Journal of Modern Applied Statistical Methods

Volume 12 | Issue 1 Article 25

5-1-2013

Bayesian Inference of Pair-Copula Constriction for Multivariate Dependency Modeling of Iran's Macroeconomic Variables

M. R. Zadkarami
ShahidChamran University, Ahvaz, Iran

O. Chatrabgoun
ShahidChamran University, Ahvaz, Iran

Bayesian Inference of Pair-Copula Constriction for Multivariate Dependency Modeling of Iran's Macroeconomic Variables

Cover Page Footnote
The authors wish to acknowledge the reviewer for his/her helpful and insightful comments to the manuscript. The authors also wish to acknowledge discussions with Dr. Gholamali Parham and Dr. Alireza Daneshkhah.

Journal of Modern Applied Statistical Methods
May 2013, Vol. 12, No. 1, 227-234

1538 – 9472/13/$95.00

Bayesian Inference of Pair-Copula Constriction for Multivariate Dependency Modeling of Iran's Macroeconomic Variables

M. R. Zadkarami O. Chatrabgoun
ShahidChamran University,
Ahvaz, Iran

Bayesian inference of pair-copula constriction (PCC) is used for multivariate dependency modeling of Iran's macroeconomics variables: oil revenue, economic growth, total consumption and investment. These constructions are based on bivariate t-copulas as building blocks and can model the nature of extreme events in bivariate margins individually. The model parameter was estimated based on Markov chain Monte Carlo (MCMC) methods. A MCMC algorithm reveals unconditional as well as conditional independence in Iran's macroeconomic variables, which can simplify resulting PCC's for these data.

Key words: Monte Carlo Markov Chain Method, pair-copula construction, vine.

Introduction

Multivariate data usually exhibit a complex pattern of dependency. Methods such as graphical model and Bayesian networks are available to investigate dependency structures in multivariate data. One increasingly popular approach for constructing high dimensional dependency is based on copulas. Copulas are multivariate distribution functions with uniform margins which allow representation of joint distribution functions as a function of marginal distributions and a copula (Sklar, 1959). Copulas are used in various fields of applied sciences, but are most widely used in economics, finance and risk management (Embrechts, et al., 2003; Patton, 2004; Nolte, 2008). The class of copulas for bivariate data is rich in comparison to the one for $d-$dimensional data with $d \geq 3$. Until recently, Gaussian and t-copulas or, more generally, elliptical copulas, have been used for multivariate data (Frahm, et al., 2003). The generalization of bivariate copulas to multivariate copulas of dimensions larger than 2 is not straightforward, however there is one simple generalization for Archimedean copulas known as exchangeable Archimedean copulas (Frey & McNeil, 2003). It should be noted that not all bivariate Archimedean copulas have a corresponding multivariate exchangeable version (Nelsen, 1999).

Approaches for constructing multivariate Archimedean copulas of more than 2 have dimensions been developed by Joe (1997), Embrechts, et al. (2003), Whelan, (2004), McNeil, et al. (2006), Savu and Trede (2006) and McNeil (2007). Joe (1996) and Bedford and Cooke (2001, 2002) constructed flexible higher-dimensional copulas by using only bivariate copulas as building blocks, which they termed vines. Kurowicka and Cooke (2006) discussed Gaussian vine constructions in details. Aas, et al. (2007) first recognized the general construction principle for deriving multivariate copulas; they used more general bivariate copulas than the Gaussian copula and applied these construction methods to financial risk data using more appropriate pair-copulas such as the bivariate t Clayton and Gumbel copulas. According to recent empirical investigations of Berg and Aas (2007) and Fischer, et al. (2007), the vine constructions based on bivariate t-

M. R. Zadkarami is a associated professor on the Faculty of Mathematics and Computer Sciences in the Statistics Department. Email him at: zadkarami@yahoo.co.uk. O. Chatrabgoun is a PhD student on the Faculty of Mathematics and Computer Sciences in the Statistics Department. Email him at: o-chatrabgoun@phdstu.scu.ac.ir.

copulas provide a better fit to multivariate financial data.

