Multivariate Reduced Rank Regression in non-Gaussian Contexts, Using Copulas *

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,rengifo}@core.ucl.ac.be

May 2004
JEL Classification codes: C35, C39.

Keywords: Multivariate Dispersion Model, multivariate statistical analysis, canonical correlations, principal component analysis.

Abstract

We propose a new procedure to perform Reduced Rank Regression (RRR) in non-Gaussian contexts, based on Multivariate Dispersion Models. Reduced-Rank Multivariate Dispersion Models (RR-MDM) generalise RRR to a very large class of distributions, which include continuous distributions like the normal, Gamma, Inverse Gaussian, and discrete distributions like the Poisson and the binomial. A multivariate distribution is created with the help of the Gaussian copula and estimation is performed using maximum likelihood. We show how this method can be amended to deal with the case of discrete data. We perform Monte Carlo simulations and show that our estimator is more efficient than the traditional Gaussian RRR. In the framework of MDM's we introduce a procedure analogous to canonical correlations, which takes into account the distribution of the data.

*The authors would like to thank Luc Bauwens and Léopold Simar for helpful discussions and suggestions. The usual disclaimers apply.

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), there is an implicit assumption of normality. Moreover, many other techniques of multivariate analysis are also implicitly based on the normality assumption. For a survey including amongst others, discriminant analysis, factor analysis, canonical correlations and principal components, see Hardle 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 (2002) analyse the case, in which data is 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 (2002), 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 (2002) also show the relation of RR-VGLM with many other classes of models, that have been proposed. Amongst them is canonical correspondence analysis of Ter Braak (1986), whose aim it is to model how a group of exogenous variables influences a table of counts.

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). An 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 (2002). 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 maximisation that looks like RRR, but with the appropriate deviance residual instead of the Gaussian residual.

The paper is organised as follows. Section 2 briefly explains Gaussian RRR, develops non-Gaussian RRR, introduces a canonical correlations for non-Gaussian data and explains how the procedures can be adapted for discrete variables. Section 3 presents some simulation results and section 4 concludes.

2 Non-Gaussian Reduced Rank Regression

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, K_1) matrix Y of an (N, K_2) matrix X of explanatory variables. Let $Y_j = (Y_{1,j}, Y_{2,j}, \ldots, Y_{K_1,j})'$ denotes the vector of the j-th observation of the K_1 variables in Y, $\mu_j = (\mu_{1,j}, \mu_{2,j}, \ldots, \mu_{K_1,j})'$ denotes the corresponding mean vector and similarly $X_j = (X_{1,j}, X_{2,j}, \ldots, X_{K_2,j})'$. We assume that $Y_j \sim \mathcal{N}(\mu_j, \Omega)$ and that μ_j is linear in X. If the model is of full rank, we have:

$$\mu_j = \mu_j^{(f)} \equiv \omega + C^{(f)} X_j \ ,$$

where $C^{(f)}$ is full rank, therefore $rank(C^{(f)}) = t = min(K_1, K_2)$. 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 RR model is

$$\mu_j = \mu_j^{(r)} \equiv \omega + A^{(r)} B^{(r)} X_j \ , \tag{2.1}$$

where $A^{(r)}$ is a (K_1, r) matrix of regression coefficients and $B^{(r)}$ a (r, K_2) matrix of factor loadings.
Estimation of the parameters is done using least-squares or maximum likelihood (MLE), resulting in the following maximisation:

$$\hat{\theta}_{RRR}^{(r)} = \underset{\theta^{(r)} \in \Theta^{(r)}}{\arg\max}\ \mathcal{L}^G \equiv log(\phi(Y, \mu_j^{(r)}, \Omega)) \ , \tag{2.2}$$

where $\theta^{(r)} = (\omega, vec(A^{(r)}), vec(B^{(r)}))$ and $\Phi(Y, \mu, \Omega)$ is the multivariate Gaussian p.d.f. of the observations Y with mean μ and variance Ω. As such, this is an underidentified system. In order to identify the model we impose some normalisation. There are several possibilities and we choose to impose $B^{(r)} = [I_r, \tilde{B}^{(r)}]$.

