# EFFICIENT ESTIMATION OF MULTIVARIATE SEMI-NONPARAMETRIC GARCH FILTERED COPULA MODELS

By

Xiaohong Chen, Zhuo Huang, Yanping Yi

August 2015

Revised October 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2215

# Efficient Estimation of Multivariate Semi-nonparametric GARCH Filtered Copula Models

Xiaohong Chen* Zhuo Huang† Yanping Yi ‡

Yale University Peking University Zhejiang University

First version: August 2015; Revised October 9, 2019

**Abstract**

This paper considers estimation of semi-nonparametric GARCH filtered copula models in which the individual time series are modelled by semi-nonparametric GARCH and the joint distributions of the multivariate standardized innovations are characterized by parametric copulas with nonparametric marginal distributions. The models extend those of Chen and Fan (2006) to allow for semi-nonparametric conditional means and volatilities, which are estimated via the method of sieves such as splines. The fitted residuals are then used to estimate the copula parameters and the marginal densities of the standardized innovations jointly via the sieve maximum likelihood (SML). We show that, even using nonparametrically filtered data, both our SML and the two-step copula estimator of Chen and Fan (2006) are still root-$n$ consistent and asymptotically normal, and the asymptotic variances of both estimators do not depend on the nonparametric filtering errors. Even more surprisingly, our SML copula estimator using the filtered data achieves the full semiparametric efficiency bound as if the standardized innovations were directly observed. These nice properties lead to simple and more accurate estimation of

*Department of Economics, Yale University, PO Box 208281, New Haven CT 06520-8281, USA. E-mail: xiaohong.chen@yale.edu. Web Page: http://cowles.econ.yale.edu/faculty/chen.htm.

†National School of Development, Peking University, Beijing,100871, China. E-mail: zhuohuang@nsd.pku.edu.cn.

‡The corresponding author. Address : School of Economics and Academy of Financial Research, Zhejiang University, Zhejiang, 310058, China. E-mail: yanping_yi@zju.edu.cn. Web Page: http://yanpingyi.weebly.com.

Value-at-Risk (VaR) for multivariate financial data with flexible dynamics, contemporaneous tail dependence and asymmetric distributions of innovations. Monte Carlo studies demonstrate that our SML estimators of the copula parameters and the marginal distributions of the standardized innovations have smaller variances and smaller mean squared errors compared to those of the two-step estimators in finite samples. A real data application is presented.

**JEL classification:** C14; C22; G32.

**Key Words:** Semi-nonparametric dynamic models; Residual copulas; Semiparametric multi-step; Residual sieve maximum likelihood; Semiparametric efficiency.

## 1 Introduction

Copula-based multivariate dynamic models have been widely used to model nonlinear dependence and financial risks among observed and/or latent series; see, e.g., Patton (2006, 2013), Cherubini et al. (2012), Zhao and Zhang (2018) and the references therein. In this paper, we consider estimation of semi-nonparametric dynamic filtered copula models, in which the dynamics of individual series are modelled as semi-nonparametric GARCH and the joint distribution of the standardized innovations of the multivariate series are characterized by parametric copulas with nonparametric marginal distributions. These models are very flexible, allowing for leverage effects, asymmetric and fat-tailed individual series, nonlinear and tail dependence among latent shocks to different financial series. They are useful in estimating and forecasting portfolio VaRs and risk managements.

There are two parts of unknown finite- and infinite-dimensional parameters associated with this class of models: **(i)** the semi-nonparametric conditional means and volatilities (semi-nonparametric GARCH) of individual observed time series; and **(ii)** the semi-nonparametric joint distributions, which consists of the copula parameters and the nonparametric marginal distributions, of the latent standardized innovations. Here the parametric copulas capture the contemporaneous dependence among the individual elements of the standardized innovations. Chen (2013) first proposed this class of models as an extension of Chen and Fan (2006) from *parametric* dynamic conditional means and volatilities of individual observed time series to a *semi-nonparametric* GARCH in part **(i)**.

This extension is important to capture the shapes of the "news impact curve" nonparametrically for individual financial series and lessen dynamic misspecification due to wrongly specified parametric functional forms of conditional means and volatilities.

In this paper, we treat semi-nonparametric functions in part (i) of the model as nuisance parameters, and focus on estimation of the copula parameters and the marginal distributions of the standardized innovations in part (ii). This is because there already exist many consistent estimators for various semi-nonparametric conditional means and volatilities using univariate time series data, such as estimators based on kernel, local polynomial regression, penalization and sieves (e.g., Fan and Yao (2003), Gao (2007), Linton and Mammen (2005), Chen and Shen (1998), Chen et al. (2014), Meister and Krei$\beta$ (2016)). For the sake of concreteness, we apply sieve estimation of the conditional mean and volatility functions in the paper. See, e.g., Yang (2006), Engle and Rangel (2008), Liu and Yang (2016) for spline GARCH estimation. We shall focus on two kinds of estimation procedures for parameters in part (ii) of the general model.

The first estimation procedure was already proposed and empirically implemented in Chen (2013): Stage 1, for each observed time series, estimate the semi-nonparametric conditional mean and volatility via the sieve quasi maximum likelihood (QML) assuming standard normal standardized innovations. Stage 2, estimate the marginal distributions of the standardized innovations via the rescaled empirical marginal distributions using the fitted residuals from Stage 1; and then estimate the copula parameters via the pseudo maximum likelihood (pseudo-ML) of the parametric copula density evaluated at the rescaled empirical marginal distributions of the fitted residuals from Stage 1. In this paper we refer to this separate estimation of marginal distributions and copula parameters as the semiparametric two-step procedure. Chen (2013) conjectured (at the end of section 5) that the asymptotic variance of the semiparametric two-step copula estimator using the sieve QML fitted residuals is the same as that of the copula estimator in Chen and Fan (2006) using parametric fitted residuals. In this paper, we show that the conjecture is indeed correct. Precisely, even using nonparametric conditional mean and GARCH fitted residuals, the semiparametric two-step copula estimator is still root-$n$ consistent and asymptotically normal, with its asymptotic variance being the same as that in Genest et al. (1995) using directly observed standardized innovations.[1]

In the paper we propose another estimation procedure: Stage 1, for each observed time series, estimate the semi-nonparametric conditional mean and volatility via the method of sieves; Stage 2, estimate the nonparametric marginal densities and the copula parameters *jointly* via the sieve maximum likelihood (SML) using the fitted residuals from Stage 1. We show that, even using nonparametrically filtered data, our joint SML copula estimator is still root-$n$ consistent and asymptotically normal, with its asymptotic variance being the same as that in Chen et al. (2006) using directly observed standardized innovations. Perhaps more surprisingly, our joint SML copula estimator using the filtered data is shown to achieve the full semiparametric efficiency bound as if the standardized innovations were directly observed.

For observed standardized innovations, the semiparametric two-step estimators of the marginals and the copula parameters in Chen (2013) and the joint SML estimators in our paper become those in Genest et al. (1995) and Chen et al. (2006) respectively. The estimator of Genest et al. (1995) is widely used but generally inefficient: The empirical marginal distributions are obviously inefficient for the marginal distributions unless the copula is independent; The two-step copula estimator is not efficient either unless the copula is independent or Gaussian (see, e.g., Klaassen and Wellner (1997)). Chen et al. (2006) has established that the joint SML estimators of the marginal distributions and the copula parameters are both efficient for semiparametric copula models with i.i.d. data. In this paper, although our joint SML copula estimator is shown to be efficient for the full semi-nonparametric dynamic model, it is unclear whether our joint SML estimator of the marginal distributions of the standardized innovations might achieve its semiparametric efficiency bound. Nevertheless, since our joint SML estimator of marginal distributions borrows information from other components of the innovations, it should be more efficient than that of Chen (2013). These nice theoretical properties lead to simple and more accurate estimation of VaRs for multivariate financial data with flexible dynamics, contemporaneous tail dependence and asymmetric distributions of innovations.

[1]In a concurrent and independent work, Neumeyer et al. (2019) also extends the result of Chen and Fan (2006) from parametric fitted residuals to the locally polynomial regression estimated nonparametric ARCH in Stage 1.

Monte Carlo studies demonstrate that our joint SML estimators of the copula parameters and the marginal distributions of the unobserved standardized innovations do have smaller variances (and smaller mean squared errors) compared to those of the semiparametric two-step estimators in finite samples ($n = 500$). For large samples ($n = 8000$), both estimators for the copula parameters perform well, while our joint SML estimators for marginal distributions are still more efficient than the empirical marginal distributions using nonparametric fitted residuals.

As a real data application, we apply the multivariate semi-nonparametric GARCH filtered Student's t-copula model to model dependence among five popular financial assets. The spline-GARCH estimates exhibit the well-known "news impact curve" (or leverage effects) in each asset. We estimate the copula parameters and the marginal distributions of the innovations using spline-GARCH fitted residuals. The joint SML and the semiparametric two-step methods produce similar estimates for copula parameters, although the joint SML gives slightly larger log-likelihood values for the copula parts. The estimated models are then used to estimate VaRs for the portfolios consisting of five assets and of paired assets. The full-sample backtesting and out-of-sample Diebold-Mariano test suggest that the VaR forecasts from the joint SML are more accurate than those from the semiparametric two-step estimates.

The rest of the paper is organized as follows: Section 2 presents the general model and the two estimation procedures. Section 3 establishes the asymptotic properties of the joint SML estimator using the semi-nonparametric GARCH filtered residuals. Section 4 presents the asymptotic properties of the semiparametric two-step estimator using the semi-nonparametric GARCH filtered residuals. Section 5 provides simulation studies and Section 6 presents an empirical application. Section 7 briefly concludes. All technical proofs and additional simulation tables are gathered into the Appendices. In this paper, we use "$\xrightarrow{\mathcal{P}}$", "$\xrightarrow{\mathcal{D}}$", and "$\rightsquigarrow$" to denote convergence in probability, convergence in distribution, and weak convergence, respectively.

## 2 The Model and Estimation Procedures

In this section we formally introduce the model and several estimation procedures.

### 2.1 Multivariate Semi-nonparametric Dynamic Filtered Copula Models

Let $\{Y_t = (Y_{1,t}, \cdots, Y_{k,t})^\intercal\}_{t=1}^n$ be the observations of a $k \times 1$ vector-valued time series. Let $\mathcal{F}^{t-1}$ denote the available information up to time $t-1$, which includes all the lagged $Y_{t-r}$ for $r \geq 1$ and other random vectors observable at time $t-1$. We assume that $\{Y_t\}_{t=1}^n$ is generated from the semi-nonparametric dynamic (in particular, GARCH) filtered copula model:

$$Y_{j,t} = \mu_{0,j}\left(\mathcal{F}^{t-1}\right) + \sigma_{0,j}\left(\mathcal{F}^{t-1}\right)\xi_{j,t}, \quad for\ j = 1, ..., k, \tag{1}$$

$$\mu_{0,j}\left(\mathcal{F}^{t-1}\right) = E\left[Y_{j,t}|\mathcal{F}^{t-1}\right] \equiv \mu_{0,j}^t, \quad \sigma_{0,j}\left(\mathcal{F}^{t-1}\right) = \sqrt{Var\left[Y_{j,t}|\mathcal{F}^{t-1}\right]} \equiv \sigma_{0,j}^t, \quad for\ j = 1, ..., k,$$

$$F_0(\xi_{1,t}, \cdots, \xi_{k,t}) = C\left(F_{0,1}(\xi_{1,t}), \cdots, F_{0,k}(\xi_{k,t}); \theta_0\right), \tag{2}$$

where the standardized multivariate innovations $\{\xi_t = (\xi_{1,t}, \cdots, \xi_{k,t})^\intercal : t \geq 1\}$ are assumed to be independent of $\mathcal{F}^{t-1}$ and are identically and independently distributed. We assume that the joint distribution $F_0(\cdot)$ of $\xi_t$ follows a semiparametric copula model (2), in which its copula function is known up to the unknown true finite dimensional parameter $\theta_0 \in \Theta$ and the true marginal distributions (and pdfs) $F_{0,j}$ (and $f_{0,j}$), $j = 1, \cdots, k$, are unspecified.

We note that Model (1)-(2) is slightly different from typical multivariate GARCH models,[2] as it implies a constant conditional correlation matrix with $Corr\left[Y_{j,t}, Y_{l,t}|\mathcal{F}^{t-1}\right] = E[\xi_{j,t}\xi_{l,t}]$ as the $(j,l)-$ element, for $j, l = 1, ..., k$, and the contemporaneous dependence of $\xi_t = (\xi_{1,t}, \cdots, \xi_{k,t})^\intercal$ is specified semi-nonparametrically in (2). Unlike typical multivariate GARCH models, (2) implies that tail dependence among $\xi_t$ depends on copulas only and is free of behaviors of marginal densities. We use the following assumption to formally summarize this class of models.

**Assumption 1.** *(i) The strictly stationary observations $\{Y_t : t \geq 1\}$ are $\beta$-mixing with $\beta(t) \leq \beta_0 t^{-\zeta}$ for some $\beta_0 > 0$ and $\zeta = \gamma - 2 > 2$, and $E\left(|Y_{j,t}|^4\right) < \infty$ for $j = 1, \cdots, k$. $\{Y_t\}$ satisfies Model (1)-(2), where the unknown true dynamic parameter $\kappa_0(\cdot) = \left[\mu_{0,1}(\cdot), \cdots, \mu_{0,k}(\cdot), \sigma_{0,1}(\cdot), \cdots, \sigma_{0,k}(\cdot)\right]^\intercal$*

[2]See, e.g., Bollerslev et al. (1988), Engle and Kroner (1995), Hafner and Preminger (2009) and others.

*are semi-nonparametrically specified;* (ii) $\{\xi_t : t \geq 1\}$ *are independent of* $\mathcal{F}^{t-1}$ *and are a random sample from the distribution* $F_0(\xi_1, \cdots, \xi_k)$ *satisfying Model* (2)*.* $F_{0,j} : \Xi_j \to [0,1]$ *is the unknown true absolutely continuous marginal distribution function of* $\xi_{j,t}$ *for* $j = 1, ..., k$*. The functional form of the copula* $C(u_1, \cdots, u_k; \theta_0) = F_0(F_{0,1}^{-1}(u_1), \cdots, F_{0,k}^{-1}(u_k))$ *is known up to the finite dimensional parameter* $\theta_0 \in \Theta$*.*

## 2.2 Estimation

There are two sets of unknown semi-nonparametric parameters associated with Model (1)-(2): **(i)** The true conditional mean functions and volatility functions $\kappa_0 = [\mu_{0,1}, \cdots, \mu_{0,k}, \sigma_{0,1}, \cdots, \sigma_{0,k}]^\intercal$; and **(ii)** the true copula parameters and the marginal distributions of innovations $\alpha_0 = (\theta_0^\intercal, F_{0,1}, \cdots, F_{0,k})^\intercal$ or $\alpha_0 = (\theta_0^\intercal, f_{0,1}, \cdots, f_{0,k})^\intercal$ with $f_{0,j}$ being the density of $F_{0,j}$ for $j = 1, ..., k$.

Let $\kappa(\cdot) = (\mu_1(\cdot), \cdots, \mu_k(\cdot), \sigma_1(\cdot), \cdots, \sigma_k(\cdot))^\intercal$ be any semi-nonparametric dynamic parameter, and denote $\kappa^t = \left(\mu_1^t, \cdots, \mu_k^t, \sigma_1^t, \cdots, \sigma_k^t\right)^\intercal$ as the realized values at $\mathcal{F}^{t-1}$, which is $\mathcal{F}^{t-1}$-measurable. Let $\alpha = (\theta^\intercal, f_1, \cdots, f_k)^\intercal$ be any parameter of the innovation processes. Then the log likelihood of $Y_t$ conditional on $\mathcal{F}^{t-1}$ and $(\kappa, \alpha)$ is

$$l(\alpha, \kappa, Y_t) \equiv \log p\left(Y_t \left| \mathcal{F}^{t-1} \right.; \alpha, \kappa\right) = \log c\left[F_1\left(\frac{Y_{1,t} - \mu_1^t}{\sigma_1^t}\right), \cdots, F_k\left(\frac{Y_{k,t} - \mu_k^t}{\sigma_k^t}\right); \theta\right] + \sum_{j=1}^{k} \log f_j\left(\frac{Y_{j,t} - \mu_j^t}{\sigma_j^t}\right) - \sum_{j=1}^{k} \log \sigma_j^t, \tag{3}$$

where $F_j$ is the corresponding cdf of $f_j$, for $j = 1, \cdots, k$.

If $(\kappa_0, \alpha_0)$ were parametrically specified, then the parameters could be estimated simultaneously by maximizing the full log (conditional) likelihood of $\{Y_t\}_{t=1}^n$: $\sum_{t=1}^n l(\alpha, \kappa, Y_t)$. For semi-nonparametric model (1)-(2), however, it is much easier to estimate $\kappa_0$ and $\alpha_0$ sequentially. In the first stage, we can estimate $\kappa_0$ via (1) by any nonparametric methods for the conditional mean and conditional variance functions. See Appendix B for sieve quasi-maximum likelihood (QML) estimation and sieve least squares estimation of $\kappa_0$. Also see the spline-GARCH regressions in Liu and Yang (2016), the local polynomial estimation of nonparametric ARCH in Neumeyer et al. (2019), the kernel estimation of semi-nonparametric ARCH in Linton and Mammen (2005), the nonparametric method in Meister and Krei$\beta$ (2016). Let $\widehat{\kappa} = (\widehat{\mu}_1, \cdots, \widehat{\mu}_k, \widehat{\sigma}_1, \cdots, \widehat{\sigma}_k)^\intercal$ denote a semi-nonparametric consistent estimator of $\kappa_0$, and

$$\widehat{\xi}_{j,t} = \frac{Y_{j,t} - \widehat{\mu}_j^t}{\widehat{\sigma}_j^t} \quad for\ j = 1, \cdots, k \quad and\ t = 1, \cdots, n$$

denote the semi-nonparametric GARCH filtered residuals. In the second stage, we could estimate $\alpha_0$ using the filtered residuals. We consider two kinds of estimation methods for $\alpha_0$ in this paper: the joint SML estimation of $\alpha_0$ using filtered residuals in Section 2.2.1; and the semiparametric two-step estimation of $\alpha_0$ in Section 2.2.2.

#### 2.2.1 Joint SML Estimation for $\alpha_0$

**SML Estimation using Semi-nonparametric GARCH Filtered Residuals**

Plugging $\left\{\widehat{\xi}_{j,t},\ 1 \le j \le k,\ 1 \le t \le n\right\}$ (or equivalently $\widehat{\kappa}$) into the full log likelihood (3), we obtain (up to a constant term)[3] :

$$l\left(\alpha, \widehat{\kappa}, Y_t\right) = \log c\left[F_1\left(\widehat{\xi}_{1,t}\right), \cdots, F_k\left(\widehat{\xi}_{k,t}\right); \theta\right] + \sum_{j=1}^{k} \log f_j\left(\widehat{\xi}_{j,t}\right). \tag{4}$$

Averaging Eq.(4) over $Y_t$ results in the estimated likelihood $\frac{1}{n}\sum_{t=1}^{n} l\left(\alpha, \widehat{\kappa}, Y_t\right)$ for $\alpha$. We propose $\widetilde{\alpha}_{sml}$ as the SML estimator of $\alpha_0 = (\theta_0^\intercal, f_{0,1}, \cdots, f_{0,k})^\intercal$ using filtered residuals :

$$\widetilde{\alpha}_{sml} = \left(\widetilde{\theta}_{sml}^\intercal, \widetilde{f}_{1,sml}, \cdots, \widetilde{f}_{k,sml}\right)^\intercal = \arg max_{\alpha \in \mathcal{A}_n} \frac{1}{n}\sum_{t=1}^{n} l\left(\alpha, \widehat{\kappa}, Y_t\right). \tag{5}$$

The sieve space $\mathcal{A}_n$ is used to approximate the infinite dimensional parameter space $\mathcal{A}$ defined in Assumption 2. See Section 3 for more details about $\mathcal{A}_n$. We note that $\widetilde{\alpha}_{sml}$ does not impose the restriction that $\widehat{\xi}_{j,t}$ has zero mean and unit variance, which is more robust to the estimation error and model misspecification. Section 3 will study the theoretical properties of $\widetilde{\alpha}_{sml}$ thoroughly.

**Infeasible SML Estimation using True Innovations**

For comparison, we introduce the infeasible SML estimation of $\alpha_0$ assuming that the true innovations

[3]With abuse of notation, we denote $l\left(\alpha, \kappa, Y_t\right) \equiv \log c\left[F_1\left(\frac{Y_{1,t}-\mu_1^t}{\sigma_1^t}\right), \cdots, F_k\left(\frac{Y_{k,t}-\mu_k^t}{\sigma_k^t}\right); \theta\right] + \sum_{j=1}^{k} \log f_j\left(\frac{Y_{j,t}-\mu_j^t}{\sigma_j^t}\right)$, except for the full model (3).

$\{\xi_t\}_{t=1}^n$ were observed (or equivalently $\kappa_0$ were known). In this case, $l(\alpha, \kappa_0, Y_t)$ (defined in Eq.(3)) is the exact log likelihood of $\xi_t$ and $\alpha$ (up to a constant term) :

$$l(\alpha, \kappa_0, Y_t) = l(\alpha, \xi_t) \equiv \log c\left[F_1(\xi_{1,t}), \cdots, F_k(\xi_{k,t}); \theta\right] + \sum_{j=1}^{k} \log f_j(\xi_{j,t}). \tag{6}$$

Denote $\widehat{\alpha}_{sml}$ as the infeasible SML estimator of $\alpha_0$ :

$$\widehat{\alpha}_{sml} = \left(\widehat{\theta}_{sml}^{\intercal}, \widehat{f}_{1,sml}, \cdots, \widehat{f}_{k,sml}\right)^{\intercal} = \arg\max_{\alpha \in \mathcal{A}_n} \frac{1}{n} \sum_{t=1}^{n} l(\alpha, \kappa_0, Y_t). \tag{7}$$

Since $\widetilde{\alpha}_{sml}$ will be compared to this infeasible estimator in terms of asymptotic variances, $\widehat{\alpha}_{sml}$ is implemented without imposing that $\xi_t$ has zero mean and unit variance. We note that $\widehat{\alpha}_{sml}$ is the SML estimator of $\alpha_0$ proposed in Chen et al. (2006) assuming i.i.d. data $\{\xi_t\}_{t=1}^n$.

### 2.2.2 Semiparametric Two-step Estimation for $\alpha_0$

**Two-step Estimation using Semi-nonparametric GARCH Filtered Residuals**

(*Step 1*) The unknown marginal $F_{0,j}$ is estimated by the rescaled empirical distribution function,

$$\widetilde{F}_{j,2s}(z_j) = \frac{1}{n+1} \sum_{t=1}^{n} \mathbb{I}\left\{\widehat{\xi}_{j,t} \leq z_j\right\}, \quad for\ j = 1, \cdots, k, \tag{8}$$

where $\left\{\widehat{\xi}_t = \left(\widehat{\xi}_{1,t}, \cdots, \widehat{\xi}_{k,t}\right)^{\intercal}, 1 \leq t \leq n\right\}$ are semi-nonparametric GARCH filtered residual series from Stage 1. The notation $\mathbb{I}$ stands for the indicator function. For $j = 1, \cdots, k$ and $t = 1, \cdots, n$, denote the transformed series as

$$\widetilde{U}_{j,t} := \widetilde{F}_{j,2s}(\widehat{\xi}_{j,t}) = \frac{1}{n+1} \sum_{s=1}^{n} \mathbb{I}\left(\widehat{\xi}_{j,s} \leq \widehat{\xi}_{j,t}\right), \quad and \quad \widetilde{U}_t = \left(\widetilde{U}_{1,t}, \cdots, \widetilde{U}_{k,t}\right)^{\intercal}. \tag{9}$$

Note that $\left\{\widetilde{U}_t, 1 \leq t \leq n\right\}$ are rescaled rank statistics and take values from $\left\{\frac{1}{n+1}, \cdots, \frac{n}{n+1}\right\}$.

(*Step 2*) $\left\{\widetilde{U}_t, 1 \leq t \leq n\right\}$ are used to estimate the copula parameter by maximizing the pseudo likelihood

$$\widetilde{\theta}_{2s} = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \log c\left(\widetilde{U}_{1,t}, \cdots, \widetilde{U}_{k,t}; \theta\right). \tag{10}$$

Section 4 will study the theoretical properties of $\widetilde{\theta}_{2s}$ and $\widetilde{F}_{j,2s}(x)$ for any fixed $x \in \Xi_j$.

**Infeasible Two-step Estimation using True Innovations**

For comparison, we also consider the infeasible two-step estimation procedure assuming that the true innovations $\{\xi_t\}_{t=1}^n$ were observed. Step 1, compute the rescaled rank statistics of $\{\xi_t\}_{t=1}^n$ as

$$\widehat{U}_{j,t} := \frac{1}{n+1}\sum_{s=1}^{n} \mathbb{I}\left(\xi_{j,s} \leq \xi_{j,t}\right), \quad \widehat{U}_t = \left(\widehat{U}_{1,t}, \cdots, \widehat{U}_{k,t}\right)^{\intercal}, \text{ for } j = 1, \cdots, k \text{ and } t = 1, \cdots, n. \tag{11}$$

Step 2: $\left\{\widehat{U}_t, 1 \leq t \leq n\right\}$ are used to estimate $\theta_0 \in \Theta$ by maximizing the pseudo likelihood

$$\widehat{\theta}_{2s} = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} \log c\left(\widehat{U}_{1,t}, \cdots, \widehat{U}_{k,t}; \theta\right). \tag{12}$$

We note that $\widehat{\theta}_{2s}$ is the rank based estimator proposed in Genest et al. (1995) assuming i.i.d. data $\{\xi_t\}_{t=1}^n$.

# 3 Asymptotic Properties of $\widetilde{\theta}_{sml}$

In this section, we first derive the asymptotic properties of $\widetilde{\alpha}_{sml} = \left(\widetilde{\theta}_{sml}^{\intercal}, \widetilde{f}_{1,sml}, \cdots, \widetilde{f}_{k,sml}\right)^{\intercal}$ (defined in Eq.(5)). And then we establish the asymptotic normality and the semiparametric efficiency of $\widetilde{\theta}_{sml}$. Let $U_t^0 = \left(U_{1,t}^0, \cdots, U_{k,t}^0\right)^{\intercal} = \left(F_{0,1}\left(\xi_{1,t}\right), \cdots, F_{0,k}\left(\xi_{k,t}\right)\right)^{\intercal}$, for $t = 1, \cdots, n$. Denote $c(U_t^0, \theta_0) = c\left[F_{0,1}\left(\xi_{1,t}\right), \cdots, F_{0,k}\left(\xi_{k,t}\right); \theta_0\right]$.

## 3.1 Asymptotic Properties of $\widetilde{\alpha}_{sml}$

### 3.1.1 Convergence Rate of $\widetilde{\alpha}_{sml}$

**Assumption 2.** *(*i*) $\theta_0 \in int(\Theta)$, where $\Theta$ is a compact subset of $\mathcal{R}^{d_\theta}$, and $c(u;\theta) > 0$ for all $u \in (0,1)^k$, $\theta \in \Theta$ ; (*ii*) for $j = 1, \cdots, k$, $f_{0,j} \in \mathcal{F}_j$, where either $\mathcal{F}_j = \{f_j = \exp(g) : g \in \Lambda^{\rho_j}(\Xi_j), \int \exp(g(x))dx = 1, \int x^2 \exp(g(x))dx < \infty\}$ or $\mathcal{F}_j = \{f_j = g^2 > 0 : g \in \Lambda^{\rho_j}(\Xi_j), \int g^2(x)dx = 1, \int x^2 g^2(x)dx < \infty\}$, $\rho_j > 1$; (*iii*) $\alpha_0 = (\theta_0^{\intercal}, f_{0,1}, \cdots, f_{0,k})^{\intercal}$ is the unique maximizer of $E\left[l\left(\alpha, \kappa_0, Y_t\right)\right]$ over $\mathcal{A} = \Theta \times \prod_{j=1}^{k} \mathcal{F}_j$.*

Let $[\rho_j] \geq 0$ be the largest integer such that $[\rho_j] < \rho_j$. A real-valued function $g$ on $\Xi_j$ is said to be $\rho_j$-smooth if it is $[\rho_j]$ times continuously differentiable on $\Xi_j$, and its $[\rho_j]$th derivative satisfies a Hölder condition with exponent $\rho_j - [\rho_j] \in (0, 1]$. We denote $\Lambda^{\rho_j}(\Xi_j)$ as the class of all real-valued functions on $\Xi_j$ which are $\rho_j$-smooth, and it is called a Hölder space. Therefore, Assumption 2 (ii) imposes smoothness condition on the unknown marginal densities. $\rho_j > 1$ implies that $\forall f_j \in \mathcal{F}_j$ is continuously differentiable. Assumption 2 (iii) is the identification condition. The infinite dimensional parameter space $\mathcal{A}$ is approximated by the sieve space $\mathcal{A}_n = \Theta \times \prod_{j=1}^{k} \mathcal{F}_{j,n}$[4]. Let $\mathcal{N}(\theta_0) = \{\theta \in \Theta : \|\theta - \theta_0\| < \epsilon\}$ be a local neighborhood of $\theta_0$, for some positive constant $\epsilon > 0$.