Estimating copula parameters is generally based on classical maximum likelihood (ML) and its variations. The most common approach is semi-parametric where the margins are fitted empirically and the dependence parameters are fitted by ML. The asymptotic properties of these semi-parametric estimates have been rigorously investigated by Genest, et al. (1995); however, confidence intervals for dependence parameters are difficult to obtain because determination of the asymptotic variance is not a simple task. Due to this, data analyses often are exclusively based on point estimates of copula parameters. Bayesian inference, or Markov chain Monte Carlo (MCMC) estimation of the parameters, provides a solution for this problem – which is not simple to solve in a classical ML framework.

Using MCMC, interval estimation of parameters can be achieved by credible interval. This is due to the MCMC algorithm introduced by Metropolis, et al. (1953) and Hastings (1970). Credible intervals for parameters of a pair-copula constriction (PCC) can simplify the PCC if they detect conditional and unconditional independency between pairs of variables. However, Bayesian literature on copulas is poor. Pitt, et al. (2006) investigated Gaussian copula regression, the main difficulty they encountered was sampling a positive definite correlation matrix. They solved the problem by employing a covariance selection prior that was introduced by Wong, et al. (2003).

Dalla Valle (2007) proposed Bayesian inference based on MCMC for multivariate Gaussian and t-copulas using the inverse Wishart distribution as a prior for the correlation matrix. The study used Bayesian inference for pair-copula constructions (PPC's) of Iran's macroeconomic variables based on bivariate t-copulas using a method similar to that used by Min and Czado (2011) for a Norwegian financial data set. Min and Czado's method allows modeling of tail dependency between two chosen margins individually, while multivariate Gaussian and t-copulas have the same tail dependency structure for any two chosen margins. A tail dependence coefficient (see Embrechts, et al., 2002) accounts for extreme events of margins occurring simultaneously, this is one of the most important characteristics of financial data because it contains information on heavy-tailedness of multivariate financial data. PCC parameters considered are association and degrees of freedom (df) parameters of bivariate t-copulas.

Copulas

Copulas are d-dimensional multivariate distributions with uniformly distributed marginal distributions on [0, 1] and are very useful for modeling a dependence structure of multivariate data. Let $X=(X_1,X_2,\dots,X_d)'$ be a d−dimensional random vector with joint distribution function $F(x_1,x_2,\dots,x_d)$ and marginal distributions

$$F_1(x_1),\ F_2(x_2),\ \dots,\ F_d(x_d).$$

According to Sklar's (1959) theorem a copula C exists such that

$$F(x_1,x_2,\dots,x_d)=C\left(F_1(x_1),F_2(x_2),\dots,F_d(x_d)\right) \tag{1}$$

and the copula $C(u_1,u_2,\dots,u_d)$ is unique if the marginal distributions are continuous. (See Joe, 1997 and Nelsen, 1999 for additional.)

The copula $C(u_1,u_2,\dots,u_d)$ of a multivariate distribution $F(x_1,x_2,\dots,x_d)$ with margins $F_1(x_1)$, $F_2(x_2),\dots,F_d(x_d)$ is given by

$$C(u_1,u_2,\dots,u_d)=C(F_1^{-1}(u_1),F_2^{-1}(u_2),\dots,F_d^{-1}(u_d))$$

and the copula density is given by

$$c(u_1,u_2,\dots,u_d)=\frac{f(F_1^{-1}(u_1),F_2^{-1}(u_2),\dots,F_d^{-1}(u_d))}{f_1\left(F_1^{-1}(u_1)\right)\dots f_d(F_d^{-1}(u_d))}$$

where$F_i^{-1}(u_i)$ is the inverse of the margins $F_i(x_i)$ for $i=1,2,\dots,d$. Using (1), the multivariate density $f(x_1,x_2,\dots,x_d)$ is a product of the corresponding copula density with marginal densities $f_i(x_i)$, $i=1,2,\dots,d$ and is given by

$$f\left(x_1,x_2,\ldots,x_d\right)= c\left(F_1\left(x_1\right),F_2\left(x_2\right),\ldots,F_d\left(x_d\right)\right).f_1\left(x_1\right)f_2\left(x_2\right)\ldots f_d\left(x_d\right)$$

thus separating the dependence structure from the marginal structure.