2.2 The general case

In the non-Gaussian case a similar procedure can be defined by replacing the multivariate Gaussian distribution by a Multivariate Dispersion Model (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 $j-th$ observation of the $i-th$ component) is distributed according to a univariate dispersion model distribution DM_i with mean $\mu_{i,j}$ and dispersion parameter γ_i:

$$Y_{i,j}|X_j \sim DM_i(\mu_{i,j}, \gamma_i) \ .$$

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

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

where $a > 0$ and d is a unit deviance. For more details we refer to Jorgensen (1997). This is a very general class of distributions, 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 [1], called copula, mapping $[0,1]^K$ into $[0,1]$, such that:

$$F(y_1,\ldots,y_K) = C(F_1(y_1),\ldots,F_K(y_K)) . \tag{2.3}$$

The joint density function is given by the product of the marginals and the copula density:

$$\frac{\partial F(y_1,\ldots,y_K)}{\partial y_1\ldots\partial y_K} = \prod_{i=1}^{K} f_i(y_i)\frac{\partial C(F_1(y_1),\ldots,F_K(y_K))}{\partial F_1(y_1)\ldots\partial F_K(y_K)} . \tag{2.4}$$

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

$$C(z_1,\ldots,z_K) = F(F_1^{-1}(z_1),\ldots,F_K^{-1}(z_K)) , \tag{2.5}$$

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_i times the Gaussian copula:

$$f(Y_j;\mu_j,\gamma,\Omega) = c(q_j,\Omega)\prod_{i=1}^{K_1} f_i(Y_{i,j},\mu_i,\gamma_i) , \tag{2.6}$$

where μ_j is as defined earlier, $\gamma = (\gamma_1,\ldots,\gamma_{K_1})$, $f_i(.,\mu_i,\gamma_i)$ is the p.d.f. corresponding to the marginals DM_i, $c(q,\Omega) =\mid \Omega \mid^{-1/2} \exp\left(\frac{1}{2}(q'(I_{K1}-\Omega^{-1})q\right)$ is the multivariate Gaussian copula and, by definition $c(q,I) = I_{k_1}$. The vector

[1] This function is unique only if the marginal distributions are continuous. The discrete case is explicitly studied in subsection 2.4

$$q_j = (\Phi^{-1}(z_{1,j}), \ldots, \Phi^{-1}(z_{K_1,j}))' \ ,$$

collects the quantiles of the $z_{i,j}$, which are the probability integral transforms (PIT) of the data under the marginal densities:

$$z_{i,j} = F_i(Y_{i,j}; \mu_i, \gamma_i) \ ,$$

and $F_i(.; \mu_i, \gamma_i)$ is the c.d.f. corresponding to the marginal models DM_i. 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, Gunther, and Tay (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. The uniformity assumption can also be tested with a Kolgomorov-Smirnov test.

Following our development, in the case of the MDM, we replace the linear mean equation (2.1) by

$$\mu_j = g\left(\omega + A^{(r)} B^{(r)} X_j\right) \ , \tag{2.7}$$

where $g(\mu_j) = (g_1(\mu_{1,j}), \ldots, g_{K_1}(\mu_{K_1,j}))$ and $g_i(.)$ is the inverse of the link function, as defined in the generalised linear models (GLM) literature. For variable i, the mean is:

$$\mu_{i,j} = g_i\left(\omega_i + A_i^{(r)} B^{(r)} X_j\right) \ ,$$

where $A_i^{(r)}$ is the i-th row of (K_1, r) matrix $A^{(r)}$.

The estimation is done using the MDM distribution instead of the multivariate normal. Denote $\hat{\theta}$ the estimator of the RRR and Ω the copula variance-covariance of the data:

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max}\ \mathcal{L}(\theta, \Omega) \ , \tag{2.8}$$

with

$$\mathcal{L}(\theta, \Omega) = \mathcal{L}(\theta) + \log(c(q, \Omega)) \ , \tag{2.9}$$

where

$$\mathcal{L}(\theta) = \sum_{j=1}^{N} \sum_{i=1}^{K_1} \log\left(f_i(Y_{i,j}, g_i(\omega_i + A_i^{(r)} B^{(r)} X_j))\right)$$

is the basis likelihood, which corresponds to the joint estimation of common parameters of the mean, under the assumption of uncorrelated marginal distributions. By definition, $\mathcal{L}(\theta, I) = \mathcal{L}(\theta)$, since $c(q, I) = I_{k_1}$.

It is clear that if the marginals are Gaussian, the whole 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. Therefore Gaussian RRR is obtained as a special case of our procedure.

Song (2000) shows in the context of panel data models for distributions in the exponential family, that the Generalised 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's 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) [2], which parallels Song (2000).

