Efficient estimation of parameters in marginals in semiparametric multivariate models

Preliminary and Incomplete – Please do not cite

Valentyn Panchenko* Artem Prokhorov†

April 29, 2009

Abstract

Recent literature on semiparametric copula models focused on the situation when the marginals are specified nonparametrically and the copula function is given a parametric form. For example, this setup is used in Chen, Fan and Tsyrennikov (2006) [Efficient Estimation of Semiparametric Multivariate Copula Models, JASA] who focus on the efficient estimation of copula parameters. We consider a reverse situation when the marginals are specified parametrically and the copula function is modelled nonparametrically. We use the method of sieve for efficient estimation of the parameters in the marginals and show the asymptotic distribution. Simulations show that the sieve MLE can be up to 40% more efficient relative to QMLE depending on the strength of dependence between marginals. An application using insurance company loss and expense data demonstrates empirical relevance of our approach.

JEL Classification: C13

Keywords: sieve MLE, copula, semiparametric efficiency

*Economics at the Australian School of Business, University of New South Wales, Sydney NSW 2052, Australia; email: valentyn.panchenko(at)unsw.edu.au

†Department of Economics, Concordia University, Montreal, PQ H3G1M8 Canada; email: artem.prokhorov(at)concordia.ca

1 Introduction

Consider an m-variate random variable Y with joint pdf $h(y_1, \ldots, y_m)$. Let $f_1(y_1), \ldots, f_m(y_m)$ denote the corresponding marginal pdf's. Assume that the marginals are known up to a parameter vector β (β collects the parameters of all marginals). The dependence structure is not given. We observe a random sample $\{\mathbf{y}_i\}_{i=1}^N = \{y_{1i}, \ldots, y_{mi}\}_{i=1}^N$. We are interested in estimating β efficiently without assuming anything about the joint distribution except for the marginals. We wish to do so using the sieve maximum likelihood estimation (see, Shen, 1997).

As a simple example consider the setting of a panel (small T, large N). We have a well specified marginal for each of T cross section (e.g., logit models, duration models, etc.) and we are interested in efficient estimation of the parameters in marginal distributions without assuming a parametric form on time dependence between them. This setting is typical for microeconomic applications. The variable of interest y_t, $t = 1, \ldots, T$, can be the duration of unemployment in period t, or the use of social services in period t. Additional motivation for this problem comes from insurance. In particular, it arises in models of survival of multiple lives, where the two or more durations are dependent (see,e.g., Frees and Valdez, 1998). In life insurance of spouses this effect is known as the "broken heart" syndrome. In finance, a similar setting arises in the so called SCOMDY models (Chen and Fan, 2006a,b), where interest is in the parameters of individual conditional distribution and innovations of the univariate GARCH models have arbitrary dependence.

We will use the well known representation of log-joint-density in terms of log-marginals and the log-copula:

$$\ln h(y_1, \ldots, y_m; \beta) = \sum_{j=1}^{m} \ln f_j(y_j; \beta) + \ln c(F_1(y_1; \beta), \ldots, F_m(y_m; \beta)), \tag{1}$$

where $c(\cdots)$ is a copula density and F_i denotes the corresponding marginal CDF's. This decomposition is due to Sklar's (1959) theorem which states that any continuous joint distribution with continuous marginals can be represented by a unique copula function of the corresponding marginal CDF's.

It is known that the parameters of the marginals can be consistently estimated by maximizing the likelihood of the marginals (by imposing independence) – this is the so called quasi maximum likelihood estimator or QMLE. However, QMLE is not efficient in the case of dependence between marginals. In the context of two-stage estimation of the parametric copula models, Joe (2005) investigated the efficiency loss of the QMLE for the parameters of the marginals relative to the full likelihood MLE. The relative efficiency loss was the largest in the case of high dependence in the model. Recently, Prokhorov and Schmidt (2008) investigated the conditions for copula redundancy, that is when using copula score does not improve the efficiency the MLE, and when pseudo MLE based on the incorrectly specified copula leads to robust estimation. Segers et al. (2008) considered the problem of efficient estimation of the nonparametric marginals in the situation when the copula is fully known.

In this paper we investigate whether we may obtain a consistent estimator of the parameters of the marginals which is more efficient then the QMLE by modelling the copula nonparametrically. The questions are how to estimate β semiparametrically, what is the semiparametric efficiency bound for estimation of β and whether we achieve this bound. There are more than one ways to do this. Two-step semiparametric estimation is one possibility (see Newey and McFadden, 1994): step one estimates c nonparametrically by, say, $\hat{c}$ and step two estimates β by some estimation method with c replaced with $\hat{c}$. Several of these methods produce efficient estimators of β (see, e.g, Severini and Wong, 1992). Here we consider simultaneous (one-step) estimation of β and $\ln c$.