PCC's for Multivariate Distributions

Pair-copula constriction (PCC) modeling structure is based on a decomposition of a multivariate density into a cascade of bivariate copulae. In other words, a vine associated with n variables is a nested set of trees where the edges of the tree, j, are the nodes of the tree $j+1$; $j = 1,\ldots,n-2$, and each tree has the maximum number of edges. A regular vine on n variables is a vine in which two edges in tree j are joined by an edge in tree $j+1$ only if these edges share a common node, $j = 1,\ldots,n-2$. There are $n(n-1)/2$ edges in a regular vine on n variables (Kurowicka& Cooke, 2006).

Bedford and Cooke Theorem

Bedford and Cooke (2001) presented the following theorem. Let $V = (T_1,\ldots, T_{n-1})$ be a regular vine on n elements, where T_1 is a connected tree with nodes $N_1 = \{1,\ldots,n\}$ and edges E_1; for $i = 2,\ldots,n-1$, T_i is a connected tree with nodes $N_i = E_{i-1}$. For each edge $e(j,k)\epsilon\, T_i$; $i = 1,\ldots,n-1$ with conditioned set $\{j,k\}$ and conditioning set D_e, let the conditional copula and copula density be $C_{jk|D_e}$ and $c_{jk|D_e}$ respectively. If the marginal distributions F_i with densities f_i; $i = 1,\ldots,n$ are given, then the vine-dependent distribution is uniquely determined and has a density given by

$$f\left(x_1,\ldots,x_n\right)= \prod_{i=1}^{n} f\left(x_i\right) \prod_{e(j,k)\in E_i} c_{jk|D_e}\left(F_{j|D_e(x_j)},F_{jk|D_e(x_k)}\right)$$

The density decomposition associated with 4 random variables $X = (X_1,\ldots,X_4)$ with a joint density function $f(x_1,\ldots,x_4)$ satisfying a copula-vine structure (this structure is called D-vine, see Kurowicka and Cooke, 2006, p. 93) as shown in Figure 1 with the marginal densities $f_1,\ldots,f_4$ is illustrated as:

$$\begin{aligned} f_{1234} = & \prod_{i=1}^{4} f\left(x_i\right)\times c_{12}\left(F\left(x_1\right),F\left(x_2\right)\right) \\ & c_{23}\left(F\left(x_2\right),F\left(x_3\right)\right)c_{34}\left(F\left(x_3\right),F\left(x_4\right)\right) \\ & \times c_{13}\left(F\left(x_1\mid x_2\right),F\left(x_3\mid x_2\right)\right) \\ & c_{24}\left(F\left(x_2\mid x_3\right),F\left(x_4\mid x_3\right)\right) \\ & c_{14}\left(F\left(x_1\mid x_2,x_3\right),F\left(x_4\mid x_2,x_3\right)\right) \end{aligned}$$

Joe (1996, p. 125) showed that the conditional distribution function $F_{U|V}\left(u|\boldsymbol{v}\right)$ appearing in the PCC are partial derivatives with respect to the second argument of the conditional copula given by

$$F_{U|V}\left(u|\boldsymbol{v}\right)=\frac{\partial C_{x,v_j|V_{-j}}\left(F\left(x\mid V_{-j}\right),F\left(v_4\mid V_{-j}\right)\right)}{\partial F\left(v_4\mid V_{-j}\right)}$$

where$C_{x,v_j|V_{-j}}(.,.)$ is a bivariate copula distribution function.

Data Adjustment and MCMC Estimation

Data analyzed in this study are four time-series data related to the Iran's macroeconomics variables: (A) oil revenue, (B) economic growth, (C) total consumption and (D) investment. These data were collected from the Islamic republic of Iran's Central Bank.

