on the number of transactions of five US department stores traded on the New York Stock Exchange during the year 1999. Sample goes from January 2nd 1999 to December 30th 1999.

highest market capitalization in order to look for signals that can influence the current positions of their investments within a given sector. Of course, this analysis can be extended to include all the assets of a given sector in order to get a complete understanding of interactions.

5. Conclusion

In this paper we introduce a set of new techniques designed to apply the reduced-rank ideas to potentially non-Gaussian data. We use to that effect the multivariate dispersion models (MDM), which provide a convenient statistical framework. We show that reduced-rank multivariate dispersion models (RR-MDM) include Gaussian reduced rank regression (RRR) as a special case, and that under small dispersion asymptotics they are equivalent to RR-VGLM of Yee and Hastie (2003). We introduce multivariate dispersion models canonical correlations (MDM-CC), a procedure similar to CC, but which takes into account the distribution of the data. Finally, we describe how our methods can be amended in the case of discrete data. We show in a Monte Carlo study that our RR-MDM yields significant gains in efficiency compared to RRR. Finally, we present an empirical application on the number of trades of five US department stores traded on the New York Stock Exchange during the year 1999. The results show that there exist a common factor among them that can be interpreted as their marginal contribution to the dissemination of sectorial information. We present a ranking of the sector leaders based on this sampled stocks and we found that this leadership is related to the market capitalization of the stocks.

Acknowledgments

The authors would like to thank Luc Bauwens, Bruce Lehmann, Pierre Giot, Daniel Peña, Eric Renault, Olivier Scaillet, Léopold Simar and two anonymous referees for helpful discussions and suggestions. This research was carried out while both authors were at the Center for Operations Research and Econometrics (CORE), at the Université Catholique de Louvain, and we wish to extend particular thanks to Luc Bauwens and members of CORE for providing a great research environment. The usual disclaimers apply.

References

Admati, A., Pfleiderer, P., 1988. A theory of intraday patterns: volume and price variability. Rev. Financial Stud. 1, 3–40.
Anderson, J.A., 1984a. Regression and ordered categorical variables. J. Roy. Statist. Soc. Ser. B 46 (1), 1–30.
Anderson, T.W., 1984b. An Introduction to Multivariate Statistical Analysis. Wiley, New York.
Bai, J., 2003. Testing parametric conditional distributions of dynamic models. Rev. Econom. Statist. 85 (3), 531–549.

Corradi, V., Swanson, N., 2006. Predictive density evaluation. In: Granger, C., Timmermann, A., Elliot, G. (Eds.), Handbook of Economic Forecasting, vol. 1. North-Holland, Amsterdam.
Denuit, M., Lambert, P., 2005. Constraints on concordance measures in bivariate discrete data. J. Multivariate Anal. 93, 40–57.
Diebold, F.X., Gunther, T.A., Tay, A.S., 1998. Evaluating density forecasts, with applications to financial risk management. Internat. Econom. Rev. 39 (4), 863–883.
Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. J. Finance 47, 577–605.
Fiocco, M., Putter, H., Van Houwelingen, J., 2005. Reduced rank proportional hazards model for competing risks. Biostatistics 6 (3), 465–478.
Fisher, R.A., 1932. Statistical Methods for Research Workers, Oliver and Boyd, Edinburgh, London.
Härdle, W., Simar, L., 2003. Applied Multivariate Statistical Analysis. Springer, Berlin.
Hasbrouck, J., Seppi, D.J., 2001. Common factors in prices, order flows and liquidity. J. Financial Econom. 59 (3), 383–411.
Heinen, A., Rengifo, E., 2007. Multivariate autoregressive modeling of time series count data using copulas. J. Emprical Finance 14 (4), 564–583.
Izenman, A.J., 1980. Assessing dimensionality in multivariate regression. In: Krishnaiah, P. (Ed.), Handbook of Statistics, vol. 1. North-Holland Publishing Company, Amsterdam, pp. 571–591.
Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall, London.
Jorgensen, B., 1987. Exponential dispersion model. J. Roy. Statist. Soc. 49 (2), 127–162.
Jorgensen, B., 1997. The Theory of Dispersion Models. Chapman & Hall, London.
Jorgensen, B., Lauritzen, S., 1998. Multivariate dispersion models. Research Report 2, Department of Statistics and Demography, Odense University, Denmark.
Machado, J., Santos Silva, J., 2005. Quantiles for counts. J. Amer. Statist. Assoc. 100 (472), 1226–1237.
McCullagh, P., Nelder, J., 1976. Generalized Linear Models. Chapman & Hall, London.
Nelsen, R.B., 1999. An Introduction to Copulas. Springer, New York.
Sklar, A., 1959. Fonctions de répartitions à *n* dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris 8, 229–231.
Song, P.X.-K., 2000. Multivariate dispersion models generated from Gaussian copulas. Scand. J. Statist. 27, 305–320.
Ter Braak, C.J.F., 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67 (5), 1167–1179.
Tso, M.K.-S., 1981. Reduced-rank regression and canonical analysis. J. Roy. Statist. Soc. Ser. B 43 (2), 183–189.
Yee, T., Hastie, T.J., 2003. Reduced-rank vector generalized linear models. Statist. Model. 3, 15–41.
Zeger, S.L., Liang, K.-Y., 1986. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42, 121–130.
Zhu, M., Hastie, T., 2003. Feature extraction for non-parametric discriminant analysis. J. Comput. Graphical Statist. 12 (1), 101–120.