**Assumption 3.** *For $j, m = 1, \cdots, k$: (i) the second-order partial derivative $\frac{\partial^2 \log c(u;\theta)}{\partial\theta\partial\theta^\intercal}$ exists and is continuous in $\mathcal{N}(\theta_0)$ and $u \in [0,1]^k$; (ii) the second-order partial derivative $\frac{\partial^2 \log c(u;\theta)}{\partial u_j \partial\theta}$ exists and is continuous in $\mathcal{N}(\theta_0)$ and $\left\{u \in [0,1]^k : 0 < u_j < 1\right\}$; (iii) the second-order partial derivative $\frac{\partial^2 \log c(u;\theta)}{\partial u_j \partial u_m}$ exists and is continuous in $\mathcal{N}(\theta_0)$ and $\left\{u \in [0,1]^k : 0 < u_j < 1,\ 0 < u_m < 1\right\}$.*

Denote $\mathbb{V}$ as the linear span of $\mathcal{A} - \{\alpha_0\}$. Under Assumption 3, for any $v = \left(v_\theta^\intercal, v_1, \cdots, v_k\right)^\intercal \in \mathbb{V}$, we have that $l(\alpha_0 + \tau v, \kappa_0, Y_t)$ is continuously differentiable in small $\tau \in [0,1]$ : $\left.\frac{dl(\alpha_0+\tau v,\kappa_0,Y_t)}{d\tau}\right|_{\tau=0} = \frac{\partial l(\alpha_0,\kappa_0,Y_t)}{\partial\alpha^\intercal}[v] = \frac{\partial \log c(U_t^0,\theta_0)}{\partial\theta^\intercal} v_\theta + \sum_{j=1}^{k} \left\{ \frac{\partial \log c(U_t^0,\theta_0)}{\partial u_j} \int_{-\infty}^{\xi_{j,t}} v_j(x)dx + \frac{v_j(\xi_{j,t})}{f_{0,j}(\xi_{j,t})} \right\}$. Define the Fisher inner product and norm on the space $\mathbb{V}$ as

$$< v, \widetilde{v} > \equiv E\left[\left(\frac{\partial l(\alpha_0, \kappa_0, Y_t)}{\partial\alpha^\intercal}[v]\right)\left(\frac{\partial l(\alpha_0, \kappa_0, Y_t)}{\partial\alpha^\intercal}[\widetilde{v}]\right)\right], \quad \|v\|^2 = < v, v >, \quad \forall v,\ \widetilde{v} \in \mathbb{V}. \tag{13}$$

The closure of $\mathbb{V}$ under $\|\cdot\|$ is $\overline{\mathbb{V}} = \left\{v = \left(v_\theta^\intercal, v_1, \cdots, v_k\right)^\intercal \in \mathcal{R}^{d_\theta} \times \prod_{j=1}^{k} \overline{\mathbb{V}}_j : \|v\| < \infty\right\}$ with

$$\overline{\mathbb{V}}_j = \left\{ v_j \in \Lambda^{\rho_j}(\Xi_j) : E\left(\frac{v_j(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right) = 0,\ E\left(\frac{v_j(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right)^2 < \infty \right\}. \tag{14}$$

Denote $\mathcal{B} = \prod_{j=1}^{k} \mathcal{M}_j \times \prod_{j=1}^{k} \mathcal{H}_j$ (see Appendix B for details) as the dynamic parameter space for $\kappa$, and $\mathcal{B}_n = \prod_{j=1}^{k} \mathcal{M}_{j,n} \times \prod_{j=1}^{k} \mathcal{H}_{j,n} \subset \mathcal{B}$ as the associated sieve space. Assume $\kappa_0 \in \mathcal{B}$, satisfying

$$\kappa_{0,j} = (\mu_{0,j}, \sigma_{0,j})^\intercal = \underset{\mu_j \in \mathcal{M}_j, \sigma_j \in \mathcal{H}_j}{\arg\min} E[\Psi(\kappa_j, Y_{j,t})], \quad for\ j = 1, \cdots, k, \tag{15}$$

[4] $\mathcal{F}_{j,n} = \left\{ f_{j,n}(x) = \left[\sum_{m=1}^{K_n} a_m A_m(x)\right]^2, \int f_{j,n}(x)dx = 1 \right\}$ if sieves are used to approximate square root density, or $\mathcal{F}_{j,n} = \left\{ f_{j,n}(x) = exp\left[\sum_{m=1}^{K_n} a_m A_m(x)\right], \int f_{j,n}(x)dx = 1 \right\}$ if sieves are used to approximate log density. See Chen (2007) for details on sieve spaces.

where $\Psi$ could be any criterion function in Stage 1, e.g. the sieve QML criterion. Let $\widehat{\kappa} \in \mathcal{B}_n$ be any semi-nonparametric estimator of $\kappa_0$ using the sample analogs of Eq.(15). Denote $\mathbb{W}$ as the linear span of $\mathcal{B} - \{\kappa_0\}$.

**Assumption 4.** $\|\widehat{\kappa} - \kappa_0\| = O_p\left(\delta_{h,n}\right) = o_p\left(n^{-1/4}\right)$.

The norm $\|\cdot\|$ on $\mathbb{W}$ (defined in Eq.(B.3)) is based on the criterion function Eq.(15). This class of norms are usually weaker than the sup or $L_2$ norm on an infinitely dimensional parameter space. For a parametric estimator $\widehat{\kappa}$, the convergence rate is well known to be $n^{-1/2}$. For a semi-nonparametric estimator, under the weak norm $\|\cdot\|$, Assumption 4 is mild. There exist many results on the convergence rates of the kernel estimates, the local polynomial estimates, the spline estimates; see e.g., Buhlmann and McNeil (2002), Linton and Mammen (2005), Liu and Yang (2016), Meister and Krei$\beta$ (2016), Neumeyer et al. (2019). Also see Appendix B.1 and B.2 for detailed illustrations of a fully nonparametric dynamic model and a semi-nonparametric GARCH model, respectively.

**Lemma 1.** *Suppose $\widehat{\alpha}_{sml}$ satisfies C.1(i). Under Assumptions 1, 2, 3, 4 and Assumptions C.2, C.3 stated in the Appendix, we have*

$$\|\widetilde{\alpha}_{sml} - \alpha_0\| = O_p\left(\widetilde{\delta}_{\alpha,n}\right) = o_p\left(n^{-1/4}\right).$$

Lemma 1 only provides a loose bound for the convergence rate, which is enough for the derivation of our main theorems. For a sharper rate, it involves more tedious calculations and proofs.

### 3.1.2 Asymptotic Normality of $\rho\left(\widetilde{\alpha}_{sml}\right)$

Let $\rho$ be a smooth functional on $\mathcal{A}$ and satisfy Assumption C.1 (iii). For any $v \in \mathbb{V}$, we denote $\frac{\partial \rho(\alpha_0)}{\partial \alpha^\intercal}[v] = \lim_{\tau \to 0} \frac{\rho(\alpha_0 + \tau v) - \rho(\alpha_0)}{\tau}$. There exists a Riesz representer $v^* \in \overline{\mathbb{V}}$, such that

$$\frac{\partial \rho(\alpha_0)}{\partial \alpha^\intercal}[v] =< v, v^* >, \forall v \in \mathbb{V}; \qquad \|v^*\|^2 = \left\|\frac{\partial \rho(\alpha_0)}{\partial \alpha^\intercal}\right\|^2 = \sup_{v \in \mathbb{V}: \|v\| > 0} \frac{\left|\frac{\partial \rho(\alpha_0)}{\partial \alpha^\intercal}[v]\right|^2}{\|v\|^2}. \tag{16}$$

To establish the asymptotic normality of $\rho\left(\widetilde{\alpha}_{sml}\right)$, we introduce the following correction term to capture the estimation error :

$$\sqrt{n}\Gamma\left(\alpha_0,\kappa_0\right)\left[v^*,\widehat{\kappa}-\kappa_0\right]\equiv\sqrt{n}\mathbb{E}_X\left[\left.\frac{\partial^2 l\left(\alpha_0+\tau_1 v^*,\kappa_0+\tau_2\left(\widehat{\kappa}-\kappa_0\right),Y_t\right)}{\partial\tau_1\partial\tau_2}\right|_{\tau_1=0\ \tau_2=0}\right]= \tag{17}$$

$$\sum_{j=1}^{k}\left\{\begin{array}{c} E\left(\begin{array}{l}\frac{\partial^2\log c(U_t^0,\theta_0)}{\partial u_j\partial\theta^\intercal}v_\theta^* f_{0,j}\left(\xi_{j,t}\right)+\frac{v_j^{*\prime}(\xi_{j,t})f_{0,j}(\xi_{j,t})-v_j^*(\xi_{j,t})f_{0,j}'(\xi_{j,t})}{f_{0,j}^2(\xi_{j,t})}\\ +\frac{\partial\log c(U_t^0,\theta_0)}{\partial u_j}v_j^*\left(\xi_{j,t}\right)+\sum_{m=1}^k\frac{\partial^2\log c(U_t^0,\theta_0)}{\partial u_m\partial u_j}\int_{-\infty}^{\xi_{m,t}}v_m^*(x)dx f_{0,j}\left(\xi_{j,t}\right)\end{array}\right)\sqrt{n}\mathbb{E}_X\left(-\frac{\widehat{\mu}_j^t-\mu_{0,j}^t}{\sigma_{0,j}^t}\right)\\ +E\left(\begin{array}{l}\frac{\partial^2\log c(U_t^0,\theta_0)}{\partial u_j\partial\theta^\intercal}v_\theta^* f_{0,j}\left(\xi_{j,t}\right)\xi_{j,t}+\frac{v_j^{*\prime}(\xi_{j,t})f_{0,j}(\xi_{j,t})-v_j^*(\xi_{j,t})f_{0,j}'(\xi_{j,t})}{f_{0,j}^2(\xi_{j,t})}\xi_{j,t}\\ +\frac{\partial\log c(U_t^0,\theta_0)}{\partial u_j}v_j^*\left(\xi_{j,t}\right)\xi_{j,t}+\sum_{m=1}^k\frac{\partial^2\log c(U_t^0,\theta_0)}{\partial u_m\partial u_j}\int_{-\infty}^{\xi_{m,t}}v_m^*(x)dx f_{0,j}\left(\xi_{j,t}\right)\xi_{j,t}\end{array}\right)\\ \times\sqrt{n}\mathbb{E}_X\left(-\frac{\widehat{\sigma}_j^t-\sigma_{0,j}^t}{\sigma_{0,j}^t}\right)\end{array}\right\},$$

where the operator $\mathbb{E}_X$ is defined in Eq.(18), $\widehat{\kappa}-\kappa_0=(\widehat{\mu}_1-\mu_{0,1},\cdots,\widehat{\mu}_k-\mu_{0,k},\widehat{\sigma}_1-\sigma_{0,1},\cdots,\widehat{\sigma}_k-\sigma_{0,k})^\intercal$, and the evaluated values of the semi-nonparametric dynamic parameters at $\mathcal{F}^{t-1}$ are abbreviated to the associated parameters with a superscript $t$. See Appendix C.1 for detailed derivation of Eq.(17). In Eq.(17), the first term in the braces equals the product of a nonrandom multiplier (determined by $v^*$) and the random term $\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}_j^t-\mu_{0,j}^t}{\sigma_{0,j}^t}\right)$, which characterizes the effect caused by estimating the conditional mean function. Similarly the second term in the braces quantifies the effect caused by estimating the volatility function. Eq.(17) is sensible only if for $j=1,\cdots,k$, the random terms

$$\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}_j^t-\mu_{0,j}^t}{\sigma_{0,j}^t}\right)=\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}_j\left(\mathcal{F}^{t-1}\right)-\mu_{0,j}\left(\mathcal{F}^{t-1}\right)}{\sigma_{0,j}\left(\mathcal{F}^{t-1}\right)}\right)\equiv\sqrt{n}\int\frac{\widehat{\mu}_j\left(x\right)-\mu_{0,j}\left(x\right)}{\sigma_{0,j}\left(x\right)}p_0(x)dx,$$
$$\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}_j^t-\sigma_{0,j}^t}{\sigma_{0,j}^t}\right)=\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}_j\left(\mathcal{F}^{t-1}\right)-\sigma_{0,j}\left(\mathcal{F}^{t-1}\right)}{\sigma_{0,j}\left(\mathcal{F}^{t-1}\right)}\right)\equiv\sqrt{n}\int\frac{\widehat{\sigma}_j\left(x\right)-\sigma_{0,j}\left(x\right)}{\sigma_{0,j}\left(x\right)}p_0(x)dx, \tag{18}$$

are well defined, where $p_0$[5] is the true (unknown) density of $\mathcal{F}^{t-1}$. Eq.(18) emphasizes that $\mathbb{E}_X$ denotes expectation, treating any plug-in estimator (e.g. $\widehat{\mu}_j\left(\cdot\right)$ and $\widehat{\sigma}_j\left(\cdot\right)$) as deterministic. $\mathbb{E}_X$ will coincide with the standard expectation, when no plug-in estimator is involved. Note that $\kappa$ could include both the finite dimensional parameter and the infinite dimensional parameter. See Appendix B.2 for an illustration where Eq.(18) is explicitly expressed in a semi-nonparametric setup.

[5]The dimension and the support of $x$, and $p_0$ depend on the model setup of the dynamic parameter, e.g. which past information is included for constructing the conditional mean and variance functions.

**Assumption 5.** *(*i*) For* $j = 1, \cdots, k$*,* $\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}_j^t - \mu_{0,j}^t}{\sigma_{0,j}^t}\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^n\left(-\frac{\partial\Psi(\kappa_{0,j},Y_{j,t})}{\partial(\mu_j,\sigma_j)^\intercal}w_{\mu_j}^*\right) + o_p(1)$*,* $\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}_j^t - \sigma_{0,j}^t}{\sigma_{0,j}^t}\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^n\left(-\frac{\partial\Psi(\kappa_{0,j},Y_{j,t})}{\partial(\mu_j,\sigma_j)^\intercal}w_{\sigma_j}^*\right) + o_p(1)$*, where* $w_{\mu_j}^*$ *and* $w_{\sigma_j}^*$ *are the associated Riesz representers ; (*ii*)* $\frac{1}{\sqrt{n}}\sum_{t=1}^n S_{\rho(\alpha)}(\alpha_0, \kappa_0, Y_t) + \sqrt{n}\Gamma(\alpha_0, \kappa_0)[v^*, \widehat{\kappa} - \kappa_0] \xrightarrow{\mathcal{D}} \mathcal{N}(0, Asyvar[\rho(\widetilde{\alpha}_{sml})])$*, where* $Asyvar[\rho(\widetilde{\alpha}_{sml})]$ *is given in Eq.*(19)*.*

$\Psi$ (defined in Eq.(15)) is the criterion function used for estimation in Stage 1. Assumption 5(i) is a standard result in the semi-nonparametric literature, since the operator $\mathbb{E}_X$ is a smooth functional of $\widehat{\mu}_j$ and $\widehat{\sigma}_j$. We justify Assumption 5(i) in Appendix B.1 for a fully nonparametric dynamic model, and in Appendix B.2 for a semi-nonparametric GARCH model, respectively. According to Eq.(17), $\sqrt{n}\Gamma(\alpha_0, \kappa_0)[v^*, \widehat{\kappa} - \kappa_0]$ is a weighted sum of the random terms $\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}_j^t - \mu_{0,j}^t}{\sigma_{0,j}^t}\right)$ and $\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}_j^t - \sigma_{0,j}^t}{\sigma_{0,j}^t}\right)$ for $j = 1, \cdots, k$, where the nonrandom weights are determined by $v^*$ (equivalently by the smooth functional $\rho$ of interest). Thus Assumption 5(ii) is implied by the triangle array CLT.

**Theorem 1.** *Under Assumptions 1, 2, 3, 4, 5 and Assumptions C.1, C.2, C.3, C.4 stated in the Appendix, we have* $\sqrt{n}[\rho(\widetilde{\alpha}_{sml}) - \rho(\alpha_0)] \xrightarrow{\mathcal{D}} \mathcal{N}(0, Asyvar[\rho(\widetilde{\alpha}_{sml})])$*, where*

$$Asyvar[\rho(\widetilde{\alpha}_{sml})] = \lim_{n\to\infty} Var\left[n^{-1/2}\sum_{t=1}^n S_{\rho(\alpha)}(\alpha_0, \kappa_0, Y_t) + \sqrt{n}\Gamma(\alpha_0, \kappa_0)[v^*, \widehat{\kappa} - \kappa_0]\right]. \tag{19}$$

$S_{\rho(\alpha)}(\alpha_0, \kappa_0, Y_t)$ *is defined in Remark 1,* $v^*$ *is the Riesz representer of* $\rho(\alpha_0)$ *defined in Eq.*(16)*, and* $\Gamma(\alpha_0, \kappa_0)[v^*, \widehat{\kappa} - \kappa_0]$ *is defined in Eq.*(17)*.*

**Remark 1.** When $\{\xi_t\}$ were observed (or $\kappa$ were known to be $\kappa_0$), $S_{\rho(\alpha)}(\alpha_0, \kappa_0, Y_t)$ is the efficient influence function[6] for $\rho(\alpha_0)$, then $Var\left[S_{\rho(\alpha)}(\alpha_0, \kappa_0, Y_t)\right]$ is the asymptotically minimum variance (Fisher's lower bound) for $\rho(\alpha_0)$ in Model (6). Furthermore, $\rho(\widehat{\alpha}_{sml})$ is semiparametrically efficient with asymptotic variance $Var\left[S_{\rho(\alpha)}(\alpha_0, \kappa_0, Y_t)\right]$ (see Theorem 1 of Chen et al. (2006)).

**Remark 2.** The correction term can also be interpreted as

$$\Gamma(\alpha_0, \kappa_0)[v^*, \widehat{\kappa} - \kappa_0] = \mathbb{E}_X\left[\frac{\partial S_{\rho(\alpha)}(\alpha_0, \kappa_0 + \tau(\widehat{\kappa} - \kappa_0), Y_t)}{\partial\tau}\right]_{\tau=0}. \tag{20}$$

[6]See the definitions of efficient influence functions and score functions in Bickel and Kwon (2001) and the references therein.

In general, $\Gamma(\alpha_0, \kappa_0)[v^*, \widehat{\kappa} - \kappa_0] \neq 0$, thus $Asyvar[\rho(\widetilde{\alpha}_{sml})] \neq Asyvar[\rho(\widehat{\alpha}_{sml})]$. Therefore, asymptotically valid standard error for $\rho(\alpha_0)$ requires a correction.

Note that $S_{\rho(\alpha)}(\alpha_0, \kappa_0, Y_t) = \frac{\partial l(\alpha_0, \kappa_0, Y_t)}{\partial \alpha^\intercal} v^*$ (see the proof of Theorem 1 or the proof of Theorem 1 in Chen et al. (2006)). Also note that $\sqrt{n}\Gamma(\alpha_0, \kappa_0)[v^*, \widehat{\kappa} - \kappa_0]$ is a linear combination of the random terms defined in Eq.(18), where the nonrandom weights are determined by $v^*$. However, in general, there is no closed-form solution for $v^*$, thus we first need a consistent estimator of it. In the following, we suggest a procedure to evaluate $Asyvar[\rho(\widetilde{\alpha}_{sml})]$ :

(*Step 1*) Following Section 3.3.1. of Chen et al. (2014), we can estimate $v^*$ by $v_T^*$ (the sieve Riesz representer). They provide closed form expressions for $v_T^*$ and $\frac{\partial l(\alpha_0, \kappa_0, Y_t)}{\partial \alpha^\intercal} v_T^*$. See Chen et al. (2014) for more details.

(*Step 2*) There are $2k$ nonrandom weights in Eq.(17). For example,

$$E\left(\begin{array}{c} \frac{\partial^2 \log c(U_t^0, \theta_0)}{\partial u_j \partial \theta^\intercal} v_\theta^* f_{0,j}(\xi_{j,t}) + \frac{v_j^{*\prime}(\xi_{j,t}) f_{0,j}(\xi_{j,t}) - v_j^*(\xi_{j,t}) f_{0,j}'(\xi_{j,t})}{f_{0,j}^2(\xi_{j,t})} \\ + \frac{\partial \log c(U_t^0, \theta_0)}{\partial u_j} v_j^*(\xi_{j,t}) + \sum_{m=1}^k \frac{\partial^2 \log c(U_t^0, \theta_0)}{\partial u_m \partial u_j} \int_{-\infty}^{\xi_{m,t}} v_m^*(x) dx f_{0,j}(\xi_{j,t}) \end{array}\right),$$

can be consistently estimated by replacing the population mean with the sample analog. And the unknown true values could be replaced by $v_T^*$ from (*Step 1*) and our SML estimates. The other $2k - 1$ weights can be estimated similarly.

(*Step 3*) The random terms defined in Eq.(18) can be expressed explicitly. See Lemmas B.1, B.2, B.3 in Appendix B for detailed descriptions.

(*Step 4*) Then $Asyvar[\rho(\widetilde{\alpha}_{sml})]$ can be evaluated by taking into account the possible autocorrelations in Eq.(19). Replace the population moments and the unknown true values with the sample analogs and our SML estimates when it is needed.

## 3.2 $\sqrt{n}$ Normality of $\widetilde{\theta}_{sml}$

In practice, the copula parameter $\theta_0$ is often of primary interest. Thus we will unfold Eqs.(19) and (20) in this section when $\rho(\alpha) = \lambda^\intercal \theta$, for any fixed $\lambda \in \mathcal{R}^{d_\theta}$ with $0 < ||\lambda|| < \infty$. Let $\theta = (\theta_1, \cdots, \theta_{d_\theta})^\intercal$.

For comparison, we first present the semiparametric efficient information of $\theta_0$ in the infeasible model (6) ($\kappa_0$ is known), with $\xi_t$'s marginal distributions completely unspecified. Let

$$\mathcal{L}_2^0([0,1]) = \{b \in \Lambda^{\rho_j}([0,1]) : \int_0^1 b(v)dv = 0, \quad \int_0^1 b^2(v)dv < \infty\}.$$

$$\text{Denote} \quad B^* = \left[b_{j,l}^*\right] \in \left\{\mathcal{L}_2^0([0,1])\right\}^{k\times d_\theta}, \quad j = 1, \cdots, k \text{ and } l = 1, \cdots, d_\theta. \tag{21}$$

$B^*$ is a $k$ by $d_\theta$ matrix, each element of which belongs to $\mathcal{L}_2^0([0,1])$. For $l = 1, \cdots, d_\theta$, each $lth$ column of $B^*$, i.e. $b_{\cdot,l}^* = \left(b_{1,l}^*, \cdots, b_{k,l}^*\right)^\intercal$ solves the following infinite-dimensional optimization problem :

$$\inf_{\left(b_{1,l},\cdots,b_{k,l}\right)\in\left\{\mathcal{L}_2^0([0,1])\right\}^k} E\left(\frac{\partial \log c\left(U_t^0,\theta_0\right)}{\partial \theta_l} - \sum_{j=1}^k \left[\frac{\partial \log c\left(U_t^0,\theta_0\right)}{\partial u_j}\int_0^{U_{j,t}^0} b_{j,l}(u)du + b_{j,l}\left(U_{j,t}^0\right)\right]\right)^2. \tag{22}$$

The efficient score of $\theta_0$ in the infeasible model (6) can be expressed in terms of $B^*$[7]:

$$S_{\theta_0}\left(U_t^0,\alpha_0,\kappa_0\right) = \frac{\partial \log c\left(U_t^0,\theta_0\right)}{\partial \theta} - \sum_{j=1}^k \left[\frac{\partial \log c\left(U_t^0,\theta_0\right)}{\partial u_j}\int_0^{U_{j,t}^0} b_{j,\cdot}^{*\intercal}(u)du + b_{j,\cdot}^{*\intercal}\left(U_{j,t}^0\right)\right], \tag{23}$$

where $b_{j,\cdot}^*$ is the $jth$ row of $B^*$. Then the semiparametric Fisher information bound for $\theta_0$ is equal to $\mathcal{I}_*(\theta_0) = E\left(S_{\theta_0}\left(U_t^0,\alpha_0,\kappa_0\right)S_{\theta_0}\left(U_t^0,\alpha_0,\kappa_0\right)^\intercal\right)$. The efficient influence function for $\rho(\alpha_0) = \lambda^\intercal\theta_0$ is $S_{\lambda^\intercal\theta}(\alpha_0,\kappa_0,Y_t) \equiv v_\theta^{*\intercal}S_{\theta_0}\left(U_t^0,\alpha_0,\kappa_0\right)$, with $v_\theta^* = \mathcal{I}_*(\theta_0)^{-1}\lambda$. Thus the asymptotically minimum variance for $\rho(\alpha_0) = \lambda^\intercal\theta_0$ is $\lambda^\intercal\mathcal{I}_*(\theta_0)^{-1}\lambda$, which is achieved by $\lambda^\intercal\widehat{\theta}_{sml}$ (Proposition 1 of Chen et al. (2006)). In the following theorem, we show that $Asyvar\left[\lambda^\intercal\widetilde{\theta}_{sml}\right] = Asyvar\left[\lambda^\intercal\widehat{\theta}_{sml}\right]$– a remarkable property.

**Assumption 6.** $c\left(U_t^0,\theta_0\right)$ *satisfies Assumption* $3'$ *in Chen et al. (2006).*

Under Assumption 6, $\mathcal{I}_*(\theta_0)$ is finite and positive definite. It is a restatement of Assumption C.1 (iii) when $\rho(\alpha) = \lambda^\intercal\theta$.

**Theorem 2.** *Under Assumptions 1, 2, 3, 4, 5, 6 and Assumptions C.1(i)(ii), C.2, C.3, C.4 stated in the Appendix, we have* $\sqrt{n}\left[\widetilde{\theta}_{sml} - \theta_0\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathcal{I}_*(\theta_0)^{-1}\right)$.

[7]There exists a one-to-one mapping between the Riesz representer $v^*$ of $\rho(\alpha) = \lambda^\intercal\theta$ and $B^*$. See page 1233 in Chen et al. (2006) or Eq.(C.4) for details.

Theorem 2 shows $\widetilde{\theta}_{sml}$ and $\widehat{\theta}_{sml}$ are asymptotically equivalent, in terms of the asymptotic variance. This property facilitates the statistical inference of the copula parameter, since no correction term needs to be computed. According to Theorem 1, we have

$$Asyvar\left[\lambda^{\intercal}\widetilde{\theta}_{sml}\right] = v_{\theta}^{*\intercal}\lim_{n\to\infty} Var\left[n^{-1/2}\sum_{t=1}^{n} S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right) + \sqrt{n}\mathbb{E}_X\left[\frac{\partial S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0 + \tau\left(\widehat{\kappa} - \kappa_0\right)\right)}{\partial \tau}\right]_{\tau=0}\right] v_{\theta}^{*}.$$

Therefore, essentially we show in the proof of Theorem 2 that

$$\sqrt{n}\mathbb{E}_X\left[\frac{\partial S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0 + \tau\left(\widehat{\kappa} - \kappa_0\right)\right)}{\partial \tau}\right]_{\tau=0} = 0. \tag{24}$$

Intuitively, the efficient score function $S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right)$ should be asymptotically orthogonal to (some version of) any marginal information, thus invariant to the local perturbation around the true $\kappa_0$.

**Remark 3. Estimation of $\mathcal{I}_*\left(\theta_0\right)$**

Procedure 1: Due to Theorem 2, the asymptotic variance of $\widetilde{\theta}_{sml}$ can be consistently estimated using Proposition 2 of Chen et al. (2006), without any correction. Essentially, one needs to replace the population moments and the unknown true values in Eqs.(22) and (23) with the sample analogs and the estimated values.

Let $\mathcal{L}_n$ be the sieve space used to approximate $\mathcal{L}_2^0$, for example the spline sieve, the polynomial sieve or the trigonometric sieve. Then $\mathcal{I}_*\left(\theta_0\right) = E\left[S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right) S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right)^{\intercal}\right]$ can be consistently estimated by

$$\begin{aligned}\widehat{\Omega}_{\theta} = \frac{1}{n}\sum_{t=1}^{n}\Bigg\{&\left(\frac{\partial \log c(\overline{U}_t, \widetilde{\theta}_{sml})}{\partial \theta} - \sum_{j=1}^{k}\left[\frac{\partial \log c(\overline{U}_t, \widetilde{\theta}_{sml})}{\partial u_j}\int_0^{\overline{U}_{j,t}} \widehat{b}_{j,\cdot}^{*\intercal}(v)dv + \widehat{b}_{j,\cdot}^{*\intercal}(\overline{U}_{j,t})\right]\right)\\ &\times\left(\frac{\partial \log c(\overline{U}_t, \widetilde{\theta}_{sml})}{\partial \theta} - \sum_{j=1}^{k}\left[\frac{\partial \log c(\overline{U}_t, \widetilde{\theta}_{sml})}{\partial u_j}\int_0^{\overline{U}_{j,t}} \widehat{b}_{j,\cdot}^{*\intercal}(v)dv + \widehat{b}_{j,\cdot}^{*\intercal}(\overline{U}_{j,t})\right]\right)^{\intercal}\Bigg\},\end{aligned} \tag{25}$$

where $\widehat{b}_{j,\cdot}^{*}$ is the $jth$ row of $\widehat{B}^{*} = \left[\widehat{b}_{j,l}^{*}\right]$ for $j = 1, \cdots, k$ and $l = 1, \cdots, d_{\theta}$. For $l = 1, \cdots, d_{\theta}$, each $lth$ column of $\widehat{B}^{*}$, i.e. $\widehat{b}_{\cdot,l}^{*} = \left(\widehat{b}_{1,l}^{*}, \cdots, \widehat{b}_{k,l}^{*}\right)^{\intercal}$ is estimated by

$$\widehat{b}_{\cdot,l}^{*} = \underset{(b_{1,l},\cdots,b_{k,l})\in\mathcal{L}_n^k}{\arg\min}\ \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial \log c(\overline{U}_t, \widetilde{\theta}_{sml})}{\partial \theta_l} - \sum_{j=1}^{k}\left[\frac{\partial \log c(\overline{U}_t, \widetilde{\theta}_{sml})}{\partial u_j}\int_0^{\overline{U}_{j,t}} b_{j,l}(v)dv + b_{j,l}(\overline{U}_{j,t})\right]\right)^2,$$

where $\overline{U}_{j,t} = \widetilde{F}_{j,sml}\left(\widehat{\xi}_{j,t}\right)$, $\widetilde{F}_{j,sml}(x) = \int_{-\infty}^{x} \widetilde{f}_{j,sml}(v)dv$, $\widetilde{f}_{j,sml}(\cdot)$ is the sieve MLE of the unknown density $f_{0,j}$ using filtered residuals (defined in Eq.(5)), for $j = 1, \cdots, k$.

Procedure 2 : One could also apply (*Step 1*) and (*Step 4*) of Remark 2 to estimate $\mathcal{I}_*(\theta_0)$.

**Remark 4.** In general, the semiparametric Fisher information bound $\mathcal{I}_*(\theta_0)$ has no closed form solution, except for the Gaussian copula. For the bivariate Gaussian copula with correlation $\theta_0$, the semiparametric information lower bound is $\left(1-\theta_0^2\right)^2$ (the lowest asymptotic variance) (see, e.g. Klaassen and Wellner (1997)). For the multivariate Gaussian copula model, the information bound is obtained in Hoff et al. (2014) and depends only on the copula parameter (Theorem 4.1. of Hoff et al. (2014)). The above results are consistent with Eq.(23) : the semiparametric information bound for the copula parameter is free of the true unknown marginal CDF $F_{0,j}$'s.

