Statistical Inference for Multivariate Residual Copula of GARCH Models

Ngai-Hang Chan[1], Jian Chen[2], Xiaohong Chen[3],
Yanqin Fan[4] and Liang Peng[2]

Summary

Recently a flexible class of semiparametric copula-based multivariate GARCH models has been proposed to quantify multivariate risks, in which univariate GARCH models are employed to capture the dynamics of individual financial series, and parametric copulas are employed to model the contemporaneous dependence among GARCH residuals with nonparametric marginals. In this paper, we address two important questions regarding statistical inference for this class of models: (1) Under what moderately mild sufficient conditions can we justify the asymptotic distribution of the pseudo maximum likelihood estimator (MLE) of the residual copula parameter stated in Chen and Fan (2006a)? (2) How do we test the correct specification of a parametric copula for the GARCH residuals? In order to answer both questions rigorously, we establish a new weighted approximation for the empirical distributions of the GARCH residuals, which is of interest in its own right. Simulation studies and real data examples are provided to examine the finite sample performance of the pseudo MLE of the residual copula parameter and the proposed goodness-of-fit test.

KEY WORDS: Copula, GARCH, Goodness-of-fit test, Pseudo maximum likelihood estimation, Residual empirical distribution.

Running Title: Multivariate Residual Copula of GARCH Models

[1]Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. Email: nhchan@sta.cuhk.edu.hk

[2]School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA. Emails: jchen@math.gatech.edu, peng@math.gatech.edu

[3]Department of Economics, New York University, 269 Mercer Street, New York, NY 10003, USA. Email: xiaohong.chen@nyu.edu

[4]Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235, USA. Email: yanqin.fan@vanderbilt.edu

1 Introduction

On June 26, 2004, governors of the G-10 central banks endorsed the publication of the revised capital accord, known as Basel II, in which the Basel Committee requires banks to adopt a more holistic approach that focuses on the interaction between the different risk categories in risk management; see McNeil, Frey and Embrechts (2005) for a succinct account of the developments of Basel II. Consequently, banks now face the important problem of how to adequately model dependence between different risk factors. Since copulas capture dependence structures among individual risk factors that are invariant to any monotonic transformation of the individual risks, they are becoming standard tools in risk management. Similarly, in Solvency II, the international Actuarial Association recommends the use of copulas to capture dependence structure of insurance portfolios . Since then, major software providers have been building various copula models to serve the industrial needs. For details on copulas, we refer to Joe (1997) and Nelsen (2005). Because individual risk series in finance and insurance are typically serially dependent, Chen and Fan (2006a) introduced a class of Semiparametric COpula-based Multivariate DYnamic (SCOMDY) models, in which the conditional mean and conditional variance of individual risk series are parametrically specified, but the joint distribution of the (standardized) innovations is semiparametrically specified as a parametric copula evaluated at the nonparametric marginals. This class of models is very flexible in capturing a wide range of temporal and contemporaneous dependence structures of multivariate (nonlinear) time series.

An important class of the SCOMDY models is the so-called semiparametric copula-based multivariate GARCH models, where a scalar GARCH model is used to capture volatility of each individual risk series and a parametric copula is used to model the contemporaneous dependence between GARCH residuals. We now formally introduce this class of SCOMDY models. Suppose that the observations $\{Y_t = (Y_{1,t}, \dots, Y_{r,t})^T\}_{t=1}^n$ follow the model (1.1)–(1.2):

$$Y_{j,t} = \sqrt{h_{j,t}}\epsilon_{j,t}, \quad h_{j,t} = c_j + \sum_{i=1}^{p_j} \alpha_{j,i} Y_{j,t-i}^2 + \sum_{i=1}^{q_j} \beta_{j,i} h_{j,t-i}, \quad j = 1, \dots, r, \tag{1.1}$$

where $\{\epsilon_t = (\epsilon_{1,t}, \dots, \epsilon_{r,t})^T\}_{t=1}^n$ is a sequence of i.i.d. random vectors with $E[\epsilon_{j,t}] =$

$0, E[(\epsilon_{j,t})^2] = 1$, and the joint distribution function F_ϵ of ϵ_t is assumed to take the semiparametric form:

$$F_\epsilon(\epsilon_1, \ldots, \epsilon_r) = C(F_{\epsilon,1}(\epsilon_1), \ldots, F_{\epsilon,r}(\epsilon_r); \theta_0), \tag{1.2}$$

where $C(x_1, \ldots, x_r; \theta)$ is a parametrized copula function up to unknown $\theta \in \Theta \subset R^m$, and for $j = 1, \ldots, r$, $F_{\epsilon,j}$ is the marginal distribution function of $\epsilon_{j,t}$, which is assumed to be continuous but otherwise unspecified. By Sklar's Theorem (see Nelsen (2005)), any multivariate distribution with continuous marginals can be uniquely represented by its copula function evaluated at its marginals. Let C_ϵ denote the unique copula corresponding to the true joint distribution F_ϵ of the GARCH residual vector ϵ_t. We call C_ϵ the residual copula, which is defined as $C_\epsilon(x_1, \ldots, x_r) = F_\epsilon(F_{\epsilon,1}^{-}(x_1), \ldots, F_{\epsilon,r}^{-}(x_r))$, where $F_{\epsilon,j}^{-}(\cdot)$ is the generalized inverse of $F_{\epsilon,j}(\cdot)$, $j = 1, \ldots, r$. Model (1.2) is tantamount to assuming that the true residual copula belongs to a parametric family: $C_\epsilon(x_1, \ldots, x_r) = C(x_1, \ldots, x_r; \theta_0)$ for some unknown $\theta_0 \in \Theta \subset R^m$.

This simple multivariate GARCH model bypasses the overparametrization issue that is encountered in many multivariate GARCH models; see Bauwens, Laurent and Rombouts (2006) for a survey on multivariate GARCH models. Although simple, the model (1.1)-(1.2) is still flexible enough to capture both dynamic and concurrent movements of multivariate risk factors, and is especially useful for modelling portfolio risks such as conditional VaR. See, e.g., Hull and White (1998) and Breymann, Dias and Embrechts (2003) for applications to exchange rate data, and Giacomini, Haerdle, Ignatieva and Spokoiny (2006) to stock data. However, before banks and insurance companies can readily apply this class of semiparametric copula-based multivariate GARCH models (1.1)-(1.2), one has to develop valid statistical inference methodologies for these models. In particular, since different parametric residual copula specifications generally imply different dependence structures among multiple risk factors, estimation of the residual copula parameter and tests for the correct specification of the parametric residual copula are of great importance. In this paper we shall address both issues.

Estimation and inference for copulas that directly couple multivariate observed variables have been pursued extensively. For example, in the context of nonparametric copulas, Fermanian, Radulovic and Wegkamp (2004) considered empirical copula estimation, while Fermanian and Scaillet (2003) and Chen and Huang (2006) proposed kernel smoothing. For parametric copulas coupled with nonparametric marginals, Genest, Ghoudi and Rivest (1995) investigated pseudo maximum likelihood estimation (MLE), while Chen, Fan and Tsyrennikov (2006) considered sieve MLE. Chen and Fan (2006b) studied the pseudo MLE and its property in estimating copulas that generate nonlinear Markov models. For i.i.d. data, Klugman and Parsa (1999), Fermanian (2005), Scaillet (2006), and Genest, Quessy and Remillard (2006) examined goodness-of-fit tests of parametric copulas. Chen and Fan (2005) developed model selection tests for multiple parametric copula comparison.

The main technical difficulty in establishing the asymptotic distribution of the pseudo MLE of the copula parameter is that the score function and its derivatives in copula-based models could blow up to infinity near the boundaries. Chen and Fan (2005, 2006b) overcome this difficulty by making use of the weak convergence of the empirical distribution function in a weighted metric. Although Chen and Fan (2006a) developed copula model selection tests for SCOMDY models, their tests rely on the asymptotic property of the pseudo MLE $\widehat{\theta}$ (see Section 2 for its definition) of the residual copula parameter θ. Crucial to the validity of their model selection tests is the surprising result that the asymptotic distribution of $\widehat{\theta}$ is not affected by the initial step estimation of the GARCH parameters. Chen and Fan (2006a) established this result by means of a heuristic argument with stringent conditions and by assuming the validity of a weighted approximation for the empirical distributions of GARCH residuals. Currently, there are sporadic results on the convergence of empirical distributions using residuals of non-linear time series; see, for example, Berkes and Horváth (2003), Horváth, Kokoszka and Teyssiére (2001) and Koul and Ling (2006). However, to the best of our knowledge, a weighted approximation result is not available for empirical distributions of residuals obtained from an initial step estimation of time series models.

In this paper, we first establish a weighted approximation for the empirical distributions of residuals of univariate GARCH models, which is important in its own right. This weighted approximation allows us to provide a rigorous justification of the limit distribution result for the pseudo MLE $\widehat{\theta}$ under moderately mild sufficient conditions; see Section 2. In addition, we develop a consistent test for the correct specification of the residual copula $C_\epsilon(x_1, \ldots, x_r)$ by a particular parametric copula class $\mathcal{C} = \{C(x_1, \ldots, x_r; \theta) : \theta \in \Theta\}$. This extends existing goodness-of-fit tests for i.i.d. data to GARCH residuals. In Section 3, we provide some simulation studies and real data examples to demonstrate finite sample properties of the pseudo MLE for θ and the goodness-of-fit test for the parametric copula. Proofs are given in the on-line version of this paper.

2 Estimation and Testing

2.1 Estimation of GARCH models

For each $j = 1, \ldots, r$, let $\gamma_j = (c_j, \alpha_{j,1}, \ldots, \alpha_{j,p_j}, \beta_{j,1}, \ldots, \beta_{j,q_j})^T$ be the true GARCH parameters associated with the model (1.1). Let $\hat{\gamma}_j = (\hat{c}_j, \hat{\alpha}_{j,1}, \ldots, \hat{\alpha}_{j,p_j}, \hat{\beta}_{j,1}, \ldots, \hat{\beta}_{j,q_j})^T$ denote the pseudo MLE of γ_j based on the sample $\{Y_{j,t}\}_{t=1}^n$, which is the MLE if $\epsilon_{j,t}$ is standard normal.