We draw heavily on the results of Chen et al. (2006). Chen et al. (2006) considered a problem which is the converse of ours. They study the sieve ML estimation when the copula has a known parametric form but the marginals are unknown. In their setting, sieves are employed to approximate univariate densities. The main difficulty of our setting is that the (log)density we are approximating is multivariate.

2 Sieve MLE

Denote the true copula density by $c_o(\mathbf{u})$, $u = (u_1, \ldots, u_m)$, and denote the true parameter vector by β_o. Let $c_o(u)$ belong to an infinite-dimensional space $\Gamma = \{c(\mathbf{u}) : [0,1]^m \rightarrow$

$[0,1]$, $\int_{[0,1]^m} c(\mathbf{u}) = 1$) and β_o belong to $B \subset R^p$. Given a finite amount of data, optimization over the infinite-dimensional space Γ is not feasible. To overcome this problem, the method of sieves can be used. Define a sequence of approximating spaces Γ_N, called sieves, such that $\bigcup_N \Gamma_N$ is dense in Γ. Optimization is then restricted to the space of the sieve (see, Shen, 1997).

Chen (2007) suggests that a convenient finite dimensional linear sieve for approximating a multivariate log-pdf on $[0,1]^m$ is a tensor product of linear univariate sieves on $[0,1]$:

$$\Gamma_N = \left\{ c_{J_N}(\mathbf{u}) = \exp\left\{ \sum_{k=1}^{J_N} a_{1k} A_k(u_1) \cdot \ldots \cdot \sum_{k=1}^{J_N} a_{mk} A_k(u_m) \right\}, \right. \tag{2}$$

$$\left. \mathbf{u} \in [0,1]^m, \int_{[0,1]^m} c_{J_N}(\mathbf{u}) d\mathbf{u} = 1 \right\}, \tag{3}$$

$$J_N \to \infty \frac{J_N}{N} \to 0, \tag{4}$$

where $\{A_k\}$ contains known basis functions and $\{a_{jk}\}$ contains the unknown sieve coefficients. Specific examples of the basis functions $A_k(u)$ include power series, trigonometric polynomials, splines, and wavelets (see, e.g., Chen, 2007). The number of sieve elements in tensor sieve is equal to J_N^m. Since in general there is no analytic solution for the MLE for the coefficients of these sieves, the practical implementation of tensor sieves is complicated.

As an alternative we suggest using Bernstein polynomials, in particular the Bernstein copula density introduced by Sancetta and Satchell (2004).

$$c_{J_N}(\mathbf{u}) = J_N^m \sum_{v_1=0}^{J_N-1} \cdots \sum_{v_m=0}^{J_N-1} \omega_v \prod_{l=1}^{m} \begin{pmatrix} J_N - 1 \\ v_l \end{pmatrix} u_l^{v_l} (1-u_l)^{J_N - v_l - 1}, \tag{5}$$

where ω_v denotes the parameters of the polynomial indexed by $v = (v_1, \ldots, v_m)$ such that $0 \le \omega_v \le 1$ and $\sum_{v_1=0}^{J_N-1} \cdots \sum_{v_m=0}^{J_N-1} \omega_v = 1$.

For the initial values of the parameters we may use the multivariate empirical density (histogram) estimator, i.e. $\omega_v = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(U_i \in H_v)$, where $U_i = (F_1(y_1), \ldots, F_m(y_m))$, $\mathbb{I}(\cdot)$ is the indicator function and

$$H_v = \left[\frac{v_1}{J_n}, \frac{v_1+1}{J_n} \right] \times \cdots \times \left[\frac{v_m}{J_n}, \frac{v_m+1}{J_n} \right]. \tag{6}$$

Note that the sieve above can be represented by a weighted sum of β-distributions. The relation between the empirical density and the MLE solution for ω still needs to be investigated. Sancetta (2007) derives the rates of convergence of Bernstein copula to the true copula. Ghosal (2001) and references therein, discusses the rate of convergence of the sieve MLE based on Bernstein polynomial (only for one-dimensional densities.)