### 3.3 Asymptotic Efficiency of $\widetilde{\theta}_{sml}$ in the Full Model

Theorem 2 and Remark 4 motivate us to investigate the semiparametric Fisher information bound for $\theta_0$ in the full model (3). We further address the semiparametric efficiency of $\widetilde{\theta}_{sml}$.

Denote $\mathcal{A}^r = \left\{\alpha \in \mathcal{A} : \int x f_j(x)dx = 0, \int x^2 f_j(x)dx = 1,\ j = 1, \cdots, k\right\}$ and $\mathbb{V}^r$ as the linear span of $\mathcal{A}^r - \{\alpha_0\}$. $\mathcal{A}^r$ imposes the restriction that the candidate density has zero mean and unity variance, otherwise we could not identify $\kappa_0$ and $f_{0,j}$'s separately. Let $\overline{\mathbb{V}^r}$ be the closure of $\mathbb{V}^r$ under the Fisher norm (13). It is easy to see that $\overline{\mathbb{V}^r} = \left\{v = \left(v_\theta^\intercal, v_1, \cdots, v_k\right)^\intercal \in \mathcal{R}^{d_\theta} \times \prod_{j=1}^{k} \overline{\mathbb{V}_j^r} : \|v\| < \infty\right\}$ with

$$\overline{\mathbb{V}_j^r} = \left\{v_j \in \overline{\mathbb{V}}_j : E\left(\frac{\xi_{j,t}^m v_j(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right) = 0,\ \text{for } m = 1, 2\right\}, \text{ where } \overline{\mathbb{V}}_j \text{ is defined in Eq.(14).} \tag{26}$$

Under Assumptions 2(ii) and 3, $\forall v = \left(v_\theta^\intercal, v_1, \cdots, v_k\right)^\intercal \in \mathbb{V}^r$ and $w = (w_{\mu,1}, \cdots, w_{\mu,k}, w_{\sigma,1}, \cdots, w_{\sigma,k})^\intercal \in \mathbb{W}$, we have that $l(\alpha_0 + \tau v, \kappa_0 + \tau w, Y_t)$ is continuously differentiable in small $\tau \in [0, 1]$ :

$$\begin{aligned}
&\left.\frac{dl(\alpha_0 + \tau v, \kappa_0 + \tau w, Y_t)}{d\tau}\right|_{\tau=0} = \frac{\partial l(\alpha_0, \kappa_0, Y_t)}{\partial \alpha^\intercal}[v] + \frac{\partial l(\alpha_0, \kappa_0, Y_t)}{\partial \kappa^\intercal}[w] \\
&= \frac{\partial \log c(U_t^0, \theta_0)}{\partial \theta^\intercal} v_\theta + \sum_{j=1}^{k} \left\{\frac{\partial \log c(U_t^0, \theta_0)}{\partial u_j} \int_{-\infty}^{\xi_{j,t}} v_j(x)dx + \frac{v_j(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right\} \\
&+ \sum_{j=1}^{k} \left[\frac{\partial \log c(U_t^0, \theta_0)}{\partial u_j} f_{0,j}(\xi_{j,t}) + \frac{f'_{0,j}(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right] \left[\frac{-w_{\mu,j}^t}{\sigma_{0,j}^t} + \frac{-\xi_{j,t} w_{\sigma,j}^t}{\sigma_{0,j}^t}\right] + \sum_{j=1}^{k} \frac{-w_{\sigma,j}^t}{\sigma_{0,j}^t}.
\end{aligned} \tag{27}$$

**Assumption 7.** *For $j = 1, \cdots, k$, (i) $\lim_{u \to 0_+} f_{0,j}\left(F_{0,j}^{-1}(u)\right)\left(1 + \left|F_{0,j}^{-1}(u)\right|\right) = 0$, and $\lim_{u \to 1_-} f_{0,j}\left(F_{0,j}^{-1}(u)\right)\left(1 + \left|F_{0,j}^{-1}(u)\right|\right) = 0$; (ii) $E\left(\frac{f'_{0,j}(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right)^2 < \infty$ ; (iii) $E\left(\frac{f'_{0,j}(\xi_{j,t})\xi_{j,t}}{f_{0,j}(\xi_{j,t})}\right)^2 < \infty$; (iv) $0 < E\left(\frac{\partial l(\alpha_0,\kappa_0,Y_t)}{\partial \alpha^\intercal}[v] + \frac{\partial l(\alpha_0,\kappa_0,Y_t)}{\partial \kappa^\intercal}[w]\right)^2 < \infty$ for $(v^\intercal, w^\intercal) \neq 0$, $v \in \mathbb{V}^r$ and $w \in \mathbb{W}$; (v) $\int \sup_{\tau \in [0, \log(\log n)\delta_n]} \left|\frac{dp\left(y|\mathcal{F}^{t-1};\alpha_0+\tau v,\kappa_0+\tau w\right)}{d\tau}\right| dy < \infty$ and $\int \sup_{\tau \in [0, \log(\log n)\delta_n]} \left|\frac{d^2p\left(y|\mathcal{F}^{t-1};\alpha_0+\tau v,\kappa_0+\tau w\right)}{d\tau^2}\right| dy < \infty$ almost surely for $v \in \mathbb{V}^r$ and $w \in \mathbb{W}$, where $p\left(y|\mathcal{F}^{t-1};\alpha,\kappa\right)$ is the conditional density of $Y_t$ given $\mathcal{F}^{t-1}$ (defined in Eq.(3)) and $\delta_n = \max\left\{\widetilde{\delta}_{\alpha,n}, \delta_{h,n}\right\}$.*

Assumption 7 (i) requires $f_{0,j}(x)$ and $xf_{0,j}(x)$ converge to zero at both tails, which is mild (see, e.g. Assumption ($F_\epsilon$) in Neumeyer et al. (2019)). Assumption 7 (ii) (iii) are conditions for location and scale models to be locally asymptotically normal (LAN) at the true parameter (see, e.g. Examples 4.1, 4.3, 4.4 in Hallin and Werker (2003) and the references therein). Assumption 7 (ii) (iii) also imply that $\left|E\left(\frac{f'_{0,j}(\xi_{j,t})\xi_{j,t}}{f_{0,j}(\xi_{j,t})}\right)\right| < \infty$ and $\left|E\left(\frac{f'^2_{0,j}(\xi_{j,t})\xi_{j,t}}{f^2_{0,j}(\xi_{j,t})}\right)\right| < \infty$ according to Cauchy–Schwarz inequality. Assumption 7 (v) is a mild condition to assure the interchange of differentiation and integration.

**Lemma 2.** *Under Assumptions 1, 2, 3, 7, we have for any $v \in \mathbb{V}^r$ and $w \in \mathbb{W}$, (i) $E\left[\left(\frac{\partial l(\alpha_0,\kappa_0,Y_t)}{\partial \alpha^\intercal}[v] + \frac{\partial l(\alpha_0,\kappa_0,Y_t)}{\partial \kappa^\intercal}[w]\right)\left(\frac{\partial l(\alpha_0,\kappa_0,Y_s)}{\partial \alpha^\intercal}[\widetilde{v}] + \frac{\partial l(\alpha_0,\kappa_0,Y_s)}{\partial \kappa^\intercal}[\widetilde{w}]\right)\right] = 0$ for $\widetilde{v} \in \mathbb{V}^r$, $\widetilde{w} \in \mathbb{W}$ and all $s < t$; (ii) $\left\{\frac{\partial l(\alpha_0,\kappa_0,Y_t)}{\partial \alpha^\intercal}[v] + \frac{\partial l(\alpha_0,\kappa_0,Y_t)}{\partial \kappa^\intercal}[w], 1 \le t \le n\right\}$ is a martingale difference sequence with respect to $\mathcal{F}^{t-1}$; (iii) $E\left(\frac{\partial l(\alpha_0,\kappa_0,Y_t)}{\partial \alpha^\intercal}[v] + \frac{\partial l(\alpha_0,\kappa_0,Y_t)}{\partial \kappa^\intercal}[w]\right)^2 = -E\left(\frac{\partial^2 l(\alpha_0,\kappa_0,Y_t)}{\partial(\alpha^\intercal,\kappa^\intercal)^\intercal \partial(\alpha^\intercal,\kappa^\intercal)}\left[(v^\intercal, w^\intercal)^\intercal, (v^\intercal, w^\intercal)^\intercal\right]\right)$.*

The second-order directional derivative in Lemma 2(iii) is defined in Eq.(C.2). Lemma 2 suggests we can define the Fisher norm on the space $\mathbb{V}^r \times \mathbb{W}$ as following :

$$\|(v^\intercal, w^\intercal)^\intercal\|_F^2 = E\left(\frac{\partial l\left(\alpha_0, \kappa_0, Y_t\right)}{\partial \alpha^\intercal}[v] + \frac{\partial l\left(\alpha_0, \kappa_0, Y_t\right)}{\partial \kappa^\intercal}[w]\right)^2, \tag{28}$$

which coincides with the norm defined in Eq.(13) on $\mathbb{V}^r \times \{0\}$. It is easy to see that the closure of $\mathbb{W}$ under this norm[8] : $\overline{\mathbb{W}}_F \subset \left\{w : E\left(\frac{w^t_{\mu,j}}{\sigma^t_{0,j}}\right)^2 < \infty,\ E\left(\frac{w^t_{\sigma,j}}{\sigma^t_{0,j}}\right)^2 < \infty,\ j = 1, \cdots, k\right\}$.

[8] The specific form of $\overline{\mathbb{W}}_F$ depends on both the model setup of the dynamic parameter and the norm (28).

**Theorem 3.** *Under Assumptions 1, 2, 3, 6, 7, furthermore, if the probability family* $\{P_{\alpha,\kappa} : \alpha \in \mathcal{A}^r, \kappa \in \mathcal{B}\}$ *for* $\{Y_t, 1 \le t \le n\}$ *is locally asymptotically normal at* $(\alpha_0, \kappa_0)$ *and*

$$\left\{ w : \frac{w^t_{\mu,j}}{\sigma^t_{0,j}} = c_{\mu,j},\ \frac{w^t_{\sigma,j}}{\sigma^t_{0,j}} = c_{\sigma,j},\ \ (c_{\mu,j}, c_{\sigma,j}) \in \mathcal{R}^2,\ \textit{for } j = 1, \cdots, k \right\} \subset \overline{\mathbb{W}}_F, \tag{29}$$

*then* (i) *the semiparametric Fisher information matrix for* $\theta_0$ *in the full model (Eq.*(3)*) equals* $\mathcal{I}_*(\theta_0)$*;* (ii) $\widetilde{\theta}_{sml}$ *satisfying conditions in Theorem 2 is semiparametrically efficient in the full model.*

Theorem 3(ii) states that $\widetilde{\theta}_{sml}$ is as efficient as the one-step full likelihood estimator of the copula parameter, which is computationally intractable. In other words, we can obtain an efficient estimator of $\theta_0$ without additional computational burden.

# 4 Semiparametric Two-step Estimation for Residual Copulas

## 4.1 Asymptotic Normality of $\widetilde{F}_{j,2s}$

For $j = 1, \cdots, k$, we consider the estimation of $\widetilde{F}_{j,2s}(x)$ for some fixed $x \in \Xi_j \subset \mathcal{R}$, where $\widetilde{F}_{j,2s}(\cdot)$ is defined in Eq.(8).

**Assumption 8.** $n^{-1/2}\sum_{t=1}^{n}\left(\mathbb{I}\{\xi_{j,t} \le x\} - F_{0,j}(x)\right) + f_{0,j}(x)\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}^t_j - \mu^t_{0,j}}{\sigma^t_{0,j}}\right) + x f_{0,j}(x)\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}^t_j - \sigma^t_{0,j}}{\sigma^t_{0,j}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V_{j,2s}(x))$*, where* $V_{j,2s}(x)$ *is given in Eq.*(30)*.*

Under Assumption 5(i), Assumption 8 is implied by the triangle array CLT and similar to Assumption 5(ii). $\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}^t_j - \mu^t_{0,j}}{\sigma^t_{0,j}}\right)$ and $\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}^t_j - \sigma^t_{0,j}}{\sigma^t_{0,j}}\right)$ have been illustrated in Lemmas B.1, B.2 and B.3.

**Theorem 4.** *Under Assumptions 1(i)(ii)(iii), 2(ii), 4, 5(i), 8, we have* $\sqrt{n}\left(\widetilde{F}_{j,2s}(x) - F_{0,j}(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V_{j,2s}(x))$*, where*

$$V_{j,2s}(x) = \lim_{n\to\infty} Var\left[\begin{array}{c} n^{-1/2}\sum_{t=1}^{n}\left(\mathbb{I}\{\xi_{j,t} \le x\} - F_{0,j}(x)\right) \\ + f_{0,j}(x)\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}^t_j - \mu^t_{0,j}}{\sigma^t_{0,j}}\right) + x f_{0,j}(x)\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}^t_j - \sigma^t_{0,j}}{\sigma^t_{0,j}}\right) \end{array}\right]. \tag{30}$$

### 4.2 Asymptotic Normality of $\widetilde{\theta}_{2s}$

To establish the asymptotic property of $\widetilde{\theta}_{2s}$, we first present the theoretical result for the empirical copula process of the semi-nonparametric GARCH filtered residuals in Lemma 3.

Denote $C_0(\cdot) = C(\cdot;\theta_0)$ as the unknown true copula function. Let $U = (U_1, \cdots, U_k)^\intercal$ be a $k$-dimensional random vector, and let $u = (u_1, \cdots, u_k)^\intercal \in [0,1]^k$ be a $k$-dimensional nonrandom vector. Denote $(U_1 \le u_1, \cdots, U_k \le u_k)$ as $(U \le u)$. Define the following empirical copula processes:

$$\widetilde{\mathbb{C}}_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \mathbb{I}\left(\widetilde{U}_t \le u\right) - C_0(u) \right\}, \; \widehat{\mathbb{C}}_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \mathbb{I}\left(\widehat{U}_t \le u\right) - C_0(u) \right\}, \; for \; u \in [0,1]^k,$$

where $\left\{\widetilde{U}_t, 1 \le t \le n\right\}$ and $\left\{\widehat{U}_t, 1 \le t \le n\right\}$ are defined in Eqs.(9) and (11), respectively. We also introduce an auxiliary empirical process:

$$\mathbb{C}_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \mathbb{I}\left(U_t^0 \le u\right) - C_0(u) \right\}, \; for \; u \in [0,1]^k; \quad \text{and} \quad \mathbb{C}_n \rightsquigarrow \mathbb{C}, \tag{31}$$

where $\mathbb{C}$ is a C-Kiefer process. It is a well-known result, see, e.g. Bickel and Wichura (1971).

**Lemma 3.** *Under Assumption 1 and Assumptions B.2, B.3, C.5, C.6, C.7 stated in the Appendix, we have* $\widetilde{\mathbb{C}}_n \rightsquigarrow \widehat{\mathbb{C}}$ *and* $\widehat{\mathbb{C}}_n \rightsquigarrow \widehat{\mathbb{C}}$, *where* $\widehat{\mathbb{C}}(u) = \mathbb{C}(u) - \sum_{j=1}^{k} \frac{\partial C_0(u)}{\partial u_j} \mathbb{C}_j(u_j)$, *for* $u \in [0,1]^k$. $\mathbb{C}$ *is defined in Eq.*(31) *and* $\mathbb{C}_j(u_j) = \mathbb{C}\left(1, \cdots, 1, u_j, 1, \cdots, 1\right)$.

**Remark 5.** Under the conditions in Lemma 3, the empirical copula process of the filtered data behaves as if the semi-nonparametric GARCH model were known:

$$\sup_{u \in [0,1]^k} \left| \widetilde{\mathbb{C}}_n(u) - \widehat{\mathbb{C}}_n(u) \right| = o_p(1). \tag{32}$$

Neumeyer et al. (2019) (Theorem 1) establishes a similar result when the temporal dependence is prefiltered using nonparametric ARCH models and the local polynomial estimation.

**Assumption 9.** *(*i*) Assumption 2 (*i*); (*ii*)* $\theta_0$ *is the unique maximizer of* $E\left[\log c\left(U_t^0; \theta\right)\right]$ *over* $\Theta$*; (*iii*) for* $i, j = 1, \cdots, d_\theta$, $\frac{\partial \log c(u;\theta)}{\partial \theta_i}$ *and* $\frac{\partial^2 \log c(u;\theta)}{\partial \theta_i \partial \theta_j}$ *are all well-defined and continuous in* $\mathcal{N}(\theta_0)$ *(defined in Assumption 3) and* $u \in [0,1]^k$*, and of uniformly bounded Hardy-Kraus variation (see Definition A.1. of Berghaus et al. (2017)); (*iv*) there exists a function* $J(u)$ *such that for each*

$\theta \in \mathcal{N}(\theta_0)$, $\max_{j=1,\cdots,k} \max_{i=1,\cdots,d_\theta} \left| \frac{\partial^2 \log c(u;\theta)}{\partial u_j \partial \theta_i} \right| \leq J(u)$, *and* $E\left(J(U_t^0)\right) < \infty$; (v) *the matrix function* $\Gamma(\theta) = E\left(-\frac{\partial^2 \log c(U_t^0;\theta)}{\partial\theta\partial\theta^\intercal}\right)$ *is continuous in* $\mathcal{N}(\theta_0)$ *and the matrix* $\Gamma_0 = \Gamma(\theta_0)$ *is positively definite.*

**Theorem 5.** *Suppose that the assumptions of Lemma 3 and Assumption 9 hold. Then, both* $\widetilde{\theta}_{2s}$ *(defined in Eq.(10)) and* $\widehat{\theta}_{2s}$ *(defined in Eq.(12)) are consistent estimators of* $\theta_0$. *Furthermore,* $\sqrt{n}\left(\widetilde{\theta}_{2s} - \theta_0\right) \xrightarrow{\mathcal{D}} N\left(0_{d_\theta}, \Gamma_0^{-1}\Sigma_0\Gamma_0^{-1}\right)$, *and* $\sqrt{n}\left(\widetilde{\theta}_{2s} - \widehat{\theta}_{2s}\right) = o_p(1)$, *where*

$$\Sigma_0 = Var\left(\frac{\partial \log c(U_t^0, \theta_0)}{\partial\theta} + \sum_{j=1}^{k} \int_{u\in[0,1]^k} \left[\mathbb{I}\left(U_{j,t} \leq u_j\right) - u_j\right] \frac{\partial^2 \log c(u;\theta_0)}{\partial\theta\partial u_j} dC(u;\theta_0)\right).$$

When semi-nonparametric GARCH filtered residuals are used, Theorem 5 shows that $\widetilde{\theta}_{2s}$ and $\widehat{\theta}_{2s}$ are asymptotically equivalent, which is a nontrivial extension to Chen and Fan (2006).

**Remark 6.** Under Theorem 5, the estimation of the asymptotic variance of $\widetilde{\theta}_{2s}$ could be simplified. A consistent estimator is defined in Proposition 3.2. of Chen and Fan (2006) (Page 133-134), and also in Eqs.(45)-(49) of Patton (2013). We thus omit the details.

# 5 Simulation Studies

To investigate the finite sample performance and the asymptotic properties of our joint SML estimates of the copula parameter and the marginals using filtered residuals, we conduct an extensive simulation study.

**Data generating process (DGP)**. A time series sample $\{(Y_{1,t}, Y_{2,t})\}_{t=1}^n$ is generated as follows:

$$Y_{j,t} = \sigma_{j,t}\ \xi_{j,t}, \quad \sigma_{j,t}^2 = m_j\left(\sum_{i=1}^{\infty} \beta_{0,j}^{i-1} v\left(Y_{j,t-i};\eta_{0,j}\right)\right) \quad for\ j = 1,2 \tag{33}$$

$$F_0(\xi_{1,t}, \xi_{2,t}) = C\left(F_{0,1}(\xi_{1,t}), F_{0,2}(\xi_{2,t});\theta_0\right), \quad E[\xi_{j,t}] = 0,\ Var[\xi_{j,t}] = 1 \quad for\ j = 1,2 \tag{34}$$

where $v(y;\eta) = y^2 + \eta y^2 1_{(y<0)}$ and $m_j(x) = \beta_{0,j}x + \omega_{0,j}(1+\gamma_{0,j} sin(x/5))/(1-\alpha_{0,j})$ for $j = 1, 2$. The true parameter values in (33), $(\omega_{0,j}, \alpha_{0,j}, \beta_{0,j}, \eta_{0,j}, \gamma_{0,j})$, $j = 1, 2$, are set to be $(0.1, 0.85, 0.05, 0.05, 0.1)$. In the first submitted version we considered another DGP by setting $\gamma_{0,j} = 0$, i.e., the function $m_j(x)$

becomes linear (in $x$) and the time series in (33) follows the standard GJR-GARCH model. The semiparametric GARCH (33) is similar to the simulation design in Liu and Yang (2016).[9] For the semiparametric bivariate distribution $F_0(\xi_1, \xi_2)$ of the standardized innovations in (34), we consider two types of marginal distributions and four types of copula functions. The unknown marginals $F_{0,j}, j = 1, 2$ are set to be: standard normal ($N(0,1)$), standardized $t(\nu)$ (std-$t(\nu)$) with the degree of freedom $\nu = 10, 7, 5$. The parametric copula functions $C(\cdot; \theta_0)$ are: Gaussian copula (zero tail dependence), Student's t-copula (symmetric tail dependence), Gumbel copula (upper tail dependence), and Clayton copula (lower tail dependence). See Appendix A for expressions of these copulas and Nelsen (2006) for their properties.

**Computing estimators**. We estimate the semi-nonparametric GARCH (33) part following the procedure of Liu and Yang (2016), in which the unknown link function $m()$ is approximated by a spline sieve. We shall report simulation results for two types of estimators of $(F_{0,1}, F_{0,2}, \theta_0)$ in (34): the joint sieve ML (SML) and the semiparametric two-step (2Step). For the joint sieve ML, we have tried the polynomial sieve ($\{x^{\frac{k}{2}}\}_{k=0}^{K_n}$) to approximate the logarithm of $f_j$, and the 4th order cardinal B-spline sieve to approximate the square root of $f_j$ (see Chen et al. (2006) for a detailed description). In our simulation studies, the first choice of sieves works slightly better. The number of sieve terms $K_n$ can be chosen according to AIC. Ideally we should choose $K_n$ according to AIC for each Monte Carlo replication. To save computational time, we choose $K_n$ for the first Monte Carlo experiment, and then fix this choice for the following experiments.

## 5.1 Simulation Results

For both large sample ($n = 8000$) and finite sample ($n = 500$) comparisons, we include four estimators of $(F_{0,1}, F_{0,2}, \theta_0)$: (1) the infeasible joint sieve ML estimates using true innovations (True-SML); (2) the infeasible two-step estimates using true innovations (True-2Step); (3) the feasible

[9] Other values of GJR-GARCH parameters and other GARCH families were also tried as simulation DGPs, and some were reported in the first submitted version. The simulation results using different semi-nonparametric GARCH models share very similar patterns in terms of estimation of $(F_{0,1}, F_{0,2}, \theta_0)$ for (34), and are no longer reported here due to the lack of space.

joint sieve ML estimates using the spline-GARCH fitted residuals (Resid-SML); and (4) the feasible two-step estimates using the spline-GARCH fitted residuals (Resid-2Step).

We report the large sample ($n = 8000$) properties of various estimators of $(F_{0,1}, F_{0,2}, \theta_0)$ based on 2000 Monte Carlo replications, and the finite sample ($n = 500$) performances based on 500 Monte Carlo replications. For each estimator of $(F_{0,1}, F_{0,2}, \theta_0)$, we compute and report their Monte Carlo sample mean (Mean), sample variance (Var), and sample mean square error (MSE). For the unknown marginal $F_{0,j}$ we estimate its values at the 1/3 quantile ($q_1$) and 2/3 quantile ($q_2$). Namely, we report $\widehat{F}_j(q_1)$, $\widehat{F}_j(q_2)$ for $j = 1, 2$. To keep the flow of the main text, we postpone all the tables summarizing simulation results to Appendix A. See Tables A.1 - A.8 for the large sample results for Gaussian, Student's t-, Clayton and Gumbel copula models with unknown marginals; and Tables A.9 - A.12 for the finite sample results.

We observe the following simulation patterns from Tables A.1 - A.8 on large sample ($n = 8000$) results. (**a**) For copula parameter estimation, the large sample variances and MSEs of the True-SML (resp. True-2Step) are very close to those of the Resid-SML (resp. Resid-2Step) estimates. (**b**) For copula parameter estimation, the large sample variances and MSEs of the Resid-SML (resp. True-SML) are slightly smaller than or close to those of the Resid-2Step (resp. True-2Step) estimates. (**c**) For marginal cdf estimation, the large sample variances and MSEs of the joint SML (True-SML, Resid-SML) estimates are smaller than those of the empirical cdfs (True-2Step, Resid-2Step) in all the models, including the semiparametric Gaussian copula model.

We also observe the following simulation patterns from Tables A.9 - A.12 on finite sample ($n = 500$) results. (**d**) For copula parameter estimation, the finite sample variances and MSEs of the True-SML (resp. True-2Step) are smaller than those of the Resid-SML (resp. Resid-2Step) estimates. (**e**) For copula parameter estimation, the finite sample variances and MSEs of the Resid-SML are smaller than those of the Resid-2Step estimates, except that all the estimates are about the same for Gaussian copula parameter. (**f**) For marginal cdf estimation, the finite sample variances and MSEs of the joint SML (True-SML, Resid-SML) estimates are much smaller than those of the empirical cdfs (True-2Step, Resid-2Step) in all the models.

The simulation results are all consistent with our theoretical conclusions. In particular, findings (**a**) and (**d**) are consistent with our theories that the asymptotic variances of the joint SML and the two-step estimators of the copula parameters are invariant to the semi-nonparametric dynamic filtering. Findings (**b**) and (**e**) are consistent with our theory that the joint SML estimates of the copula parameters using fitted residuals are still semiparametrically efficient. Interestingly, while the semiparametric two-step copula estimators are inefficient except for the Gaussian copula parameter, the efficiency loss of the two-step copula estimator is mild in large samples for copula models with little asymmetric tail dependence. Findings (**c**) and (**f**) show that, when the two series are dependent (through copulas), the information of the dependence structure improves the efficiency in estimating the marginal distributions (also see Chen et al. (2006)).

When calculating VaRs in real applications, it is important to have more efficient and accurate estimation of both the copula parameters and the marginal distributions. The simulation findings encourage the use of the joint sieve MLE in calculating VaRs in Section 6.

# 6 An Empirical Application

In this section, we apply our multivariate semi-nonparametric GARCH-filtered copula model to investigate dependence among the daily returns of five popular asset classes: S&P500 Stock Index (S&P500), Nasdaq100 Stock Index (NAS100), Barclays U.S. Corporate High Yield Bond Index (HYB), Barclays Capital U.S. MBS Index (MBS), and S&P GSCI commodity index (GSCI). Our dataset spans an 11-year period from January 2007 to December 2017, a total of $T = 2476$ trading days. We use the S&P500 to represent the whole stock market in the U.S., the NAS100 to represent technology stocks, the HYB to represent credit assets, the MBS to represent investment-grade mortgage-backed assets, and the GSCI to represent commodity assets.

The return equations for the S&P500 and other assets are specified respectively as

$$\text{S\&P500}: Y_{1,t} = c_1 + \rho_1 Y_{1,t-1} + \sigma_{1,t}\xi_{1,t} \tag{35}$$

$$\text{Others}: Y_{j,t} = c_j + \rho_j Y_{j,t-1} + \beta_j Y_{1,t-1} + \sigma_{j,t}\xi_{j,t}, \quad \text{for } j = 2, 3, 4, 5. \tag{36}$$

In the first submitted version we applied the semiparametric GARCH model (33) of Liu and Yang (2016) for the volatility part. Our theoretical results hold for general semi-nonparametric GARCH filtering in the first stage. For robustness check, here we follow Example 2.3 of Chen (2013) and specify semi-nonparametric volatility as

$$\sigma_{j,t}^2 = \omega_j + \theta_j \sigma_{j,t-1}^2 + h_j(\sigma_{j,t-1}\xi_{j,t-1}), \quad \text{for } j = 1, 2, 3, 4, 5,$$

where $E(\xi_{j,t}) = 0$ and $E(\xi_{j,t}^2) = 1$. The news impact functions $h_j(\cdot)$ and the marginal distribution of $\xi_{j,t}$ are unspecified. $(\xi_{1,t}, \xi_{2,t}, \xi_{3,t}, \xi_{4,t}, \xi_{5,t})'$ are independent across time and the joint distribution is modelled by a semiparametric copula model.

**Semi-nonparametric GARCH filtration**. Following Chen (2013), we first estimate each set of univariate conditional mean and GARCH parameters via spline quasi-maximum likelihood (QMLE), where each unknown $h_j(\cdot)$ is approximated via cubic B-spline sieves excluding a constant term.[10] We then obtain filtered residuals $\{\widehat{\xi}_{j,t}\}$ for each time series.

The summary statistics of the raw returns and standardized semi-nonparametric GARCH filtered residuals are presented in Table 1. After filtering, the standardized series are less fat-tailed and autocorrelated. It is clear that the residual series all follow non-Gaussian distributions with negative skewness and leptokurtosis. The S&P500 is positively correlated with NAS100, HYB, and GSCI across the whole sample, but it has a substantial negative sample correlation (-0.213) with MBS. This negative correlation is mostly driven by interest rates because low interest rates are associated with high stock returns and low mortgage returns.

The semi-nonparametric GARCH estimates exhibit the well-known "news impact curve" (or leverage effects) in each asset. For illustration, we plot the news impact curves of the S&P500 estimated from the spline-GARCH and the standard linear GARCH (1,1) model in Figure 1. Obviously, the spline-GARCH predicts more volatility for negative return shocks and less volatility for positive return shocks than standard GARCH (1,1).

**Estimation of multivariate copula parameters**. We examine the dependence among the

[10] In the empirical analysis, we use five B-splines basis functions.