Assume $Y \sim MED(\mu, \gamma)$ and define $var(Y_j) = \gamma_j \nu(\mu_j)$ and $\mu = g(\eta)$, where $\eta = \omega + A^{(r)} B^{(r)} X_j$ is the linear reduced rank predictor. Then, under small dispersion asymptotics and Theorem 2 of Song (2000), the loglikelihood (2.9) can be approximated by:

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

where $V_i = diag(\nu(\mu_{i,1}), \ldots, \nu(\mu_{i,K1}))$. The first order condition of this with respect to θ, the vector containing the reduced-rank coefficients, is given as:

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

where $\Sigma = \gamma\ V_i^{\frac{1}{2}}\ \Omega\ V_i^{\frac{1}{2}}$. Finally, this can be written as:

$$\sum_{i=1}^{N}\left(\frac{\partial \eta_i'}{\partial \mu_i}\right) W(Y_i-\mu_i)X_j = 0\ , \tag{2.10}$$

where $W^{-1} = (\frac{\partial \eta_i'}{\partial \mu_i})'\Sigma(\frac{\partial \eta_i'}{\partial \mu_i})$ is the variance of the adjusted dependent variable, as defined in the GLM literature. Equation (2.10) is the first order condition of Yee and Hastie (2003).

2.3 Canonical Correlations and Principal Component Analysis

Canonical Correlations (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 summarise the dependence between the two sets of variables. Principal Component Analysis (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 and PCA and RRR under the assumption of Gaussian errors. In particular Tso (1981)

[2] The dispersion model is an exponential dispersion model with mean μ_i and dispersion γ_i when $d(y_i, \mu_i, \gamma_i) = a(\mu_i) + b(y_i) + c(\mu_i)y_i$, for given functions a, b and c.

shows that maximum likelihood estimation of a reduced-rank model under normal errors is equivalent to CC. Izenman (1980) works in the context of the following estimation procedure:

$$\hat{\theta}_{RRR} = \underset{\theta \in \Theta}{\arg\min} \sum_{j=1}^{T} [(Y_j - \omega - ABX_j)' \Omega^{-1} (Y_j - \omega - ABX_j)] \,, \tag{2.11}$$

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 (2.11) is equivalent to maximum likelihood estimation 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 a 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 RR-MDM.

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

$$\hat{\theta}_\Omega = \underset{\theta \in \Theta}{\arg\min} \; \mathcal{L}(\theta, \Omega) \,. \tag{2.12}$$

By analogy to the Gaussian RRR, we want Ω to be Ω_{YY}, the unconditional copula covariance matrix of the dependent variable. 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 probability integral transform (PIT) $z^0_{i,j} = F^0_i(Y_{i,j}, \mu^0_j)$ of the raw data. As we want to consider the unconditional variance covariance of the dependent variables (i.e. without considering the impact of the explanatory variables X), 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_i by F^0_i. 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 follows this distribution, the assumption should be tested, for instance with the density forecast evaluation techniques of Diebold, Gunther, and Tay (1998) and if it is found to be satisfactory we can proceed. However, if the data is assumed to follow a distribution $F^0_i(., \gamma_i)$, which depends on some unknown parameter γ_i, 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:

$$\hat{\Omega}_{YY} = Var[q^0] = \frac{1}{N} \sum_{j=1}^{N} q^0_j q^{0\prime}_j \,,$$

where the probability integral transform of the data is:

$$q_j^0 = (q_{1,j}^0, \ldots, q_{K_1,j}^0)$$

$$q_{i,j}^0 = \Phi^{-1}(z_{i,j}^0) \ ,$$

$$z_{i,j}^0 = F_i^0(Y_{i,j})$$

where F_i^0 is the c.d.f. of the i-th variable under the distribution characterized only by the mean, which corresponds to the assumption that the variables Y depend on a constant only (without variables X):

$$Y_{i,j} \sim DM_i(\mu_i^0, \gamma_i) \ ,$$

where μ_i^0 is the constant mean of the variable Y_i. 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 $r(Y, \mu) = d^{\frac{1}{2}}(Y, \mu)$, where $d(Y, \mu)$ is the deviance of GLM models (see McCullagh and Nelder (1976)), which takes the place of the Pearson residual in a density which looks otherwise very much like the multivariate normal. The loglikelihood can be written as:

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

where $\Sigma = diag(\gamma_i)\ \Omega\ diag(\gamma_i)$. 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, the deviance residual takes the form $r = \frac{Y-\mu}{\mu}$, 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 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 (2003) who use continuousation 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 (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 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, 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 $\mathbf{N}$, such that $f_y = P(Y = y), y \in \chi$, continuoused by U, then