We can now write the sieve for $\Theta = B \times \Gamma$ as $\Theta_N = B \times \Gamma_N$. If we further let $\theta = (\beta', c)'$, the sieve MLE can be written as

$$\hat{\theta} = \arg\max_{\theta \in \Theta_N} \sum_{i=1}^{N} \ln h(\mathbf{y}_i; \theta) \tag{7}$$

Assumption 1 *(identification)* $\beta_o \in int(B) \subset R^p$*, B is compact and there exists a unique θ_o which maximizes $E[\ln h(\mathbf{Y}_i; \theta)]$ over $\Theta = B \times \Gamma$*

It is common in nonparametrics to restrict the class of considered functions by a certain smoothness property. Let g denote a real-valued, J times continuously differentiable function on $[0,1]^m$ whose J-th derivative satisfies the following condition for some $K > 0$ and $r \in (J, J+1]$:

$$|D^J g(x) - D^J g(y)| \leq K|x - y|_E^{r-J}, \text{for all } x, y \in [0,1]^m, \tag{8}$$

where $D^\alpha = \frac{\partial^\alpha}{\partial x_1^{\alpha_1} \ldots \partial x_m^{\alpha_m}}, \alpha = \alpha_1 + \ldots + \alpha_m$ is the differential operator, and $|x|_E = (x'x)^{1/2}$ is the Euclidean norm. Then g is said to belong to the Hölder class on $[0,1]^m$, denoted $\Lambda^r([0,1]^m)$. It is also called r-smooth on $[0,1]^m$. Linear sieves are known to approximate r-smooth functions well (see, e.g., Chen, 2007).

Assumption 2 *(smoothness - Hölder class for copula; differentiability for marginals)* $\Gamma = \{c = \exp(g) : g \in \Lambda^r([0,1]^m), r > 1/2, \int c(u)du = 1\}$ *and* $\ln f_j(y_j; \beta), j = 1, \ldots, m$*, are twice differentiable w.r.t.* β

Now consider the pathwise derivative of the loglikelihood in direction $\nu = (\nu'_\beta, \nu_\gamma)' \in V$, where V is the linear span of $\Theta - \{\theta_o\}$.

$$\begin{aligned}\dot{l}(\theta_o)[\nu] &\equiv \lim_{t\to 0} \frac{\ln h(\theta+t\nu,y)-\ln h(\theta,y)}{t}\Big|_{\theta=\theta_o} \\ &= \frac{\partial \ln h(\theta_o,y)}{\partial\theta'}[\nu] \\ &= \sum_{j=1}^{m}\left\{\frac{\partial \ln f_j(y_j,\beta_o)}{\partial\beta'} + \frac{1}{c(F_1(y_1,\beta_o),\ldots,F_m(y_m,\beta_o))}\ \frac{\partial c(u_1,\ldots,u_m)}{u_j}\Big|_{u_k=F_k(y_k,\beta_o)}\ \frac{\partial F_j(y_j,\beta_o)}{\partial\beta'}\right\}\nu_\beta \\ &\quad + \frac{1}{c(F_1(y_1,\beta_o),\ldots,F_m(y_m,\beta_o))}\nu_\gamma(u_1,\ldots,u_m)\end{aligned}$$

On space V, define the Fisher inner product $\langle\cdot,\cdot\rangle \equiv E\left[\dot{l}(\theta_o)[\cdot]\dot{l}(\theta_o)[\cdot]\right]$ and the Fisher norm $||\nu|| \equiv \sqrt{\langle\nu,\nu\rangle}$, where expectation is with respect to the true density h.

We will estimate β by sieve MLE. As a more general problem, Chen et al. (2006) consider the plug-in estimator where a sieve MLE estimator is used as an argument of an arbitrary functional. They study the asymptotic properties of the functional and note that they depend on smoothness of the functional and the rate of convergence of the sieve MLE estimator. In our setting, the functional has the simple linear form $\lambda'\beta$, where $\lambda \in R^p$ and each element of λ is greater than zero and less than infinity. Then $\lambda'\beta$ can be viewed as a smooth functional of θ and the results of Chen et al. (2006) will apply.

Following Chen et al. (2006), we will look for the Riesz representer $\nu^* \in \bar{V}$, where $\bar{V}$ is the closure of V. The Riesz representer should satisfy these two conditions:

$$\lambda'(\beta-\beta_o) = \langle\theta-\theta_o,\nu^*\rangle, \text{for any } \theta-\theta_o \in \bar{V}$$

and

$$||\nu^*||^2 \equiv \sup_{\nu\neq 0,\nu\in\bar{V}} \frac{|\lambda'\nu_\beta|^2}{||\nu||^2} < \infty$$

Then, under certain assumptions, the efficiency bound for β is $||\nu^*||^2$.