**Table 1:** Summary Statistics of Raw Returns and Semi-nonparametric GARCH Filtered Residuals

**Panel A: Summary Statistics of Raw Returns**

| | N | Mean | Std | Skew | Kurt | Min | Q1 | Median | Q3 | Max | AR1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| S&P500 | 2746 | 0.020 | 1.270 | -0.432 | 13.781 | -9.474 | -0.402 | 0.057 | 0.554 | 10.420 | -0.129 |
| NAS100 | 2746 | 0.044 | 1.345 | -0.425 | 10.191 | -11.122 | -0.470 | 0.098 | 0.674 | 10.364 | -0.092 |
| HYB | 2746 | 0.026 | 0.334 | -1.811 | 29.932 | -4.847 | -0.071 | 0.047 | 0.145 | 2.744 | 0.441 |
| MBS | 2746 | 0.013 | 0.194 | 0.206 | 9.906 | -1.324 | -0.074 | 0.014 | 0.108 | 1.694 | 0.024 |
| GSCI | 2746 | -0.031 | 1.499 | -0.321 | 6.505 | -8.653 | -0.758 | 0.012 | 0.757 | 7.214 | -0.049 |

**Panel B: Correlation Matrix of Raw Returns**

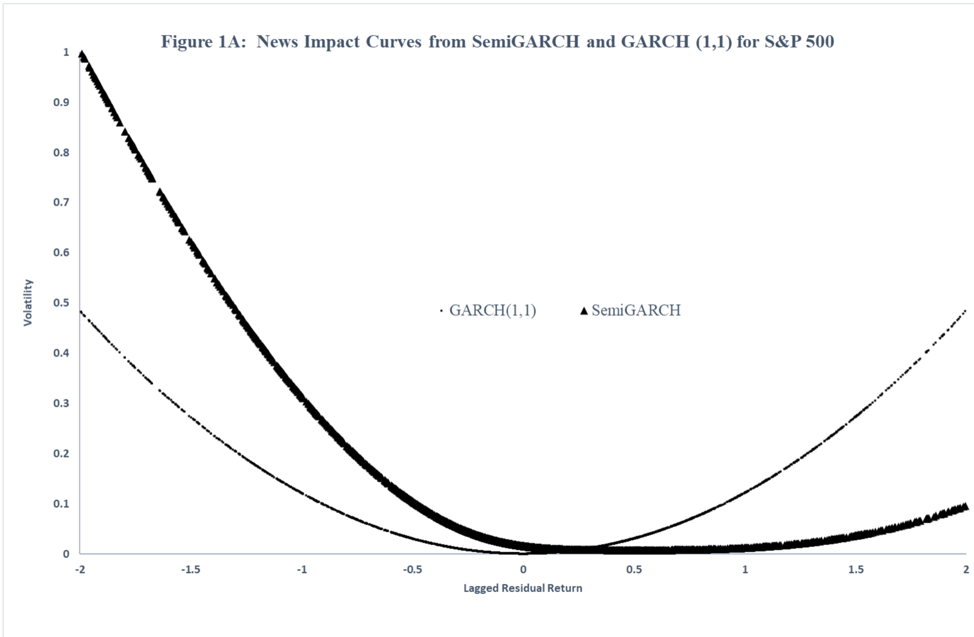
| | S&P500 | NAS100 | HYB | MBS | GSCI |
|---|---|---|---|---|---|
| S&P500 | 1 | 0.930 | 0.330 | -0.213 | 0.390 |
| NAS100 | 0.930 | 1 | 0.289 | -0.217 | 0.324 |
| HYB | 0.330 | 0.289 | 1 | 0.051 | 0.308 |
| MBS | -0.213 | -0.217 | 0.051 | 1 | -0.124 |
| GSCI | 0.390 | 0.324 | 0.308 | -0.124 | 1 |

**Panel C: Summary Statistics of Standardized SemiGARCH Filtered Residuals**

| | N | Mean | Std | Skew | Kurt | Min | Q1 | Median | Q3 | Max | AR1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| S&P500 | 2745 | 0.007 | 0.994 | -0.684 | 5.326 | -6.549 | -0.503 | 0.075 | 0.617 | 3.701 | 0.046 |
| NAS100 | 2745 | 0.002 | 0.992 | -0.601 | 4.805 | -5.625 | -0.499 | 0.073 | 0.597 | 4.204 | 0.015 |
| HYB | 2745 | 0.018 | 0.989 | -0.263 | 5.073 | -4.760 | -0.525 | 0.055 | 0.596 | 5.513 | 0.020 |
| MBS | 2745 | -0.021 | 1.001 | -0.216 | 5.128 | -6.015 | -0.608 | 0.012 | 0.616 | 6.309 | 0.025 |
| GSCI | 2745 | 0.006 | 1.001 | -0.308 | 4.331 | -5.877 | -0.579 | 0.030 | 0.656 | 3.931 | 0.018 |

**Panel D: Correlation Matrix of Standardized SemiGARCH Filtered Residuals**

| | S&P500 | NAS100 | HYB | MBS | GSCI |
|---|---|---|---|---|---|
| S&P500 | 1 | 0.901 | 0.447 | -0.249 | 0.327 |
| NAS100 | 0.901 | 1 | 0.378 | -0.221 | 0.253 |
| HYB | 0.447 | 0.378 | 1 | -0.008 | 0.294 |
| MBS | -0.249 | -0.221 | -0.008 | 1 | -0.145 |
| GSCI | 0.327 | 0.253 | 0.294 | -0.145 | 1 |



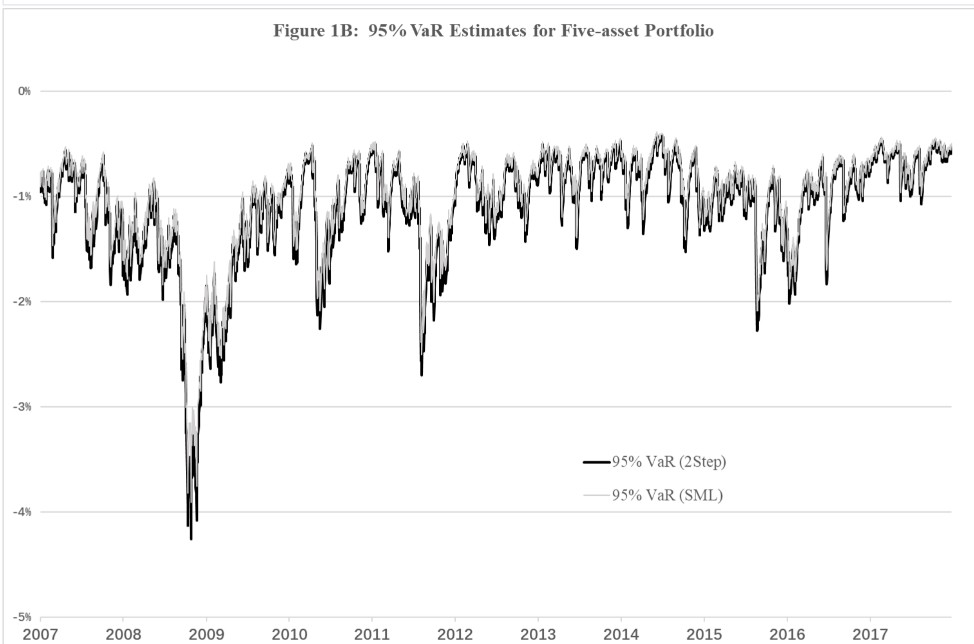



**Figure 1:** News Impact Curves and Value-at-Risk Estimates

five assets with multivariate Student's t-copula by applying the two-step and joint SML estimation methods to the spline-GARCH filtered standardized residuals.

We present estimates and standard errors of the 5-dimension Student's t-copula parameters and log-likelihood values for the copula part in Table 2. Several findings can be obtained from the table. First, parameter estimates from the joint SML and two-step methods are very close. Second, the joint SML method gives higher log-likelihood value for the copula part, which suggests a better fit to the dependence in the real data. Third, the estimated correlation coefficients of Student's t-copula are close to the sample correlations of spline-GARCH filtered standardized residuals, while the estimated tail dependence is small. Hence we suspect that Gaussian copula could also be used for the 5 assets during this sample period. [11]

**Full-sample VaR estimation**. To further compare the estimates from the two methods, we consider the Value-at-Risk estimation of investment portfolios. We examine five portfolios. The first portfolio consists of above five assets with equal weights. The remaining four portfolios consist of two assets with equal weights: the S&P500 and each from HYB, MBS or GSCI. The portfolios are rebalanced on a daily basis to keep equal weights as designed.

We define $(1-\alpha)$ $VaR_t$ to be the $\alpha$ conditional quantile of the portfolio's return at time $t$ based on information set at time $t-1$. We calculate the estimate of $VaR_t$ by plugging the estimated parameters into the model and simulating the whole conditional distribution of the return for each day. We first use the estimated parameters from the Student's t-copula in Table 2, and simulate many draws of $u_t = (u_{1,t}, ..., u_{5,t})$ from the estimated copula. Next we obtain the draws of $\epsilon_t = (\epsilon_{1,t}, ..., \epsilon_{5,t}) = (\hat{F}_1^{-1}(u_{1,t}), ..., \hat{F}_5^{-1}(u_{5,t}))$, where $\hat{F}_j$, $j = 1, ..., 5$, are estimated marginal distributions. Then we can easily obtain simulated returns by using the return equations defined in (35) and (36). The return of the equal-weighted portfolios can be obtained by taking average of the individual asset returns. The $\alpha$ quantile of the simulated portfolios returns is the $(1-\alpha)$ $VaR_t$.

[11] We also estimate bivariate mixture of Student's t-copula and Clayton copula for four pairs of assets: the S&P500 and each of the other four assets. The estimation results show that bivariate Student's t-copula fits well for this data set. The estimated bivariate correlations are very close to the correlation matrix estimated from multivariate copula.

**Table 2:** Parameter Estimation for Five-asset Student's t-copula

| **2Step Method** | | | | | | |
|---|---|---|---|---|---|---|
| | | S&P500 | NAS100 | HYB | MBS | GSCI |
| | S&P500 | 1 | 0.899 | 0.414 | -0.277 | 0.332 |
| | | | (0.001) | (0.021) | (0.040) | (0.025) |
| | NAS100 | 0.899 | 1 | 0.351 | -0.242 | 0.256 |
| Correlation | | (0.001) | | (0.024) | (0.038) | (0.028) |
| Matrix | HYB | 0.414 | 0.351 | 1 | -0.014 | 0.282 |
| | | (0.021) | (0.024) | | (0.029) | (0.025) |
| | MBS | -0.277 | -0.242 | -0.014 | 1 | -0.160 |
| | | (0.040) | (0.038) | (0.029) | | (0.028) |
| | GSCI | 0.332 | 0.256 | 0.282 | -0.160 | 1 |
| | | (0.025) | (0.028) | (0.025) | (0.028) | |
| $1/\nu$ | 0.096 | | | | | |
| | (0.008) | | | | | |
| logC | 2945.46 | | | | | |

| **SML Method** | | | | | | |
|---|---|---|---|---|---|---|
| | | S&P500 | NAS100 | HYB | MBS | GSCI |
| | S&P500 | 1 | 0.903 | 0.418 | -0.277 | 0.333 |
| | | | (0.003) | (0.016) | (0.017) | (0.017) |
| | NAS100 | 0.903 | 1 | 0.357 | -0.242 | 0.259 |
| Correlation | | (0.003) | | (0.018) | (0.019) | (0.018) |
| Matrix | HYB | 0.418 | 0.357 | 1 | -0.014 | 0.284 |
| | | (0.016) | (0.018) | | (0.020) | (0.019) |
| | MBS | -0.277 | -0.242 | -0.014 | 1 | -0.159 |
| | | (0.017) | (0.019) | (0.020) | | (0.020) |
| | GSCI | 0.333 | 0.259 | 0.284 | -0.159 | 1 |
| | | (0.017) | (0.018) | (0.019) | (0.020) | |
| $1/\nu$ | 0.097 | | | | | |
| | (0.009) | | | | | |
| logC | 3000.38 | | | | | |

Obviously, the VaR estimates depend on the parameters of both the marginal distributions and the copula function. For each portfolio, the two 95% $VaR_t$ series estimated from two-step and joint SML methods are very close, with the estimates from the two-step being slightly lower. The plots of VaR series from the two methods for the five-asset portfolio are presented in Figure 1.

**Backtesting of full-sample VaR estimates**. To evaluate the accuracy of VaR estimates in the full sample, we conduct backtesting using two popular statistical tests. The first is the proportion of failures (POF) test proposed by Kupiec (1995). It is a likelihood ratio test to assess if the observed proportion of failures (realized return lower than VaR) in the sample is consistent with the VaR confidence level, with the null hypothesis $H_0 : \Pr(r_t < VaR_t) = 1 - \alpha$. The second test is the conditional coverage (CC) test proposed by Christoffersen (1998), which is a likelihood ratio test to assess both the proportion of failures and the independence of failures over consecutive time periods, with the null hypothesis $H_0 : \Pr(r_t < VaR_t | I_{t-1}) = 1 - \alpha$. In Panel A of Table 3, we present the observed POF and the p-values of the two tests for 99% and 95% $VaR$ estimates from the two-step and joint SML methods. We find that the observed POF are on average more consist to the specified VaR confidence levels for the joint SML method. The p-values from the two tests suggest that both of the two VaR estimates are reasonably good for most cases and it is more likely to reject the null hypothesis for VaR estimates from the two-step method, compared with the joint SML method.

**Predictive accuracy of out-of-sample VaR forecasts**. We compare the predictive accuracy of out-of-sample VaR forecasts by using the two-step and joint SML methods. First, we set the window size to be 500 and obtain the rolling-window VaR forecasts on a daily basis. Then we perform the Diebold-Mariano (DM) test of Diebold and Mariano (1995) to compare the predictive accuracy using the asymmetric loss function based on the error terms $e_t = r_t - VaR_t$ to evaluate VaR forecasts with the $(1 - \alpha)$ confidence level, $L_\alpha(e_t) = (\alpha - \mathbb{1}(e_t < 0))e_t$, which is the quantile loss function. The difference of the two loss function from the two VaR estimates is defined as $d_t = L_\alpha(e_{t,1}) - L_\alpha(e_{t,2})$, for $j = 1, 2$ representing 2Step and SML. The null hypothesis of the test is that the two sequences of VaR forecasts from the two-step and joint SML are equally good in terms

**Table 3:** Evaluation of Full-sample and Out-of-sample VaR

**Panel A: Backtesting of Full-sample VaR Estimates**

| 99% VaR | | | | | | |
|---|---|---|---|---|---|---|
| | 2Step | | | SML | | |
| | POF(%) | p_POF | p_CC | POF(%) | p_POF | p_CC |
| SP500-NAS | 0.656 | 0.053 | 0.137 | 0.984 | 0.933 | 0.762 |
| SP500-HYB | 0.692 | 0.086 | 0.202 | 0.729 | 0.134 | 0.280 |
| SP500-MBS | 0.729 | 0.134 | 0.280 | 0.729 | 0.134 | 0.280 |
| SP500-GSCI | 0.875 | 0.500 | 0.645 | 0.984 | 0.933 | 0.762 |
| Five-asset | 0.692 | 0.086 | 0.202 | 1.203 | 0.301 | 0.392 |

| 95% VaR | | | | | | |
|---|---|---|---|---|---|---|
| | 2Step | | | SML | | |
| | POF(%) | p_POF | p_CC | POF(%) | p_POF | p_CC |
| SP500-NAS | 4.045 | 0.018 | 0.024 | 4.519 | 0.240 | 0.096 |
| SP500-HYB | 4.446 | 0.175 | 0.319 | 4.373 | 0.124 | 0.166 |
| SP500-MBS | 4.337 | 0.103 | 0.230 | 4.373 | 0.124 | 0.260 |
| SP500-GSCI | 4.592 | 0.320 | 0.436 | 4.446 | 0.175 | 0.319 |
| Five-asset | 4.300 | 0.085 | 0.133 | 5.029 | 0.944 | 0.717 |

**Panel B: Diebold-Mariano Test of Out-of-sample VaR Forecasts**

| 99% VaR | | | | |
|---|---|---|---|---|
| | POF_2Step(%) | POF_SML(%) | DM_stat | p_value |
| SP500-NAS | 0.802 | 0.891 | 1.579 | 0.114 |
| SP500-HYB | 0.535 | 0.624 | 3.115 | 0.002 |
| SP500-MBS | 0.713 | 0.713 | 1.757 | 0.079 |
| SP500-GSCI | 0.846 | 1.069 | 0.721 | 0.471 |

| 95% VaR | | | | |
|---|---|---|---|---|
| | POF_2Step(%) | POF_SML(%) | DM_stat | p_value |
| SP500-NAS | 3.608 | 3.920 | 1.972 | 0.049 |
| SP500-HYB | 3.786 | 4.009 | 2.972 | 0.003 |
| SP500-MBS | 4.098 | 4.098 | 1.380 | 0.168 |
| SP500-GSCI | 4.187 | 4.365 | 0.799 | 0.424 |

of the defined loss function, that is, $H_0 : E[d_t] = 0$. The DM statistic is calculated as the sample average of $d_t$ divided by its standard error. A positive DM statistic suggests that the VaR forecasts from the joint SML method are superior.

Due to the computational burden of rolling window estimation, we only conduct the out-of-sample test for the four pairs of bivariate portfolios. In Panel B of Table 3, we present the observed POF and DM test statistics and the corresponding p-values. The table reveals that the VaR forecasts from the joint SML provide better POF, and the DM statistics also favor VaR forecasts from the joint SML method.

## 7 Conclusion

The class of semiparametric copula-based multivariate dynamic models proposed in Chen and Fan (2006) has gained popularity in financial econometrics due to its flexible modelling of multivariate nonlinear risks. In this paper, we first extend their models to allow for semi-nonparametric conditional means and volatilities, and show that their semiparametric two-step estimators of residual copula parameters are still root-$n$ asymptotically normal with the asymptotic variances unaffected by the nonparametric filtering. In addition, we propose a new joint sieve maximum likelihood method using filtered residuals, and show that this procedure leads to semiparametric efficient estimation of the residual copula parameters, whose asymptotic variances do not depend on the nonparametric filtering either. This remarkable property greatly simplifies the accurate inference on residual copula parameters. Our theoretical results are consistent with the findings in the Monte Carlo studies.

Given the nice asymptotic properties of the two types of residual copula estimators discovered in this paper, one could easily extend the pseudo-likelihood ratio copula model selection tests developed in Chen and Fan (2006) from parametric dynamic models to semi-nonparametric dynamic models. The details are not presented here due to the length of the paper.

## Acknowledgements

The authors would like to thank Editor Zhengjun Zhang and two anonymous referees for their constructive comments that have improved the presentation of the paper. Earlier versions had been presented at the 2016 FERM, the 2017 China Meeting of the Econometric Society, the 2018 FERM and other conferences. Xiaohong Chen thanks the Cowles Foundation for its research support. Zhuo Huang acknowledges financial support from the National Natural Science Foundation of China (71671004). Yanping Yi acknowledges financial support from the National Natural Science Foundation of China (71973122).

## References

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6):1795–1843.

Akritas, M. G. and Keilegom, I. V. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, 28(3):549–567.

Berghaus, B., Bücher, A., and Volgushev, S. (2017). Weak convergence of the empirical copula process with respect to weighted metrics. *Bernoulli*, 23(1):743–772.

Bickel, P. and Wichura, M. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Annals of Mathematical Statistics*, 42(5):1656–1670.

Bickel, P. J. and Kwon, J. (2001). Inference for semiparametric models : some questions and an answer. *Statistica Sinica*, 11:863–960.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131.

Buhlmann, P. and McNeil, A. J. (2002). An algorithm for nonparametric garch modelling. *Computational Statistics & Data Analysis*, 40:665–683.

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, 18:17–39.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6B, chapter 76, pages 5549–5632. Elsevier.

Chen, X. (2013). *Penalized Sieve Estimation and Inference of Seminonparametric Dynamic Models : A Selective Review*, volume III of *Advances in Economics and Econometrics*, chapter 13, pages 485–544. Cambridge University Press.

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

Chen, X., Fan, Y., and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association*, 101(475):1228–1240.

Chen, X. and Liao, Z. (2014). Sieve m inference of irregular parameters. *Journal of Econometrics*, 182(1):70–86.

Chen, X., Liao, Z., and Sun, Y. (2014). Sieve inference on possibly misspecified semi-nonparametric time series models. *Journal of Econometrics*, 178:639–658.

Chen, X., Linton, O., and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, 71(5):1591–1608.

Chen, X. and Shen, X. (1998). Sieve extremum estimates for weakly dependent data. *Econometrica*, 66(2):289–314.

Chen, X., Wu, W. B., and Yi, Y. (2009). Efficient estimation of copula-based semiparametric markov models. *The Annals of Statistics*, 37(6B):4214–4253.

Cherubini, U., Gobbi, F., Mulinacci, S., and Romagnoli, S. (2012). *Dynamic Copula Methods in Finance*. Chichester: Wiley.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Dedecker, J. and Louhichi, S. (2002). Maximal inequalities and empirical central limit theorems. In Dehling, H., Mikosch, T., and Sorensen, M., editors, *Empirical Process Techniques for Dependent Data*, pages 137–159. Birkhäuser, Boston, MA.

Dette, H., Pardo-Fernández, J. C., and Keilegom, I. V. (2009). Goodness-of-fit tests for multiplicative models with dependent data. *Scandinavian Journal of Statistics*, 36(4):782–799.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Doukhan, P., Massart, P., and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 31(2):393–427.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric Theory*, 11(1):122–150.

Engle, R. F. and Rangel, J. G. (2008). The spline-garch model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies*, 21(3):1187–1222.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Models*. Springer, New York.

Francq, C. and Zakoian, J.-M. (2019). *GARCH models : structure, statistical inference and financial applications*. Wiley, second edition edition.

Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman & Hall/CRC, London.

Genest, C., Ghoudi, K., and Remillard, B. (2007). Rank-based extensions of the brock, dechert, and scheinkman test. *Journal of the American Statistical Association*, 102(480):1363–1376.

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

Hafner, C. M. and Preminger, A. (2009). On asymptotic theory for multivariate garch models. *Journal of Multivariate Analysis*, 100(9):2044–2054.

Hallin, M. and Werker, B. J. (2003). Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli*, 9(1):137–165.

Hoff, P. D., Niu, X., and Wellner, J. A. (2014). Information bounds for gaussian copulas. *Bernoulli*, 20(2):604–622.

Klaassen, C. A. J. and Wellner, J. A. (1997). Efficient estimation in the bivariate normal copula model : normal margins are least favourable. *Bernoulli*, 3(1):55–77.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3(2):73–84.

Linton, O. and Mammen, E. (2005). Estimating semiparametric arch($\infty$) models by kernel smoothing methods. *Econometrica*, 73(3):771–836.

Liu, R. and Yang, L. (2016). Spline estimation of a semiparametric garch model. *Econometric Theory*, 32(4):1023–1054.

Meister, A. and Krei$\beta$, J.-P. (2016). Statistical inference for nonparametric garch models. *Stochastic Processes and their Applications*, 126(10):3009–3040.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition edition.

Neumeyer, N., Omelka, M., and Hudecova, S. (2019). A copula approach for dependence modeling in multivariate nonparametric time series. *Journal of Multivariate Analysis*, 171:139–162.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Patton, A. J. (2013). *Copula Methods for Forecasting Multivariate Time Series*, volume 2B of *Handbook of Economic Forecasting*, chapter 16, pages 899–960. Elsevier.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Yang, L. (2006). A semiparametric garch model for foreign exchange volatility. *Journal of Econometrics*, 130(2):365–384.

Zhao, Z. and Zhang, Z. (2018). Semiparametric dynamic max-copula model for multivariate time series. *Journal of the Royal Statistical Society, Series B*, 80:409–432.

Appendix

This appendix contains three parts. Part A presents all the tables for the simulation study, as well as robustness checks for the empirical findings using semi-nonparametric GARCH filtering of Liu and Yang (2016) in the first stage. Part B provides sufficient conditions for Assumptions 4 and 5(i) in semi-nonparametric dynamic models. Part C contains additional assumptions and proofs for theoretical results in Sections 3 and 4.

# A Monte Carlo Results and Robustness Checks

We first recall the expressions and basic properties of the four copula distribution functions used in our Monte Carlo studies.

The bivariate Gaussian copula function with correlation $\theta = \rho$ is:

$$C(u_1, u_2; \rho) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right] dydx, \quad |\rho| < 1,$$

where $\Phi(\cdot)$ is CDF of the standard normal distribution. The Gaussian copula has no tail dependence.

The bivariate Student's t-copula function, for $\theta = (\nu, \rho)'$, $|\rho| < 1$, $\nu \in (1, \infty]$ is

$$\begin{aligned} C(u_1, u_2; \theta) &= \mathbf{t}_{\nu,\rho}(Q(\nu, u_1), Q(\nu, u_2)) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{Q(\nu,u_1)} \int_{-\infty}^{Q(\nu,u_2)} \left\{1 + \frac{x^2+y^2-2\rho xy}{\nu(1-\rho^2)}\right\}^{-\frac{\nu+2}{2}} dydx, \end{aligned}$$

where $\mathbf{t}_{\nu,\rho}(\cdot,\cdot)$ is the bivariate Student's t-distribution with mean zeros, the correlation matrix has the off-diagonal element $\rho$, and degrees of freedom $\nu$; $Q(\nu, \cdot)$ is the quantile function of a univariate Student's t-distribution with mean zero, and degrees of freedom $\nu$. The Student's t-copula has Kendall's tau $\tau = \frac{2}{\pi} \arcsin \rho$, and symmetric tail dependence: $\lambda_L = \lambda_U = 2t_{\nu+1}(-\sqrt{(\nu+1)(1-\rho)/(1+\rho)})$. The Student's t-copula becomes the Gaussian copula in the limit when $\nu \to \infty$.

The bivariate Clayton copula function is

$$C(u_1, u_2; \theta) = \left[u_1^{-\theta} + u_2^{-\theta} - 1\right]^{-1/\theta}, \ 0 \leq \theta < \infty.$$

The Clayton copula has Kendall's tau $\tau = \frac{\theta}{2+\theta}$, and lower tail dependence coefficient $\lambda_L = 2^{-1/\theta}$, but no upper tail dependence. The Clayton copula becomes the independence copula $C_I(u_1, u_2) = u_1 u_2$ in the limit when $\theta \to 0$.

The bivariate Gumbel copula function is

$$C(u_1, u_2; \theta) = \exp(-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta}), \quad 1 \leq \theta < \infty.$$

The Gumbel copula has Kendall's tau $\tau = 1 - \frac{1}{\theta}$, and upper tail dependence coefficient $\lambda_U = 2 - 2^{1/\theta}$, but no lower tail dependence. The Gumbel copula becomes the independence copula in the limit when $\theta \to 1$.

Tables A.1 - A.8 present large sample ($n = 8000$) simulation results for two kinds of estimation methods for copula parameters and marginal cdfs of semiparametric residual copula models with

**Table A.1:** Gaussian copula with unknown margins: Estimation of copula parameters. $n = 8000$, MC= 2000. Reported Var and MSE are the true values multiplied by 1000.

| | | True-2Step | Resid-2Step | True-SML | Resid-SML |
|---|---|---|---|---|---|
| $\rho = 0.7$ | Mean | 0.7003 | 0.6999 | 0.7000 | 0.6996 |
| $F_{0,1} = F_{0,2} = N(0,1)$ | Var | 0.0342 | 0.0344 | 0.0342 | 0.0343 |
| | MSE | 0.0343 | 0.0344 | 0.0342 | 0.0345 |
| $\rho = 0.9$ | Mean | 0.9000 | 0.8996 | 0.8999 | 0.8996 |
| $F_{0,1} = F_{0,2} = N(0,1)$ | Var | 0.0047 | 0.0047 | 0.0047 | 0.0047 |
| | MSE | 0.0047 | 0.0049 | 0.0047 | 0.0049 |
| $\rho = 0.7$ | Mean | 0.7003 | 0.6997 | 0.7007 | 0.7001 |
| $F_{0,1} = F_{0,2} = \text{std-}t(5)$ | Var | 0.0342 | 0.0343 | 0.0359 | 0.0359 |
| | MSE | 0.0343 | 0.0343 | 0.0363 | 0.0359 |
| $\rho = 0.9$ | Mean | 0.9000 | 0.8995 | 0.9003 | 0.8998 |
| $F_{0,1} = F_{0,2} = \text{std-}t(5)$ | Var | 0.0047 | 0.0047 | 0.0053 | 0.0053 |
| | MSE | 0.0047 | 0.0050 | 0.0054 | 0.0053 |

Gaussian, Student's t-, Clayton and Gumbel copulas respectively. Tables A.9 - A.12 present finite sample ($n = 500$) simulation results for the two kinds of estimation methods for semiparametric residual copula models with Student's t-, Clayton and Gumbel copulas respectively. In Table A.13 and Table A.14, we present empirical results with Semi-nonparametric GARCH Filtering of Liu and Yang (2016) in the first stage as a robustness check. The empirical results are very close to the ones reported in the main text already.