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

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 continuoused 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 continuoused 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 continuoused 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 (2003), in the context of quantile regression for counts. Our estimator is therefore the average over M uniform draws of the continuoused estimates:

$$\hat{\theta}_{cont} = \frac{1}{M} \sum_{l=1}^{M} \hat{\theta}_{cont}^{(l)} \ , \tag{2.13}$$

where $\hat{\theta}_{cont}^{(l)}$ is the estimate of the parameters obtained with the l-th 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 three-dimensional data set consisting of a $(500, 3)$ matrix of normally distributed explanatory variables X and a three-dimensional matrix Y, generated from the distribution $DM(\mu, \gamma)$, which is allowed to depend on a dispersion parameter γ, according to:

$$Y_i | X_i \sim DM(\mu_i, \gamma_i) \ ,$$

where the conditional mean is:

$$\mu_i = g(\omega + CX_i) \ ,$$

where $C = A^{(r)}B^{(r)}$ is given and where $g(.)$ 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 we also use the exponential function, which avoids the possibility of having a negative mean. 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 normal 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 three results: 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 $\|C\| = trace(C'C)^{1/2}$ and we define:

$$Bias = \left\| \frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} \hat{C}^i - C \right\|$$
$$Efficiency = \|V\|$$
$$MSE = \|W\| \ ,$$

where $V = \left(\frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} (\hat{C}_{i,j}^{(k)} - \bar{C}_{i,j}^{(k)})^2\right)$ and $W = \left(\frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} (\hat{C}_{i,j}^{(k)} - C_{i,j}^{(k)})^2\right)$.
Finally and in order to be sure that the result is not driven by some outliers, we present the number of times (in percentage) that the our models are better than the competing ones.

In terms of computational time, it takes more or less 6 hours to do the Reduced Rank Gamma Regression (MDM-RRGR) and about 14 hours to do the Reduced Rank Poisson Regression (MDM-RRPR). This happens because in the last case we perform three different cases: the simple MDM-RRPR, the MDM-RRPR continuoused with a simple uniform draw and the MDM-RRPR continuoused averaging over 20 uniform draws of the continuoused estimates. Using the normalization defined above, we have not had numerical problems in the optimization procedure[3].

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 Reduced Rank Gamma Regression (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.

Table 1: **Bias, Efficiency and Mean Squared Error of MDM-RRGR.**

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 $Bias = \left\| \frac{1}{N_{sim}} \sum_{N_{sim}}^{i=1} \hat{C}^i - C \right\|$, $Efficiency = \|V\|$ and $MSE = \|W\|$ of the parameter estimates and where $V = \left(\frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} (\hat{C}_{i,j}^{(k)} - \bar{C}_{i,j}^{(k)})^2\right)$ and $W = \left(\frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} (\hat{C}_{i,j}^{(k)} - C_{i,j}^{(k)})^2\right)$. Moreover, it presents the number of times (in percentage) that the model $MDM - RRGR$ is better than the competing models.

Model	Bias	Eff	MSE	RRR	RRR-link
RRR	0.38	0.67	0.68		
$RRR - link$	0.04	0.70	0.70	0.52	
$MDM - RRGR$	0.04	0.50	0.50	0.74	0.72

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 continuousation 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 continuousation (RRPR). Then, we estimate a RRPR

[3] All the simulations where done using a Pentium IV 2.6 Ghz

model with a simple continuousation, i.e. we use a single uniform random variable $[0,1]$ ($M = 1$ in 2.13, 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 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 continuousation has only a small impact on the quality of our estimates, as it improves results by $5-10\%$. 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 continuousation respectively 100%, 83% and 96% of the time, and the RRPR-cont-20 outperforms RRR, RRR-link, RRPR without continuousation and RRPR-cont respectively 100%, 91%, 99% and 82% of the time, which establishes that averaging over continuoused estimates does help.