Similar to Berkes and Horváth (2003), for $q_j \geq p_j$, define

$$\begin{cases} d_{j,0}(\gamma_j) = c_j/(1 - \beta_{j,1} - \cdots - \beta_{j,q_j}) \\ d_{j,1}(\gamma_j) = \alpha_{j,1} \\ d_{j,2}(\gamma_j) = \alpha_{j,2} + \beta_{j,1} d_{j,1}(\gamma_j) \\ \cdots \\ d_{j,p_j}(\gamma_j) = \alpha_{j,p_j} + \beta_{j,1} d_{j,p_j-1}(\gamma_j) + \cdots + \beta_{j,p_j-1} d_{j,1}(\gamma_j) \\ d_{j,p_j+1}(\gamma_j) = \beta_{j,1} d_{j,p_j}(\gamma_j) + \cdots + \beta_{j,p_j} d_{j,1}(\gamma_j) \\ \cdots \\ d_{j,q_j}(\gamma_j) = \beta_{j,1} d_{j,q_j-1}(\gamma_j) + \cdots + \beta_{j,q_j-1} d_{j,1}(\gamma_j); \end{cases}$$

for $q_j < p_j$, define

$$\begin{cases} d_{j,0}(\gamma_j) = c_j/(1-\beta_{j,1}-\cdots-\beta_{j,q_j}) \\ d_{j,1}(\gamma_j) = \alpha_{j,1} \\ d_{j,2}(\gamma_j) = \alpha_{j,2} + \beta_{j,1} d_{j,1}(\gamma_j) \\ \cdots \\ d_{j,q_j+1}(\gamma_j) = \alpha_{j,q_j+1} + \beta_{j,1} d_{j,q_j}(\gamma_j) + \cdots + \beta_{j,q_j} d_{j,1}(\gamma_j) \\ \cdots \\ d_{j,p_j}(\gamma_j) = \alpha_{j,p_j} + \beta_{j,1} d_{j,p_j-1}(\gamma_j) + \cdots + \beta_{j,q_j} d_{j,p_j-q_j}(\gamma); \end{cases}$$

for $i > \max(p_j, q_j)$, define

$$d_{j,i}(\gamma_j) = \beta_{j,1} d_{j,i-1}(\gamma_j) + \beta_{j,2} d_{j,i-2}(\gamma_j) + \cdots + \beta_{j,q_j} d_{j,i-q_j}(\gamma_j).$$

Set $w_{j,k}(\gamma_j) = d_{j,0}(\gamma_j) + \sum_{i=1}^{\infty} d_{j,i}(\gamma_j) Y_{j,k-i}^2$ and

$$\begin{aligned} \Gamma_j \quad &= \{u = (u_1, \ldots, u_{p_j+q_j+1})^T : u > 0, u_{p_j+2} + \cdots + u_{p_j+q_j+1} \le \Delta_0^* < 1, \\ & 0 < \Delta_1^* \le \min(u_1, \ldots, u_{p_j+q_j+1}) \le \max(u_1, \ldots, u_{p_j+q_j+1}) \le \Delta_2^*, q_j \Delta_1^* < \Delta_0^*\}. \end{aligned}$$

Remark 1. When $E\epsilon_{j,1}^4 < \infty$ and $\gamma_j \in \Gamma_j$, it follows from Berkes and Horváth (2003, equations 1.8 and 3.4) that

$$\begin{cases} h_{j,k} = w_{j,k}(\gamma_j) \\ \hat{\gamma}_j - \gamma_j = \frac{1}{n}\sum_{t=1}^{n} (\epsilon_{j,t}^2 - 1) A_j^{-1} \frac{w'_{j,t}(\gamma_j)}{w_{j,t}(\gamma_j)} + o_p(n^{-1/2}), \end{cases} \tag{2.1}$$

where $A_j = E\{\frac{w'_{j,1}(\gamma_j)}{w_{j,1}(\gamma_j)} (\frac{w'_{j,1}(\gamma_j)}{w_{j,1}(\gamma_j)})^T\}$ and

$$w'_{j,t}(\gamma_j) = (\frac{\partial}{\partial c_j} w_{j,t}(\gamma_j), \frac{\partial}{\partial \alpha_{j,1}} w_{j,t}(\gamma_j), \ldots, \frac{\partial}{\partial \beta_{j,q_j}} w_{j,t}(\gamma_j))^T.$$

Put $\hat{w}_{j,1}(\gamma_j) = 1$ and $\hat{w}_{j,k}(\gamma_j) = d_{j,0}(\gamma_j) + \sum_{i=1}^{k-1} d_{j,i}(\gamma_j) Y_{j,k-i}^2$ for $2 \le k \le n$. Then we can estimate ϵ_t by

$$\hat{\epsilon}_t = (\hat{\epsilon}_{1,t}, \ldots, \hat{\epsilon}_{r,t})^T = (\frac{Y_{1,t}}{\sqrt{\hat{w}_{1,t}(\hat{\gamma}_j)}}, \ldots, \frac{Y_{r,t}}{\sqrt{\hat{w}_{r,t}(\hat{\gamma}_j)}})^T. \tag{2.2}$$

2.2 Estimation of residual copula parameters

We estimate the true marginal distribution of $\epsilon_{j,t}$, $F_{\epsilon,j}(x)$, by

$$\hat{F}_{\epsilon,j}(x) = \frac{1}{n-\nu+1}\sum_{t=\nu}^{n} I(\hat{\epsilon}_{j,t} \le x),$$

where $\nu = \nu(n)$ is an integer. We then estimate the residual copula parameter θ by $\hat{\theta}$, the pseudo MLE based on the pseudo sample

$$\{(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}), \ldots, \hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t}))^T\}_{t=\nu}^{n}, \tag{2.3}$$

i.e.,

$$\begin{aligned} \hat{\theta} &= \arg\max_\theta \tfrac{1}{n-\nu+1} \textstyle\sum_{t=\nu}^{n} \log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}), \ldots, \hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t}); \theta) \\ &:= \arg\max_\theta l_n(\theta), \end{aligned}$$

where $c(x_1, \ldots, x_r; \theta) = \frac{\partial^r}{\partial x_1 \ldots \partial x_r} C(x_1, \ldots, x_r; \theta)$ is the copula density function. This estimation approach was employed by Genest, Ghoudi and Rivest (1995) for independent data and by Chen and Fan (2006a) for dependent data.

2.3 Weighted approximation for residual empirical distributions

Let U be a Gaussian process with

$$EU(x) = 0, \quad E\{U(x)U(y)\} = \prod_{i=1}^{r}\{x_i \wedge y_i\} - \prod_{i=1}^{r}\{x_i y_i\},$$

where $x = (x_1, \ldots, x_r)^T$ and $y = (y_1, \ldots, y_r)^T$.

The following Conditions 1 and 2 are imposed for the study of empirical process and weighted empirical process of the estimated residuals of GARCH models.

1. For $j = 1, \ldots, r$, $\gamma_j \in \Gamma_j$, $E\epsilon_{j,1}^4 < \infty$, and there exists $\mu > 0$ such that $\lim_{t\to 0} t^{-\mu} P(\epsilon_{j,1}^2 \le t) = 0$;

2. For $j=1,\ldots,r$, the support of $\epsilon_{j,t}$ is $(-\infty,\infty)$, $F_{\epsilon,j}$ has continuous density $F'_{\epsilon,j}$, and there exist $\beta_3\in(0,1/4)$ and $\Delta_3>0$ such that

$$\sup_s \sup_{|x-1|\le\Delta_3} \frac{sF'_{\epsilon,j}(sx)}{\{F_{\epsilon,j}(s)(1-F_{\epsilon,j}(s))\}^{\beta_3}}<\infty$$

for $j=1,\ldots,r$.

THEOREM **2.1.** *Suppose that Conditions 1–2 hold and $\nu/\log n\to\infty$, $\nu/n\to 0$ as $n\to\infty$. Then*

$$\sup_x \frac{|\sqrt{n-\nu+1}\{\hat F_{\epsilon,j}(x)-F_{\epsilon,j}(x)\}-U((1,\ldots,1,F_{\epsilon,j}(x),1,\ldots,1)^T)-\frac{1}{2}xF'_{\epsilon,j}(x)\tau_j|}{\{F_{\epsilon,j}(x)(1-F_{\epsilon,j}(x))\}^{\beta_3}} = o_p(1), \tag{2.4}$$

where $(U(T_r(x_1,\ldots,x_r)),\tau_1,\ldots,\tau_r)^T$ is a vector valued Gaussian process with zero mean and covariance structure

$$\begin{aligned} & E(\tau_j\tau_i) \\ = \;& E\{(\epsilon^2_{j,1}-1)(\epsilon^2_{i,1}-1)\}E\{(\tfrac{w'_{j,1}(\gamma_j)}{w_{j,1}(\gamma_j)})^T A_j^{-1}E(\tfrac{w'_{j,1}(\gamma_j)}{w_{j,1}(\gamma_j)})(\tfrac{w'_{i,1}(\gamma_i)}{w_{i,1}(\gamma_i)})^T A_i^{-1}E(\tfrac{w'_{i,1}(\gamma_i)}{w_{i,1}(\gamma_i)})\},\end{aligned} \tag{2.5}$$

$$\begin{aligned} & E\{U(T_r(x_1,\ldots,x_r))\tau_j\} \\ = \;& E\{(\epsilon^2_{j,1}-1)I(F_1(\epsilon_{1,1})\le x_1,\ldots,F_r(\epsilon_{r,1})\le x_r)\}E\{(\tfrac{w'_{j,1}(\gamma_j)}{w_{j,1}(\gamma_j)})^T\}A_j^{-1}E\{\tfrac{w'_{j,t}(\gamma_j)}{w_{j,t}(\gamma_j)}\},\end{aligned} \tag{2.6}$$

and $T_r(x_1,\ldots,x_r)$ is defined immediately after Theorem 2.2 below.

Remark 2. Theorem 2.1 establishes a weighted approximation for the empirical distributions of the residuals $(\epsilon_{j,t})$ in GARCH models. Later on, we shall use Theorem 2.1 to derive the asymptotic distributions of the pseudo MLE of the residual copula parameter θ and of the goodness-of-fit test statistic for testing the parametric specification of the residual copula. For (unweighted) approximation to empirical process of squared residuals $(\epsilon^2_{j,t})$ in ARCH and GARCH models, we refer to Horváth, Kokoszka and Teyssiére (2001) and Berkes and Horváth (2003).

2.4 Asymptotic properties of pseudo MLE of θ

Let θ_0 denote the true value of θ and assume the following consistency conditions of $\hat\theta$.

Conditions C1–C4:

1. $\log c(x_1, \ldots, x_r; \theta)$ is a continuous function of θ for each $(x_1, \ldots, x_r)^T \in [0,1]^r$;

2. Θ is a compact subset of R^m;

3. $E \sup_{\theta \in \Theta} |\log c(F_{\epsilon,1}(\epsilon_{1,1}), \ldots, F_{\epsilon,r}(\epsilon_{r,1}); \theta)| < \infty$;

4. For any $\Delta_0 \in (0, 1/2)$ and $\Delta_1 \in (1/2, 1)$, there exist $\beta_0 \in (0,1)$, $M_0 > 0$, $\beta_1 > 0$ and $M_1 > 0$ such that

$$\sup_{\theta \in \Theta} |\log c(x_1, \ldots, x_r; \theta)| \leq M_0 \{\wedge_{i=1}^r x_i\}^{-\beta_0}$$

for $\wedge_{i=1}^r x_i \leq \Delta_0$,

$$\sup_{\theta \in \Theta} |\log c(x_1, \ldots, x_r; \theta)| \leq M_0 \{1 - \vee_{i=1}^r x_i\}^{-\beta_0}$$

for $\vee_{i=1}^r x_i \geq \Delta_1$ and

$$\sup_{\theta \in \Theta} |\log c(x_1, \ldots, x_r; \theta) - \log c(y_1, \ldots, y_r; \theta)| < M_1 \sum_{i=1}^r |x_i - y_i|^{\beta_1}$$

for $\frac{\Delta_0}{2} < \wedge_{i=1}^r x_i \leq \vee_{i=1}^r x_i < \Delta_1 + \frac{1-\Delta_1}{2}$ and $\frac{\Delta_0}{2} < \wedge_{i=1}^r y_i \leq \vee_{i=1}^r y_i < \Delta_1 + \frac{1-\Delta_1}{2}$.