As in Chen et al. (2006), write

$$\begin{aligned}\sup_{\nu\neq 0,\nu\in\bar{V}} \frac{|\nu_\beta|^2}{||\nu||^2} &= \sup_{\nu\neq 0,\nu\in\bar{V}}\left\{|\lambda'\nu_\beta|^2\left(E\left[\dot{l}(\theta_o)[\nu]^2\right]\right)^{-1}\right\} \\ &= \lambda'\left(ES_\beta S'_\beta\right)^{-1}\lambda,\end{aligned} \tag{9}$$

where

$$\begin{aligned}S'_\beta &= \sum_{j=1}^{m}\left\{\frac{\partial \ln f_j(y_j,\beta_o)}{\partial\beta'} + \left(\frac{1}{c(\mathbf{u})}\ \frac{\partial c(u_1,\ldots,u_m)}{u_j}\right)\Big|_{u_k=F_k(y_k,\beta_o)}\ \frac{\partial F_j(y_j,\beta_o)}{\partial\beta'}\right\} \\ &\quad + \frac{1}{c(F_1(y_1,\beta_o),\ldots,F_m(y_m,\beta_o))}g^*(u_1,\ldots,u_m)\end{aligned} \tag{10}$$

and $(g_1^*, \ldots, g_p^*)$, which belong to the product space of square integrable zero-mean functions on $[0,1]^m$, are the solutions to the following infinite-dimensional optimization problem for $q = 1, \ldots, p$:

$$\inf_{g_q} E\left[\sum_{j=1}^{m}\left\{\frac{\partial \ln f_j(y_j,\beta_o)}{\partial \beta_q} + \left(\frac{1}{c(\mathbf{u})}\frac{\partial c(\mathbf{u})}{u_j}\right)\Bigg|_{u_k=F_k(y_k,\beta_o)}\frac{\partial F_j(y_j,\beta_o)}{\partial \beta_q}\right\}\right. \tag{11}$$

$$\left. + \frac{1}{c(\mathbf{u})}\Bigg|_{u_k=F_k(y_k,\beta_o)} g_q(u_1,\ldots,u_m)\right]^2 \tag{12}$$

Assumption 3 *(nonsingular information) Assume that $ES_\beta S'_\beta$ is nonsingular.*

Assumption 4 *(convergence of sieve MLE) Assume that $||\hat{\theta} - \theta_o|| = O_P(\delta_N)$ for $(\delta_N)^w = o(N^{-1/2})$ and there exists $\Pi_N \nu^* \in V_N - \{\theta_o\}$ such that $\delta_N ||\Pi_N \nu^* - \nu^*|| = o(N^{-1/2})$.*

Theorem 1 *Under Assumptions 1-4 and regularity conditions of Chen et al. (2006) (Assumptions 5 and 6), $\sqrt{N}(\hat{\beta} - \beta_o) \Rightarrow N(0, (E[S_\beta S'_\beta])^{-1})$*

Given the consistent SML estimates $\hat{\beta}$ and $\hat{c}$, the asymptotic variance can be estimated consistently in a sieve minimization problem as follows

$$\hat{V} = \arg\min_{g_q \in \mathbf{A}_N}\left[\sum_{j=1}^{m}\left\{\frac{\partial \ln f_j(y_j,\hat{\beta})}{\partial \beta_q} + \left(\frac{1}{\hat{c}(\hat{\mathbf{u}})}\frac{\partial \hat{c}(\hat{u}_1,\ldots,\hat{u}_m)}{u_j}\right)\Bigg|_{\hat{u}_k=F_k(y_k,\hat{\beta})}\frac{\partial F_j(y_j,\hat{\beta})}{\partial \beta_q}\right\}\right.$$

$$\left. + \frac{1}{\hat{c}(F_1(y_1,\hat{\beta}),\ldots,F_m(y_m,\hat{\beta}))} g_q(\hat{u}_1,\ldots,\hat{u}_m)\right],$$

where $q = 1, \ldots, p$ and $\mathbf{A}_N$ is one of the sieve spaces discussed above.

3 Simulations

Our initial simulations with linear tensor sieves, including splines, polynomials, and trigonometric polynomials, exhibit slow convergence rates. In contrast, using Bernstein polynomials, we were able to obtain the convergence within reasonable time. We therefore present the results for the latter sieve.