First it is necessary to remove serial correlation of the four time series, that is, the observation of each variable must be independent over time. Hence, the serial correlation in the conditional mean and the conditional variance are modeled by an AR(1) and a GARCH(1,1) model (Bollerslev, 1986), respectively. For time series i, the model for log-return x_i is $x_{i,t}=c_i+\alpha_i x_{i,t-1}+\sigma_{i,t}z_{i,t}$, $E\left[z_{i,t}\right]=0$, $\sigma_{i,t}^2=\alpha_{i,0}+a_i\varepsilon_{i,t-1}^2+b_i\sigma_{i,t-1}^2$ where $\varepsilon_{i,t-1}=\sigma_{i,t}+z_{i,t}$ (Aas, et al., 2009). Table 1 shows the analyses performed on the standard residuals, z_i.

Figure 1: Vine Structure with 4 Elements

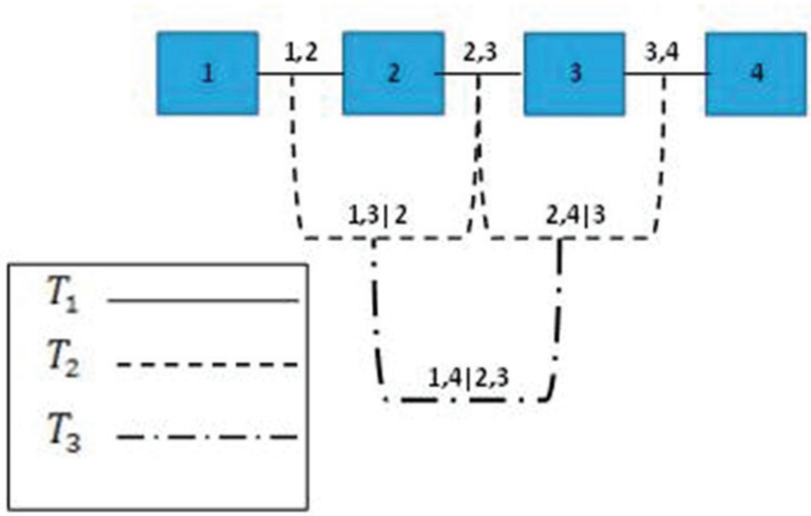


Table 1: Data Adjusted after Removing Serial Correlation from Data

Data	Model	Distribution Error
Oil Revenue	ARMA(1,1)-GARCH(1,1)	Skewed Normal
Economic Growth	GARCH(1,1)	Skewed t-student
Total Consumption	ARMA(1,1)-GARCH(1,1)	t-student
Investment	AR(1)-GARCH(1,1)	Student

Considering a regular vine (as shown in Figure 1), the Bayesian inference for these variables can be carried out. According the vine structure, the formula for these variables is

$$c\left(u_A, u_B, u_C, u_D\right) = c_{AB}.c_{BC}.c_{CD}..c_{AC|B}.c_{BD|C}.c_{AD|BC}$$

The building pair-copulas of the PCC model (2) are now specified as bivariate t-copulas; however, the methodology is generic and applies more widely. Further it is assumed that the margins of X are uniform. This is motivated by the standard semi-parametric copula estimation procedure suggested by Genest, et al. (1995) where approximate uniform margins are obtained by applying the empirical probability integral transformation to multivariate data.

The bivariate t-copula (Embrechts, et al., 2003) has 2 parameters: the association parameter ρ ∈ (-1,1) and the df parameter ϑ ∈ (0,∞) and its density is given by

$$c\left(u_1,u_2|\vartheta,\rho\right)=$$

$$\frac{\Gamma(\frac{\vartheta+1}{2})\Gamma(\frac{\vartheta}{2})}{\sqrt{1-\rho^2}\,[\Gamma\left(\frac{\vartheta+1}{2}\right)]^2}$$

$$.\frac{(\left[1+\frac{(t_\vartheta^{-1}(u_1))^2}{\vartheta}\right]\left[1+\frac{(t_\vartheta^{-1}(u_2))^2}{\vartheta}\right])^{\frac{\vartheta+1}{2}}}{(1+\frac{(t_\vartheta^{-1}(u_1))^2+(t_\vartheta^{-1}(u_2))^2-2\rho t_\vartheta^{-1}\left(u_1\right).t_\vartheta^{-1}(u_2)}{\vartheta(1-\rho^2)})^{\frac{\vartheta+1}{2}}}$$

wheret_ϑ^(-1) (.) is a quantile function of a t-distribution with ϑ degrees of freedom.