Note that Conditions C1–C3 are standard conditions for consistency of MLE based on i.i.d. data. Condition C4 is similar to that imposed by Genest, Ghoudi and Rivest (1995) and Chen and Fan (2005) for i.i.d. data; it controls the speed of divergence of logarithm of the copula density at the boundaries, and is satisfied by all the commonly used copula densities.

Theorem **2.2.** *Suppose that Conditions 1 and C1–C4 hold, and $\nu(n)/n \to 0$ as $n \to \infty$. Then $\hat{\theta} \xrightarrow{p} \theta_0$ as $n \to \infty$.*

Remark 3. Since the approximation rate between the estimated residuals $\hat{\epsilon}_{j,t}$ and the residuals $\epsilon_{j,t}$ is poor for small t, we employ ν in the preceding Theorem to obtain a better approximation rate. The same idea was employed in Hall and Yao (2003) for deriving the limit distribution of pseudo-MLE for GARCH models.

Before we state the asymptotic normality result, we introduce some additional notations. Put

$$\dot{c}(x_1,\ldots,x_r;\theta) = (\frac{\partial}{\partial\theta_1}c(x_1,\ldots,x_r;\theta),\ldots,\frac{\partial}{\partial\theta_m}c(x_1,\ldots,x_r;\theta))^T,$$

$$\delta(x_1,\ldots,x_r;\theta) = \frac{\dot{c}(x_1,\ldots,x_r;\theta)}{c(x_1,\ldots,x_r;\theta)},$$

and for $i = 1,\ldots,r$,

$$\delta_i(x_1,\ldots,x_r;\theta) = \frac{\partial}{\partial x_i}\delta(x_1,\ldots,x_r;\theta).$$

Define

$$C_1(x_1) = x_1,$$

$$C_i(x_i|x_1,\ldots,x_{i-1}) = P(F_{\epsilon,i}(\epsilon_{i,1}) \le x_i | F_{\epsilon,1}(\epsilon_{1,1}) = x_1,\ldots,F_{\epsilon,i-1}(\epsilon_{i-1,1}) = x_{i-1})$$

for $i = 2,\ldots,r$,

$$T_i(x_1,\ldots,x_i) = (C_1(x_1), C_2(x_2|x_1),\ldots,C_i(x_i|x_1,\ldots,x_{i-1}))$$

for $i = 1,\ldots,r$, and

$$\Sigma(\theta) = (E\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log c(F_{\epsilon,1}(\epsilon_{1,1}),\ldots,F_{\epsilon,r}(\epsilon_{r,1});\theta)\})_{1\le i,j\le m}.$$

We impose the following additional regularity conditions for asymptotic normality.

Conditions N1–N6:

1. For $j = 2,\ldots,r$, the function $C_j(x_j|T_{j-1}^{-}(x_1,\ldots,x_{j-1}))$ is differentiable with respect to $x_1,\ldots,x_{j-1}$ over the interior of $[0,1]^{j-1}$ and

$$\sum_{i=1}^{j-1}\int_{[0,1]^{j-1}} |\frac{\partial}{\partial x_i}C_j(x_j|T_{j-1}^{-}(x_1,\ldots,x_{j-1}))|\,dx_1\cdots dx_{j-1} \le M_2 \in (0,\infty);$$

2. There exists $\beta_2 \in (0,1/2)$ such that

$$\sup_{0\le x_1,\ldots,x_r\le 1}\prod_{i=1}^{r}(x_i)^{\beta_2}(1-x_i)^{\beta_2}|\delta(T_r^{-}(x_1,\ldots,x_r);\theta_0)| < \infty,$$

and

$$\int \prod_{i=1}^{r} (x_i)^{\beta_2}(1-x_i)^{\beta_2} |\, d\delta(T_r^{-}(x_1,\ldots,x_r;\theta_0))| < \infty;$$

3. For any $\Delta_4 \in (0,1/2)$ and $\Delta_5 \in (1/2,1)$, there exist $\beta_4 \in (0,\beta_2)$, $M_4 > 0$, $\beta_5 > 0$ and $M_5 > 0$ such that

$$|\delta_j(x_1,\ldots,x_r;\theta_0)| \le M_4 x_j \{\wedge_{i=1}^{r} x_i\}^{-\beta_4}$$

for $\wedge_{i=1}^{r} x_i \le \Delta_4$,

$$|\delta_j(x_1,\ldots,x_r;\theta_0)| \le M_4 (1-x_j)\{1-\vee_{i=1}^{r} x_i\}^{-\beta_4}$$

for $\vee_{i=1}^{r} x_i \ge \Delta_5$ and

$$|\delta(x_1,\ldots,x_r;\theta_0) - \delta(y_1,\ldots,y_r;\theta_0)| < M_5 \sum_{i=1}^{r} |x_i - y_i|^{\beta_5}$$

for $\frac{\Delta_4}{2} < \wedge_{i=1}^{r} x_i \le \vee_{i=1}^{r} x_i < \Delta_5 + \frac{1-\Delta_5}{2}$ and $\frac{\Delta_4}{2} < \wedge_{i=1}^{r} y_i \le \vee_{i=1}^{r} y_i < \Delta_5 + \frac{1-\Delta_5}{2}$;

4. For $1 \le i,j \le m$, $\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log c(x_1,\ldots,x_r;\theta)$ is a continuous function of θ in an open neighborhood of θ_0 for each $(x_1,\ldots,x_r)^T \in [0,1]^r$;

5. There exists an open neighborhood Θ_0 of θ_0 such that

$$E \sup_{\theta\in\Theta_0} |\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log c(F_{\epsilon,1}(\epsilon_{1,1}),\ldots,F_{\epsilon,r}(\epsilon_{r,1});\theta)| < \infty$$

for $1 \le i,j \le m$;

6. For any $\Delta_6 \in (0,1/2)$ and $\Delta_7 \in (1/2,1)$, there exist $\beta_6 \in (0,1)$, $M_6 > 0$, $\beta_7 > 0$ and $M_7 > 0$ such that

$$\sup_{\theta\in\Theta_0} |\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log c(x_1,\ldots,x_r;\theta)| \le M_6 \{\wedge_{i=1}^{r} x_i\}^{-\beta_6}$$

for $\wedge_{i=1}^{r} x_i \le \Delta_6$,

$$\sup_{\theta\in\Theta_0} |\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log c(x_1,\ldots,x_r;\theta)| \le M_6 \{1-\vee_{i=1}^{r} x_i\}^{-\beta_6}$$

for $\vee_{i=1}^{r} x_i \ge \Delta_7$ and

$$\sup_{\theta\in\Theta_0} |\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log c(x_1,\ldots,x_r;\theta) - \frac{\partial^2}{\partial\theta_i \partial\theta_j} \log c(y_1,\ldots,y_r;\theta)| < M_7 \sum_{i=1}^{r} |x_i - y_i|^{\beta_7}$$

for $\frac{\Delta_6}{2} < \wedge_{i=1}^{r} x_i \le \vee_{i=1}^{r} x_i < \Delta_7 + \frac{1-\Delta_7}{2}$ and $\frac{\Delta_6}{2} < \wedge_{i=1}^{r} y_i \le \vee_{i=1}^{r} y_i < \Delta_7 + \frac{1-\Delta_7}{2}$.

Note that conditions N4-N5 are standard for proving asymptotic normality of MLE based on i.i.d. data. Conditions N1-N2 are imposed by Csörgő and Révész (1975) for multivariate empirical processes. Conditions N3 and N6 are similar to the ones imposed by Genest, Ghoudi and Rivest (1995) and Chen and Fan (2005) for asymptotic normality based on i.i.d. data; they are employed to control the speed of divergence of partial derivatives of the logarithm of the copula density at the boundaries, and are again satisfied by all the commonly used copula densities.

THEOREM **2.3.** *Suppose that Conditions 1-2, C1–C4 and N1–N6 hold, and*

$$\nu(n)/\log n \to \infty, \quad \nu(n)/n \to 0 \quad \textit{as} \quad n \to \infty.$$

Then

$$\begin{aligned}
& \sqrt{n}(\hat{\theta} - \theta_0) \\
\stackrel{d}{\to} \; & -\Sigma^{-1}(\theta_0)\{\textstyle\int \delta(T_r^-(x_1, \ldots, x_r); \theta_0)\, dU(x_1, \ldots, x_r) \\
& + \textstyle\sum_{i=1}^r \int \delta_i(x_1, \ldots, x_r; \theta_0) U((1, \ldots, 1, x_i, 1, \ldots, 1)^T) c(x_1, \ldots, x_r; \theta_0)\, dx_1 \ldots dx_r \\
= \; & Z.
\end{aligned}$$

Remark 4. Theorem 2.3 states that under moderately mild sufficient conditions, the limit distribution of $\hat{\theta}$ is independent of the GARCH filtering. In Chen and Fan (2006a), they obtained the normal limit distribution by means of heuristic arguments under stringent conditions and by assuming the existence of the weighted approximation for the empirical distributions of GARCH residuals. Since the variance given in Chen and Fan (2006a) is expressed in terms of a conditional distribution, it seems difficult to check if the limit in Theorem 2 equals to that of Chen and Fan (2006a). It is known, however, that both limits are normal distribution and are independent of the parameter estimation in GARCH models. Owing to this independence of GARCH models, we can therefore employ a parametric bootstrap method to estimate he variance of $\hat{\theta}$ to construct confidence intervals for θ_0. In the simulation study below, we illustrate the independence property of GARCH models and examine the accuracy of the parametric bootstrap method for constructing confidence intervals.

