Economics Letters 123 (2014) 257–261

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

An algorithm for constructing high dimensional distributions from distributions of lower dimension☆

Stanislav Anatolyev[a,*], Renat Khabibullin[b], Artem Prokhorov[c,d]

[a] New Economic School, Nakhimovsky Pr., 47, office 1721(3), Moscow 117418, Russia
[b] Barclays Capital, Four Winds Plaza, Bolshaya Gruzinskaya Street 71, Moscow 123056, Russia
[c] University of Sydney Business School and CIREQ, Sydney NSW 2006, Australia
[d] St.Petersburg State University, St.Petersburg 199034, Russia

HIGHLIGHTS

- When distribution is highly dimensional the maximum likelihood procedure is non-operational.
- The proposed sequential procedure constructs a multivariate distribution using lower-dimensional ones.
- The procedure provides excellent fit in financial applications.

ARTICLE INFO

Article history:
Received 18 January 2014
Received in revised form
21 February 2014
Accepted 26 February 2014
Available online 5 March 2014

JEL classification:
C13

Keywords:
Pseudo-likelihood
Multivariate distribution
Copulas

ABSTRACT

We propose a new sequential procedure for estimating multivariate distributions in cases when conventional maximum likelihood has too many parameters and is therefore inaccurate or non-operational. The procedure constructs a multivariate distribution and its pseudo-likelihood sequentially, in each step using lower-dimensional distributions with a small number of parameters. In an application, the procedure provides excellent fit when the dimension is moderate, and remains operational when the conventional method fails.

1. Introduction

Consider the problem of constructing a high dimensional distribution. As an example, suppose we wish to estimate a d-dimensional Student-t distribution. The problem has at least $d(d-1)/2$ parameters. The conventional approach is to construct a joint log-density from this d-dimensional distribution and use it in a maximum likelihood (ML) routine. However, for large d and moderate sample sizes, the likelihood is highly unstable, Hessians are near singular, estimates are inaccurate, and global convergence is hard to achieve.

One solution is to use copulas which have tighter parameterizations. However, the functional form of such copulas limits the nature of dependence they can accommodate (Nelsen, 2006, Section 4.6). Another solution is to use 'vine copulas' (Aas et al., 2009) when the d-variate density is decomposed into a product of up to $d(d-1)/2$ bivariate densities. However, there are still $O(d^2)$ parameters in the joint likelihood; in addition, the required ordering of components is rarely available, especially in the time series context. Yet another alternative is to use the factor copula approach (Oh and Patton, 2013). However, the joint density obtained lacks a closed form; in addition, it is unclear whether the convolution of distributions imposed by the factor copula covers all classes of joint distributions one may wish to model.

The proposed method replaces the initial estimation problem with a sequence of bivariate problems. The procedure can be thought of as recovering the joint distribution from the distributions of all lower-dimensional sub-vectors comprising the original random vector. This provides sufficient flexibility as there

☆ Helpful discussions with Eric Renault, Tommaso Proietti, Dmytro Matsypura and Oleg Prokopyev are gratefully acknowledged. We also thank the editor and referee for speedy reviewing and useful suggestions.

* Corresponding author. Tel.: +7 495 1291700; fax: +7 495 1293722.
E-mail addresses: sanatoly@nes.ru (S. Anatolyev), sir.renat@gmail.com (R. Khabibullin), artem.prokhorov@sydney.edu.au (A. Prokhorov).

http://dx.doi.org/10.1016/j.econlet.2014.02.022
0165-1765/

are more degrees of freedom in choosing a parameterization in each step. The proposed estimator can be viewed as a traditional pseudo maximum likelihood estimator, but it is more flexible and works reasonably well in situations when the traditional ML fails.

2. The algorithm

In this section we describe the proposed algorithm, while in the next section we discuss its asymptotic properties.

Step 1. Estimate the marginals by fitting a suitable parametric distribution $\widehat{F}_j = F(\widehat{\theta}_j)$ for each $j = 1, \ldots, d$. This step involves d estimation problems.