The conditional distribution function for x = u_1 and a scalar v = u_2 takes the form

$$h\left(u_1|u_2,\rho,\vartheta\right)=$$

$$t_{\vartheta+1}(\frac{t_\vartheta^{-1}\left(u_1\right)-\rho t_\vartheta^{-1}\left(u_2\right)}{\sqrt{\frac{\left(\vartheta+\left(t_\vartheta^{-1}\left(u_2\right)\right)^2\right)\left(1-\rho^2\right)}{\vartheta+1}}})$$

and is called the h-function for the t-copula with parameters ρ and ϑ (Aas, et al. 2007). The parameters of the model used for this study are:

$$\theta=(\rho_{1,2},\vartheta_{1,2},\rho_{2,3},\vartheta_{2,3},\rho_{3,4},\vartheta_{3,4},\rho_{1,3|2},$$
$$\vartheta_{1,3|2},\rho_{2,4|3},\vartheta_{2,4|3},\rho_{1,4|2,3},\vartheta_{1,4|2,3})$$

Because a Bayesian approach was followed, the statistical model must be completed by specifying the prior distributions for all model parameters. A uniform $(-1,1)$ prior is specified for the association parameter ρ of a t-copula pair and a uniform $(1,U)$ prior for the corresponding df parameter ϑ because, in general, little prior information is available. Here the lower cut value 1 was chosen instead of 0 to avoid numerical instabilities in evaluating a quantile function of the bivariate t-distribution. The upper cut value U can be chosen by the data analyst to assess the closeness to the bivariate Gaussian copula. Finally, it was assumed that prior distributions for ρ and ϑ are independent within each pair and independent over all pairs.

For estimating the MCMC parameters the package bivariate t distribution in *Winbugs*14 software was used (in other tree of the vine structure this is conducting using an h-function). By pre-specified prior distribution for ρ and prior distribution $U(1{,}100)$ for ϑ and 80,000 a Metropolis-Hasting iteration algorithm MCMC estimation of the parameter can be obtained. The results of the Bayesian estimation and MlE are summarized in Table 2.

Based on results (see Table 2 and Figure 2) it can be concluded that the association parameter is unimodal and symmetric, the difference among mode, mean and median is negligible, and the degree of freedom is asymmetric. Based on the 95% credible interval, the dependency structure among variables can be simplified as:

$$c\left(u_A,u_B,u_C,u_D\right)=c_{AB}.c_{BC}.c_{CD}.c_{BD|C}.c_{AD|BC}$$

Conclusion

Bayesian inference provides solutions for many difficult problems that are not simple to solve in a classical ML framework. This study shows how identifying unconditional as well as conditional independence in macroeconomic variables can simplify resulting PCC's. Results show that the independence between oil revenue and total consumption given economic growth in these data is significant.

Acknowledgement

The authors wish to acknowledge the reviewer for his/her helpful and insightful comments to the manuscript. The authors also wish to acknowledge discussions with Dr. Gholamali Parham and Dr. Alireza Daneshkhah.

References

Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2007). Pair-copula constructions of multiple dependence. To appear in *Insurance: Mathematics and Economics*. DOI 10.1016/j.insmatheco.2007.02.001.

Bedford, T., & Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, *32*(*1-4*), 245-268.

Table (2): Bayesian Estimation of Parameters

Parameter	2.5%Quantile	Estimated Median	97.5% Quantile	Estimated Post. Mean	Estimated Post. Mode	MLE
$\rho_{1,2}$	0.1849	0.4431	0.7	0.4431	0.4422	0.4465
$\vartheta_{1,2}$	0.2475	0.3743	0.6568	0.3941	0.3655	0.3422
$\rho_{2,3}$	3.415	4.315	5.222	4.316	4.315	4.316
$\vartheta_{2,3}$	1.014	1.476	2.42	1.538	1.451	1.443
$\rho_{3,4}$	0.7843	1.398	2.016	1.399	1.389	1.401
$\vartheta_{3,4}$	0.6903	1.006	1.648	1.047	1.001	1.017
$\rho_{1,3\|2}$	-0.9819	-0.01785	0.9422	-0.01772	-0.0187	-0.0188
$\vartheta_{1,3\|2}$	0.1661	0.5109	1.168	0.5518	0.491	0.489
$\rho_{2,4\|3}$	0.8218	1.223	1.623	1.223	1.24	1.24
$\vartheta_{2,4\|3}$	0.9594	2.954	6.757	3.191	2.02	2.08
$\rho_{1,4\|2,3}$	-0.7009	-0.4408	-0.1818	-0.4408	-0.441	-0.441
$\vartheta_{1,4\|2,3}$	2.285	7.031	16.09	7.596	6.642	6.831