2.5 A goodness-of-fit test of residual copulas

The results established in the preceding subsection assume the correct specification of the residual copula by the parametric copula class $\mathcal{C} = \{C(x_1, \ldots, x_r; \theta) : \theta \in \Theta\}$. In this subsection, we propose a consistent test for this assumption. Let

$$H_0 : P\left(C_\epsilon(\epsilon_1, \ldots, \epsilon_r) = C(\epsilon_1, \ldots, \epsilon_r; \theta_0)\right) = 1 \text{ for some } \theta_0 \in \Theta$$

and

$$H_1 : P\left(C_\epsilon(\epsilon_1, \ldots, \epsilon_r) = C(\epsilon_1, \ldots, \epsilon_r; \theta)\right) < 1 \text{ for all } \theta \in \Theta.$$

Define the empirical estimator of $C_\epsilon(x_1, \ldots, x_r)$ as

$$\hat{C}_\epsilon(x_1, \ldots, x_r) = \frac{1}{n} \sum_{t=1}^{n} I(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}) \le x_1, \ldots, \hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t}) \le x_r).$$

Then our test statistic is

$$T_n = \int \{\hat{C}_\epsilon(x_1, \ldots, x_r) - C(x_1, \ldots, x_r; \hat{\theta})\}^2 c(x_1, \ldots, x_r; \hat{\theta})\, dx_1 \ldots dx_r,$$

where $\hat{\theta}$ is the pseudo MLE of θ_0 under H_0. Denote

$$\dot{C}(x_1, \ldots, x_r; \theta) = (\frac{\partial}{\partial \theta_1} C(x_1, \ldots, x_r; \theta), \ldots, \frac{\partial}{\partial \theta_m} C(x_1, \ldots, x_r; \theta))^T,$$

THEOREM **2.4.** *Assume conditions of Theorem 2.3 hold. Further, suppose*

$$\begin{cases} \max_{1 \le i \le r} \sup_{0 \le x \le 1} F'_{\epsilon,i}(F^-_{\epsilon,i}(x)) F^-_{\epsilon,i}(x) < \infty \\ \sup_{\theta \in \Theta_0} \sup_{0 \le x_1, \ldots, x_r \le 1} |\dot{C}(x_1, \ldots, x_r; \theta)| < \infty \\ \sup_{0 \le x_1, \ldots, x_r \le 1} \sum_{i=1}^r \frac{\partial}{\partial x_i} C(x_1, \ldots, x_r; \theta_0) < \infty. \end{cases} \tag{2.7}$$

Then under H_0,

$$\begin{aligned} nT_n \xrightarrow{d} & \int \{U(T_r(x_1, \ldots, x_r)) + \textstyle\sum_{j=1}^r \frac{\partial}{\partial x_j} C(x_1, \ldots, x_r; \theta_0) \times \\ & \times U((1, \ldots, 1, x_j, 1, \ldots, 1)^T) - Z^T \dot{C}(x_1, \ldots, x_r; \theta_0)\}^2 c(x_1, \ldots, x_r; \theta_0)\, dx_1 \ldots dx_r, \end{aligned}$$

where Z *is given in Theorem 2.3.*

Remark 5. As in the estimation (see Remark 4), the asymptotic distribution of the test statistic under H_0 is also independent of GARCH filtering. This motivates us to employ a parametric bootstrap method to obtain the critical point of the test instead of simulating one from the limit distribution. The details are given in the real data examples below.

3 Simulations and Real Data Examples

We generate 1,000 random samples each with a sample size $n = 500$ from model (1.1) with residual copula specified as the mixture copula

$$C(u_1, \ldots, u_r; \theta_1, \theta_2, \lambda) = \lambda C_1(u_1, \ldots, u_r; \theta_1) + (1-\lambda) C_2(u_1, \ldots, u_r; \theta_2),$$

where

$$C_1(u_1, \ldots, u_r; \theta_1) = \{\sum_{i=1}^{r} u_i^{-\theta_1} - r + 1\}^{-1/\theta_1}, \theta_1 > 0$$

and

$$C_2(u_1, \ldots, u_r; \theta_2) = \exp\{-[\sum_{i=1}^{r} (-\log(u_i))^{\theta_2}]^{1/\theta_2}\}, \theta_2 \geq 1.$$

To demonstrate the property that the pseudo MLE is independent of the parameter estimation in GARCH models, we consider $\theta_1 = 3.0$, $\theta_2 = 2.0$, $\lambda = 0.3$ or 0.7, the marginal distributions of ϵ_t are $N(0,1)$, $r = 3$ and GARCH model as either $c_1 = c_2 = c_3 = 1$, $\alpha_{1,1} = 0.2$, $\alpha_{2,1} = 0.3$, $\alpha_{3,1} = 0.4$, $\beta_{1,1} = \beta_{2,1} = \beta_{3,1} = 0.2$ (case (i)) or $c_1 = c_2 = c_3 = 0.2$, $\alpha_{1,1} = \alpha_{2,1} = \alpha_{3,1} = 0.2$, $\beta_{1,1} = 0.6$, $\beta_{2,1} = 0.5$, $\beta_{3,1} = 0.4$ (case (ii)). The average and the corresponding standard deviation of the proposed pseudo maximum likelihood estimator are reported in Table 1, along with the true values of the model parameters. This table shows that the proposed method works reasonably well and the proposed pseudo maximum likelihood estimator is independent of the GARCH filtering, as indicated by Theorem 2.3. We also observe that the standard deviation of $\hat{\theta}_1$ for the case $\lambda = 0.3$ is larger than that for the case $\lambda = 0.7$, and the standard deviation of $\hat{\theta}_2$ for the case $\lambda = 0.3$ is smaller than that for the case $\lambda = 0.7$. This observation is in line with the role of the parameter λ in the mixture copula.

For constructing confidence intervals, we draw 400 random samples each with size $n = 500$ from $C(x_1, x_2, x_3; \hat{\theta}_1, \hat{\theta}_2, \hat{\lambda})$ for each random sample. And for each bootstrap sample, we compute the bootstrap version of the MLE, say $\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\lambda}^*$. Use these 400 bootstrap MLE's, we then estimate the variance of $\hat{\theta}_1, \hat{\theta}_2, \hat{\lambda}$ so that confidence intervals can be obtained. Based on 1,000 random samples, the coverage probabilities for $\theta_1, \theta_2, \lambda$ with level 0.9 are 0.903, 0.928, 0.899 for case (i) with $\lambda = 0.3$; 0.900, 0.926, 0.893 for case (i) with $\lambda = 0.7$; 0.908, 0.909, 0.890 for case (ii) with $\lambda = 0.3$;

0.898, 0.920, 0.903 for case (ii) with $\lambda = 0.7$. These numbers show that the proposed parametric bootstrap method works well, i.e., the simulation study further confirms the property of independence of GARCH Filtering given in the Theorem 2.3.

Next we apply the proposed estimate and test to two real data sets. The first one includes 2,275 daily log-returns of S&P 500 index, Cisco System and Intel Corporation from January 2, 1991 to December 31, 1999; see Figure 1 and Fan, Wang and Yao (2006). The second data set contains 2,635 daily log-returns of stock prices of Nortel, Lucent and Cisco from April 4, 1996 to September 22, 2006; see Figure 2. Here, we fit the mixture copula $C(x_1, x_2, x_3; \theta_1, \theta_2, \lambda)$ to the residuals from filtering a GARCH(1,1) for each series; see Tables 2 – 5 for parameter estimates. As mentioned in Remark 3, we employ the following parametric bootstrap method to obtain p-values of the test.

We draw 200 random samples with size 2,275 for the first data set and size 2,635 for the second data set from $C(x_1, x_2, x_3; \hat{\theta}_1, \hat{\theta}_2, \hat{\lambda})$. Based on each sample, we compute the bootstrap version of T_n, say T_n^*. Hence we have $T_n^*(1), \ldots, T_n^*(200)$, and the p-value is calculated as $\frac{1}{200}\sum_{i=1}^{200} I(T_n^*(i) \geq T_n)$, see Tables 3 and 5. The p-values in Tables 3 and 5 clearly reject the mixture copula for both data sets. Since this mixture copula is mainly designed to capture the two tails of a data set, it is less effective in capturing both the tail part as well as the middle part of a data set. Therefore, seeking a more flexible parametric residual copula model to capture the whole span of the data set constitutes the next challenging task.

Acknowledgments. We would like to thank the Co-editors, an associate editor and a reviewer for their helpful comments. Chan and Peng's research was supported in part by HKSAR-RGC Earmarked Grants 400305, 400306 and NSF Grant SES-0631608. Chen and Fan's research was supported in part by NSF Grant SES-0318091 and SES-0317984 respectively.

References

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press. Harvard, Boston.

Berkes, I. and Horváth, L. (2003). Limit r esults for the empirical process of squared residuals in GARCH models. *Stochastic Processes and their Applications* **105**, 271–298.

Bauwens, J., Laurent, S. and Rombouts, J.V.K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* **21**, 79–110.

Breymann, W., Dias, A. and Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in Finance. *Quantitative Finance* **3**, 1–16.

Chen, S. and Huang, T. (2006). Nonparametric estimation of copula functions for dependence modeling. *Technical Report.* Department of Statistics, Iowa State University, Ames.

Chen, X. and Fan, Y. (2005). Pseudo-likelihood ratio tests for model selection in semiparametric multivariate copula models. *Canadian Journal of Statistics* **33**, 389–414.

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

Chen, X. and Fan, Y. (2006b). Semiparametric estimation of copula-based time series models. *Journal of Econometrics* **130**, 307–335.

Bivariate Failure-Time Data. Forthcoming in *Econometric Theory* ???.

Chen, X., Fan, Y. and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *J. Amer. Statist. Assoc.* **101**, 1228–1240.

Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31**, 1852–1884.

Csörgő, M. and Révész, P. (1975). A strong approximation of the multivariate empirical process. *Studia Scientiarum Mathematicarum Hungarica* **10**, 427–434.

Fan, J., Wang, M. and Yao, Q. (2006). Modeling multivariate volatilities via conditionally uncorrelated components. *Technical Report.* Department of Statistics, London School of Economics, London.

Fermanian, J.D. (2005). Goodness-of-fit tests for copulas. *Journal of Multivariate Analysis* **95**, 119–152.

Fermanian,J.D., Radulovic, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10**, 847–860.

Fermanian, J.D. and Scaillet, O. (2003). Nonparametric estimation of copulas for time series. *Journal of Risk* **5**, 25–54.

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

Genest, C., Quessy, J.-F. and Rmillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics* **33**, 337–366.

Giacomini, E., Haerdle, W., Ignatieva, E. and Spokoiny, V. (2006). Inhomogeneous dependency modelling with time varying copulae. *Technical Report.* CASE, Humboldt University of Berlin, Berlin.

Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models. *Econometrica* **71**, 285–317.

Hull, J. and White, A. (1998). Value at Risk when daily changes in market variables are not normally distributed. *Journal of Derivatives* **5**, 9–19.

Horváth, L., Kokoszka, P. and Teyssiére, G. (2001). Empirical process of the squared residuals of an ARCH sequence. *Ann. Statist.* **29**, 445–469.

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

Klugman, S.A. and Parsa, R. (1999). Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics* **24**, 139–148.

Koul, H.L. and Ling, S. (2006). Fitting an error distribution in some heteroscedastic time series models. *Ann. Statist.* **34**, 994–1012.

McNeil, A.J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management*, Princeton University Press, Princeton, New Jersey.

Nelsen, R.B. (2005). *An Introduction to Copulas, 2nd Ed.*, Springer, New York.

Scaillet, O. (2006). Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters. *Journal of Multivariate Analysis* **97**, in press.

Shorack, G.R. and Wellner, J.A. (1985). *Empirical Processes with Applications to Statistics*, Wiley, New York.

Vervaat, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrs. verw. Geb.* **23**, 249-253.

Table 1: Estimation results for mixture copula $C(u_1, u_2, u_3; \theta_1, \theta_2, \lambda)$. Standard deviations are given in parenthesis.