Table 2: **Bias, Efficiency and Mean Squared Error of MDM-RRPR.**

The model estimate is based on Poisson marginals and multivariate gaussian copula. We have applied our procedure without considering the discreteness of the marginal distributions (RRPR) and we have developed a simple continuousation procedure (RRPR-cont) and we have developed a Monte Carlo simulation on the continuoused estimates (RRPR-cont-20). Our last estimator is therefore the average over $M = 20$ uniform draws of the continuoused estimates: $\hat{\theta}_{cont} = \frac{1}{M}\sum_{l=1}^{M}\hat{\theta}_{cont}^{(l)}$. Where $\hat{\theta}_{cont}^{(l)}$ is the estimate of the parameters obtained with the $l-th$ random uniform draw $U^{(l)}$. 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 $Bias = \left\|\frac{1}{N_{sim}}\sum_{N_{sim}}^{i=1}\hat{C}^i - C\right\|$, $Efficiency = \|V\|$ and $MSE = \|W\|$ of the parameter estimates and where $V = \left(\frac{1}{N_{sim}}\sum_{k=1}^{N_{sim}}(\hat{C}_{i,j}^{(k)} - \bar{C}_{i,j}^{(k)})^2\right)$ and $W = \left(\frac{1}{N_{sim}}\sum_{k=1}^{N_{sim}}(\hat{C}_{i,j}^{(k)} - C_{i,j}^{(k)})^2\right)$. Moreover, it presents the number of times (in percentage) that the model $RRPR-cont$ and $RRPR-cont-20$ are better than the competing models, including the $RRPR$.

Model	Bias	Eff	MSE	RRR	RRR-link	RRPR	RRPR-cont
RRR	9.15	3.28	34.64				
$RRR-link$	1.21	0.45	0.40	1.00			
$RRPR$	2.38	2.90	3.99	1.00	0.11		
$RRPR-cont$	0.19	0.39	0.39	1.00	0.83	0.96	
$RRPR-cont-20$	0.18	0.35	0.36	1.00	0.91	0.99	0.82

4 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.

References

Anderson, J. A., 1984a, Regression and ordered categorical variables, *Journal of the Royal Statistical Society, B* 46, 1–30.

Anderson, Thomas W., 1984b, *An Introduction to Multivariate Statistical Analysis* (John Wiley and Sons: New York).

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.

Diebold, F.X., T.A. Gunther, and A.S. Tay, 1998, Evaluating density forecasts, with applications to financial risk management, *International Economic Review* pp. 863–883.

Fisher, R. A., 1932, *Statistical Methods for Research Workers*.

Hardle, W., and L. Simar, 2003, *Applied Multivariate Statistical Analysis* (Springer: UK).

Heinen, A., and E. Rengifo, 2003, Multivariate modeling of time series count data: An autoregressive conditional poisson model, CORE Discussion Paper 2003/25.

Izenman, Alan Julian, 1980, Assessing dimensionality in multivariate regression, in P.R. Krishnaiah, ed.: *Handbook of Statistics*, vol. 1 . pp. 571–591 (North-Holland Publishing Company).

Joe, Harry, 1997, *Multivariate Models and Dependence Concepts* (Chapman and Hall: London).

Jorgensen, Bent, 1987, Exponential dispersion model, *Journal of the Royal Statistical Society* 49, 127–162.

——— , 1997, *The Theory of Dispersion Models* (Chapman and Hall: London).

Jorgensen, B., and S.L. Lauritzen, 1998, Multivariate dispersion models, Research Report 2, Department of Statistics and Demography, Odense University, Denmark.

Machado, J., and J.M.C. Santos Silva, 2003, Quantiles for counts, Working Paper, ISEG, Universidade Técnica de Lisboa.

McCullagh, P., and J.A. Nelder, 1976, *Generalized Linear Models* (Chapman and Hall: London).

Nelsen, Roger B, 1999, *An introduction to Copulas* (Springer: New York).

Sklar, A, 1959, Fonctions de répartitions à n dimensions et leurs marges, *Public Institute of Statistics of the Univeristy of Paris* 8, 229–231.

Song, Peter Xue-Kun, 2000, Multivariate dispersion models generated from gaussian copulas, *Scandinavian Journal of Statistics* 27, 305–320.

Ter Braak, Cajo J. F., 1986, Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis, *Ecology* 67, 1167–1179.

Tso, M. K-S., 1981, Reduced-rank regression and canonical analysis, *Journal of the Royal Statistical Society, B* 43, 183–189.

Yee, Thomas, and Trevor Hastie, 2002, Reduced-rank vector generalized linear models, Working Paper, University of Auckland.

Yee, T.W., and T. J Hastie, 2003, Reduced-rank vector generalized linear models, *Statistical Modelling* 3, 15–41.

Zeger, Scott L., and Kung-Yee Liang, 1986, Longitudinal data analysis for discrete and continuous outcomes, *Biometrics* 42, 121–130.

Zhu, Mu, and Trevor Hastie, 2002, Feature extraction for non-parametric discriminant analysis, Working Paper, University of Waterloo.