One of the practical problems we face is the choice of the degree of polynomials J_N in finite samples. While some asymptotic results on the rate of convergence and its dependence on J_N are available, they are not informative in the finite sample situation. The literature on sieves suggest using typical model selection techniques, such as BIC, AIC or cross-validation. However, the theoretical implications of using these techniques in the context of sieves are not explored.

The DGP we use in simulations is similar to Joe (2005) who studied asymptotic relative efficiency (ARE) which is the ratio of the asymptotic variance of the QMLE to Full MLE estimators of the parameters of the marginals. Joe (2005) finds that the ARE depends on the specification of the marginals and copula. In particular, the higher is the dependence implied by the copula, the lower is the ARE of the QMLE. We take the case where the ARE is the lowest and investigate whether we may improve the efficiency of the QMLE by using the semiparametric sieve MLE technique.

We consider bivariate DGP with the marginals which are exponentially distributed with the both mean parameters set to 0.5. The dependence is modelled by the Plackett copula with the parameter 0.002, which implies that we are close the lower Frechet bound. Joe (2005) reports ARE=0.064 of the QMLE in this specific case. In the simulation we use correctly specified marginals up to the 2 parameters to be estimated, while the copula function is modelled using the Bernstein polynomials sieve. We use the BIC to determine the degree of the sieve J_N. The number of observations $N = 1000$. BIC is minimized for $J_N = 8$. Thus we have to estimate 64 nuisance parameters of the sieve and 2 parameters of the marginals. The optimization is complicated by the restrictions on the parameters of the sieve and the parameters of the marginals. We used standard constrained maximization routine in Matlab. Because of time constraints we used only 100 simulation runs. This will be extended in the future. We report the simulated mean of the Sieve MLE, QMLE and Full MLE estimators, their simulated variance and the simulated relative efficiency (RE) of the QMLE with respect to the Sieve MLE.

The result suggests that in this specific situation we were able to improve the efficiency relatively to the QMLE. Please, note that it seems that there is some evidence of downward bias in the estimates based on Sieve MLE for $J_N = 8$. Therefore, we try $J_N = 9$ in which case the bias seems to become smaller and the variance is also improved. This may suggest

	μ_1 SMLE	QMLE	FMLE	μ_2 SMLE	QMLE	FMLE
$J_N = 8$						
Mean	0.488679	0.501630	0.499924	0.488633	0.498784	0.499509
Var	0.000126	0.000194	0.000012	0.000159	0.000233	0.000012
RE	0.649485			0.682403		
$J_N = 9$						
Mean	0.489800	0.501900	0.499951	0.489600	0.498300	0.500001
Var	0.000118	0.000194	0.000012	0.000153	0.000234	0.000012
RE	0.607530			0.653698		

that BIC may not be the optimal procedure to select J_N.

4 Application from insurance

We demonstrate the use of the sieve MLE with an insurance application. We have data on 1,500 insurance claims. For each claim, we have the amount of claim payment, or loss, (Y_1) and the amount of claim-related expenses (Y_2). The claim-related expenses known as ALAE (allocated loss adjustment expense) include the insurance company expenses attributable to an individual claim, e.g. the lawyers' fees and claim investigation expenses. The claim amount variable is censored – there is a dummy indicating whether a given claim amount has reached the policy limit. For details of the data set, see Frees and Valdez (1998).

The claim amount and ALAE are assumed to be distributed according to the Pareto distribution with parameters (λ_1, θ_1) and (λ_2, θ_2), respectively. The Pareto distribution has the following CDF:

$$F(Y) = 1 - \left(\frac{\lambda + Y}{\lambda}\right)^{-\theta} \tag{13}$$

Interest lies in efficient estimation of the marginal distribution parameters, allowing for strong dependence between the claim amount and ALAE.

Because Y_1 is censored, the log-likelihood contribution of censored observations is not $\ln \frac{\partial F(y_1)}{\partial y_1}$ but $\ln(Prob(Y_1 > y_1)) = \theta_1(\ln(\lambda_1) - \ln(\lambda_1 + y_1))$. Similarly when constructing the sieve approximation we use the joint likelihood that accounts for censoring of some observations on Y_1. This correction is given in Eq. (4.3) of Frees and Valdez (1998).

In the table below, we report our sieve ML (SML) estimates and standard errors along with the QMLE assuming independence between Y_1 and Y_2. For reference, we also include the full ML (FML) estimates and standard errors reported by Frees and Valdez (1998). The full MLE are estimates based on a fully parametric model, in which the Frank copula with dependence parameter α is used to construct a bivariate distribution with the Pareto marginals.