Bedford, T., & Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, *30*(*4*), 1031-1068.

Berg, D., & Aas, K. (2007). *Models for construction of multivariate dependence*. Discussion Paper SAMBA/23/07.

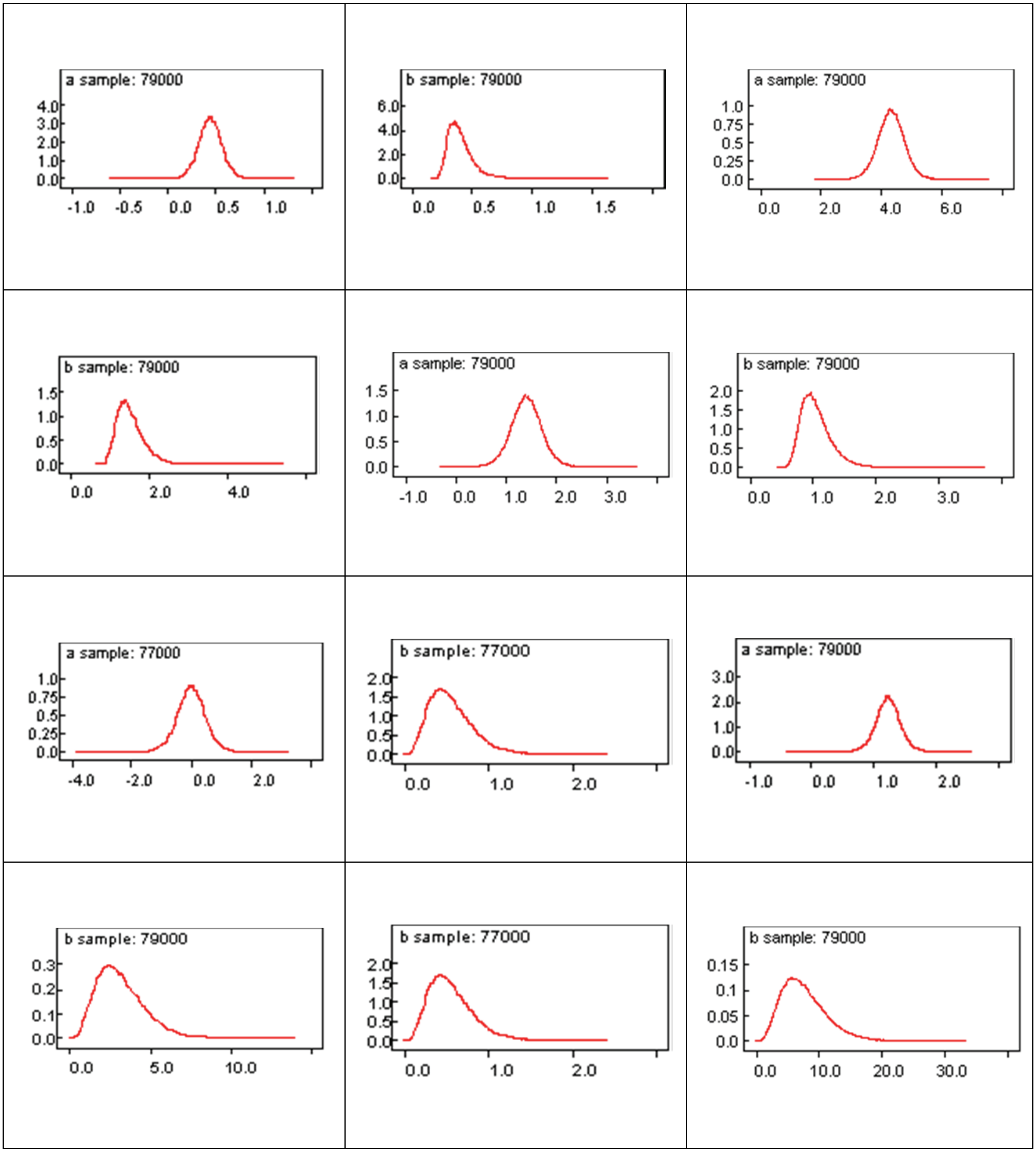
Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*, 307-327.