	Case (i) $\lambda = 0.3$	Case (i) $\lambda = 0.7$	Case (ii) $\lambda = 0.3$	Case (ii) $\lambda = 0.7$
c_1	1.033 (0.292)	1.111 (0.306)	0.241 (0.117)	0.260 (0.122)
$\alpha_{1,1}$	0.206 (0.073)	0.215 (0.071)	0.209 (0.065)	0.219 (0.062)
$\beta_{1,1}$	0.207 (0.181)	0.197 (0.177)	0.566 (0.149)	0.563 (0.137)
c_2	1.062 (0.269)	1.129 (0.278)	0.237 (0.109)	0.254 (0.110)
$\alpha_{2,1}$	0.309 (0.083)	0.327 (0.081)	0.208 (0.068)	0.217 (0.067)
$\beta_{2,1}$	0.195 (0.146)	0.188 (0.135)	0.457 (0.173)	0.456 (0.174)
c_3	1.071 (0.245)	1.124 (0.247)	0.219 (0.086)	0.236 (0.091)
$\alpha_{3,1}$	0.408 (0.089)	0.437 (0.085)	0.206 (0.070)	0.216 (0.067)
$\beta_{3,1}$	0.192 (0.114)	0.192 (0.108)	0.381 (0.191)	0.375 (0.185)
θ_1	2.956 (0.504)	2.959 (0.242)	2.984 (0.537)	2.981 (0.250)
θ_2	2.064 (0.091)	2.166 (0.163)	2.061 (0.102)	2.171 (0.195)
λ	0.298 (0.064)	0.693 (0.061)	0.295 (0.065)	0.690 (0.060)

Table 2: Parameter estimates for GARCH(1,1) of the daily log-returns of S&P 500 index, stock prices of Cisco and Intel.

	S&P 500 index	Cisco systems	Intel Corporation
c	0.0096	0.3272	0.1336
α	0.0636	0.0737	0.0186
β	0.9247	0.8879	0.9597

Table 3: Copula parameter estimates and test statistic of the daily log-returns of S&P 500 index, stock prices of Cisco and Intel.

	$C_2(;\theta_1,\theta_2,\lambda)$
Parameter estimation	(1.0549, 1.4618, 0.4716)
Test statistic nT_n	0.1806
P-value	0.000

Table 4: Parameter estimates for GARCH(1,1) of the daily log-returns of stock prices of Nortel, Lucent and Cisco.

	Nortel	Lucent	Cisco
c	7.0×10^{-6}	1.0×10^{-5}	8.0×10^{-6}
α	0.0360	0.0436	0.0627
β	0.9609	0.9504	0.9301

Table 5: Copula parameter estimates and test statistic of the daily log-returns of tock prices of Nortel, Lucent and Cisco.

	$C_2(;\theta_1,\theta_2,\lambda)$
Parameter estimation	(1.05839, 1.3456, 0.4284)
Test statistic nT_n	0.2301
P-value	0.000

4 Proofs

Proof of Theorem 2.1. For $u = (u_1, \ldots, u_r)^T$, put

$$\hat{\eta}_{j,t}(u) = \hat{w}_{j,t}^{1/2}(\gamma_j + (n-\nu+1)^{-1/2}u)/\sqrt{h_{j,t}},$$

$$\eta_{j,t}(u) = w_{j,t}^{1/2}(\gamma_j + (n-\nu+1)^{-1/2}u)/\sqrt{h_{j,t}},$$

$$\xi_{j,t}^{(1)}(s,u) = I(\epsilon_{j,t} \le s\hat{\eta}_{j,t}(u)) - F_{\epsilon,j}(s\hat{\eta}_{j,t}(u)) - \{I(\epsilon_{j,t} \le s\eta_{j,t}(u)) - F_{\epsilon,j}(s\eta_{j,t}(u))\},$$

$$\xi_{j,t}^{(2)}(s,u) = I(\epsilon_{j,t} \le s\eta_{j,t}(u)) - F_{\epsilon,j}(s\eta_{j,t}(u)) - \{I(\epsilon_{j,t} \le s) - F_{\epsilon,j}(s)\},$$

$$\xi_{j,t}(s,u) = \xi_{j,t}^{(1)}(s,u) + \xi_{j,t}^{(2)}(s,u),$$

$$S_j(s,u) = \sum_{t=\nu}^{n} \xi_{j,t}(s,u).$$

Following the proof of Theorem 2.2 of Berkes and Horváth (2003), we only need to show that for any $A > 0$ and $\epsilon > 0$ there exist $\gamma = \gamma(\epsilon)$, $\delta = \delta(\epsilon)$ and $N = N(\epsilon)$ such that for $n \ge N$

$$P\Big(\sup_{|u|\le A} \sup_{0<F(t)\le(\gamma n\log n)^{-1}} \frac{|S_j(s,u)|}{\{F_{\epsilon,j}(s)(1-F_{\epsilon,j}(s))\}^{\beta_3}} \ge \epsilon\sqrt{n-\nu+1}\Big) \le \epsilon, \tag{4.1}$$

$$P\Big(\sup_{|u|\le A} \sup_{(\gamma n\log n)^{-1}\le F(t)\le\delta} \frac{|S_j(s,u)|}{\{F_{\epsilon,j}(s)(1-F_{\epsilon,j}(s))\}^{\beta_3}} \ge \epsilon\sqrt{n-\nu+1}\Big) \le \epsilon, \tag{4.2}$$

$$P\Big(\sup_{|u|\le A} \sup_{0<1-F(t)\le(\gamma n\log n)^{-1}} \frac{|S_j(s,u)|}{\{F_{\epsilon,j}(s)(1-F_{\epsilon,j}(s))\}^{\beta_3}} \ge \epsilon\sqrt{n-\nu+1}\Big) \le \epsilon, \tag{4.3}$$

$$P\Big(\sup_{|u|\le A} \sup_{(\gamma n\log n)^{-1}\le 1-F(t)\le\delta} \frac{|S_j(s,u)|}{\{F_{\epsilon,j}(s)(1-F_{\epsilon,j}(s))\}^{\beta_3}} \ge \epsilon\sqrt{n-\nu+1}\Big) \le \epsilon. \tag{4.4}$$

Similar to the proof of Lemma 6.1 of Berkes and Horváth (2003), there exists a constant $C(A) > 0$ such that for any s, $x > 0$ and $|u| \le A$

$$P(|S_j(s,u)| \ge x\sqrt{n-\nu+1}) \le \frac{C(A)}{x^4(n-\nu+1)}. \tag{4.5}$$

Using Lemmas 4.1-4.3 of Berkes and Horváth (2003) or (5.42) and (5.43) of Hall and Yao (2003), we have

$$\sup_{|u|\le A} \sup_{\nu\le t\le n} |\hat{\eta}_{j,t}(u) - 1| = o_p(1). \tag{4.6}$$

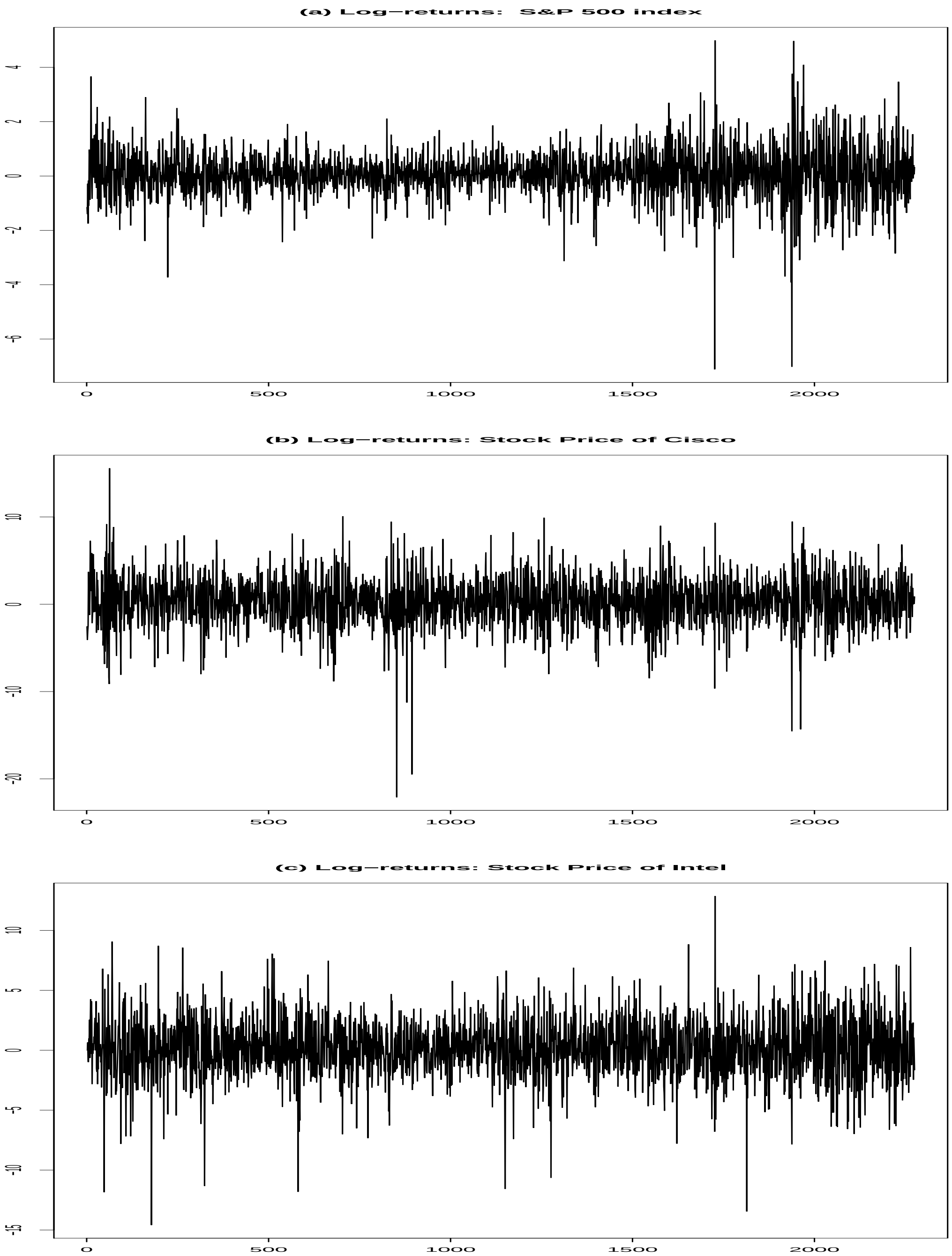

Figure 1: Daily Log-returns of S&P 500 index (a), Stock Price of Cisco Systems (b) and Stock Price of Intel Corporation (c) from January 2, 1991 to December 31, 1999.