	QML Est. (St.Er.)	SML Est.	FML Est. (St.Er.)
λ_1	14,443.46	14,348.89	14,558
	(1,280.53)		(1,390)
θ_1	1.15	1.19	1.12
	(0.06)		(0.07)
λ_2	15,132.63	15,133.76	16,678
	(1,619.73)		(1,824)
θ_2	2.22	2.22	2.31
	(0.18)		(0.19)
α			-3.16
			(0.17)
LogL	-31950.80	-31698.93	

Our SML estimates are very close to the QMLE. Both QMLE and SMLE differ noticeably from the FML estimates, which is consistent only under correct specification of the bivariate distribution. No assumption on correct parametric specification of the joint distribution is involved in consistency arguments for the QMLE and SMLE. To obtain the SMLE, we used the cosine sieve with five elements in the sieve ($J_N = 5$). The choice is arbitrary. More work is needed on finding an optimal rule for this. Also, we do not yet have a stable estimate for the SMLE asymptotic variance.

5 Concluding Remarks

We have proposed an efficient semiparametric estimator of marginal distribution parameters. This is a sieve maximum likelihood estimator based on a finite-dimensional approximation of the unspecified part of the joint distribution. As such, the estimator inherits the costs and benefits of the multivariate sieve MLE. The major benefit permitted by sieve MLE is the increased asymptotic efficiency compared to quasi-MLE. We showed that the efficiency gains are non-trivial. In some simulations the relative efficiency with respect to QMLE was about 0.6 (a 40% improvement).

The gains come at an increased computational expense. The MLE convergence is slow for the traditional sieves we considered. We found that the Bernstein polynomial is preferred to other sieves in simulations. The running times are greater than QMLE or full MLE assuming a parametric copula family but they are still reasonable (at least for the two dimensional problems we consider). Moreover, simulations reveal a downward bias in SMLE, which seems to be caused by the sieve approximation error – it decreases as the number of sieve elements increases.

A simple alternative to the proposed method is a fully parametric ML estimation problem. Although simpler computationally, it imposes an assumption on the dependence structure, which, if violated, renders the ML estimates inconsistent. In this respect, the semiparametric approach is more robust but clearly no more efficient than any parametric alternative.

Important questions to address are the optimal choice of J_N, the choice of sieve and estimation of SMLE asymptotic variance.

References

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, vol. 6, 5549–5632.

Chen, X. and Y. Fan (2006a): "Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification," *Journal of Econometrics*, 135, 125–154.

——— (2006b): "Estimation of copula-based semiparametric time series models," *Journal of Econometrics*, 130, 307–335.

Chen, X., Y. Fan, and V. Tsyrennikov (2006): "Efficient Estimation of Semiparametric Multivariate Copula Models," *Journal of the American Statistical Association*, 101, 1228–1240.

Frees, E. and E. Valdez (1998): "Understanding relationships using copulas," *North American Actuarial Journal*, 2, 1–25.

Ghosal, S. (2001): "Convergence rates for density estimation with Bernstein polynomials," *Annals of Statistics*, 29, 1264–1280.

Joe, H. (2005): "Asymptotic efficiency of the two-stage estimation method for copula-based models," *Journal of Multivariate Analysis*, 94, 401–419.

Newey, W. and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. Engle and D. McFadden, vol. IV, 2113–2241.

Prokhorov, A. and P. Schmidt (2008): "Robustness, redundancy and validity of copulas in likelihood models," Working paper.

Sancetta, A. (2007): "Nonparametric estimation of distributions with given marginals via Bernstein-Kantorovich polynomials: L1 and pointwise convergence theory," *Journal of Multivariate Analysis*, 98, 1376–1390.

Sancetta, A. and S. Satchell (2004): "The Bernstein Copula And Its Applications To Modeling And Approximations Of Multivariate Distributions," *Econometric Theory*, 20, 535–562.

Segers, J., R. v. d. Akker, and B. Werker (2008): "Improving Upon the Marginal Empirical Distribution Functions when the Copula is Known," Discussion Paper 2008-40, Tilburg University, Center for Economic Research.

Severini, T. A. and W. H. Wong (1992): "Profile Likelihood and Conditionally Parametric Models," *The Annals of Statistics*, 20, 1768–1802.

Shen, X. (1997): "On Methods of Sieves and Penalization," *The Annals of Statistics*, 25, 2555–2591.

Sklar, A. (1959): "Fonctions de repartition a n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Universite de Paris*, 8, 229–231.