Step 2. Using the $\widehat{F}_j$'s, estimate a bivariate distribution $\widehat{F}_{ij} = C^{(2)}(\widehat{F}_i, \widehat{F}_j; \widehat{\theta}_{ij})$ for each pair (i, j), where $C^{(2)}$ denotes a bivariate copula. There are $d(d-1)$ estimation problems in this step.

Step 3. Using the $\widehat{F}_j$'s and $\widehat{F}_{ij}$'s, estimate a trivariate distribution $C^{(3)}(\widehat{F}_i, \widehat{F}_{jk}; \widehat{\theta}_{ijk})$, for each combination of i and (j, k), where $C^{(3)}$ is a suitable *compounding function* capturing dependence between each element i and each disjoint pair (j, k). There are $d(d-1)(d-2)/2$ such combinations. Now, average $(\widehat{F}_i, \widehat{F}_{jk})$ over permutations of (i, j, k):

$$\widehat{F}_{ijk} = \frac{C^{(3)}\left(\widehat{F}_i, \widehat{F}_{jk}; \widehat{\theta}_{ijk}\right) + C^{(3)}\left(\widehat{F}_j, \widehat{F}_{ik}; \widehat{\theta}_{jik}\right) + C^{(3)}\left(\widehat{F}_k, \widehat{F}_{ij}; \widehat{\theta}_{kij}\right)}{3}.$$

Step m. Using the $\widehat{F}_j$'s and $\widehat{F}_{i_1,\ldots,j-1,j+1,\ldots,i_m}$, estimate an m-dimensional distribution of each m-tuple. There are $d!/(d-m)!(m-1)!$ possible combinations of $\widehat{F}_i$'s with disjoint $(m-1)$-variate marginals. Let $i_1 < i_2 < \cdots < i_m$, then obtain a model average estimate of the distribution for the $(i_1, i_2, \ldots, i_m)$-th m-tuple:

$$\widehat{F}_{i_1 i_2 \ldots i_m} = \frac{1}{m} \sum_{l=1}^{m} C^{(m)}\left(\widehat{F}_l, \widehat{F}_{i_1,\ldots,l-1,l+1,\ldots,i_m}; \widehat{\theta}_{l,i_1,\ldots,l-1,l+1,\ldots,i_m}\right),$$

where $C^{(m)}$ is an m-th order compounding function which is set to be a suitable asymmetric bivariate copula.

Step d. Estimate the d-variate distribution:

$$\widehat{F}_{12\ldots d} = \frac{1}{d} \sum_{l=1}^{d} C^{(d)}\left(\widehat{F}_l, \widehat{F}_{1,\ldots,l-1,l+1,\ldots,d}; \widehat{\theta}_{l,1,\ldots,l-1,l+1,\ldots,d}\right),$$

where $C^{(d)}$ is a d-th order compounding function. There are d such functions to be estimated.

3. Asymptotic properties

Let $\widehat{\boldsymbol{\theta}}$ contain all $\hat{\theta}$'s from Steps 1 to d. Then, by the Sklar (1959) theorem, the distribution $\widehat{F}_{12\ldots d}(x_1, \ldots, x_d)$ implies a d-copula $K(u_1, \ldots, u_d; \widehat{\boldsymbol{\theta}})$ and the corresponding estimator of density $\widehat{f}_{12\ldots d}(x_1, \ldots, x_d)$ implies a d-copula density $k(u_1, \ldots, u_d; \widehat{\boldsymbol{\theta}})$.[1] There is no guarantee that the m-th order compounding functions are also m-copulas, $m = 3, \ldots, d$, unless we use a compatible copula family.[2] However, the resulting estimator $\widehat{F}_{12\ldots d}$ is a continuous, non-decreasing, bounded d-variate function with range $[0, 1]$, which *is* a distribution and thus implies a d-copula. The following result gives explicit formulas for the copula (density) implied by our estimator.