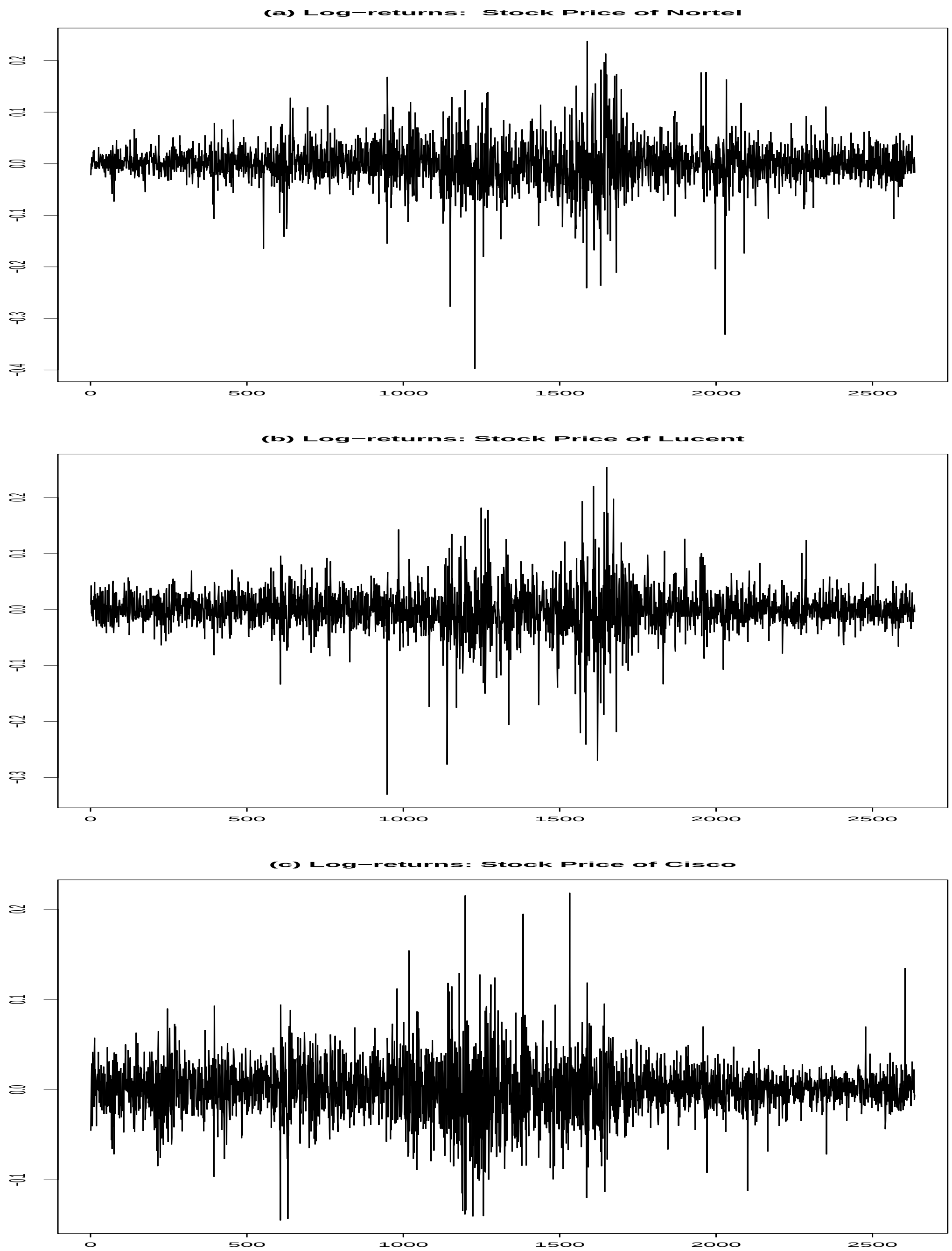

Figure 2: Daily Log-returns of stock prices of Nortel (a), Lucent (b) and Cisco (c) from April 4, 1996 to September 22, 2006.

It is known that

$$P(\sup_{0\le F(s)\le(\gamma n\log n)^{-1}} \frac{1}{F_{\epsilon,j}^{1/2}(s)} |(n-\nu+1)^{-1}\sum_{t=\nu}^{n} I(\epsilon_{j,t}\le s) - F_{\epsilon,j}(s)| \ge \epsilon\sqrt{n-\nu+1}) \le \epsilon \tag{4.7}$$

for n large enough.

By Condition 2 and (6.8)-(6.10) of Berkes and Horváth (2003),

$$\begin{aligned} P(\sup_{|u|\le A} \quad & \sup_{0<F(s)\le(\gamma n\log n)^{-1}} \{F_{\epsilon,j}(s)(1-F_{\epsilon,j}(s))\}^{-\beta_3} \\ & |(n-\nu+1)^{-1}\textstyle\sum_{t=\nu}^{n}\{F_{\epsilon,j}(s\hat{\eta}_{j,t}(u)) - F_{\epsilon,j}(s)\}| \\ & \ge \epsilon\sqrt{n-\nu+1}) \quad \le \quad \epsilon \end{aligned} \tag{4.8}$$

for n large enough. Using Condition 2, (4.6), (4.7) and (4.8), we can show that

$$\begin{aligned} P(\sup_{|u|\le A} \quad & \sup_{0<F(t)\le(\gamma n\log n)^{-1}} \{F_{\epsilon,j}(s)(1-F_{\epsilon,j}(s))\}^{-\beta_3} \\ |(n-\nu+1)^{-1}\textstyle\sum_{t=\nu}^{n} & \{I(\epsilon_{j,t}\le s\hat{\eta}_{j,t}(u)) - F_{\epsilon,j}(s\hat{\eta}_{j,t}(u))\}| \ge \epsilon\sqrt{n-\nu+1}) \\ & \le \epsilon \end{aligned} \tag{4.9}$$

for n large enough. Hence (4.1) can be proved by using (4.7) - (4.9).

By (4.5),

$$P(\sup_{(\gamma n\log n)^{-1}\le s\le\delta} \frac{|S_j(s,u)|}{\{F_{\epsilon,j}(s)(1-F_{\epsilon,j}(s))\}^{\beta_3}} \ge \epsilon\sqrt{n-\nu+1}) \le \epsilon \tag{4.10}$$

as n large enough. Like the proof of Lemma 6.3 of Berkes and Horváth (2003), we can show (4.2) by using (4.10). Similarly we can show (4.3) and (4.4). Hence the theorem.

Proof of Theorem 2.2. We only consider the case $\nu=1$ since the other cases can be dealt with similarly. Write

$$F_{\epsilon,j}(\hat{\epsilon}_{j,t})I(\epsilon_{j,t}\ge 0) = F_{\epsilon,j}(\epsilon_{j,t}\{\frac{\hat{w}_{j,t}(\hat{\gamma}_j)}{w_{j,t}(\gamma_j)}\}^{1/2})I(\epsilon_{j,t}\ge 0).$$

Similar to the proof of Lemma 5.1 of Berkes and Horváth (2003), we can show that

$$\begin{cases} \frac{1}{n}\sum_{t=1}^{n} |F_{\epsilon,j}(\hat{\epsilon}_{j,t}) - F_{\epsilon,j}(\epsilon_{j,t})|I(\epsilon_{j,t}\ge 0) \xrightarrow{p} 0 \\ \frac{1}{n}\sum_{t=1}^{n} |F_{\epsilon,j}(\hat{\epsilon}_{j,t}) - F_{\epsilon,j}(\epsilon_{j,t})|I(\epsilon_{j,t}< 0) \xrightarrow{p} 0, \end{cases}$$

i.e.,

$$\frac{1}{n}\sum_{t=1}^{n}|F_{\epsilon,j}(\hat{\epsilon}_{j,t})-F_{\epsilon,j}(\epsilon_{j,t})|\xrightarrow{p}0. \tag{4.11}$$

As in the proof of Theorem 2.1 of Berkes and Horváth (2003), we can show that

$$\lim_{n\to\infty}\sup_{-\infty<t<\infty}|\hat{F}_{\epsilon,i}(t)-F_{\epsilon,i}(t)|=0 \quad \text{a.s.} \tag{4.12}$$

for $i=1,\ldots,r$.

Write

$$\begin{aligned}
& \tfrac{1}{n}\textstyle\sum_{t=1}^{n}\{\log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta)-\log c(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta)\} \\
= & \tfrac{1}{n}\textstyle\sum_{t=1}^{n}\log\{c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta)\}I(\wedge_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})\le\frac{\Delta_0}{2}) \\
& +\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\log\{c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta)\}I(\vee_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})\ge\Delta_1+\frac{1-\Delta_1}{2}) \\
& -\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\log\{c(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta)\}I(\wedge_{j=1}^{r}F_{\epsilon,j}(\epsilon_{j,t})\le\frac{\Delta_0}{2}) \\
& -\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\log\{c(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta)\}I(\vee_{j=1}^{r}F_{\epsilon,j}(\epsilon_{j,t})\ge\Delta_1+\frac{1-\Delta_1}{2}) \\
& +\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\{\log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta)-\log c(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta)\}\times \\
& \times I(\frac{\Delta_0}{2}<\wedge_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})<\vee_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})<\Delta_1+\frac{1-\Delta_1}{2})\times \\
& \times I(\frac{\Delta_0}{2}<\wedge_{j=1}^{r}F_{\epsilon,j}(\epsilon_{j,t})<\vee_{j=1}^{r}F_{\epsilon,j}(\epsilon_{j,t})<\Delta_1+\frac{1-\Delta_1}{2}) \\
& +\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\{\log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta)\}\times \\
& \times I(\frac{\Delta_0}{2}<\wedge_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})<\vee_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})<\Delta_1+\frac{1-\Delta_1}{2})I(\frac{\Delta_0}{2}\ge\wedge_{j=1}^{r}F_{\epsilon,j}(\epsilon_{j,t})) \\
& +\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\{\log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta)\}\times \\
& \times I(\frac{\Delta_0}{2}<\wedge_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})<\vee_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})<\Delta_1+\frac{1-\Delta_1}{2})I(\vee_{j=1}^{r}F_{\epsilon,j}(\epsilon_{j,t})\ge\Delta_1+\frac{1-\Delta_1}{2}) \\
& -\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\{\log c(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta)\}\times \\
& \times I(\frac{\Delta_0}{2}\ge\wedge_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}))I(\frac{\Delta_0}{2}<\wedge_{j=1}^{r}F_{\epsilon,j}(\epsilon_{j,t})<\vee_{j=1}^{r}F_{\epsilon,j}(\epsilon_{j,t})<\Delta_1+\frac{1-\Delta_1}{2}) \\
& -\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\{\log c(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta)\}\times \\
& \times I(\vee_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})\ge\Delta_1+\frac{1-\Delta_1}{2})I(\frac{\Delta_0}{2}<\wedge_{j=1}^{r}F_{\epsilon,j}(\epsilon_{j,t})<\vee_{j=1}^{r}F_{\epsilon,j}(\epsilon_{j,t})<\Delta_1+\frac{1-\Delta_1}{2}) \\
= & I_1+\cdots+I_9.
\end{aligned}$$

By Condition C4, as $\Delta_0^*\to 0$,

$$\begin{aligned}
|I_1| & \le \tfrac{M_0}{n}\textstyle\sum_{t=1}^{n}\{\wedge_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})\}^{-\beta_0}I(\wedge_{j=1}^{r}\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})\le\Delta_0^*) \\
& \le \tfrac{M_0}{n}\textstyle\sum_{t=1}^{n}\sum_{j=1}^{r}\{\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})\}^{-\beta_0}I(\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t})\le\Delta_0^*) \\
& = \tfrac{M_0}{n}\textstyle\sum_{t=1}^{n}\sum_{j=1}^{r}\{\frac{t}{n}\}^{-\beta_0}I(t/n\le\Delta_0^*) \\
& = O(\{\Delta_0^*\}^{1-\beta_0})
\end{aligned}$$

and

$$\begin{aligned}|I_3| &\le \tfrac{M_0}{n}\textstyle\sum_{t=1}^n \{\wedge_{j=1}^r F_{\epsilon,j}(\epsilon_{j,t})\}^{-\beta_0} I(\wedge_{j=1}^r F_{\epsilon,j}(\epsilon_{j,t}) \le \Delta_0^*)\\ &\le \tfrac{M_0}{n}\textstyle\sum_{t=1}^n \sum_{j=1}^r \{F_{\epsilon,j}(\epsilon_{j,t})\}^{-\beta_0} I(F_{\epsilon,j}(\epsilon_{j,t}) \le \Delta_0^*)\\ &= O_p(\{\Delta_0^*\}^{1-\beta_0}).\end{aligned}$$