**Table A.2:** Gaussian copula with unknown margins: Estimation of marginal distributions. $n = 8000$, MC= 2000. $F_{0,j}(q_1) = 1/3$, $F_{0,j}(q_2) = 2/3$ for $j = 1, 2$. Reported Var and MSE are the true values multiplied by 1000.

| | | True-2Step | | | | Resid-2Step | | | | True-SML | | | | Resid-SML | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | |
| | True value | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| $\rho = 0.7$ | Mean | 0.3333 | 0.6666 | 0.3332 | 0.6665 | 0.3334 | 0.6665 | 0.3333 | 0.6664 | 0.3348 | 0.6682 | 0.3347 | 0.6680 | 0.3348 | 0.6682 | 0.3347 | 0.6680 |
| | Var | 0.0265 | 0.0286 | 0.0268 | 0.0272 | 0.0251 | 0.0270 | 0.0245 | 0.0263 | 0.0177 | 0.0178 | 0.0181 | 0.0171 | 0.0161 | 0.0163 | 0.0160 | 0.0156 |
| | MSE | 0.0266 | 0.0286 | 0.0268 | 0.0273 | 0.0251 | 0.0271 | 0.0245 | 0.0263 | 0.0199 | 0.0202 | 0.0201 | 0.0189 | 0.0183 | 0.0185 | 0.0178 | 0.0175 |
| $\rho = 0.9$ | Mean | 0.3333 | 0.6666 | 0.3333 | 0.6666 | 0.3334 | 0.6665 | 0.3334 | 0.6665 | 0.3350 | 0.6684 | 0.3349 | 0.6683 | 0.3350 | 0.6684 | 0.3350 | 0.6683 |
| | Var | 0.0265 | 0.0286 | 0.0271 | 0.0267 | 0.0251 | 0.0270 | 0.0252 | 0.0254 | 0.0177 | 0.0178 | 0.0177 | 0.0172 | 0.0160 | 0.0161 | 0.0156 | 0.0157 |
| | MSE | 0.0266 | 0.0286 | 0.0271 | 0.0267 | 0.0251 | 0.0271 | 0.0252 | 0.0254 | 0.0203 | 0.0207 | 0.0201 | 0.0198 | 0.0188 | 0.0190 | 0.0183 | 0.0182 |
| | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | |
| | True value | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| $\rho = 0.7$ | Mean | 0.3333 | 0.6666 | 0.3332 | 0.6665 | 0.3335 | 0.6664 | 0.3334 | 0.6664 | 0.3314 | 0.6686 | 0.3315 | 0.6686 | 0.3319 | 0.6682 | 0.3318 | 0.6683 |
| | Var | 0.0265 | 0.0286 | 0.0268 | 0.0272 | 0.0281 | 0.0312 | 0.0286 | 0.0301 | 0.0069 | 0.0069 | 0.0075 | 0.0075 | 0.0105 | 0.0105 | 0.0106 | 0.0107 |
| | MSE | 0.0266 | 0.0286 | 0.0268 | 0.0273 | 0.0281 | 0.0313 | 0.0286 | 0.0302 | 0.0107 | 0.0107 | 0.0110 | 0.0111 | 0.0127 | 0.0127 | 0.0131 | 0.0133 |
| $\rho = 0.9$ | Mean | 0.3333 | 0.6666 | 0.3333 | 0.6666 | 0.3335 | 0.6664 | 0.3335 | 0.6664 | 0.3316 | 0.6684 | 0.3317 | 0.6683 | 0.3320 | 0.6680 | 0.3319 | 0.6681 |
| | Var | 0.0265 | 0.0286 | 0.0271 | 0.0267 | 0.0281 | 0.0312 | 0.0285 | 0.0291 | 0.0076 | 0.0076 | 0.0075 | 0.0075 | 0.0102 | 0.0102 | 0.0102 | 0.0101 |
| | MSE | 0.0266 | 0.0286 | 0.0271 | 0.0267 | 0.0281 | 0.0313 | 0.0285 | 0.0291 | 0.0105 | 0.0105 | 0.0102 | 0.0103 | 0.0120 | 0.0120 | 0.0122 | 0.0121 |

**Table A.3:** Student's t-copula with unknown margins: Estimation of copula parameters. $n = 8000$, MC= 2000. Reported Var and MSE of $\rho$ are the true values multiplied by 1000.

| | | True-2Step | | Resid-2Step | | True-SML | | Resid-SML | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\rho$ | $\nu$ | $\rho$ | $\nu$ | $\rho$ | $\nu$ | $\rho$ | $\nu$ |
| $(\rho, \nu) = (0.7, 5)$ | Mean | 0.7000 | 5.0005 | 0.6997 | 5.0281 | 0.6999 | 5.0325 | 0.6996 | 5.0495 |
| $F_{0,1} = F_{0,2} =$ | Var | 0.0467 | 0.1398 | 0.0467 | 0.1423 | 0.0465 | 0.1327 | 0.0465 | 0.1343 |
| $N(0,1)$ | MSE | 0.0467 | 0.1398 | 0.0468 | 0.1431 | 0.0465 | 0.1337 | 0.0467 | 0.1368 |
| $(\rho, \nu) = (0.7, 5)$ | Mean | 0.7000 | 5.0005 | 0.6996 | 5.0258 | 0.7000 | 5.0327 | 0.6996 | 5.0523 |
| $F_{0,1} = F_{0,2} =$ | Var | 0.0467 | 0.1398 | 0.0467 | 0.1422 | 0.0467 | 0.1398 | 0.0469 | 0.1425 |
| std-$t(10)$ | MSE | 0.0467 | 0.1398 | 0.0468 | 0.1429 | 0.0467 | 0.1409 | 0.0470 | 0.1452 |
| $(\rho, \nu) = (0.7, 5)$ | Mean | 0.7000 | 5.0005 | 0.6996 | 5.0251 | 0.7002 | 5.0486 | 0.6999 | 5.0684 |
| $F_{0,1} = F_{0,2} =$ | Var | 0.0467 | 0.1398 | 0.0468 | 0.1428 | 0.0469 | 0.1431 | 0.0474 | 0.1449 |
| std-$t(7)$ | MSE | 0.0467 | 0.1398 | 0.0469 | 0.1434 | 0.0470 | 0.1455 | 0.0474 | 0.1496 |
| $(\rho, \nu) = (0.9, 5)$ | Mean | 0.8999 | 5.0167 | 0.8996 | 5.0651 | 0.8999 | 5.0339 | 0.8996 | 5.0715 |
| $F_{0,1} = F_{0,2} =$ | Var | 0.0067 | 0.1465 | 0.0068 | 0.1510 | 0.0067 | 0.1355 | 0.0067 | 0.1394 |
| $N(0,1)$ | MSE | 0.0068 | 0.1468 | 0.0069 | 0.1552 | 0.0067 | 0.1366 | 0.0068 | 0.1446 |
| $(\rho, \nu) = (0.9, 5)$ | Mean | 0.8999 | 5.0167 | 0.8996 | 5.0566 | 0.9000 | 5.0353 | 0.8997 | 5.0604 |
| $F_{0,1} = F_{0,2} =$ | Var | 0.0067 | 0.1465 | 0.0068 | 0.1508 | 0.0067 | 0.1439 | 0.0068 | 0.1477 |
| std-$t(10)$ | MSE | 0.0068 | 0.1468 | 0.0069 | 0.1541 | 0.0067 | 0.1451 | 0.0069 | 0.1513 |
| $(\rho, \nu) = (0.9, 5)$ | Mean | 0.8999 | 5.0167 | 0.8996 | 5.0528 | 0.9001 | 5.0588 | 0.8998 | 5.0799 |
| $F_{0,1} = F_{0,2} =$ | Var | 0.0067 | 0.1465 | 0.0068 | 0.1505 | 0.0069 | 0.1511 | 0.0069 | 0.1499 |
| std-$t(7)$ | MSE | 0.0068 | 0.1468 | 0.0070 | 0.1533 | 0.0069 | 0.1545 | 0.0070 | 0.1563 |

**Table A.4:** Student's t-copula with unknown margins: Estimation of marginal distributions. $n = 8000$, MC= 2000. $F_{0,j}(q_1) = 1/3$, $F_{0,j}(q_2) = 2/3$ for $j = 1, 2$. Reported Var and MSE are the true values multiplied by 1000.

| | | True-2Step | | | | Resid-2Step | | | | True-SML | | | | Resid-SML | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | |
| | True value | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| $\rho = 0.7$ | Mean | 0.3333 | 0.6666 | 0.3333 | 0.6664 | 0.3334 | 0.6665 | 0.3334 | 0.6663 | 0.3348 | 0.6682 | 0.3347 | 0.6680 | 0.3348 | 0.6681 | 0.3347 | 0.6680 |
| $\nu = 5$ | Var | 0.0264 | 0.0278 | 0.0271 | 0.0275 | 0.0252 | 0.0261 | 0.0251 | 0.0261 | 0.0173 | 0.0174 | 0.0172 | 0.0169 | 0.0161 | 0.0157 | 0.0156 | 0.0155 |
| | MSE | 0.0264 | 0.0278 | 0.0271 | 0.0276 | 0.0252 | 0.0261 | 0.0251 | 0.0262 | 0.0193 | 0.0197 | 0.0192 | 0.0188 | 0.0183 | 0.0178 | 0.0175 | 0.0173 |
| $\rho = 0.9$ | Mean | 0.3333 | 0.6666 | 0.3333 | 0.6665 | 0.3334 | 0.6665 | 0.3334 | 0.6664 | 0.3350 | 0.6684 | 0.3349 | 0.6683 | 0.3350 | 0.6683 | 0.3350 | 0.6683 |
| $\nu = 5$ | Var | 0.0264 | 0.0278 | 0.0273 | 0.0268 | 0.0252 | 0.0261 | 0.0255 | 0.0252 | 0.0173 | 0.0175 | 0.0169 | 0.0172 | 0.0162 | 0.0158 | 0.0158 | 0.0157 |
| | MSE | 0.0264 | 0.0278 | 0.0273 | 0.0269 | 0.0252 | 0.0261 | 0.0255 | 0.0253 | 0.0200 | 0.0204 | 0.0194 | 0.0199 | 0.0191 | 0.0185 | 0.0184 | 0.0183 |
| | | $F_{0,1}$ =std-$t(10)$ | | $F_{0,2}$ =std-$t(10)$ | | $F_{0,1}$ =std-$t(10)$ | | $F_{0,2}$ =std-$t(10)$ | | $F_{0,1}$ =std-$t(10)$ | | $F_{0,2}$ =std-$t(10)$ | | $F_{0,1}$ =std-$t(10)$ | | $F_{0,2}$ =std-$t(10)$ | |
| | True value | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| $\rho = 0.7$ | Mean | 0.3333 | 0.6666 | 0.3333 | 0.6664 | 0.3334 | 0.6665 | 0.3334 | 0.6663 | 0.3330 | 0.6670 | 0.3331 | 0.6669 | 0.3332 | 0.6668 | 0.3332 | 0.6668 |
| $\nu = 5$ | Var | 0.0264 | 0.0278 | 0.0271 | 0.0275 | 0.0258 | 0.0271 | 0.0259 | 0.0269 | 0.0051 | 0.0051 | 0.0053 | 0.0053 | 0.0042 | 0.0042 | 0.0042 | 0.0042 |
| | MSE | 0.0264 | 0.0278 | 0.0271 | 0.0276 | 0.0258 | 0.0271 | 0.0259 | 0.0271 | 0.0052 | 0.0052 | 0.0053 | 0.0053 | 0.0042 | 0.0042 | 0.0042 | 0.0042 |
| $\rho = 0.9$ | Mean | 0.3333 | 0.6666 | 0.3333 | 0.6665 | 0.3334 | 0.6665 | 0.3334 | 0.6664 | 0.3330 | 0.6670 | 0.3330 | 0.6670 | 0.3331 | 0.6669 | 0.3331 | 0.6669 |
| $\nu = 5$ | Var | 0.0264 | 0.0278 | 0.0273 | 0.0268 | 0.0258 | 0.0271 | 0.0262 | 0.0259 | 0.0046 | 0.0046 | 0.0048 | 0.0048 | 0.0035 | 0.0035 | 0.0035 | 0.0035 |
| | MSE | 0.0264 | 0.0278 | 0.0273 | 0.0269 | 0.0258 | 0.0271 | 0.0262 | 0.0260 | 0.0047 | 0.0048 | 0.0049 | 0.0049 | 0.0036 | 0.0036 | 0.0036 | 0.0035 |
| | | $F_{0,1}$ =std-$t(7)$ | | $F_{0,2}$ =std-$t(7)$ | | $F_{0,1}$ =std-$t(7)$ | | $F_{0,2}$ =std-$t(7)$ | | $F_{0,1}$ =std-$t(7)$ | | $F_{0,2}$ =std-$t(7)$ | | $F_{0,1}$ =std-$t(7)$ | | $F_{0,2}$ =std-$t(7)$ | |
| | True value | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| $\rho = 0.7$ | Mean | 0.3333 | 0.6666 | 0.3333 | 0.6664 | 0.3334 | 0.6664 | 0.3334 | 0.6663 | 0.3324 | 0.6675 | 0.3326 | 0.6674 | 0.3326 | 0.6674 | 0.3327 | 0.6674 |
| $\nu = 5$ | Var | 0.0264 | 0.0278 | 0.0271 | 0.0275 | 0.0265 | 0.0277 | 0.0266 | 0.0277 | 0.0053 | 0.0052 | 0.0051 | 0.0054 | 0.0055 | 0.0055 | 0.0056 | 0.0053 |
| | MSE | 0.0264 | 0.0278 | 0.0271 | 0.0276 | 0.0265 | 0.0277 | 0.0266 | 0.0279 | 0.0060 | 0.0059 | 0.0057 | 0.0060 | 0.0060 | 0.0060 | 0.0060 | 0.0058 |
| $\rho = 0.9$ | Mean | 0.3333 | 0.6666 | 0.3333 | 0.6665 | 0.3334 | 0.6664 | 0.3334 | 0.6663 | 0.3324 | 0.6675 | 0.3325 | 0.6674 | 0.3326 | 0.6674 | 0.3326 | 0.6674 |
| $\nu = 5$ | Var | 0.0264 | 0.0278 | 0.0273 | 0.0268 | 0.0265 | 0.0277 | 0.0269 | 0.0265 | 0.0052 | 0.0051 | 0.0049 | 0.0049 | 0.0049 | 0.0048 | 0.0048 | 0.0047 |
| | MSE | 0.0264 | 0.0278 | 0.0273 | 0.0269 | 0.0265 | 0.0277 | 0.0269 | 0.0266 | 0.0060 | 0.0058 | 0.0055 | 0.0055 | 0.0054 | 0.0054 | 0.0054 | 0.0053 |

**Table A.5:** Clayton copula with unknown margins: Estimation of copula parameters. $n = 8000$, MC= 2000. Reported Var and MSE are the true values multiplied by 1000.

| | | True-2Step | Resid-2Step | True-SML | Resid-SML |
|---|---|---|---|---|---|
| $\theta = 1$ | Mean | 1.0014 | 0.9993 | 1.0037 | 1.0017 |
| $F_{0,1} = F_{0,2} = N(0,1)$ | Var | 0.0008 | 0.0008 | 0.0007 | 0.0007 |
| | MSE | 0.0008 | 0.0008 | 0.0007 | 0.0007 |
| $\theta = 3$ | Mean | 2.9997 | 2.9834 | 3.0117 | 2.9952 |
| $F_{0,1} = F_{0,2} = N(0,1)$ | Var | 0.0035 | 0.0035 | 0.0032 | 0.0032 |
| | MSE | 0.0035 | 0.0037 | 0.0033 | 0.0032 |
| $\theta = 5$ | Mean | 4.9961 | 4.9499 | 5.0201 | 4.9740 |
| $F_{0,1} = F_{0,2} = N(0,1)$ | Var | 0.0078 | 0.0080 | 0.0075 | 0.0077 |
| | MSE | 0.0078 | 0.0105 | 0.0079 | 0.0084 |
| $\theta = 1$ | Mean | 1.0014 | 0.9988 | 1.0018 | 0.9994 |
| $F_{0,1} = F_{0,2} = \text{std-}t(5)$ | Var | 0.0008 | 0.0008 | 0.0007 | 0.0007 |
| | MSE | 0.0008 | 0.0008 | 0.0007 | 0.0007 |
| $\theta = 3$ | Mean | 2.9997 | 2.9789 | 3.0125 | 2.9938 |
| $F_{0,1} = F_{0,2} = \text{std-}t(5)$ | Var | 0.0035 | 0.0035 | 0.0031 | 0.0033 |
| | MSE | 0.0035 | 0.0040 | 0.0032 | 0.0034 |
| $\theta = 5$ | Mean | 4.9961 | 4.9340 | 5.0251 | 4.9703 |
| $F_{0,1} = F_{0,2} = \text{std-}t(5)$ | Var | 0.0078 | 0.0088 | 0.0073 | 0.0086 |
| | MSE | 0.0078 | 0.0132 | 0.0079 | 0.0095 |

**Table A.6:** Clayton copula with unknown margins: Estimation of marginal distributions. $n = 8000$, MC= 2000. $F_{0,j}(q_1) = 1/3$, $F_{0,j}(q_2) = 2/3$ for $j = 1, 2$. Reported Var and MSE are the true values multiplied by 1000.

| | | True-2Step | | | | Resid-2Step | | | | True-SML | | | | Resid-SML | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | |
| | True value | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| $\theta = 1$ | Mean | 0.3334 | 0.6668 | 0.3332 | 0.6666 | 0.3336 | 0.6666 | 0.3334 | 0.6665 | 0.3348 | 0.6682 | 0.3346 | 0.6679 | 0.3349 | 0.6680 | 0.3346 | 0.6678 |
| | Var | 0.0274 | 0.0264 | 0.0271 | 0.0277 | 0.0264 | 0.0247 | 0.0258 | 0.0256 | 0.0174 | 0.0179 | 0.0179 | 0.0179 | 0.0166 | 0.0163 | 0.0166 | 0.0166 |
| | MSE | 0.0274 | 0.0264 | 0.0271 | 0.0277 | 0.0264 | 0.0247 | 0.0258 | 0.0256 | 0.0197 | 0.0201 | 0.0195 | 0.0194 | 0.0189 | 0.0182 | 0.0183 | 0.0180 |
| $\theta = 3$ | Mean | 0.3334 | 0.6668 | 0.3334 | 0.6666 | 0.3336 | 0.6666 | 0.3336 | 0.6665 | 0.3350 | 0.6683 | 0.3349 | 0.6682 | 0.3349 | 0.6679 | 0.3348 | 0.6678 |
| | Var | 0.0274 | 0.0264 | 0.0266 | 0.0278 | 0.0264 | 0.0247 | 0.0254 | 0.0258 | 0.0173 | 0.0166 | 0.0172 | 0.0165 | 0.0161 | 0.0159 | 0.0156 | 0.0155 |
| | MSE | 0.0274 | 0.0264 | 0.0266 | 0.0278 | 0.0264 | 0.0247 | 0.0254 | 0.0259 | 0.0202 | 0.0193 | 0.0198 | 0.0187 | 0.0186 | 0.0175 | 0.0177 | 0.0169 |
| $\theta = 5$ | Mean | 0.3334 | 0.6668 | 0.3334 | 0.6666 | 0.3336 | 0.6666 | 0.3336 | 0.6665 | 0.3351 | 0.6684 | 0.3350 | 0.6682 | 0.3347 | 0.6674 | 0.3346 | 0.6674 |
| | Var | 0.0274 | 0.0264 | 0.0270 | 0.0279 | 0.0264 | 0.0247 | 0.0257 | 0.0261 | 0.0173 | 0.0160 | 0.0169 | 0.0158 | 0.0160 | 0.0158 | 0.0153 | 0.0153 |
| | MSE | 0.0274 | 0.0264 | 0.0270 | 0.0279 | 0.0264 | 0.0247 | 0.0257 | 0.0261 | 0.0203 | 0.0188 | 0.0196 | 0.0182 | 0.0178 | 0.0164 | 0.0169 | 0.0158 |
| | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | |
| | True value | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| $\theta = 1$ | Mean | 0.3334 | 0.6668 | 0.3332 | 0.6666 | 0.3337 | 0.6665 | 0.3335 | 0.6664 | 0.3315 | 0.6684 | 0.3314 | 0.6685 | 0.3320 | 0.6680 | 0.3318 | 0.6682 |
| | Var | 0.0274 | 0.0264 | 0.0271 | 0.0277 | 0.0300 | 0.0286 | 0.0300 | 0.0297 | 0.0081 | 0.0079 | 0.0081 | 0.0080 | 0.0127 | 0.0127 | 0.0116 | 0.0114 |
| | MSE | 0.0274 | 0.0264 | 0.0271 | 0.0277 | 0.0301 | 0.0286 | 0.0300 | 0.0298 | 0.0114 | 0.0111 | 0.0117 | 0.0114 | 0.0146 | 0.0145 | 0.0139 | 0.0138 |
| $\theta = 3$ | Mean | 0.3334 | 0.6668 | 0.3334 | 0.6666 | 0.3337 | 0.6665 | 0.3337 | 0.6664 | 0.3321 | 0.6679 | 0.3320 | 0.6680 | 0.3324 | 0.6676 | 0.3323 | 0.6677 |
| | Var | 0.0274 | 0.0264 | 0.0266 | 0.0278 | 0.0300 | 0.0286 | 0.0293 | 0.0299 | 0.0108 | 0.0106 | 0.0101 | 0.0102 | 0.0159 | 0.0159 | 0.0153 | 0.0151 |
| | MSE | 0.0274 | 0.0264 | 0.0266 | 0.0278 | 0.0301 | 0.0286 | 0.0294 | 0.0300 | 0.0124 | 0.0121 | 0.0119 | 0.0120 | 0.0169 | 0.0168 | 0.0165 | 0.0162 |
| $\theta = 5$ | Mean | 0.3334 | 0.6668 | 0.3334 | 0.6666 | 0.3337 | 0.6665 | 0.3337 | 0.6664 | 0.3325 | 0.6674 | 0.3325 | 0.6675 | 0.3326 | 0.6674 | 0.3325 | 0.6674 |
| | Var | 0.0274 | 0.0264 | 0.0270 | 0.0279 | 0.0301 | 0.0286 | 0.0297 | 0.0301 | 0.0127 | 0.0125 | 0.0128 | 0.0126 | 0.0183 | 0.0181 | 0.0180 | 0.0179 |
| | MSE | 0.0274 | 0.0264 | 0.0270 | 0.0279 | 0.0302 | 0.0286 | 0.0298 | 0.0302 | 0.0133 | 0.0130 | 0.0134 | 0.0133 | 0.0188 | 0.0187 | 0.0186 | 0.0185 |

**Table A.7:** Gumbel copula with unknown margins: Estimation of copula parameters. $n = 8000$, MC= 2000. Reported Var and MSE are the true values multiplied by 1000.

| | | True-2Step | Resid-2Step | True-SML | Resid-SML |
|---|---|---|---|---|---|
| $\theta = 2.5$ | Mean | 2.5005 | 2.4963 | 2.4969 | 2.4933 |
| $F_{0,1} = F_{0,2} = N(0,1)$ | Var | 0.00091 | 0.00091 | 0.00087 | 0.00087 |
| | MSE | 0.00091 | 0.00092 | 0.00088 | 0.00092 |
| $\theta = 5$ | Mean | 4.9973 | 4.9732 | 4.9926 | 4.9736 |
| $F_{0,1} = F_{0,2} = N(0,1)$ | Var | 0.0041 | 0.0041 | 0.0040 | 0.0040 |
| | MSE | 0.0041 | 0.0048 | 0.0040 | 0.0047 |
| $\theta = 2.5$ | Mean | 2.5005 | 2.4947 | 2.5059 | 2.5004 |
| $F_{0,1} = F_{0,2} = \text{std-}t(5)$ | Var | 0.00091 | 0.00090 | 0.00092 | 0.00094 |
| | MSE | 0.00091 | 0.00093 | 0.00095 | 0.00094 |
| $\theta = 5$ | Mean | 4.9973 | 4.9642 | 5.0185 | 4.9907 |
| $F_{0,1} = F_{0,2} = \text{std-}t(5)$ | Var | 0.0041 | 0.0041 | 0.0046 | 0.0053 |
| | MSE | 0.0041 | 0.0054 | 0.0049 | 0.0054 |

**Table A.8:** Gumbel copula with unknown margins: Estimation of marginal distributions. $n = 8000$, MC= 2000. $F_{0,j}(q_1) = 1/3$, $F_{0,j}(q_2) = 2/3$ for $j = 1, 2$. Reported Var and MSE are the true values multiplied by 1000.

| | | True-2Step | | | | Resid-2Step | | | | True-SML | | | | Resid-SML | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | | $F_{0,1} = N(0,1)$ | | $F_{0,2} = N(0,1)$ | |
| | True value | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| $\theta = 2.5$ | Mean | 0.3333 | 0.6665 | 0.3334 | 0.6667 | 0.3334 | 0.6664 | 0.3334 | 0.6665 | 0.3349 | 0.6683 | 0.3349 | 0.6683 | 0.3349 | 0.6682 | 0.3349 | 0.6682 |
| | Var | 0.0282 | 0.0273 | 0.0282 | 0.0271 | 0.0264 | 0.0260 | 0.0264 | 0.0258 | 0.0180 | 0.0172 | 0.0182 | 0.0177 | 0.0162 | 0.0161 | 0.0166 | 0.0165 |
| | MSE | 0.0282 | 0.0274 | 0.0282 | 0.0271 | 0.0264 | 0.0261 | 0.0264 | 0.0258 | 0.0206 | 0.0197 | 0.0207 | 0.0203 | 0.0188 | 0.0185 | 0.0191 | 0.0188 |
| $\theta = 5$ | Mean | 0.3334 | 0.6666 | 0.3334 | 0.6667 | 0.3336 | 0.6665 | 0.3334 | 0.6665 | 0.3350 | 0.6684 | 0.3350 | 0.6684 | 0.3352 | 0.6682 | 0.3352 | 0.6682 |
| | Var | 0.0278 | 0.0271 | 0.0282 | 0.0271 | 0.0257 | 0.0259 | 0.0264 | 0.0258 | 0.0185 | 0.0175 | 0.0183 | 0.0179 | 0.0166 | 0.0162 | 0.0166 | 0.0164 |
| | MSE | 0.0278 | 0.0271 | 0.0282 | 0.0271 | 0.0257 | 0.0260 | 0.0264 | 0.0258 | 0.0215 | 0.0204 | 0.0211 | 0.0207 | 0.0200 | 0.0185 | 0.0199 | 0.0187 |
| | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | | $F_{0,1}$ =std-$t(5)$ | | $F_{0,2}$ =std-$t(5)$ | |
| | True value | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| $\theta = 2.5$ | Mean | 0.3333 | 0.6665 | 0.3334 | 0.6667 | 0.3335 | 0.6664 | 0.3335 | 0.6665 | 0.3317 | 0.6683 | 0.3318 | 0.6682 | 0.3320 | 0.6680 | 0.3320 | 0.6681 |
| | Var | 0.0282 | 0.0273 | 0.0282 | 0.0271 | 0.0298 | 0.0298 | 0.0301 | 0.0291 | 0.0086 | 0.0086 | 0.0088 | 0.0087 | 0.0119 | 0.0122 | 0.0125 | 0.0123 |
| | MSE | 0.0282 | 0.0274 | 0.0282 | 0.0271 | 0.0298 | 0.0299 | 0.0301 | 0.0291 | 0.0112 | 0.0113 | 0.0112 | 0.0112 | 0.0137 | 0.0140 | 0.0143 | 0.0143 |
| $\theta = 5$ | Mean | 0.3334 | 0.6666 | 0.3334 | 0.6667 | 0.3336 | 0.6664 | 0.3335 | 0.6665 | 0.3323 | 0.6677 | 0.3323 | 0.6677 | 0.3325 | 0.6675 | 0.3325 | 0.6676 |
| | Var | 0.0278 | 0.0271 | 0.0282 | 0.0271 | 0.0290 | 0.0293 | 0.0301 | 0.0291 | 0.0113 | 0.0115 | 0.0116 | 0.0117 | 0.0163 | 0.0164 | 0.0163 | 0.0163 |
| | MSE | 0.0278 | 0.0271 | 0.0282 | 0.0271 | 0.0291 | 0.0293 | 0.0301 | 0.0291 | 0.0124 | 0.0126 | 0.0127 | 0.0128 | 0.0170 | 0.0171 | 0.0170 | 0.0172 |

**Table A.9:** Finite sample performance : Gaussian copula with unknown margins. $n = 500$, MC= 500. $F_{0,j}(q_1) = 1/3$, $F_{0,j}(q_2) = 2/3$ for $j = 1, 2$. Reported Var and MSE are the true values multiplied by 1000.

| Estimation of copula parameter | | True-2Step | Resid-2Step | True-SML | Resid-SML |
|---|---|---|---|---|---|
| Gaussian (0.7) | Mean | 0.7029 | 0.6949 | 0.7016 | 0.6935 |
| $F_{0,1} = F_{0,2} = std - t(5)$ | Var | 0.6029 | 0.6239 | 0.6175 | 0.6435 |
| | MSE | 0.6113 | 0.6497 | 0.6199 | 0.6861 |
| Gaussian (0.9) | Mean | 0.9006 | 0.8922 | 0.9012 | 0.8929 |
| $F_{0,1} = F_{0,2} = std - t(5)$ | Var | 0.0793 | 0.1009 | 0.0799 | 0.1018 |
| | MSE | 0.0796 | 0.1616 | 0.0814 | 0.1526 |

| Estimation of marginal distribution | | True-2Step | | Resid-2Step | | True-SML | | Resid-SML | |
|---|---|---|---|---|---|---|---|---|---|
| | True value | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| Gaussian (0.7) | Mean | 0.3335 | 0.6664 | 0.3365 | 0.6637 | 0.3331 | 0.6669 | 0.3361 | 0.6640 |
| $F_{0,1} = F_{0,2} = std - t(5)$ | Var | 0.5005 | 0.4382 | 0.4800 | 0.5055 | 0.0636 | 0.0619 | 0.0718 | 0.0742 |
| $F_{0,1}$ | MSE | 0.5006 | 0.4383 | 0.4903 | 0.5142 | 0.0636 | 0.0620 | 0.0795 | 0.0814 |
| | Mean | 0.3344 | 0.6669 | 0.3384 | 0.6630 | 0.3328 | 0.6673 | 0.3367 | 0.6633 |
| $F_{0,2}$ | Var | 0.4545 | 0.4751 | 0.4674 | 0.4999 | 0.0587 | 0.0577 | 0.0675 | 0.0694 |
| | MSE | 0.4556 | 0.4752 | 0.4928 | 0.5130 | 0.0589 | 0.0581 | 0.0790 | 0.0809 |
| Gaussian (0.9) | Mean | 0.3335 | 0.6664 | 0.3365 | 0.6637 | 0.3335 | 0.6665 | 0.3350 | 0.6651 |
| $F_{0,1} = F_{0,2} = std - t(5)$ | Var | 0.5005 | 0.4382 | 0.4800 | 0.5055 | 0.0561 | 0.0542 | 0.0597 | 0.0614 |
| $F_{0,1}$ | MSE | 0.5006 | 0.4383 | 0.4903 | 0.5142 | 0.0561 | 0.0542 | 0.0625 | 0.0639 |
| | Mean | 0.3342 | 0.6663 | 0.3375 | 0.6628 | 0.3334 | 0.6666 | 0.3356 | 0.6645 |
| $F_{0,2}$ | Var | 0.4768 | 0.4707 | 0.4651 | 0.5083 | 0.0562 | 0.0556 | 0.0585 | 0.0601 |
| | MSE | 0.4775 | 0.4708 | 0.4823 | 0.5232 | 0.0563 | 0.0556 | 0.0637 | 0.0648 |

**Table A.10:** Finite sample performance: Student's t-copula with unknown margins. $n = 500$, MC= 500. $F_{0,j}(q_1) = 1/3$, $F_{0,j}(q_2) = 2/3$ for $j = 1, 2$. Reported Var and MSE are the true values multiplied by 1000.