Dalla Valle, L. (2007). Bayesian copulae distributions with application to operational risk management. To appear in *Methodology and Computing in Applied Probability*. DOI 10.1007/s11009-007- 9067-x.

Embrechts, P., Lindskog, F., & McNeil, A. J. (2003). Modelling dependence with copulas and applications to risk management. In *Handbook of heavy tailed distributions in finance*, 31, 307-327. Amsterdam: Elsevier/North-Holland.

Embrechts, P., McNeil, A. J., & Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In *Risk management: Value at risk and beyond*, M. Dempster (Ed.), 176-223. Cambridge, MA: Cambridge University Press.

Figure 2: Plot of the Bayesian Estimation of the Parameters



Fischer, M., Kock, C., Schluter, S., & Weigert, F. (2007).*An empirical analysis of multivariate copula models*.Discussion Paper, http://www.statistik.wiso.unierlangen.de/ lehrstuhl/mfischer/mfwork.html.

Frahm, G., Junker, M., &Szimayer, A. (2003). Elliptical copulas: Applicability and limitations. *Statistics & Probability Letters*, *63*(*3*), 275-286.

Frey, R., & McNeil, A. J. (2003). Dependent defaults in models of portfolio credit risk. *Journal of Risk*, *6*(*1*), 59-92.

Genest, C., Ghoudi, K., &Rivest, L-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, *82*(*3*), 543-552.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*, 97-109.

Joe, H. (1996). Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. In *Distributions with fixed marginals and related topics*, L. Ruschendorf, B. Schweizer & M. D. Taylor (Eds.), 120-141. Hayward: Institute of Mathematical Statistics.

Joe, H. (1997). *Multivariate models and dependence concepts, volume 73: Monographs on statistics and applied probability*. London: Chapman & Hall.

Kurowicka, D., & Cooke, R. (2006). *Uncertainty analysis with high dimensional dependence modelling*. Chichester: John Wiley & Sons.

McNeil, A. J. (2007). Sampling nested Archimedean copulas. To appear in *Journal of Statistical Computation and Simulation*.

McNeil, A. J., Frey, R., & Embrechts, P. (2006). *Quantitative risk management: Concepts, techniques and tools*. Princeton, NJ: Princeton University Press.

Metropolis, N., et al. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*, 1087-1092.

Min, A., & Czado, C. (2011). Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, *8*(*4*), 511-546.

Nelsen, R. B. (1999). *An introduction to copulas, volume 139: Lecture notes in statistics*. New York, NY: Springer-Verlag.

Nolte, I. (2008). Modeling a multivariate transaction process.*Journal of Financial Econometrics*, *6*, 143-170.

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, *2*, 130-168.

Pitt, M., Chan, D., & Kohn, R. (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika*, *93*(*3*), 537-554.

Savu, C., & Trede, M. (2006). Hierarchical Archimedean copulas. In *International Conference on High Frequency Finance*, 109-116. Konstanz, Germany, May.

Sklar, M. (1959). Fonctions de r´epartition `a n dimensions et leursmarges. *Publications de l'Institut de Statistique de l'Université de Paris*, *8*, 229-231.

Whelan, N. (2004). Sampling from Archimedean copulas. *Quantitative Finance*, *4*(*3*), 339-352.

Wong, F., Carter, C. K., & Kohn, R. (2003). Efficient estimation of covariance selection models. *Biometrika*, *90*(*4*), 809-830.