Similarly,

$$I_2 = O(\{1-\Delta_1^*\}^{1-\beta_0}) \quad \text{and} \quad I_4 = O_p(\{1-\Delta_1^*)\}^{1-\beta_0})$$

as $\Delta_1^* \to 1$. By (4.11), (4.12) and Condition C4,

$$\begin{aligned}I_5 &\le \tfrac{M_1}{n}\textstyle\sum_{t=1}^n \sum_{j=1}^r |\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) - F_{\epsilon,j}(\epsilon_{j,t})|^{\beta_1}\\ &\le \tfrac{M_1}{n}\textstyle\sum_{t=1}^n \sum_{j=1}^r |\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) - F_{\hat{\epsilon},j}(\epsilon_{j,t})|^{\beta_1}\\ &+\tfrac{M_1}{n}\textstyle\sum_{t=1}^n \sum_{j=1}^r |F_{\hat{\epsilon},j}(\epsilon_{j,t}) - F_{\epsilon,j}(\epsilon_{j,t})|^{\beta_1}\\ &= o_p(1).\end{aligned}$$

Note that

$$\begin{aligned}|I_6| &= |\tfrac{1}{n}\textstyle\sum_{t=1}^n \{\log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta)\} I(\vee_{j=1}^r |\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) - F_{\epsilon,j}(\epsilon_{j,t})| > \tfrac{\Delta_0}{2})\times\\ &\times I(\tfrac{\Delta_0}{2} < \wedge_{j=1}^r \hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) < \vee_{j=1}^r \hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) < \Delta_1 + \tfrac{1-\Delta_1}{2}) I(\tfrac{\Delta_0}{2} \ge \wedge_{j=1}^r F_{\epsilon,j}(\epsilon_{j,t}))\\ &+\tfrac{1}{n}\textstyle\sum_{t=1}^n \{\log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta)\} I(\vee_{j=1}^r |\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) - F_{\epsilon,j}(\epsilon_{j,t})| \le \tfrac{\Delta_0}{2})\times\\ &\times I(\tfrac{\Delta_0}{2} < \wedge_{j=1}^r \hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) < \vee_{j=1}^r \hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) < \Delta_1 + \tfrac{1-\Delta_1}{2}) I(\tfrac{\Delta_0}{2} \ge \wedge_{j=1}^r F_{\epsilon,j}(\epsilon_{j,t}))|\\ &\le \tfrac{1}{n}\textstyle\sum_{t=1}^n |\log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta)| I(\vee_{j=1}^r |\hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) - F_{\epsilon,j}(\epsilon_{j,t})| > \tfrac{\Delta_0}{2})\times\\ &\times I(\tfrac{\Delta_0}{2} < \wedge_{j=1}^r \hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) < \vee_{j=1}^r \hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) < \Delta_1 + \tfrac{1-\Delta_1}{2}) I(\tfrac{\Delta_0}{2} \ge \wedge_{j=1}^r F_{\epsilon,j}(\epsilon_{j,t}))\\ &+\tfrac{1}{n}\textstyle\sum_{t=1}^n |\log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta)| I(\wedge_{j=1}^r \hat{F}_{\epsilon,j}(\hat{\epsilon}_{j,t}) \le \Delta_0).\end{aligned}$$

It follows from (4.11), (4.12) and Condition C4 that

$$I_6 = o_p(1) \quad \text{as} \quad \Delta_0 \to 0.$$

Similarly,

$$I_7 = o_p(1) \quad \text{as} \quad \Delta_1 \to 1, \quad I_8 = o_p(1) \quad \text{as} \quad \Delta_0 \to 0, \quad I_9 = o_p(1) \quad \text{as} \quad \Delta_1 \to 1.$$

Therefore,

$$\sup_{\theta\in\Theta} |\frac{1}{n}\sum_{t=1}^n \{\log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta) - \log c(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta)\}| = o_p(1). \tag{4.13}$$

By Theorem 4.2.1 of Amemiya (1985) and Conditions C1-C3,

$$\sup_{\theta\in\Theta}|\frac{1}{n}\sum_{t=1}^{n}\log c(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta)-E\{\log c(F_{\epsilon,1}(\epsilon_{1,1}),\ldots,F_{\epsilon,r}(\epsilon_{r,1});\theta)\}|\xrightarrow{p}0. \quad (4.14)$$

Thus, by (4.13) and (4.14),

$$\sup_{\theta\in\Theta}|\frac{1}{n}\sum_{t=1}^{n}\log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta)-E\{\log c(F_{\epsilon,1}(\epsilon_{1,1}),\ldots,F_{\epsilon,r}(\epsilon_{r,1});\theta)|\xrightarrow{p}0. \quad (4.15)$$

It follows from Jensen's inequality that

$$E\frac{\log c(F_{\epsilon,1}(\epsilon_{1,1}),\ldots,F_{\epsilon,r}(\epsilon_{r,1});\theta)}{\log c(F_{\epsilon,1}(\epsilon_{1,1}),\ldots,F_{\epsilon,r}(\epsilon_{r,1});\theta_0)}<\log E\frac{c(F_{\epsilon,1}(\epsilon_{1,1}),\ldots,F_{\epsilon,r}(\epsilon_{r,1});\theta)}{c(F_{\epsilon,1}(\epsilon_{1,1}),\ldots,F_{\epsilon,r}(\epsilon_{r,1});\theta_0)}\quad\text{for}\quad\theta\neq\theta_0,$$

i.e.,

$$E\log c(F_{\epsilon,1}(\epsilon_{1,1}),\ldots,F_{\epsilon,r}(\epsilon_{r,1});\theta)<E\log c(F_{\epsilon,1}(\epsilon_{1,1}),\ldots,F_{\epsilon,r}(\epsilon_{r,1});\theta_0)\quad\text{for}\quad\theta\neq\theta_0. \quad (4.16)$$

Thus the theorem follows from (4.15), (4.16) and Theorem 4.1.1 of Amemiya (1985).

Before proving Theorem 2.3, we need a lemma which generalizes Lemma 5.1 of Claeskens and Van Keilegom (2003) to the higher dimensional case. Let $U_1=(U_{1,1},\ldots,U_{r,1})^T,\ldots,U_n=(U_{1,n},\ldots,U_{r,n})^T$ denote independent random vectors with multivariate uniform distribution on $[0,1]^r$. For $x=(x_1,\ldots,x_r)^T$, define the empirical process

$$U_n(x)=\sqrt{n-\nu+1}\{\frac{1}{n-\nu+1}\sum_{t=\nu}^{n}I(U_{1,t}\leq x_1,\ldots,U_{r,t}\leq x_r)-\prod_{i=1}^{r}x_i\}.$$

Let $x^{(i_1,\ldots,i_k)}$ denote $x=(x_1,\ldots,x_r)^T$ with x_j replaced by 1 when j is not one of $i_1,\ldots,i_k$, $x^{(i_1,\ldots,i_k)*}$ denote $x=(x_1,\ldots,x_r)^T$ with x_j replaced by 1 when j is one of $i_1,\ldots,i_k$, and $x_j^{(i_1,\ldots,i_k)*}$ denote the j-th element of $x^{(i_1,\ldots,i_k)*}$. Put

$$V_n(x_1,\ldots,x_r)=\sum_{k=1}^{r}(-1)^k\sum_{i_1\neq\ldots\neq i_k}\{\prod_{j=1}^{r}x_j^{(i_1,\ldots,i_k)*}\}U_n(x^{(i_1,\ldots,i_k)})$$

and

$$V(x_1,\ldots,x_r)=\sum_{k=1}^{r}(-1)^k\sum_{i_1\neq\ldots\neq i_k}\{\prod_{j=1}^{r}x_j^{(i_1,\ldots,i_k)*}\}U(x^{(i_1,\ldots,i_k)}).$$

LEMMA **4.1.** *For any* $\beta \in (0,1/2)$,

$$\sup_{x=(x_1,\ldots,x_r)^T\in[0,1]^r} \frac{|V_n(x)-V(x)|}{\prod_{i=1}^r (x_i)^\beta (1-x_i)^\beta} = o_p(1). \qquad (4.17)$$

Proof. By induction, this lemma can be proved by following the proof of Lemma 5.1 of Claeskens and Van Keilegom (2003).

Proof of Theorem 2.3. Define

$$\begin{aligned}\beta_n(x_1,\ldots,x_r) &= \sqrt{n-\nu+1}\{\tfrac{1}{n-\nu+1}\textstyle\sum_{t=\nu}^n I(F_{\epsilon,1}(\epsilon_{1,t})\le x_1,\ldots,F_{\epsilon,r}(\epsilon_{r,t})\le x_r)\\ &\quad -C(x_1,\ldots,x_r;\theta)\}.\end{aligned}$$

Write

$$\begin{aligned}&\tfrac{1}{\sqrt{n-\nu+1}}\textstyle\sum_{t=\nu}^n \delta(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta_0)\\
=\ & \textstyle\int \delta(x_1,\ldots,x_r)\, d\beta_n(x_1,\ldots,x_r;\theta_0)\\
&+\textstyle\int \delta(x_1,\ldots,x_r;\theta_0)c(x_1,\ldots,x_r;\theta_0)\,dx_1\ldots dx_r\\
=\ & \textstyle\int \delta(T_r^-(x_1,\ldots,x_r);\theta_0)\,d\beta_n(T_r^-(x_1,\ldots,x_r))\\
\overset{d}{=}\ & \textstyle\int \delta(T_r^-(x_1,\ldots,x_r);\theta_0)\,dU_n(x_1,\ldots,x_r)\\
=\ & \textstyle\int \delta(T_r^-(x_1,\ldots,x_r);\theta_0)\,d\{U_n(x_1,\ldots,x_r)-U(x_1,\ldots,x_r)\}\\
&+\textstyle\int \delta(T_r^-(x_1,\ldots,x_r);\theta_0)\,dU(x_1,\ldots,x_r)\\
=\ & (-1)^r\textstyle\int \delta(T_r^-(x_1,\ldots,x_r);\theta_0)\,d\{V_n(x_1,\ldots,x_r)-V(x_1,\ldots,x_r)\}\\
&+(-1)^{r+1}\textstyle\int \delta(T_r^-(x_1,\ldots,x_r);\theta_0)\,d\{\sum_{k=1}^{r-1}(-1)^k\sum_{i_1\ne\ldots\ne i_k}(\prod_{j=1}^r x_j^{(i_1,\ldots,i_k)*})\\
&(U_n(x^{(i_1,\ldots,i_k)})-U(x^{(i_1,\ldots,i_k)}))\}+\textstyle\int \delta(T_r^-(x_1,\ldots,x_r);\theta_0)\,dU(x_1,\ldots,x_r)\\
=\ & I_1+I_2+I_3.\end{aligned} \qquad (4.18)$$

By Lemma 4.1 (i.e., (4.17)) and Condition N2,

$$I_1 = \int \{V_n(x_1,\ldots,x_r)-V(x_1,\ldots,x_r)\}\,d\delta(T_r^-(x_1,\ldots,x_r);\theta_0) = o_p(1). \qquad (4.19)$$