| | | True-2Step | | Resid-2Step | | True-SML | | Resid-SML | |
|---|---|---|---|---|---|---|---|---|---|
| Estimation of copula parameter | | $\rho$ | $\nu$ | $\rho$ | $\nu$ | $\rho$ | $\nu$ | $\rho$ | $\nu$ |
| $(\rho, \nu) = (0.9, 5)$ | Mean | 0.9001 | 5.4547 | 0.8941 | 6.3343 | 0.9008 | 5.5357 | 0.8945 | 6.1518 |
| $F_{0,1} = F_{0,2} = std - t(10)$ | Var | 0.1144 | 3.5887 | 0.1294 | 9.3873 | 0.1140 | 3.7805 | 0.1286 | 7.6620 |
| | MSE | 0.1144 | 3.7954 | 0.1648 | 11.1678 | 0.1146 | 4.0674 | 0.1590 | 8.9887 |
| $(\rho, \nu) = (0.9, 5)$ | Mean | 0.9001 | 5.4547 | 0.8938 | 6.2530 | 0.9008 | 5.5277 | 0.8943 | 6.0691 |
| $F_{0,1} = F_{0,2} = std - t(7)$ | Var | 0.1144 | 3.5887 | 0.1296 | 9.2709 | 0.1137 | 3.7144 | 0.1291 | 8.0912 |
| | MSE | 0.1144 | 3.7954 | 0.1679 | 10.8408 | 0.1144 | 3.9929 | 0.1618 | 9.2341 |
| Estimation of marginal distribution | | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| $(\rho, \nu) = (0.9, 5)$ | Mean | 0.3338 | 0.6658 | 0.3362 | 0.6634 | 0.3337 | 0.6662 | 0.3342 | 0.6657 |
| $F_{0,1} = F_{0,2} = std - t(10)$ | Var | 0.4687 | 0.4283 | 0.4301 | 0.4353 | 0.0682 | 0.0672 | 0.0435 | 0.0425 |
| $F_{0,1}$ | MSE | 0.4689 | 0.4291 | 0.4383 | 0.4461 | 0.0683 | 0.0674 | 0.0443 | 0.0433 |
| | Mean | 0.3337 | 0.6661 | 0.3369 | 0.6633 | 0.3335 | 0.6664 | 0.3344 | 0.6656 |
| $F_{0,2}$ | Var | 0.4825 | 0.4690 | 0.4125 | 0.4677 | 0.0711 | 0.0704 | 0.0429 | 0.0433 |
| | MSE | 0.4826 | 0.4694 | 0.4252 | 0.4792 | 0.0711 | 0.0705 | 0.0440 | 0.0445 |
| $(\rho, \nu) = (0.9, 5)$ | Mean | 0.3338 | 0.6658 | 0.3363 | 0.6632 | 0.3336 | 0.6662 | 0.3344 | 0.6654 |
| $F_{0,1} = F_{0,2} = std - t(7)$ | Var | 0.4687 | 0.4283 | 0.4360 | 0.4462 | 0.0662 | 0.0663 | 0.0487 | 0.0491 |
| $F_{0,1}$ | MSE | 0.4689 | 0.4291 | 0.4449 | 0.4583 | 0.0663 | 0.0664 | 0.0498 | 0.0506 |
| | Mean | 0.3337 | 0.6661 | 0.3370 | 0.6632 | 0.3335 | 0.6664 | 0.3346 | 0.6653 |
| $F_{0,2}$ | Var | 0.4825 | 0.4690 | 0.4155 | 0.4764 | 0.0692 | 0.0681 | 0.0465 | 0.0473 |
| | MSE | 0.4826 | 0.4694 | 0.4293 | 0.4885 | 0.0692 | 0.0682 | 0.0481 | 0.0490 |

**Table A.11:** Finite sample performance: Estimation of copula parameters with unknown margins. $n = 500$, MC= 500. Reported Var and MSE are the true values multiplied by 1000.

| | | True-2Step | Resid-2Step | True-SML | Resid-SML |
|---|---|---|---|---|---|
| Clayton(3) | Mean | 3.0248 | 2.7770 | 3.0468 | 2.8127 |
| $F_{0,1} = F_{0,2} = std - t(5)$ | Var | 0.0559 | 0.0550 | 0.0406 | 0.0419 |
| | MSE | 0.0565 | 0.1047 | 0.0428 | 0.0770 |
| Clayton(5) | Mean | 5.0015 | 4.3589 | 5.0706 | 4.5060 |
| $F_{0,1} = F_{0,2} = std - t(5)$ | Var | 0.1291 | 0.1566 | 0.0894 | 0.1227 |
| | MSE | 0.1291 | 0.5677 | 0.0944 | 0.3668 |
| Gumbel(2.5) | Mean | 2.5081 | 2.4268 | 2.5113 | 2.4373 |
| $F_{0,1} = F_{0,2} = std - t(5)$ | Var | 0.0150 | 0.0132 | 0.0139 | 0.0128 |
| | MSE | 0.0151 | 0.0186 | 0.0141 | 0.0167 |
| Gumbel(5) | Mean | 4.9751 | 4.5314 | 5.0346 | 4.6465 |
| $F_{0,1} = F_{0,2} = std - t(5)$ | Var | 0.0672 | 0.0725 | 0.0655 | 0.0664 |
| | MSE | 0.0678 | 0.2921 | 0.0667 | 0.1914 |

**Table A.12:** Finite sample performance: Estimation of marginal distributions with unknown margins. $n = 500$, MC= 500. $F_{0,1} = F_{0,2} =$ std-$t(5)$; $F_{0,j}(q_1) = 1/3$, $F_{0,j}(q_2) = 2/3$ for $j = 1, 2$. Reported Var and MSE are the true values multiplied by 1000.

| | | True-2Step | | Resid-2Step | | True-SML | | Resid-SML | |
|---|---|---|---|---|---|---|---|---|---|
| | True value | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 1/3 | 2/3 |
| Clayton | Mean | 0.3359 | 0.6664 | 0.3388 | 0.6628 | 0.3333 | 0.6665 | 0.3349 | 0.6652 |
| $\theta = 3$ | Var | 0.4432 | 0.4297 | 0.4451 | 0.4897 | 0.0472 | 0.0467 | 0.0662 | 0.0667 |
| $F_{0,1}$ | MSE | 0.4498 | 0.4297 | 0.4750 | 0.5046 | 0.0472 | 0.0468 | 0.0687 | 0.0689 |
| | Mean | 0.3337 | 0.6674 | 0.3374 | 0.6642 | 0.3330 | 0.6668 | 0.3345 | 0.6656 |
| | Var | 0.4132 | 0.4329 | 0.4565 | 0.4597 | 0.0440 | 0.0412 | 0.0668 | 0.0652 |
| $F_{0,2}$ | MSE | 0.4134 | 0.4334 | 0.4731 | 0.4660 | 0.0442 | 0.0412 | 0.0681 | 0.0662 |
| Clayton | Mean | 0.3359 | 0.6664 | 0.3388 | 0.6628 | 0.3335 | 0.6664 | 0.3344 | 0.6660 |
| $\theta = 5$ | Var | 0.4432 | 0.4297 | 0.4451 | 0.4897 | 0.0451 | 0.0453 | 0.0623 | 0.0637 |
| $F_{0,1}$ | MSE | 0.4498 | 0.4297 | 0.4750 | 0.5046 | 0.0451 | 0.0454 | 0.0634 | 0.0641 |
| | Mean | 0.3344 | 0.6674 | 0.3375 | 0.6643 | 0.3331 | 0.6667 | 0.3340 | 0.6664 |
| | Var | 0.4185 | 0.4528 | 0.4576 | 0.4695 | 0.0445 | 0.0431 | 0.0612 | 0.0590 |
| $F_{0,2}$ | MSE | 0.4196 | 0.4534 | 0.4747 | 0.4750 | 0.0446 | 0.0431 | 0.0617 | 0.0591 |
| Gumbel | Mean | 0.3334 | 0.6685 | 0.3369 | 0.6648 | 0.3324 | 0.6676 | 0.3352 | 0.6648 |
| $\theta = 2.5$ | Var | 0.4152 | 0.4658 | 0.4278 | 0.5096 | 0.0533 | 0.0536 | 0.0638 | 0.0646 |
| $F_{0,1}$ | MSE | 0.4152 | 0.4691 | 0.4402 | 0.5133 | 0.0541 | 0.0545 | 0.0673 | 0.0682 |
| | Mean | 0.3340 | 0.6676 | 0.3376 | 0.6640 | 0.3326 | 0.6674 | 0.3352 | 0.6647 |
| | Var | 0.4204 | 0.4339 | 0.4636 | 0.4710 | 0.0547 | 0.0569 | 0.0698 | 0.0710 |
| $F_{0,2}$ | MSE | 0.4209 | 0.4348 | 0.4817 | 0.4781 | 0.0553 | 0.0575 | 0.0734 | 0.0748 |
| Gumbel | Mean | 0.3331 | 0.6672 | 0.3370 | 0.6636 | 0.3327 | 0.6674 | 0.3341 | 0.6656 |
| $\theta = 5$ | Var | 0.4212 | 0.4198 | 0.4309 | 0.4731 | 0.0457 | 0.0451 | 0.0524 | 0.0519 |
| $F_{0,1}$ | MSE | 0.4212 | 0.4201 | 0.4444 | 0.4827 | 0.0461 | 0.0455 | 0.0529 | 0.0529 |
| | Mean | 0.3340 | 0.6676 | 0.3376 | 0.6640 | 0.3328 | 0.6672 | 0.3339 | 0.6657 |
| | Var | 0.4204 | 0.4339 | 0.4580 | 0.4689 | 0.0482 | 0.0483 | 0.0574 | 0.0582 |
| $F_{0,2}$ | MSE | 0.4209 | 0.4348 | 0.4760 | 0.4759 | 0.0485 | 0.0486 | 0.0578 | 0.0591 |

**Table A.13:** Parameter Estimation for Five-asset Student's t-copula with SemiGARCH Filtering of Liu and Yang (2016)

| **2Step Method** | | | | | | |
|---|---|---|---|---|---|---|
| | | S&P500 | NAS100 | HYB | MBS | GSCI |
| | S&P500 | 1 | 0.901 | 0.415 | -0.279 | 0.332 |
| | | | (0.001) | (0.021) | (0.040) | (0.025) |
| | NAS100 | 0.901 | 1 | 0.350 | -0.242 | 0.254 |
| Correlation | | (0.001) | | (0.024) | (0.038) | (0.028) |
| Matrix | HYB | 0.415 | 0.350 | 1 | -0.017 | 0.279 |
| | | (0.021) | (0.024) | | (0.029) | (0.025) |
| | MBS | -0.279 | -0.242 | -0.017 | 1 | -0.158 |
| | | (0.040) | (0.038) | (0.029) | | (0.028) |
| | GSCI | 0.332 | 0.254 | 0.279 | -0.158 | 1 |
| | | (0.025) | (0.028) | (0.025) | (0.028) | |
| $1/\nu$ | 0.0945 | | | | | |
| | (0.0083) | | | | | |
| logC | 2972.54 | | | | | |

| **SML Method** | | | | | | |
|---|---|---|---|---|---|---|
| | | S&P500 | NAS100 | HYB | MBS | GSCI |
| | S&P500 | 1 | 0.904 | 0.421 | -0.279 | 0.338 |
| | | | (0.003) | (0.016) | (0.017) | (0.017) |
| | NAS100 | 0.904 | 1 | 0.355 | -0.241 | 0.260 |
| Correlation | | (0.003) | | (0.018) | (0.019) | (0.018) |
| Matrix | HYB | 0.421 | 0.355 | 1 | -0.016 | 0.285 |
| | | (0.016) | (0.018) | | (0.020) | (0.019) |
| | MBS | -0.279 | -0.241 | -0.016 | 1 | -0.158 |
| | | (0.017) | (0.019) | (0.020) | | (0.020) |
| | GSCI | 0.338 | 0.260 | 0.285 | -0.158 | 1 |
| | | (0.017) | (0.018) | (0.019) | (0.020) | |
| $1/\nu$ | 0.0940 | | | | | |
| | (0.0085) | | | | | |
| logC | 3039.20 | | | | | |

**Table A.14:** Evaluation of Full sample and Out-of-sample Value-at-Risk Forecasts with Semi-GARCH Filtering of Liu and Yang (2016)

**Panel A: Backtesting of Full-sample Value-at-Risk Forecasts**

| 99% VaR | | | | | | |
|---|---|---|---|---|---|---|
| | 2Step | | | SML | | |
| | POF(%) | p_POF | p_CC | POF(%) | p_POF | p_CC |
| S&P500-NAS | 0.692 | 0.087 | 0.202 | 0.802 | 0.280 | 0.217 |
| S&P500-HYB | 0.692 | 0.087 | 0.202 | 0.765 | 0.198 | 0.371 |
| S&P500-MBS | 0.875 | 0.500 | 0.645 | 0.838 | 0.381 | 0.561 |
| S&P500-GSCI | 0.911 | 0.635 | 0.710 | 0.911 | 0.635 | 0.710 |
| Five-asset | 0.620 | 0.031 | 0.088 | 1.020 | 0.915 | 0.745 |
| 95% VaR | | | | | | |
| | 2Step | | | SML | | |
| | POF(%) | p_POF | p_CC | POF(%) | p_POF | p_CC |
| S&P500-NAS | 4.118 | 0.029 | 0.033 | 4.555 | 0.278 | 0.241 |
| S&P500-HYB | 4.300 | 0.085 | 0.133 | 4.264 | 0.070 | 0.117 |
| S&P500-MBS | 3.571 | 0.000 | 0.001 | 3.608 | 0.000 | 0.002 |
| S&P500-GSCI | 4.082 | 0.023 | 0.054 | 4.227 | 0.057 | 0.102 |
| Five-asset | 4.300 | 0.085 | 0.133 | 5.029 | 0.944 | 0.717 |

**Panel B: D-M Test of Out-of-sample Value-at-Risk Forecasts**

| 99% VaR | | | | |
|---|---|---|---|---|
| | POF_2Step(%) | POF_Sieve(%) | DM_stat | p-value |
| S&P500-NAS | 0.67 | 0.76 | 2.98 | 0.00 |
| S&P500-HYB | 0.67 | 0.76 | 1.79 | 0.07 |
| S&P500-MBS | 0.62 | 0.71 | 2.94 | 0.00 |
| S&P500-GSCI | 0.71 | 0.80 | 1.04 | 0.30 |
| 95% VaR | | | | |
| | POF_2Step(%) | POF_Sieve(%) | DM_stat | p-value |
| S&P500-NAS | 3.56 | 4.05 | 1.05 | 0.29 |
| S&P500-HYB | 3.38 | 3.56 | 1.69 | 0.09 |
| S&P500-MBS | 3.65 | 3.61 | 1.26 | 0.21 |
| S&P500-GSCI | 4.36 | 4.45 | 2.36 | 0.02 |

# B Semi-nonparametric GARCH Models : Asymptotic Results

In this part, we justify Assumption 4 in Eqs.(B.8) and (B.9). We verify Assumption 5(i) in Lemma B.1 when $\kappa$ is fully nonparametric, and in Lemmas B.2, B.3 when $\kappa$ is semi-nonparametric. Different semi-nonparametric GARCH models might restrict their choices of $\mathcal{B}$ and $\mathcal{B}_n$ to certain classes (see, e.g. Linton and Mammen (2005), Yang (2006)). We emphasize the fact that our main theorems do not rely on the specific form of the GARCH filters, as long as the dynamic model is correctly specified and Assumptions 4, 5(i) are satisfied. For a concrete example, see Appendix B.2.

## B.1 Justification of Assumption 5(i) when $\kappa$ is Nonparametric

Assume $\kappa_0 = (\mu_{0,1}(\cdot), \cdots, \mu_{0,k}(\cdot), \sigma_{0,1}(\cdot), \cdots, \sigma_{0,k}(\cdot))^\intercal \in \mathcal{B}$ and $\widehat{\kappa} = (\widehat{\mu}_1(\cdot), \cdots, \widehat{\mu}_k(\cdot), \widehat{\sigma}_1(\cdot), \cdots, \widehat{\sigma}_k(\cdot))^\intercal \in \mathcal{B}_n$. For each $j = 1, \cdots, k$, let $\mathbb{W}_{j,\mu} \times \mathbb{W}_{j,\sigma}$ be the linear span of $\mathcal{M}_j \times \mathcal{H}_j - \{\mu_{0,j}, \sigma_{0,j}\}$ and suppose

$$(\mu_{0,j}, \sigma_{0,j}) = \underset{\mu_j \in \mathcal{M}_j, \sigma_j \in \mathcal{H}_j}{\arg\min} E\left[\frac{1}{2}\left(\frac{Y_{j,t} - \mu_j^t}{\sigma_j^t}\right)^2 + \log \sigma_j^t\right]. \tag{B.1}$$

We estimate $\mu_j$ and $\sigma_j$ using the sample analogs of Eq.(B.1).

Denote $\Psi(\mu_j, \sigma_j, Y_{j,t}) = \frac{1}{2}\left(\frac{Y_{j,t} - \mu_j^t}{\sigma_j^t}\right)^2 + \log \sigma_j^t$. We note that $\Psi(\mu_j, \sigma_j, Y_{j,t})$ is twice continuously differentiable in $\mathcal{M}_j \times \mathcal{H}_j$ if $\forall \sigma_j(\cdot) \in \mathcal{H}_j$ is strictly positive on the support. $\forall (w_{\mu,j}, w_{\sigma,j}), (\widetilde{w}_{\mu,j}, \widetilde{w}_{\sigma,j}) \in \mathbb{W}_{j,\mu} \times \mathbb{W}_{j,\sigma}$, the first-order and second-order directional derivatives of $\Psi$ are $\frac{\partial \Psi(\mu_j, \sigma_j, Y_{j,t})}{\partial (\mu_j, \sigma_j)^\intercal}(w_{\mu,j}, w_{\sigma,j})^\intercal = -\frac{Y_{j,t} - \mu_j^t}{\sigma_j^t}\frac{w_{\mu,j}^t}{\sigma_j^t} - \left(\frac{Y_{j,t} - \mu_j^t}{\sigma_j^t}\right)^2 \frac{w_{\sigma,j}^t}{\sigma_j^t} + \frac{w_{\sigma,j}^t}{\sigma_j^t}$; $\frac{\partial^2 \Psi(\mu_j, \sigma_j, Y_{j,t})}{\partial (\mu_j, \sigma_j)^\intercal \partial (\mu_j, \sigma_j)}[(w_{\mu,j}, w_{\sigma,j})^\intercal, (\widetilde{w}_{\mu,j}, \widetilde{w}_{\sigma,j})^\intercal] = \frac{w_{\mu,j}^t}{\sigma_j^t}\frac{\widetilde{w}_{\mu,j}^t}{\sigma_j^t} + 3\left(\frac{Y_{j,t} - \mu_j^t}{\sigma_j^t}\right)^2 \frac{w_{\sigma,j}^t}{\sigma_j^t}\frac{\widetilde{w}_{\sigma,j}^t}{\sigma_j^t} - \frac{w_{\sigma,j}^t}{\sigma_j^t}\frac{\widetilde{w}_{\sigma,j}^t}{\sigma_j^t} + 2\left(\frac{Y_{j,t} - \mu_j^t}{\sigma_j^t}\right)\frac{w_{\mu,j}^t}{\sigma_j^t}\frac{\widetilde{w}_{\sigma,j}^t}{\sigma_j^t} + 2\left(\frac{Y_{j,t} - \mu_j^t}{\sigma_j^t}\right)\frac{\widetilde{w}_{\mu,j}^t}{\sigma_j^t}\frac{w_{\sigma,j}^t}{\sigma_j^t}$. We define the norm and the inner product (induced by the criterion function) on $\mathbb{W}_{j,\mu} \times \mathbb{W}_{j,\sigma}$ :

$$\begin{aligned}
\langle (w_{\mu,j}, w_{\sigma,j})^\intercal, (\widetilde{w}_{\mu,j}, \widetilde{w}_{\sigma,j})^\intercal \rangle_j &\equiv E\left(\frac{\partial^2 \Psi(\mu_{0,j}, \sigma_{0,j}, Y_{j,t})}{\partial (\mu_j, \sigma_j)^\intercal \partial (\mu_j, \sigma_j)}[(w_{\mu,j}, w_{\sigma,j})^\intercal, (\widetilde{w}_{\mu,j}, \widetilde{w}_{\sigma,j})^\intercal]\right) \\
&= E\left(\frac{w_{\mu,j}^t}{\sigma_{0,j}^t}\frac{\widetilde{w}_{\mu,j}^t}{\sigma_{0,j}^t}\right) + 2E\left(\frac{w_{\sigma,j}^t}{\sigma_{0,j}^t}\frac{\widetilde{w}_{\sigma,j}^t}{\sigma_{0,j}^t}\right), \\
\|(w_{\mu,j}, w_{\sigma,j})^\intercal\|_j^2 &= E\left(\frac{w_{\mu,j}^t}{\sigma_{0,j}^t}\right)^2 + 2E\left(\frac{w_{\sigma,j}^t}{\sigma_{0,j}^t}\right)^2.
\end{aligned} \tag{B.2}$$

Define the norm for $w = (w_{\mu,1}, \cdots, w_{\mu,k}, w_{\sigma,1}, \cdots, w_{\sigma,k})^\intercal \in \mathbb{W}$ as :

$$||w||^2 = \sum_{j=1}^{k} \|(w_{\mu,j}, w_{\sigma,j})^\intercal\|_j^2 . \tag{B.3}$$

Let $\overline{\mathbb{W}}_{j,\mu}$, $\overline{\mathbb{W}}_{j,\sigma}$ and $\overline{\mathbb{W}}$ be the closures of $\mathbb{W}_{j,\mu}$, $\mathbb{W}_{j,\sigma}$ and $\mathbb{W}$ under the Fisher norms.

We are interested in two linear functionals: $(i)$ $\rho_{j,\mu}(\mu_j, \sigma_j) = E\left(\frac{\mu_j^t}{\sigma_{0,j}^t}\right)$, $\forall(\mu_j, \sigma_j) \in \mathcal{M}_j \times \mathcal{H}_j$. It is easy to see $w_{\mu_j}^* = (\sigma_{0,j}, 0)^\intercal$ is the unique Riesz representer of $\rho_{j,\mu}$ under the norm defined by Eq.(B.2); $(ii)$ $\rho_{j,\sigma}(\mu_j, \sigma_j) = E\left(\frac{\sigma_j^t}{\sigma_{0,j}^t}\right)$, $\forall(\mu_j, \sigma_j) \in \mathcal{M}_j \times \mathcal{H}_j$. $w_{\sigma_j}^* = \left(0, \frac{1}{2}\sigma_{0,j}\right)^\intercal$ is the unique Riesz representer of $\rho_{j,\sigma}$, under the norm defined by Eq.(B.2). In the following lemma, we verify Assumption 5(i) for $\sqrt{n}\left(\rho_{j,\mu}(\widehat{\mu}_j, \widehat{\sigma}_j) - \rho_{j,\mu}(\mu_{0,j}, \sigma_{0,j})\right)$ and $\sqrt{n}\left(\rho_{j,\sigma}(\widehat{\mu}_j, \widehat{\sigma}_j) - \rho_{j,\sigma}(\mu_{0,j}, \sigma_{0,j})\right)$, which are used to establish the main theorems of this paper.

Let $\epsilon_n = o\left(n^{-1/2}\right)$. Denote $\Pi_n(\cdot)$ as the projection of any infinitely dimensional space (e.g., $\mathcal{M}_j$, $\mathcal{H}_j$) to its sieve space (e.g., $\mathcal{M}_{j,n}$, $\mathcal{H}_{j,n}$). Denote $\mathcal{N}_{\kappa_j,n} = \{(\mu_j, \sigma_j) \in \mathcal{M}_{j,n} \times \mathcal{H}_{j,n} : ||\,(\mu_j - \mu_{0,j}, \sigma_j - \sigma_{0,j})^\intercal\,||_j \leq O\left(\log(\log n)\delta_{h,n}\right)\}$ and $\mathcal{N}_{\sigma_j,n} = \{\sigma_j \in \mathcal{H}_{j,n} : (\mu_j, \sigma_j) \in \mathcal{N}_{\kappa_j,n}\}$. Let $\widehat{\kappa}_j = (\widehat{\mu}_j, \widehat{\sigma}_j)^\intercal$. For $\kappa_j \in \mathcal{N}_{\kappa_j,n}$, we consider local alternative values $\kappa_j \pm \epsilon_n \Pi_n\left(w_{\mu_j}^*\right)$ and $\kappa_j \pm \epsilon_n \Pi_n\left(w_{\sigma_j}^*\right)$. Define $\mu_n(g) = \frac{1}{n}\sum_{t=1}^{n}[g\left(Y_{j,t}, \mathcal{F}^{t-1}\right) - \mathbb{E}_X g\left(Y_{j,t}, \mathcal{F}^{t-1}\right)]$.

**Assumption B.1.** *(*i*)* $(\mu_{0,j}, \sigma_{0,j})$ *is the unique minimizer of Eq.*(B.1) *over* $\mathcal{M}_j \times \mathcal{H}_j \subset \Lambda^{\gamma_j}\left(\textit{support}\left(\mathcal{F}^{t-1}\right)\right)$, $\gamma_j > \frac{1}{2}$*; (*ii*)* $E\left[\left(\sigma_{0,j}^t\right)^4\right] < \infty$, $E\left[\xi_{j,t}^4\right] < \infty$ *and* $\inf \sigma_{0,j} > 0$*; (*iii*)* $\forall \sigma_j \in \mathcal{N}_{\sigma_j,n}$ *is strictly positive on the support ; (*iv*)* $\widehat{\kappa}_j$ *satisfies Assumption 4; (*v*)* $\overline{\mathbb{W}}_{j,\mu} \times \overline{\mathbb{W}}_{j,\sigma} \subset \{(w_{\mu,j}, w_{\sigma,j}) : E\left(\frac{w_{\mu,j}^t}{\sigma_{0,j}^t}\right)^2 < \infty,\ E\left(\frac{w_{\sigma,j}^t}{\sigma_{0,j}^t}\right)^2 < \infty\}$*; (*vi*)* $\sigma_{0,j} \in \overline{\mathbb{W}}_{j,\mu}$, $\sigma_{0,j} \in \overline{\mathbb{W}}_{j,\sigma}$*, and there exist* $\Pi_n^\mu(\sigma_{0,j}) \in \mathbb{W}_{j,\mu}^n$, $\Pi_n^\sigma(\sigma_{0,j}) \in \mathbb{W}_{j,\sigma}^n$ *such that* $E\left(\frac{\Pi_n^\mu(\sigma_{0,j})^t - \sigma_{0,j}^t}{\sigma_{0,j}^t}\right)^2 = o(n^{-1/2})$ *and* $E\left(\frac{\Pi_n^\sigma(\sigma_{0,j})^t - \sigma_{0,j}^t}{\sigma_{0,j}^t}\right)^2 = o(n^{-1/2})$*; (*vii*) uniformly over* $\kappa_j \in \mathcal{N}_{\kappa_j,n}$ *and* $w^* \in \{w_{\mu_j}^*, w_{\sigma_j}^*\}$, $E\left\{\Psi\left(\kappa_j \pm \epsilon_n \Pi_n\left(w^*\right), Y_{j,t}\right) - \Psi\left(\kappa_{0,j}, Y_{j,t}\right) - \Psi\left(\kappa_j, Y_{j,t}\right) + \Psi\left(\kappa_{0,j}, Y_{j,t}\right)\right\} = \frac{1}{2}\left\|\kappa_j \pm \epsilon_n \Pi_n\left(w^*\right) - \kappa_{0,j}\right\|_j^2 - \frac{1}{2}\left\|\kappa_j - \kappa_{0,j}\right\|_j^2 + O\left(\epsilon_n^2\right)$*; (*viii*) uniformly over* $\kappa_j \in \mathcal{N}_{\kappa_j,n}$ *and* $w^* \in \{w_{\mu_j}^*, w_{\sigma_j}^*\}$, $\mu_n\left(\frac{\partial \Psi(\kappa_j, Y_{j,t})}{\partial(\mu_j, \sigma_j)^\intercal}\Pi_n\left(w^*\right) - \frac{\partial \Psi(\kappa_{0,j}, Y_{j,t})}{\partial(\mu_j, \sigma_j)^\intercal}\Pi_n\left(w^*\right)\right) = o_p\left(n^{-1/2}\right)$.

Assumption B.1 (iv) requires $\left\{Y_{j,t}, \mu_j^t, \sigma_j^t,\ 1 \leq t \leq n\right\}$ satisfy certain mixing conditions over the parameter space $\mathcal{M}_j \times \mathcal{H}_j$ (see, e.g. Carrasco and Chen (2002), Linton and Mammen (2005), Yang (2006), Francq and Zakoian (2019)). Under Assumption B.1 (vi), $\Pi_n\left(w_{\mu_j}^*\right) = \left(\Pi_n^\mu\left(\sigma_{0,j}\right), 0\right)^\intercal$

and $\Pi_n\left(w^*_{\sigma_j}\right) = \left(0, \frac{1}{2}\Pi^\sigma_n(\sigma_{0,j})\right)^\intercal$ with $\left\|\Pi_n\left(w^*_{\mu_j}\right) - w^*_{\mu_j}\right\|_j = o(n^{-1/4})$ and $\left\|\Pi_n\left(w^*_{\sigma_j}\right) - w^*_{\sigma_j}\right\|_j = o(n^{-1/4})$. Assumption B.1 (vii) characterizes the local quadratic behavior of the population criterion, while (viii) makes the stochastic equicontinuity condition, which are standard assumptions in the sieve method literature.