Proposition 1. *Let $\widehat{F}_m^{-1}(u_m)$, $m = 1, \ldots, d$, denote the inverse of the marginal cdf $\widehat{F}_m$ from Step 1 and let $\widehat{f}_m$ denote the pdf corresponding to $\widehat{F}_m$. Then, the copula implied by $\widehat{F}_{12\ldots d}$ can be written as follows:*

$$K(u_1, \ldots, u_d; \widehat{\boldsymbol{\theta}}) = \widehat{F}_{12\ldots d}(\widehat{F}_1^{-1}(u_1), \ldots, \widehat{F}_d^{-1}(u_d)),$$
$$k(u_1, \ldots, u_d; \widehat{\boldsymbol{\theta}}) = \frac{\widehat{f}_{12\ldots d}(\widehat{F}_1^{-1}(u_1), \ldots, \widehat{F}_d^{-1}(u_d))}{\prod_{m=1}^{d} \widehat{f}_m(\widehat{F}_m^{-1}(u_m))}.$$

It is clear from Proposition 1 that our algorithm provides an estimate of a flexible parametric d-variate *pseudo*-copula.[3] So the asymptotic properties of our estimator are basically the well-studied properties of copula-based *pseudo*- or *quasi*-ML estimator (Joe, 2005; Prokhorov and Schmidt, 2009). The following proposition summarizes these results, without proof.

Proposition 2. *Asymptotically the estimator $\widehat{\boldsymbol{\theta}}$ minimizes the Kullback–Leibler divergence criterion,*

$$\widehat{\boldsymbol{\theta}} \xrightarrow{p} \arg\min_{\boldsymbol{\theta}} \mathbb{E} \ln \frac{c(u_1, \ldots, u_d)}{k(u_1, \ldots, u_d; \boldsymbol{\theta})},$$

where c is the true copula density and expectation is with respect to the true distribution. Furthermore, under standard regularity conditions, $\widehat{\boldsymbol{\theta}}$ is consistent and asymptotically normal. If the true copula belongs to the family $k(u_1, \ldots, u_d; \boldsymbol{\theta})$, it is consistent for the true value of $\boldsymbol{\theta}$. If the copula family is misspecified, the convergence is to a pseudo-true value of $\boldsymbol{\theta}$, which minimizes the Kullback–Leibler distance.

Fundamentally, our algorithm uses the following form of the joint distribution:

$$H(x_1, \ldots, x_d) = C^{(d)}(F_d(x_d), C^{(d-1)}(F_{d-1}(x_{d-1}), \ldots)),$$

where marginals are ordered in an arbitrary way. For example, $C^{(3)}$ can be formed as $C^{(3)}\left(F_1, C^{(2)}(F_2, F_3)\right)$, or as $C^{(3)}\left(F_2, C^{(2)}(F_1, F_3)\right)$, etc. Since no single ordering is preferred we apply model averaging to combine them. This is a central question in the literature on combining multiple prediction densities (Geweke and Amisano, 2011), where optimal weights, also known as scoring rules, are worked out in the context of information theory. As an example, define $c_j^{(3)}$ as $c_j^{(3)} \equiv c^{(3)}(F_j, C_k^{(2)})$, where $j, k = 1, 2, 3, j \neq k$ and $C_k^{(2)} \equiv C^{(2)}(F_k, F_l), l \neq k, l \neq j$. Then, it is possible in principle to obtain the optimal weights ω_j's as solutions to the following problem:

$$\max_{\omega_l : \sum \omega_j = 1} \sum_{\text{sample}} \ln \sum_j \omega_j c_j^{(3)}.$$

Such scoring rules make ω_j's a function of $c_j^{(3)}$'s and may be worth pursuing in large samples. However, it has been noted in this literature that, in finite samples, a simple average often performs better due to the error from estimating ω's (Stock and Watson, 2004). Moreover, in our setting, the optimal weights would need to be solved for in each step, imposing a heavy computational burden.

[1] We denote the implied copula distribution and density functions by K and k, respectively, to distinguish them from the true copula distribution $C(u_1, \ldots, u_d)$ and true copula density $c(u_1, \ldots, u_d)$.

[2] There are several impossibility results concerning construction of high dimensional copulas by using lower dimensional copulas as argument of bivariate copulas (Quesada-Molina and Rodriguez-Lallena, 1994).

[3] Here by *pseudo*-copula we mean a possibly misspecified copula function. The same term is sometimes used in reference to the empirical copula obtained using univariate empirical cdf's.