Iteratively applying the same trick to $U_n(x^{(i_1,\ldots,i_k)})-U(x^{(x_1,\ldots,i_k)})$, we can show that

$$I_2 = o_p(1). \qquad (4.20)$$

For any $x=(x_1,\ldots,x_r)^T$, define $D_x=\{(a_1,\ldots,a_r): a_i\le x_i, i=1,\ldots,r\}$. It follows from Condition N1 and Theorem 1 of Csörgő and Révész (1975) that

$$\sup_x |\beta_n(x)-U(T_rD_x)| = o(1) \quad a.s. \qquad (4.21)$$

Using Theorem 2.1, Condition N3 and Taylor expansion, we can show that

$$\begin{aligned}
& \tfrac{1}{\sqrt{n-\nu+1}} \textstyle\sum_{t=\nu}^{n} \{\delta(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta_0) - \delta(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta_0)\} \\
= \ & \tfrac{1}{\sqrt{n-\nu+1}} \textstyle\sum_{t=\nu}^{n}\sum_{i=1}^{r} \delta_i(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta_0)\{\hat{F}_{\epsilon,i}(\hat{\epsilon}_{i,t}) - F_{\epsilon,i}(\hat{\epsilon}_{i,t})\}\{1+o_p(1)\} \\
= \ & \tfrac{1}{\sqrt{n-\nu+1}} \textstyle\sum_{t=\nu}^{n}\sum_{i=1}^{r} \delta_i(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta_0)\{\hat{F}_{\epsilon,i}(\hat{\epsilon}_{i,t}) - F_{\epsilon,i}(\hat{\epsilon}_{i,t})\}\{1+o_p(1)\} \\
& +\tfrac{1}{\sqrt{n-\nu+1}} \textstyle\sum_{t=\nu}^{n}\sum_{i=1}^{r} \delta_i(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta_0)\{F_{\epsilon,i}(\hat{\epsilon}_{i,t}) - F_{\epsilon,i}(\epsilon_{i,t})\}\{1+o_p(1)\} \\
= \ & \tfrac{1}{n-\nu+1} \textstyle\sum_{t=\nu}^{n}\sum_{i=1}^{r} \delta_i(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta_0)\{U((1,\ldots,1,F_{\epsilon,i}(\hat{\epsilon}_{i,t}),1,\ldots,1)^T) \\
& +\tfrac{1}{2}\hat{\epsilon}_{i,t}F'_{\epsilon,t}(\hat{\epsilon}_{i,t})\sqrt{n-\nu+1}(\hat{\gamma}_i-\gamma_i)^T E(\tfrac{w'_{i,1}(\gamma_i)}{w_{i,1}(\gamma_i)})\}\{1+o_p(1)\} \\
& +\tfrac{1}{n-\nu+1} \textstyle\sum_{t=\nu}^{n}\sum_{i=1}^{r} \delta_i(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta_0)\times \\
& \times F'_{\epsilon,i}(\epsilon_{i,t})\sqrt{n-\nu+1}\{\hat{\epsilon}_{i,t}-\epsilon_{i,t}\}\{1+o_p(1)\} \\
= \ & \tfrac{1}{n-\nu+1} \textstyle\sum_{t=\nu}^{n}\sum_{i=1}^{r} \delta_i(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta_0)\{U((1,\ldots,1,F_{\epsilon,i}(\epsilon_{i,t}),1,\ldots,1)^T) \\
& +\tfrac{1}{2}\epsilon_{i,t}F'_{\epsilon,t}(\epsilon_{i,t})\sqrt{n-\nu+1}(\hat{\gamma}_i-\gamma_i)^T E(\tfrac{w'_{i,1}(\gamma_i)}{w_{i,1}(\gamma_i)})\}\{1+o_p(1)\} \\
& -\tfrac{1}{n-\nu+1} \textstyle\sum_{t=\nu}^{n}\sum_{i=1}^{r} \delta_i(F_{\epsilon,1}(\epsilon_{1,t}),\ldots,F_{\epsilon,r}(\epsilon_{r,t});\theta_0)\tfrac{1}{2}\epsilon_{i,t}F'_{\epsilon,i}(\epsilon_{i,t})\sqrt{n-\nu+1}(\hat{\gamma}_i-\gamma_i) \\
& \tfrac{w'_{i,t}(\gamma_i)}{w_{i,t}(\gamma_i)}\{1+o_p(1)\} \\
= \ & \textstyle\sum_{i=1}^{r}\int \delta_i(x_1,\ldots,x_r;\theta_0)U((1,\ldots,1,x_i,1,\ldots,1)^T)\times \\
& \times c(x_1,\ldots,x_r;\theta_0)\,dx_1\ldots dx_r + o_p(1).
\end{aligned} \tag{4.22}$$

Using Conditions N4-N6 and similar arguments in proving (4.15), we can show that

$$\begin{aligned}
& \sup_{\theta\in\Theta_0}|\tfrac{1}{n-\nu+1}\textstyle\sum_{t=\nu}^{n}\tfrac{\partial^2}{\partial\theta_i\partial\theta_j}\log c(\hat{F}_{\epsilon,1}(\hat{\epsilon}_{1,t}),\ldots,\hat{F}_{\epsilon,r}(\hat{\epsilon}_{r,t});\theta) \\
& -E\{\tfrac{\partial^2}{\partial\theta_i\partial\theta_j}\log c(F_{\epsilon,1}(\epsilon_{1,1}),\ldots,F_{\epsilon,r}(\epsilon_{r,1});\theta)\}| \xrightarrow{p} 0
\end{aligned} \tag{4.23}$$

for $1 \le i,j \le m$. Hence the theorem follows from (4.18), (4.19), (4.20), (4.22), (4.23) and Theorems 4.2.1 and 4.1.3 of Amemiya (1985).

Proof of Theorem 2.4. It follows from Theorem 2.1 that

$$\begin{aligned}
& \sup_{x_j}|\sqrt{n-\nu+1}\{\hat{F}_{\epsilon,j}(F^-_{\epsilon,j}(x_j))-x_j\} - U((1,\ldots,1,x_j,1,\ldots,1)^T) \\
& -\tfrac{1}{2}F^-_{\epsilon,j}(x_j)F'_{\epsilon,j}(F^-_{\epsilon,j}(x_j))\sqrt{n-\nu+1}(\hat{\gamma}_j-\gamma_j)^T E(\tfrac{w'_{j,1}(\gamma_j)}{w_{j,1}(\gamma_j)})| = o_p(1).
\end{aligned} \tag{4.24}$$

Applying Lemma 1 of Vervaat (1972) to (4.24), we have

$$\begin{aligned}
& \sup_{x_j}|\sqrt{n-\nu+1}\{F_{\epsilon,j}(\hat{F}^-_{\epsilon,j}(x_j))-x_j\} + U((1,\ldots,1,x_j,1,\ldots,1)^T) \\
& +\tfrac{1}{2}F^-_{\epsilon,j}(x_j)F'_{\epsilon,j}(F^-_{\epsilon,j}(x_j))\sqrt{n-\nu+1}(\hat{\gamma}_j-\gamma_j)^T E(\tfrac{w'_{j,1}(\gamma_j)}{w_{j,1}(\gamma_j)})| = o_p(1),
\end{aligned} \tag{4.25}$$

that is,

$$\sup_{x_j} |\sqrt{n-\nu+1}F'_{\epsilon,j}(F^-_{\epsilon,j}(x_j))\{\hat{F}^-_{\epsilon,j}(x_j) - F^-_{\epsilon,j}(x_j)\} + U((1,\ldots,1,x_j,1,\ldots,1)^T)$$
$$+\frac{1}{2}F^-_{\epsilon,j}(x_j)F'_{\epsilon,j}(F^-_{\epsilon,j}(x_j))\sqrt{n-\nu+1}(\hat{\gamma}_j-\gamma_j)^T E(\frac{w'_{j,1}(\gamma_j)}{w_{j,1}(\gamma_j)})| = o_p(1). \tag{4.26}$$

Write

$$\begin{aligned}
& \hat{C}_\epsilon(x_1,\ldots,x_r) - C(x_1,\ldots,x_r;\theta_0) \\
= \; & \frac{1}{n-\nu+1}\sum_{t=\nu}^n \{I(F_{\epsilon,1}(\epsilon_{1,t}) \le F_{\epsilon,1}(\hat{F}^-_{\epsilon,1}(x_1)\sqrt{\tfrac{\hat{w}_{1,t}(\hat{\gamma}_1)}{w_{1,t}(\gamma_1)}}),\ldots, \\
& F_{\epsilon,r}(\epsilon_{r,t}) \le F_{\epsilon,r}(\hat{F}^-_{\epsilon,r}(x_r)\sqrt{\tfrac{\hat{w}_{r,t}(\hat{\gamma}_r)}{w_{r,t}(\gamma_r)}})) \\
& -C(F_{\epsilon,1}(\hat{F}^-_{\epsilon,1}(x_1)\sqrt{\tfrac{\hat{w}_{1,t}(\hat{\gamma}_1)}{w_{1,t}(\gamma_1)}}),\ldots,F_{\epsilon,r}(\hat{F}^-_{\epsilon,r}(x_r)\sqrt{\tfrac{\hat{w}_{r,t}(\hat{\gamma}_r)}{w_{r,t}(\gamma_r)}});\theta_0)\} \\
& +\frac{1}{n-\nu+1}\sum_{t=\nu}^n \{C(F_{\epsilon,1}(\hat{F}^-_{\epsilon,1}(x_1)\sqrt{\tfrac{\hat{w}_{1,t}(\hat{\gamma}_1)}{w_{1,t}(\gamma_1)}}),\ldots,F_{\epsilon,r}(\hat{F}^-_{\epsilon,r}(x_r)\sqrt{\tfrac{\hat{w}_{r,t}(\hat{\gamma}_r)}{w_{r,t}(\gamma_r)}});\theta_0) \\
& -C(x_1,\ldots,x_r;\theta_0),
\end{aligned} \tag{4.27}$$

$$\begin{aligned}
& F_{\epsilon,j}(\hat{F}^-_{\epsilon,j}(x_j)\sqrt{\tfrac{\hat{w}_{j,t}(\hat{\gamma}_j)}{w_{j,t}(\gamma_j)}}) - x_j \\
= \; & \{\hat{F}^-_{\epsilon,j}(x_j)\sqrt{\tfrac{\hat{w}_{j,t}(\hat{\gamma}_j)}{w_{j,t}(\gamma_j)}} - F^-_{\epsilon,j}(x_j)\}F'_{\epsilon,j}(F^-_{\epsilon,j}(x_j))(1+o(1))
\end{aligned} \tag{4.28}$$

and

$$\begin{aligned}
& \sqrt{\tfrac{\hat{w}_{j,t}(\hat{\gamma}_j)}{w_{j,t}(\gamma_j)}} - 1 \\
= \; & \frac{1}{2}(\hat{\gamma}_j-\gamma_j)^T\frac{w'_{j,t}(\gamma_j)}{w_{j,t}(\gamma_j)}(1+o_p(1))
\end{aligned} \tag{4.29}$$

as t and n large. Hence the theorem follows from (4.26) - (4.29), Theorem 2.3, (2.7) and Taylor expansions.