**Lemma B.1.** *For $j = 1, \cdots, k$, under Assumption B.1, as $n \to \infty$,*

$$\text{(i)} \quad \sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}^t_j - \mu^t_{0,j}}{\sigma^t_{0,j}}\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \xi_{j,t} + o_p(1) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1);$$

$$\text{(ii)} \quad \sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}^t_j - \sigma^t_{0,j}}{\sigma^t_{0,j}}\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{1}{2}\left(\xi^2_{j,t} - 1\right) + o_p(1) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4}E\left[\xi^4_{j,t} - 1\right]\right).$$

## Proof of Lemma B.1

By definition of $\widehat{\kappa}_j$ and $\kappa_{0,j}$, it is easy to show that (see similar proofs in e.g. Chen and Shen (1998), Chen et al. (2006)) : $0 \le \frac{1}{n}\sum_{t=1}^n\left[\Psi\left(\widehat{\kappa}_j \pm \epsilon_n\Pi_n(w^*), Y_{j,t}\right) - \Psi\left(\widehat{\kappa}_j, Y_{j,t}\right)\right] = \pm\epsilon_n\mu_n\left(\frac{\partial\Psi(\kappa_{0,j},Y_{j,t})}{\partial(\mu_j,\sigma_j)^\intercal}w^*\right) \pm \epsilon_n\left\langle\widehat{\kappa}_j - \kappa_{0,j}, w^*\right\rangle_j + \epsilon_n \times o_p\left(n^{-1/2}\right)$, where $w^* \in \{w^*_{\mu_j}, w^*_{\sigma_j}\}$. Therefore,

$$\sqrt{n}\left\langle\widehat{\kappa}_j - \kappa_{0,j}, w^*\right\rangle_j = \sqrt{n}\mu_n\left(-\frac{\partial\Psi(\kappa_{0,j}, Y_{j,t})}{\partial(\mu_j,\sigma_j)^\intercal}w^*\right) + o_p(1). \tag{B.4}$$

Part $(i)$ : $w^* = w^*_{\mu_j} = (\sigma_{0,j}, 0)^\intercal$. Thus $\sqrt{n}\left\langle\widehat{\kappa}_j - \kappa_{0,j}, w^*\right\rangle_j = \sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}^t_j - \mu^t_{0,j}}{\sigma^t_{0,j}}\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \xi_{j,t} + o_p(1) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Part $(ii)$ : $w^* = w^*_{\sigma_j} = \left(0, \frac{1}{2}\sigma_{0,j}\right)^\intercal$. Thus $\sqrt{n}\left\langle\widehat{\kappa}_j - \kappa_{0,j}, w^*\right\rangle_j = \sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}^t_j - \sigma^t_{0,j}}{\sigma^t_{0,j}}\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{1}{2}\left(\xi^2_{j,t} - 1\right) + o_p(1) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4}E\left[\xi^4_{j,t} - 1\right]\right)$. □

### B.2 An Illustration : Semi-nonparametric GARCH Filter using Liu and Yang (2016)

In this section, we verify Assumptions 4 and 5(i) in the theoretical framework of Liu and Yang (2016). The following univariate semi-nonparametric GARCH models are used to prefilter the temporal dependence of each $\{Y_{j,t}\}_{t=1}^n$ for $j = 1, \cdots, k$ :

$$Y_{j,t} = \sigma_{0,j}\left(\mathcal{F}^{t-1}\right)\xi_{j,t} = m_j^{1/2}(X_{j,t})\xi_{j,t}, \; X_{j,t} = \sum_{i=1}^\infty \beta^{i-1}_{0,j}v\left(Y_{j,t-i};\eta_{0,j}\right), \; t = 1, \cdots, n.$$

Define $X_{j,\gamma_j,t} = \sum_{i=1}^{\infty} \beta_j^{i-1} v\left(Y_{j,t-i}; \eta_j\right)$, where $\gamma_j = (\beta_j, \eta_j)^\intercal$ is the finite-dimensional parameter for the semi-nonparametric GARCH model. $\gamma_{0,j} = (\beta_{0,j}, \eta_{0,j})^\intercal \in \Gamma_j \equiv [\beta_{j,1}, \beta_{j,2}] \times [\eta_{j,1}, \eta_{j,2}]$ is the unknown true value for $\gamma_j$, and $X_{j,t} = X_{j,\gamma_{0,j},t}$ is the true conditioning variable. The form of $v(\cdot,\cdot)$ is known. The functional forms of $m_j(\cdot)$, for $j = 1, \cdots, k$ are left unspecified[12], and can be estimated using kernel estimation or sieve estimation.

Because each $X_{j,\gamma_j,t}$ takes value on $(0,\infty)$, a common transformation has been done to $X_{j,\gamma_j,t}$, so that $B$ spline regression can be applied to the transformed variables. For $\forall j$, $\forall t$, $\gamma_j \in \Gamma_j$, define the transformed variable: $U_{j,\gamma_j,t} = G\left(X_{j,\gamma_j,t}\right) = \frac{G_{\gamma_{j,1}}\left(X_{j,\gamma_j,t}\right) + G_{\gamma_{j,2}}\left(X_{j,\gamma_j,t}\right)}{2}$, where $\gamma_{j,1} = (\beta_{j,1}, \eta_{j,1})^\intercal$, $\gamma_{j,2} = (\beta_{j,2}, \eta_{j,2})^\intercal$, $G_{\gamma_{j,1}}$ and $G_{\gamma_{j,2}}$ are CDFs of $X_{j,\gamma_{j,1},t}$ and $X_{j,\gamma_{j,2},t}$ respectively. It is noted that $X_{j,\gamma_j,t}, U_{j,\gamma_j,t} \in \mathcal{F}^{t-1}$. In particular, for the unknown true value $\gamma_{0,j}$, the transformed variable is $U_{j,t} = U_{j,\gamma_{0,j},t} = G\left(X_{j,\gamma_{0,j},t}\right) = G\left(X_{j,t}\right) = \frac{G_{\gamma_{j,1}}(X_{j,t}) + G_{\gamma_{j,2}}(X_{j,t})}{2}$, $\forall j$ and $\forall t$. Denote

$$h_{j,\gamma_j}(u) =: E\left(Y_{j,t}^2 \mid U_{j,\gamma_j,t} = u\right), \quad h_{j,\gamma_j}(U_{j,\gamma_j,t}) =: E\left(Y_{j,t}^2 \mid U_{j,\gamma_j,t}\right),$$

$$h_j\left(U_{j,t}\right) =: h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t}) = E\left(Y_{j,t}^2 \mid U_{j,\gamma_{0,j},t}\right) = m_j\left(X_{j,t}\right). \tag{B.5}$$

We assume the unknown function $h_{j,\gamma_j} \in \mathcal{H}_j$, $\forall \gamma_j \in \Gamma_j$, for $j = 1, \cdots, k$. $\mathcal{H}_j$ $(j = 1, \cdots, k)$ denotes the space of functions on $[0,1]$ satisfying certain smoothness conditions.

The conditional variance function $m_j(\cdot)$ is regressed using $B$ spline, although other choices of sieves can also be applied. Let $G^{(2)}$ denote the space of cubic spline functions on $[0,1]$ used for estimation. For $j = 1, \cdots, k$, given a realization of the $j$th time series $\{Y_{j,t}\}_{t=1}^n$, define for $\forall \gamma_j \in \Gamma_j$ the cubic spline estimator of $h_{j,\gamma_j}(\cdot)$ : $\widehat{h}_{j,\gamma_j}(\cdot) = \arg min_{h \in G^{(2)}} \frac{1}{n''} \sum_{t=n'+1}^{n} \left\{Y_{j,t}^2 - h\left(U_{j,\gamma_j,t}\right)\right\}^2$. The estimator of $\gamma_j$ is $\widehat{\gamma}_j = argmin_{\gamma_j \in \Gamma_j} \frac{1}{n''} \sum_{t=n'+1}^{n} \left\{Y_{j,t}^2 - \widehat{h}_{j,\gamma_j}(U_{j,\gamma_j,t})\right\}^2$, with $n'' = n - n'$, where the first $n'$ data points are not used in the above estimator for implementation reasons (refer to Liu and Yang (2016) for detailed explanation).

To summarize, the semi-nonparametric GARCH parameter $\kappa = \left(\gamma_1^\intercal, \cdots, \gamma_k^\intercal, \psi_1, \cdots, \psi_k\right)^\intercal$, $\gamma_j \in \Gamma_j$ and $\psi_j \in \mathcal{H}_j$, $j = 1, \cdots, k$. The unknown true value $\kappa_0 = (\gamma_{0,1}^\intercal, \cdots, \gamma_{0,k}^\intercal, h_{1,\gamma_{0,1}}, \cdots, h_{k,\gamma_{0,k}})^\intercal$.

[12]For parametric GARCH models, the functional forms of $m_j(\cdot)$, for $j = 1, \cdots, k$ are known. When calculating the effects of first-stage estimation error, the parts caused by estimating $m_j's$ disappear.

The parameter space $\mathcal{B} = \prod_{j=1}^{k} \Gamma_j \times \prod_{j=1}^{k} \mathcal{H}_j$ and $\mathcal{B}_n = \prod_{j=1}^{k} \Gamma_j \times \prod_{j=1}^{k} \mathcal{H}_j^n$, where $\mathcal{H}_j^n$ is the space of cubic spline functions. Let $\mathbb{W}$ be the linear span of $\mathcal{B} - \{\kappa_0\}$. For each $j$, define $Q(h, Y_{j,t}, U_{j,t}) = \left[Y_{j,t}^2 - h(U_{j,t})\right]^2, \forall h \in \mathcal{H}_j$. Therefore $h_j = h_{j,\gamma_{0,j}} = argmin_{h \in \mathcal{H}_j} E[Q(h, Y_{j,t}, U_{j,t})]$ and $\widehat{h}_{j,\gamma_{0,j}} = argmin_{h \in G^{(2)}} \frac{1}{n''} \sum_{t=n'+1}^{n} Q(h, Y_{j,t}, U_{j,t})$. We define a pseudo-metric on $\mathcal{H}_j$ induced by the criterion function :

$$\frac{1}{2}||h - h_j||_{h_j}^2 = E[Q(h, Y_{j,t}, U_{j,t})] - E[Q(h_j, Y_{j,t}, U_{j,t})] = E[h(U_{j,t}) - h_j(U_{j,t})]^2, \quad \forall h \in \mathcal{H}_j. \tag{B.6}$$

Let $\mathbb{W}_j$ be the Hilbert space generated by $\mathcal{H}_j - \{h_j\}$, equipped with the inner product

$$< w_{h_j}, \widetilde{w}_{h_j} >_{h_j} = 2E\left[w_{h_j}(U_{j,t})\, \widetilde{w}_{h_j}(U_{j,t})\right], \quad \forall w_{h_j}, \widetilde{w}_{h_j} \in \mathbb{W}_j. \tag{B.7}$$

Then define the norm for $w = (w_{\gamma_1}^{\intercal}, \cdots, w_{\gamma_k}^{\intercal}, w_{h_1}, \cdots, w_{h_k})^{\intercal} \in \mathbb{W}$ as

$$||w||^2 = \sum_{j=1}^{k} ||w_{\gamma_j}||_E^2 + \sum_{j=1}^{k} ||w_{h_j}||_{h_j}^2, \quad \text{where } ||\cdot||_E \text{ denotes the corresponding Euclidean norm.}$$

Under Assumption B.2, by choosing a different number of interior knots, $n^{-\frac{5}{12}} \log n \leq \delta_{h_j,n} \leq n^{-\frac{2}{5}} (\log n)^{4/5}$ (see Assumption (A6) and Proposition 2 in Liu and Yang (2016)), and we have

$$\sup_{u \in [0,1]} \left|\widehat{h}_{j,\gamma_{0,j}}(u) - h_j(u)\right| = O_p\left(\delta_{h_j,n}\right) = o_p\left(n^{-1/4}\right). \tag{B.8}$$

For simplicity, we assume $\delta_{h,n} = \delta_{h_j,n}$ for all $j$. Obviously $||h - h_j||_{h_j} \leq \sqrt{2} \sup_{u \in [0,1]} |h(u) - h_j(u)|$, $\forall h \in \mathcal{H}_j$. Thus we have $||\widehat{h}_{j,\gamma_{0,j}} - h_j||_{h_j} \leq O_p(\delta_{h,n}) = o_p\left(n^{-1/4}\right)$, due to Eq.(B.8). Then we have

$$||\widehat{\kappa} - \kappa_0|| = O_p(\delta_{h,n}) = o_p\left(n^{-1/4}\right), \tag{B.9}$$

which can be used to prove Lemma 1 and Lemma B.3. In the current setup, $\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}_j^t - \mu_{0,j}^t}{\sigma_{0,j}^t}\right) = 0$ and $\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}_j^t - \sigma_{0,j}^t}{\sigma_{0,j}^t}\right) = \frac{1}{2}E\left[\frac{\partial \log h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{\partial \gamma_j^{\intercal}}\right]\sqrt{n}\left(\widehat{\gamma}_j - \gamma_{0,j}\right) + \frac{1}{2}\sqrt{n}\mathbb{E}_X\left[\frac{\widehat{h}_{j,\gamma_{0,j}}\left(U_{j,\gamma_{0,j},t}\right) - h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}\right]$.

**Assumption B.2.** *The data generating process of each individual time series satisfies Assumptions (A1) - (A6) in Liu and Yang (2016).*

**Lemma B.2.** *(Theorem 2 of Liu and Yang (2016)) For $j = 1, \cdots, k$, under Assumption B.2, as $n \to \infty$, $\sqrt{n}(\widehat{\gamma}_j - \gamma_{0,j}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} S(U_{j,t})\left(\xi_{j,t}^2 - 1\right) + o_p(1) \xrightarrow{\mathcal{D}} \Upsilon_j \sim N(0, \Sigma(\gamma_{0,j}))$.*

$S(U_{j,t})$ is a random vector with the same dimension as $\gamma_j$. See page 31-32 of Liu and Yang (2016) for more details. Lemma B.2 is used to quantify the estimation errors caused by estimating the parametric GARCH parts.

Now we consider a specific linear functional on $\mathcal{H}_j$ : $\forall h \in \mathcal{H}_j$, $\rho_j(h) = E\left(\frac{h(U_{j,t})}{h_j(U_{j,t})}\right)$. In particular, $\rho_j(h_j) = E\left(\frac{h_j(U_{j,t})}{h_j(U_{j,t})}\right) = 1$. In the following lemma, we verify that $\rho_j\left(\widehat{h}_{j,\gamma_{0,j}}\right)$ satisfies Assumption 5 (i), which is used to establish the main theorems of this paper. We have $\rho_j(h) - \rho_j(h_j) = E\left(\frac{h(U_{j,t})-h_j(U_{j,t})}{h_j(U_{j,t})}\right) =< w^*_{h_j}, h - h_j >_{h_j}$, where $w^*_{h_j}(U_{j,t}) = \frac{1}{2h_j(U_{j,t})}$ is the unique Riesz representer of $\rho_j$, under the norm defined by Eqs.(B.6) and (B.7).

**Assumption B.3.** *(*i*)* $\sup_{\left\{h\in\mathcal{H}^n_j:||h-h_j||_{h_j}\leq O\left(\log(\log n)\delta_{h,n}\right)\right\}} \mu_n\left\{\Pi_n\left(w^*_{h_j}(U_{j,t})\right)\times[h(U_{j,t})-h_j(U_{j,t})]\right\} = o_p\left(\frac{1}{\sqrt{n}}\right)$*; (*ii*)* $\sup_{\left\{h\in\mathcal{H}^n_j:||h-h_j||_{h_j}\leq O\left(\log\log(n)\delta_{h,n}\right)\right\}} E\left\{\left[\Pi_n\left(w^*_{h_j}(U_{j,t})\right)-w^*_{h_j}(U_{j,t})\right]\times[h(U_{j,t})-h_j(U_{j,t})]\right\} = o\left(\frac{1}{\sqrt{n}}\right)$*; (*iii*)* $\mu_n\left\{\left(Y^2_{j,t}-h_j(U_{j,t})\right)\left[\Pi_n\left(w^*_{h_j}(U_{j,t})\right)-w^*_{h_j}(U_{j,t})\right]\right\} = o_p\left(\frac{1}{\sqrt{n}}\right)$*; (*iv*)* $E(h_j(U_{j,t}))^2 < \infty$ *and* $\inf_{u\in[0,1]} h_j(u) > 0$.

**Lemma B.3.** *For $j = 1, \cdots, k$, under Assumptions B.2 and B.3, as $n \to \infty$,*

$$\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{h}_{j,\gamma_{0,j}}(U_{j,t})-h_j(U_{j,t})}{h_j(U_{j,t})}\right) = n^{-1/2}\sum_{t=1}^{n}\left(-\frac{\partial Q(h_j,Y_{j,t},U_{j,t})}{\partial h}w^*_{h_j}\right)+o_p(1)$$
$$= n^{-1/2}\sum_{t=1}^{n}\left(\xi^2_{j,t}-1\right)+o_p(1)\xrightarrow{\mathcal{D}}\mathbb{H}_j\sim N\left(0,Var\left[\xi^2_{j,t}-1\right]\right),$$

*where* $-\frac{\partial Q(h_j,Y_{j,t},U_{j,t})}{\partial h}w^*_{h_j} = \frac{Y^2_{j,t}-h_j(U_{j,t})}{h_j(U_{j,t})} = \xi^2_{j,t}-1$.

Lemma B.3 is used to quantify the estimation errors caused by estimating the nonparametric GARCH parts. Assumptions B.2 and B.3 only impose restrictions on each individual time series $\{Y_{j,t}, X_{j,t}\}^n_{t=1}$. Proof of Lemma B.3 is similar to that of Lemma B.1, thus it is omitted.

# C Proofs of Lemmas and Theorems

Let $\alpha_{0n} \in \arg\min_{\alpha\in\overline{\mathcal{A}}_n}\|\alpha-\alpha_0\|$. Denote $\Pi_n(\cdot)$ as the projection of $\mathbb{V}$ to $\mathcal{A}_n - \{\alpha_{0n}\}$.

## C.1 Proofs for Theoretical Results in Section 3

**Derivation of Eq.**(17)

Under Assumptions 2 (ii) and 3, for any $v = \left(v_\theta^\intercal, v_1, \cdots, v_k\right)^\intercal \in \mathbb{V}$ and $w = (w_{\mu,1}, \cdots, w_{\mu,k}, w_{\sigma,1}, \cdots, w_{\sigma,k})^\intercal \in \mathbb{W}$, we have that $l\left(\alpha_0 + \tau_1 v, \kappa_0 + \tau_2 w, Y_t\right)$ is continuously differentiable in small $\tau_1, \tau_2 \in [0,1]$ :

$$\left.\frac{\partial^2 l\left(\alpha_0 + \tau_1 v, \kappa_0 + \tau_2 w, Y_t\right)}{\partial\tau_1\partial\tau_2}\right|_{\tau_1=0\ \tau_2=0} =$$

$$\sum_{j=1}^{k}\left\{\begin{array}{c}\left(\begin{array}{l}\frac{\partial^2 \log c(U_t^0,\theta_0)}{\partial u_j\partial\theta^\intercal} v_\theta f_{0,j}\left(\xi_{j,t}\right) + \frac{v_j'(\xi_{j,t}) f_{0,j}(\xi_{j,t}) - v_j(\xi_{j,t}) f_{0,j}'(\xi_{j,t})}{f_{0,j}^2(\xi_{j,t})} \\ + \frac{\partial \log c(U_t^0,\theta_0)}{\partial u_j} v_j\left(\xi_{j,t}\right) + \sum_{m=1}^{k} \frac{\partial^2 \log c(U_t^0,\theta_0)}{\partial u_m \partial u_j} \int_{-\infty}^{\xi_{m,t}} v_m(x)dx f_{0,j}\left(\xi_{j,t}\right)\end{array}\right) \\ \times\left[\frac{-w_{\mu,j}^t}{\sigma_{0,j}^t} + \frac{-\xi_{j,t} w_{\sigma,j}^t}{\sigma_{0,j}^t}\right]\end{array}\right\}.$$

To evaluate the asymptotic effect of prefiltering on the joint estimation of semiparametric multivariate copula models, we define a functional $\Gamma\left(\alpha_0, \kappa_0\right)[\cdot,\cdot]$ on $\mathbb{V}\times\mathbb{W}$ as

$$\Gamma\left(\alpha_0, \kappa_0\right)[v,w] \equiv \mathbb{E}_X\left[\left.\frac{\partial^2 l\left(\alpha_0 + \tau_1 v, \kappa_0 + \tau_2 w, Y_t\right)}{\partial\tau_1\partial\tau_2}\right|_{\tau_1=0\ \tau_2=0}\right] = \tag{C.1}$$

$$\sum_{j=1}^{k}\left\{\begin{array}{l} E\left(\begin{array}{l}\frac{\partial^2 \log c(U_t^0,\theta_0)}{\partial u_j\partial\theta^\intercal} v_\theta f_{0,j}\left(\xi_{j,t}\right) + \frac{v_j'(\xi_{j,t}) f_{0,j}(\xi_{j,t}) - v_j(\xi_{j,t}) f_{0,j}'(\xi_{j,t})}{f_{0,j}^2(\xi_{j,t})} \\ + \frac{\partial \log c(U_t^0,\theta_0)}{\partial u_j} v_j\left(\xi_{j,t}\right) + \sum_{m=1}^{k} \frac{\partial^2 \log c(U_t^0,\theta_0)}{\partial u_m \partial u_j} \int_{-\infty}^{\xi_{m,t}} v_m(x)dx f_{0,j}\left(\xi_{j,t}\right)\end{array}\right) E\left(\frac{-w_{\mu,j}^t}{\sigma_{0,j}^t}\right) \\ +E\left(\begin{array}{l}\frac{\partial^2 \log c(U_t^0,\theta_0)}{\partial u_j\partial\theta^\intercal} v_\theta f_{0,j}\left(\xi_{j,t}\right)\xi_{j,t} + \frac{v_j'(\xi_{j,t}) f_{0,j}(\xi_{j,t}) - v_j(\xi_{j,t}) f_{0,j}'(\xi_{j,t})}{f_{0,j}^2(\xi_{j,t})}\xi_{j,t} \\ + \frac{\partial \log c(U_t^0,\theta_0)}{\partial u_j} v_j\left(\xi_{j,t}\right)\xi_{j,t} + \sum_{m=1}^{k} \frac{\partial^2 \log c(U_t^0,\theta_0)}{\partial u_m \partial u_j} \int_{-\infty}^{\xi_{m,t}} v_m(x)dx f_{0,j}\left(\xi_{j,t}\right)\xi_{j,t}\end{array}\right) E\left(\frac{-w_{\sigma,j}^t}{\sigma_{0,j}^t}\right)\end{array}\right\},$$

for $v \in \mathbb{V}$ and $w \in \mathbb{W}$. The second equality holds under the assumption that $\xi_t$ is independent of $\mathcal{F}^{t-1}$. Under Assumption C.2 (iii), $\Gamma\left(\alpha_0, \kappa_0\right)[\cdot,\cdot]$ is a bounded bilinear functional on $\mathbb{V}\times\mathbb{W}$. When $\Gamma\left(\alpha_0, \kappa_0\right)[\cdot,\cdot]$ is evaluated at $[v^*, \widehat{\kappa} - \kappa_0]$, we obtain Eq.(17) and the correction term in Theorem 1.

**Assumption C.1.** *(*i*) $\|\widehat{\alpha}_{sml} - \alpha_0\| = O_p(\delta_{\alpha,n}) = o_p(n^{-1/4})$ ; (*ii*) there exists $\Pi_n v^* \in \mathcal{A}_n - \{\alpha_{0n}\}$ such that $\|\Pi_n v^* - v^*\| = O(n^{-1/4})$; (*iii*) the smooth functional $\rho$ satisfies Assumption 3 in Chen et al. (2006) with the smoothness parameter $\omega \geq 2$.*

Assumption C.1 is a restatement of Assumptions 3 and 4 in Chen et al. (2006). If the smooth functional $\rho$ is linear, then $\omega = \infty$ (e.g. the copula parameter and the marginal CDF function evaluated at a point); for more general smooth functionals, $\omega = 2$.

**Assumption C.2.** *(*i*)* $\forall\kappa \in \mathcal{B}_n$ *with* $||\kappa-\kappa_0|| \leq O\left(\log(\log n)\delta_{h,n}\right)$, $\left\{Y_t, \kappa^t,\ 1 \leq t \leq n\right\}$ *is a strictly stationary* $\beta$*-mixing sequence with* $\beta(t) \leq \beta_0 t^{-\zeta}$ *for some* $\beta_0 > 0$, $\zeta = \gamma - 2 > 2$; *(*ii*) for any small* $\epsilon > 0$, *there exist a constant* $s \in (0,2)$ *and a measurable function* $U_n(\cdot)$ *such that* $\sup_{\left\{\alpha\in\mathcal{A}_n,\kappa\in\mathcal{B}_n:||\alpha-\alpha_0||\leq\epsilon,||\kappa-\kappa_0||\leq O\left(\log(\log n)\delta_{h,n}\right)\right\}} |l(\alpha,\kappa,Y_t) - l(\alpha,\kappa_0,Y_t)| \leq U_n(Y_t)\,||\kappa-\kappa_0||^s$, *with* $(\delta_{h,n})^s = o(n^{-1/4})$, $sup_n E\left[U_n(Y_t)\right]^\gamma \leq M_1$, *for* $M_1 > 0$ *and* $\gamma > 2$; *(*iii*) for all* $v \in \mathbb{V}$, *and* $w \in \mathbb{W}$, *there exists* $0 < M_2 < \infty$, *such that* $\left|E\left(\frac{\partial^2 l(\alpha_0,\kappa_0,Y_t)}{\partial\alpha\partial\kappa^\intercal}[v,w]\right)\right| \leq M_2||v||||w||$.

Assumption C.2(ii) imposes some smoothness condition on $l(\alpha,\kappa,Y_t)$ with respect to $\kappa$ in the decaying neighborhood $||\kappa-\kappa_0|| \leq O\left(\log(\log n)\delta_{h,n}\right)$, where $||\cdot||$ is defined in Eq.(B.3) (see similar assumptions in e.g. Chen and Shen (1998)). Assumption C.2(iii) imposes that $\Gamma(\alpha_0,\kappa_0)[\cdot,\cdot]$ is a bounded bilinear functional.

Define the Kullback-Leibler equivalent metric $K_l(\alpha,\kappa) \equiv \mathbb{E}_X\left[l(\alpha,\kappa,Y_t) - l(\alpha_0,\kappa_0,Y_t)\right]$.

**Assumption C.3.** *(Local behavior of criterion)*

*(*i*) Let* $\eta_n = n^{-\tau}$ *with* $1/8 \leq \tau \leq 1/4$. *Uniformly over* $\alpha \in \mathcal{A}_n$, $\kappa \in \mathcal{B}_n$ *with* $\|\alpha-\alpha_0\| \leq o(\eta_n)$, $||\kappa-\kappa_0|| \leq O\left(\log(\log n)\delta_{h,n}\right)$ *:*

$$K_l(\alpha,\kappa) = -\frac{||\alpha-\alpha_0||^2}{2} + \Gamma(\alpha_0,\kappa_0)[\alpha-\alpha_0,\kappa-\kappa_0] + E\left(\frac{1}{2}\frac{\partial^2 l(\alpha_0,\kappa_0,Y_t)}{\partial\kappa\partial\kappa^\intercal}[\kappa-\kappa_0,\kappa-\kappa_0]\right) + o\left(n^{-1/4}\eta_n\right).$$

*(*ii*) Uniformly over* $\alpha \in \mathcal{A}_n$ *and* $\kappa \in \mathcal{B}_n$ *with* $||\alpha-\alpha_0|| \leq o\left(n^{-/8}\right)$, $||\kappa-\kappa_0|| \leq O\left(\log(\log n)\delta_{h,n}\right)$, *the following stochastic equicontinuity holds :*

$$\mu_n\left\{l(\alpha,\kappa,Y_t) - l(\alpha_0,\kappa,Y_t) - \left[l(\alpha,\kappa_0,Y_t) - l(\alpha_0,\kappa_0,Y_t)\right]\right\} = o_p\left(n^{-1/2}\right).$$

**Assumption C.4.** *(Local behavior of criterion in the direction of the Riesz representer)*

*(*i*) Let* $\epsilon_n = o\left(\frac{1}{\sqrt{n}}\right)$. *Uniformly over* $\alpha \in \mathcal{A}_n$ *and* $\kappa \in \mathcal{B}_n$ *with* $||\alpha-\alpha_0|| \leq O\left(\log(\log n)\widetilde{\delta}_{\alpha,n}\right)$, $||\kappa-\kappa_0|| \leq O\left(\log(\log n)\delta_{h,n}\right)$ *:*

$$K_l(\alpha,\kappa) - K_l(\alpha \pm \epsilon_n\Pi_n(v^*),\kappa) = \frac{||\alpha \pm \epsilon_n\Pi_n(v^*) - \alpha_0||^2}{2} - \frac{||\alpha-\alpha_0||^2}{2} \mp \epsilon_n\Gamma(\alpha_0,\kappa_0)[\Pi_n(v^*),\kappa-\kappa_0] + O\left(\epsilon_n^2\right).$$

(ii) $\sup_{\left\{\alpha\in\mathcal{A},\kappa\in\mathcal{B}:||\alpha-\alpha_0||\leq O\left(\log(\log n)\widetilde{\delta}_{\alpha,n}\right),\ ||\kappa-\kappa_0||\leq O\left(\log(\log n)\delta_{h,n}\right)\right\}} \mu_n\left(\left[\frac{\partial l(\alpha,\kappa,Y_t)}{\partial\alpha^{\intercal}} - \frac{\partial l(\alpha_0,\kappa_0,Y_t)}{\partial\alpha^{\intercal}}\right]\Pi_n\left(v^*\right)\right) = o_p\left(n^{-1/2}\right).$

Assumptions C.3 (i) and C.4 (i) characterize the local quadratic behavior of the Kullback-Leibler equivalent metric. Assumptions C.3 (ii) and C.4 (ii) make the stochastic equicontinuity conditions. Assumptions C.3 and C.4 are very common regularity assumptions in the sieve method literature, see, e.g. Chen and Shen (1998), Chen and Liao (2014) and the references therein. They are easily satisfied if the log-likelihood function is twice continuously differentiable around true $\alpha_0, \kappa_0$.

## Proof of Lemma 1

Let $L_n(\alpha) = \frac{1}{n}\sum_{t=1}^{n} l(\alpha,\kappa_0,Y_t)$ and $\widetilde{L_n}(\alpha) = \frac{1}{n}\sum_{t=1}^{n} l(\alpha,\widehat{\kappa},Y_t)$, where $l(\alpha,\kappa_0,Y_t)$ is defined in Eq.(6) and $l(\alpha,\widehat{\kappa},Y_t)$ is defined in Eq.(4). We will first establish (*L.1.*) and (*L.2.*).

(*L.1.*) $\widetilde{L_n}(\alpha) - L_n(\alpha) = o_p\left(n^{-1/4}\right)$ uniformly over $\alpha\in\mathcal{A}_n$.

$$\left|\widetilde{L_n}(\alpha) - L_n(\alpha)\right| = \frac{1}{n}\left|\sum_{t=1}^{n}\left(l(\alpha,\widehat{\kappa},Y_t) - l(\alpha,\kappa_0,Y_t)\right)\right| \leq \frac{1}{n}\sum_{t=1}^{n} U_n(Y_t)\,||\widehat{\kappa}-\kappa_0||^s = o_p\left(n^{-1/4}\right),$$

due to Assumption C.2 (ii).

(*L.2.*) We will show that $\widetilde{L_n}(\alpha) - \widetilde{L_n}(\alpha_0) - [L_n(\alpha) - L_n(\alpha_0)] = o_p\left(n^{-1/4}\eta_n\right)$ uniformly over $\alpha\in\mathcal{A}_n$ with $\|\alpha-\alpha_0\|\leq o(\eta_n)$, where $\eta_n = n^{-\tau}$ with $1/8\leq\tau\leq 1/4$.

Let $\alpha\in\mathcal{A}_n$ with $\|\alpha-\alpha_0\|\leq o(\eta_n)$ and $\kappa\in\mathcal{B}_n$ with $||\kappa-\kappa_0||\leq O(\delta_{h,n})$ :

$$\widetilde{L_n}(\alpha) - \widetilde{L_n}(\alpha_0) - [L_n(\alpha) - L_n(\alpha_0)]$$

$$= \mu_n\{l(\alpha,\widehat{\kappa},Y_t) - l(\alpha_0,\widehat{\kappa},Y_t)\} - \mu_n\{l(\alpha,\kappa_0,Y_t) - l(\alpha_0,\kappa_0,Y_t)\}$$

$$+\mathbb{E}_X[l(\alpha,\widehat{\kappa},Y_t)] - E[l(\alpha_0,\kappa_0,Y_t)] + E[l(\alpha_0,\kappa_0,Y_t)] - \mathbb{E}_X[l(\alpha_0,\widehat{\kappa},Y_t)] + E[l(\alpha_0,\kappa_0,Y_t)] - E[l(\alpha,\kappa_0,Y_t)]$$

$$= \mathbb{E}_X[l(\alpha,\widehat{\kappa},Y_t)] - E[l(\alpha_0,\kappa_0,Y_t)] - (\mathbb{E}_X[l(\alpha_0,\widehat{\kappa},Y_t)] - E[l(\alpha_0,\kappa_0,Y_t)]) - (E[l(\alpha,\kappa_0,Y_t)] - E[l(\alpha_0,\kappa_0,Y_t)])$$

$+\, o_p\left(n^{-1/2}\right)$, where the last equality holds uniformly due to Assumption C.3(ii).

Under Assumption C.3(i), we have:

$$E\left[l\left(\alpha, \kappa, Y_t\right)\right]-E\left[l\left(\alpha_0, \kappa_0, Y_t\right)\right]=E\left(\frac{1}{2} \frac{\partial^2 l\left(\alpha_0, \kappa_0, Y_t\right)}{\partial \alpha \partial \alpha^{\intercal}}\left[\alpha-\alpha_0, \alpha-\alpha_0\right]\right)$$
$$+E\left(\frac{1}{2} \frac{\partial^2 l\left(\alpha_0, \kappa_0, Y_t\right)}{\partial \kappa \partial \kappa^{\intercal}}\left[\kappa-\kappa_0, \kappa-\kappa_0\right]\right)+E\left(\frac{\partial^2 l\left(\alpha_0, \kappa_0, Y_t\right)}{\partial \alpha \partial \kappa^{\intercal}}\left[\alpha-\alpha_0, \kappa-\kappa_0\right]\right)+o\left(n^{-1/4} \eta_n\right).$$

Therefore, uniformly over $\alpha \in \mathcal{A}_n$ with $\|\alpha-\alpha_0\| \leq o\left(\eta_n\right)$ and $\kappa \in \mathcal{B}_n$ with $||\kappa-\kappa_0|| \leq O\left(\delta_{h,n}\right)$ :

$$\mathbb{E}_X\left[l\left(\alpha, \widehat{\kappa}, Y_t\right)\right]-E\left[l\left(\alpha_0, \kappa_0, Y_t\right)\right]-\left(\mathbb{E}_X\left[l\left(\alpha_0, \widehat{\kappa}, Y_t\right)\right]-E\left[l\left(\alpha_0, \kappa_0, Y_t\right)\right]\right)-\left(E\left[l\left(\alpha, \kappa_0, Y_t\right)\right]-E\left[l\left(\alpha_0, \kappa_0, Y_t\right)\right]\right)$$
$$=\mathbb{E}_X\left(\frac{\partial^2 l\left(\alpha_0, \kappa_0, Y_t\right)}{\partial \alpha \partial \kappa^{\intercal}}\left[\alpha-\alpha_0, \widehat{\kappa}-\kappa_0\right]\right)+o_p\left(n^{-1/4} \eta_n\right)=o_p\left(n^{-1/4} \eta_n\right), \text{ by Assumption C.2 (iii).}$$

Thus we have (*L.2.*). Then we could use almost the same proof of Theorem 3.1 in Ai and Chen (2003) to show $\|\widetilde{\alpha}_{sml}-\alpha_0\|=o_p(n^{-1/4})$ , since we already have (*L.1.*), (*L.2.*), and $\|\widehat{\alpha}_{sml}-\alpha_0\|=O_p(\delta_{\alpha,n})=o_p(n^{-1/4})$ (Assumption C.1(i)). □

## Proof of Lemma 2

For $v, \widetilde{v} \in \mathbb{V}^r$ and $w, \widetilde{w} \in \mathbb{W}$, we denote the second-order directional derivative of $l\left(\alpha_0, \kappa_0, Y_t\right)$ as

$$\frac{\partial^2 l\left(\alpha_0, \kappa_0, Y_t\right)}{\partial\left(\alpha^{\intercal}, \kappa^{\intercal}\right)^{\intercal} \partial\left(\alpha^{\intercal}, \kappa^{\intercal}\right)}\left[(v^{\intercal}, w^{\intercal})^{\intercal},(\widetilde{v}^{\intercal}, \widetilde{w}^{\intercal})^{\intercal}\right] \equiv \left.\left.\frac{d^2 l\left(\alpha_0+\tau_1 v+\tau_2 \widetilde{v}, \kappa_0+\tau_1 w+\tau_2 \widetilde{w}, Y_t\right)}{d \tau_1 d \tau_2}\right|_{\tau_1=0}\right|_{\tau_2=0}. \tag{C.2}$$

The proof is similar to that of Lemma 4.1. in Chen et al. (2009), thus it is omitted. □

## Proof of Theorem 1

Define $r\left[\alpha, \alpha_0, \widehat{\kappa}, Y_t\right]=l\left(\alpha, \widehat{\kappa}, Y_t\right)-l\left(\alpha_0, \widehat{\kappa}, Y_t\right)-\frac{\partial l\left(\alpha_0, \widehat{\kappa}, Y_t\right)}{\partial \alpha^{\intercal}}\left[\alpha-\alpha_0\right]$. Then, by definition of $\widetilde{\alpha}_{sml}$, we have

$$\begin{aligned}
0 \leq & \frac{1}{n} \sum_{t=1}^n\left[l\left(\widetilde{\alpha}_{sml}, \widehat{\kappa}, Y_t\right)-l\left(\widetilde{\alpha}_{sml} \pm \epsilon_n \Pi_n\left(v^*\right), \widehat{\kappa}, Y_t\right)\right] \\
= & \mu_n\left[l\left(\widetilde{\alpha}_{sml}, \widehat{\kappa}, Y_t\right)-l\left(\widetilde{\alpha}_{sml} \pm \epsilon_n \Pi_n\left(v^*\right), \widehat{\kappa}, Y_t\right)\right]+\mathbb{E}_X\left[l\left(\widetilde{\alpha}_{sml}, \widehat{\kappa}, Y_t\right)-l\left(\widetilde{\alpha}_{sml} \pm \epsilon_n \Pi_n\left(v^*\right), \widehat{\kappa}, Y_t\right)\right] \\
= & \mp \epsilon_n \mu_n\left[\frac{\partial l\left(\alpha_0, \widehat{\kappa}, Y_t\right)}{\partial \alpha^{\intercal}} \Pi_n\left(v^*\right)\right]+\mu_n\left[r\left(\widetilde{\alpha}_{sml}, \alpha_0, \widehat{\kappa}, Y_t\right)-r\left(\widetilde{\alpha}_{sml} \pm \epsilon_n \Pi_n\left(v^*\right), \alpha_0, \widehat{\kappa}, Y_t\right)\right] \\
& +\mathbb{E}_X\left[l\left(\widetilde{\alpha}_{sml}, \widehat{\kappa}, Y_t\right)-l\left(\widetilde{\alpha}_{sml} \pm \epsilon_n \Pi_n\left(v^*\right), \widehat{\kappa}, Y_t\right)\right],
\end{aligned} \tag{C.3}$$

where $\epsilon_n$ is defined in Assumption C.4(i).

**(*S.1.*) We first show**

$\mathbb{E}_X\left[l\left(\widetilde{\alpha}_{sml},\widehat{\kappa},Y_t\right)-l\left(\widetilde{\alpha}_{sml}\pm\epsilon_n\Pi_n\left(v^*\right),\widehat{\kappa},Y_t\right)\right]=\pm\epsilon_n<\widetilde{\alpha}_{sml}-\alpha_0,\ v^*>\mp\epsilon_n\Gamma\left(\alpha_0,\kappa_0\right)\left[v^*,\widehat{\kappa}-\kappa_0\right]+\epsilon_n o_p\left(n^{-1/2}\right)$.

$$
\begin{aligned}
&\mathbb{E}_X\left[l\left(\widetilde{\alpha}_{sml},\widehat{\kappa},Y_t\right)-l\left(\widetilde{\alpha}_{sml}\pm\epsilon_n\Pi_n\left(v^*\right),\widehat{\kappa},Y_t\right)\right]=K_l\left(\widetilde{\alpha}_{sml},\widehat{\kappa}\right)-K_l\left(\widetilde{\alpha}_{sml}\pm\epsilon_n\Pi_n\left(v^*\right),\widehat{\kappa}\right)\\
=&\frac{||\widetilde{\alpha}_{sml}\pm\epsilon_n\Pi_n\left(v^*\right)-\alpha_0||^2}{2}-\frac{||\widetilde{\alpha}_{sml}-\alpha_0||^2}{2}\mp\epsilon_n\Gamma\left(\alpha_0,\kappa_0\right)\left[\Pi_n\left(v^*\right),\widehat{\kappa}-\kappa_0\right]+O_p\left(\epsilon_n^2\right)\\
=&\frac{||\widetilde{\alpha}_{sml}\pm\epsilon_n\Pi_n\left(v^*\right)-\alpha_0||^2}{2}-\frac{||\widetilde{\alpha}_{sml}-\alpha_0||^2}{2}\mp\epsilon_n\Gamma\left(\alpha_0,\kappa_0\right)\left[v^*,\widehat{\kappa}-\kappa_0\right]+\epsilon_n o_p\left(n^{-1/2}\right)\\
=&\pm\epsilon_n<\widetilde{\alpha}_{sml}-\alpha_0,\ v^*>\pm\epsilon_n<\widetilde{\alpha}_{sml}-\alpha_0,\ \Pi_n\left(v^*\right)-v^*>+\frac{1}{2}\epsilon_n^2||\Pi_n\left(v^*\right)||^2\\
&\mp\epsilon_n\Gamma\left(\alpha_0,\kappa_0\right)\left[v^*,\widehat{\kappa}-\kappa_0\right]+\epsilon_n o_p\left(n^{-1/2}\right)\\
=&\pm\epsilon_n<\widetilde{\alpha}_{sml}-\alpha_0,\ v^*>\mp\epsilon_n\Gamma\left(\alpha_0,\kappa_0\right)\left[v^*,\widehat{\kappa}-\kappa_0\right]+\epsilon_n o_p\left(n^{-1/2}\right),
\end{aligned}
$$

where the second equality is due to Assumption C.4(i), the third equality is due to Assumptions C.1(ii),C.2(iii) and 4, the last equality holds because of Assumption C.1 (ii) and Lemma 1.

**(*S.2.*) We next show**

$\mp\epsilon_n\mu_n\left[\frac{\partial l(\alpha_0,\widehat{\kappa},Y_t)}{\partial\alpha^\intercal}\Pi_n\left(v^*\right)\right]=\mp\epsilon_n\mu_n\left(\frac{\partial l(\alpha_0,\kappa_0,Y_t)}{\partial\alpha^\intercal}v^*\right)+\epsilon_n o_p\left(n^{-1/2}\right)$.

$$
\begin{aligned}
&\mu_n\left[\frac{\partial l\left(\alpha_0,\widehat{\kappa},Y_t\right)}{\partial\alpha^\intercal}\Pi_n\left(v^*\right)\right]=\mu_n\left(\frac{\partial l\left(\alpha_0,\kappa_0,Y_t\right)}{\partial\alpha^\intercal}v^*\right)+\mu_n\left(\frac{\partial l\left(\alpha_0,\kappa_0,Y_t\right)}{\partial\alpha^\intercal}\left(\Pi_n\left(v^*\right)-v^*\right)\right)\\
&+\mu_n\left(\frac{\partial l\left(\alpha_0,\widehat{\kappa},Y_t\right)}{\partial\alpha^\intercal}\Pi_n\left(v^*\right)-\frac{\partial l\left(\alpha_0,\kappa_0,Y_t\right)}{\partial\alpha^\intercal}\Pi_n\left(v^*\right)\right)=\mu_n\left(\frac{\partial l\left(\alpha_0,\kappa_0,Y_t\right)}{\partial\alpha^\intercal}v^*\right)+o_p\left(n^{-1/2}\right).
\end{aligned}
$$

The last equality is implied by Chebyshev inequality, i.i.d. data, $||\Pi_n\left(v^*\right)-v^*||=o(1)$ and Assumption C.4 (ii).

**(*S.3.*) We then show**

$\mu_n\left[r\left(\widetilde{\alpha}_{sml},\alpha_0,\widehat{\kappa},Y_t\right)-r\left(\widetilde{\alpha}_{sml}\pm\epsilon_n\Pi_n\left(v^*\right),\alpha_0,\widehat{\kappa},Y_t\right)\right]=\epsilon_n o_p\left(n^{-1/2}\right).$

$$\begin{aligned}
&\mu_n\left[r\left(\widetilde{\alpha}_{sml},\alpha_0,\widehat{\kappa},Y_t\right)-r\left(\widetilde{\alpha}_{sml}\pm\epsilon_n\Pi_n\left(v^*\right),\alpha_0,\widehat{\kappa},Y_t\right)\right]\\
&=\mu_n\left[l\left(\widetilde{\alpha}_{sml},\widehat{\kappa},Y_t\right)-l\left(\widetilde{\alpha}_{sml}\pm\epsilon_n\Pi_n\left(v^*\right),\widehat{\kappa},Y_t\right)\right]\pm\epsilon_n\mu_n\left[\frac{\partial l\left(\alpha_0,\widehat{\kappa},Y_t\right)}{\partial\alpha^\intercal}\Pi_n\left(v^*\right)\right]\\
&=\mp\epsilon_n\mu_n\left[\left(\frac{\partial l\left(\overline{\alpha},\widehat{\kappa},Y_t\right)}{\partial\alpha^\intercal}-\frac{\partial l\left(\alpha_0,\widehat{\kappa},Y_t\right)}{\partial\alpha^\intercal}\right)\Pi_n\left(v^*\right)\right]=\epsilon_n o_p\left(n^{-1/2}\right),
\end{aligned}$$

where $\overline{\alpha}\in\mathcal{A}$ is between $\widetilde{\alpha}_{sml}$ and $\widetilde{\alpha}_{sml}\pm\epsilon_n\Pi_n\left(v^*\right)$. The last equality is implied by Assumption C.4(ii).

Under Eq.(C.3), (*S.1.*)–(*S.3.*), we have

$$\begin{aligned}
&0\leq\mp\epsilon_n\mu_n\left(\frac{\partial l\left(\alpha_0,\kappa_0,Y_t\right)}{\partial\alpha^\intercal}v^*\right)\mp\epsilon_n\Gamma\left(\alpha_0,\kappa_0\right)\left[v^*,\widehat{\kappa}-\kappa_0\right]\pm\epsilon_n<\widetilde{\alpha}_{sml}-\alpha_0,\ v^*>+\epsilon_n o_p\left(n^{-1/2}\right)\\
&\sqrt{n}<\widetilde{\alpha}_{sml}-\alpha_0,\ v^*>=\sqrt{n}\mu_n\left(\frac{\partial l\left(\alpha_0,\kappa_0,Y_t\right)}{\partial\alpha^\intercal}v^*\right)+\sqrt{n}\Gamma\left(\alpha_0,\kappa_0\right)\left[v^*,\widehat{\kappa}-\kappa_0\right]+o_p(1).
\end{aligned}$$

Together with Assumptions 5, C.1 and Lemma 1, we get

$$\begin{aligned}
&\sqrt{n}\left[\rho\left(\widetilde{\alpha}_{sml}\right)-\rho\left(\alpha_0\right)\right]=\sqrt{n}<\widetilde{\alpha}_{sml}-\alpha_0,\ v^*>+o_p(1)\\
&=\sqrt{n}\mu_n\left(\frac{\partial l\left(\alpha_0,\kappa_0,Y_t\right)}{\partial\alpha^\intercal}v^*\right)+\sqrt{n}\Gamma\left(\alpha_0,\kappa_0\right)\left[v^*,\widehat{\kappa}-\kappa_0\right]+o_p(1)\\
&\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\ \lim_{n\to\infty}Var\left[\sqrt{n}\left(\frac{\sum_{t=1}^n S_{\rho(\alpha)}\left(\alpha_0,\kappa_0,Y_t\right)}{n}+\Gamma\left(\alpha_0,\kappa_0\right)\left[v^*,\widehat{\kappa}-\kappa_0\right]\right)\right]\right).\quad\square
\end{aligned}$$

## Proof of Theorem 2

We could rewrite the efficient score defined in Eq.(23) as

$$S_{\theta_0}\left(U_t^0,\alpha_0,\kappa_0\right)=\frac{\partial\log c\left(U_t^0,\theta_0\right)}{\partial\theta}-\sum_{j=1}^k\left[\frac{\partial\log c\left(U_t^0,\theta_0\right)}{\partial u_j}\int_{-\infty}^{\xi_{j,t}}g_{j,\cdot}^{*\intercal}(x)dx+\frac{g_{j,\cdot}^{*\intercal}\left(\xi_{j,t}\right)}{f_{0,j}(\xi_{j,t})}\right],\qquad\text{(C.4)}$$

where $g_{j,\cdot}^*(x)=b_{j,\cdot}^*\left(F_{0,j}(x)\right)f_{0,j}(x)$, for all $x\in\Xi_j$ and $j=1,\cdots,k$. $b_{j,\cdot}^*$'s are defined in Eqs.(21), (22) and (23).

In this proof, we show that Eq.(24) is valid, i.e. $\sqrt{n}\mathbb{E}_X\left[\frac{\partial S_{\theta_0}\left(U_t^0,\alpha_0,\kappa_0+\tau(\widehat{\kappa}-\kappa_0)\right)}{\partial\tau}\right]_{\tau=0}=0$, which is a $d_\theta$ by 1 random vector. To avoid tedious expressions, we assume $\mu_{0,j}$ is known to be zero, for

$j=1,\cdots,k$. The extension to $\mu_{0,j}\neq 0$ can be trivially replicated following the current proof. Then the $l$th component of Eq.(24) is $\sqrt{n}\Gamma\left(\alpha_0,\kappa_0\right)\left[\overset{*}{v}_l,\widehat{\kappa}-\kappa_0\right]$, with $\overset{*}{v}_l=\left[e_l,-g^*_{1,l},\cdots,-g^*_{k,l}\right]^\intercal$ and $e_l$ is the $l$th row of the $d_\theta$ by $d_\theta$ identity matrix, i.e. :

$$\sum_{j=1}^{k}\left\{\begin{array}{c}E\left(\begin{array}{l}\frac{\partial^2\log c(U^0_t,\theta_0)}{\partial u_j\partial\theta_l}f_{0,j}\left(\xi_{j,t}\right)\left(-\xi_{j,t}\right)+\frac{g^{*\prime}_{j,l}(\xi_{j,t})f_{0,j}(\xi_{j,t})-g^*_{j,l}(\xi_{j,t})f'_{0,j}(\xi_{j,t})}{f^2_{0,j}(\xi_{j,t})}\xi_{j,t}\\+\frac{\partial\log c(U^0_t,\theta_0)}{\partial u_j}g^*_{j,l}\left(\xi_{j,t}\right)\xi_{j,t}+\sum_{m=1}^{k}\frac{\partial^2\log c(U^0_t,\theta_0)}{\partial u_m\partial u_j}\int_{-\infty}^{\xi_{m,t}}g^*_{m,l}(x)dxf_{0,j}\left(\xi_{j,t}\right)\xi_{j,t}\end{array}\right)\\ \times\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}^t_j-\sigma^t_{0,j}}{\sigma^t_{0,j}}\right)\end{array}\right\},$$

for $l=1,\cdots,d_\theta$, according to Eq.(17). We show that for $l=1,\cdots,d_\theta$ and $j=1,\cdots,k$,

$$E\left(\begin{array}{l}\frac{\partial^2\log c(U^0_t,\theta_0)}{\partial u_j\partial\theta_l}f_{0,j}\left(\xi_{j,t}\right)\left(-\xi_{j,t}\right)+\frac{g^{*\prime}_{j,l}(\xi_{j,t})f_{0,j}(\xi_{j,t})-g^*_{j,l}(\xi_{j,t})f'_{0,j}(\xi_{j,t})}{f^2_{0,j}(\xi_{j,t})}\xi_{j,t}\\+\frac{\partial\log c(U^0_t,\theta_0)}{\partial u_j}g^*_{j,l}\left(\xi_{j,t}\right)\xi_{j,t}+\sum_{m=1}^{k}\frac{\partial^2\log c(U^0_t,\theta_0)}{\partial u_m\partial u_j}\int_{-\infty}^{\xi_{m,t}}g^*_{m,l}(x)dxf_{0,j}\left(\xi_{j,t}\right)\xi_{j,t}\end{array}\right)=0, \tag{C.5}$$

which is sufficient for Eq.(24). In the proof, for simplicity, we assume $d_\theta=1$. The extension to $d_\theta\geq 2$ is straightforward. Denote $g^*_{j,1}$ as $g^*_j$ for easier presentation, $j=1,\cdots,k$.

$\{\xi_t=(\xi_{1,t},\cdots,\xi_{k,t})^\intercal\}_{t=1}^n$ is a random sample satisfying Assumption 1. The true unconditional variance of $\xi_{j,t}$, $\sigma^2_{0j}=1$ for $j=1,\cdots,k$. We construct a few tool models for $\{\xi_t=(\xi_{1,t},\cdots,\xi_{k,t})^\intercal\}_{t=1}^n$, in all of which the copula function is known apart from a finite dimensional parameter $\theta_0$. They are different in terms of the knowledge of marginal distributions.

1. The marginal distributions of $\xi_t$ are completely unknown. The likelihood function is

$$l\left(\alpha,\xi_t\right)=\log c\left[F_1\left(\xi_{1,t}\right),\cdots,F_k\left(\xi_{k,t}\right);\theta\right]+\sum_{j=1}^{k}logf_j\left(\xi_{j,t}\right).$$

2. For $j=1,\cdots,k$, the marginal distribution $F_{j,\sigma_j,\tau}(s)=F_{0,j}\left(\frac{s}{\sigma_j}\right)-\tau\int_{-\infty}^{\frac{s}{\sigma_j}}g^*_j(x)dx$ is known except for the variance $\sigma^2_j$ and the parameter $\tau$. The corresponding density is $f_{j,\sigma_j,\tau}(s)=f_{0,j}\left(\frac{s}{\sigma_j}\right)\frac{1}{\sigma_j}-\tau g^*_j(\frac{s}{\sigma_j})\frac{1}{\sigma_j}$. The likelihood function is

$$l\left(\theta,\sigma,\tau,\xi_t\right)=\log c\left[F_{1,\sigma_1,\tau}\left(\xi_{1,t}\right),\cdots,F_{k,\sigma_k,\tau}\left(\xi_{k,t}\right);\theta\right]+\sum_{j=1}^{k}logf_{j,\sigma_j,\tau}\left(\xi_{j,t}\right). \tag{C.6}$$

The true value of $\sigma$ is $\sigma_0=(1,\cdots,1)$, while the true value of $\tau$ is $\tau_0=0$.

3. The least favorable parametric sub-model: for $j = 1, \cdots, k$, the marginal distribution $H_{j,\tau}(s) = F_{0,j}(s) - \tau \int_{-\infty}^{s} g_j^*(x)dx$ is known except the parameter $\tau$. The likelihood function is (C.6) evaluated at $\sigma = \sigma_0 = (1, \cdots, 1)$, i.e. $l(\theta, \sigma_0, \tau, \xi_t)$.

Therefore, in terms of the efficiency bound for $\theta_0$, model 3 $\geq$ model 2 $\geq$ model 1. As shown in Chen et al. (2006), model 3 is the least favorable parametric submodel for model 1, thus they are all equal in terms of efficiency bound for $\theta_0$.

Because model 3 is the least favorable parametric submodel, by construction, it is easy to show

$$S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right) = \frac{\partial l\left(\theta_0, \sigma_0, \tau_0, \xi_t\right)}{\partial \theta} + \frac{\partial l\left(\theta_0, \sigma_0, \tau_0, \xi_t\right)}{\partial \tau}, \tag{C.7}$$

where $S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right)$ is defined in Eq.(C.4). The efficient influence function of $\theta_0$ in model 1 is, $\left\{E\left[S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right)\right]^2\right\}^{-1} S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right)$. Due to the definition of the efficient score, we must have

$$E\left[S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right) \frac{\partial l\left(\theta_0, \sigma_0, \tau_0, \xi_t\right)}{\partial \sigma}\right] = 0.$$

Plugging Eq.(C.7) into this equation, we get

$$E\left[\left(\frac{\partial l\left(\theta_0, \sigma_0, \tau_0, \xi_t\right)}{\partial \theta} + \frac{\partial l\left(\theta_0, \sigma_0, \tau_0, \xi_t\right)}{\partial \tau}\right) \frac{\partial l\left(\theta_0, \sigma_0, \tau_0, \xi_t\right)}{\partial \sigma}\right] = 0.$$

Then because of the information identity, we can further get

$$E\left[\frac{\partial^2 l\left(\theta_0, \sigma_0, \tau_0, \xi_t\right)}{\partial \theta \partial \sigma} + \frac{\partial^2 l\left(\theta_0, \sigma_0, \tau_0, \xi_t\right)}{\partial \tau \partial \sigma}\right] = 0. \tag{C.8}$$

A straightforward calculation shows, for $j = 1, \cdots, k$

$$\begin{aligned}
&\frac{\partial^2 l\left(\theta_0, \sigma_0, \tau_0, \xi_t\right)}{\partial \theta \partial \sigma_j} + \frac{\partial^2 l\left(\theta_0, \sigma_0, \tau_0, \xi_t\right)}{\partial \tau \partial \sigma_j} \\
&= \begin{array}{l} \frac{\partial^2 \log c(U_t^0, \theta_0)}{\partial u_j \partial \theta} f_{0,j}\left(\xi_{j,t}\right)\left(-\xi_{j,t}\right) + \frac{g_j^{*\prime}(\xi_{j,t}) f_{0,j}(\xi_{j,t}) - g_j^*(\xi_{j,t}) f_{0,j}'(\xi_{j,t})}{f_{0,j}^2(\xi_{j,t})} \xi_{j,t} \\ + \frac{\partial \log c(U_t^0, \theta_0)}{\partial u_j} g_j^*\left(\xi_{j,t}\right) \xi_{j,t} + \sum_{m=1}^{k} \frac{\partial^2 \log c(U_t^0, \theta_0)}{\partial u_m \partial u_j} \int_{-\infty}^{\xi_{m,t}} g_m^*(x) dx f_{0,j}\left(\xi_{j,t}\right) \xi_{j,t}. \end{array}
\end{aligned} \tag{C.9}$$

Eqs.(C.8) and (C.9) together prove Eq.(C.5). Consequently, Eq.(24) is valid. Therefore, we have

$$Asyvar\left[\lambda^{\mathsf{T}} \widetilde{\theta}_{sml}\right] = v_\theta^{*\mathsf{T}} \lim_{n \to \infty} Var\left[n^{-1/2} \sum_{t=1}^{n} S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right)\right] v_\theta^* = Asyvar\left[\lambda^{\mathsf{T}} \widehat{\theta}_{sml}\right]. \quad \square$$

## Proof of Theorem 3

$(i)$ The semiparametric Fisher information matrix for $\theta_0$ in the full model (Eq.(3)) equals the variance of the projection residual of the $\theta$-part score $\frac{\partial \log c(U_t^0,\theta_0)}{\partial\theta^\intercal}$ onto the tangent space generated by any other parameters– the marginals and the dynamic parameters. We essentially need to solve the following infinite-dimensional optimization problems for $l = 1, \cdots, d_\theta$ :

$$\inf_{v_{\cdot,l}\in\prod_{j=1}^k \overline{\mathbb{V}_j^r},\ w_{\cdot,l}\in\overline{\mathbb{W}}_F} E\left(\frac{\partial \log c(U_t^0,\theta_0)}{\partial\theta_l} - \sum_{j=1}^k\left\{\frac{\partial \log c(U_t^0,\theta_0)}{\partial u_j}\int_{-\infty}^{\xi_{j,t}} v_{j,l}(x)dx + \frac{v_{j,l}(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right\}\right. \tag{C.10}$$
$$\left. + \sum_{j=1}^k \frac{w_{\sigma,jl}^t}{\sigma_{0,j}^t} + \sum_{j=1}^k\left[\frac{\partial \log c(U_t^0,\theta_0)}{\partial u_j} f_{0,j}(\xi_{j,t}) + \frac{f_{0,j}'(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right]\left[\frac{w_{\mu,jl}^t}{\sigma_{0,j}^t} + \frac{\xi_{j,t}w_{\sigma,jl}^t}{\sigma_{0,j}^t}\right]\right)^2,$$

where $v_{\cdot,l} = (v_{1,l},\cdots,v_{k,l})^\intercal$ and $w_{\cdot,l} = (w_{\mu,1l},\cdots,w_{\mu,kl},w_{\sigma,1l},\cdots,w_{\sigma,kl})^\intercal$. Because $\xi_t$ and $U_t^0$ are independent of $\mathcal{F}^{t-1}$, if $w_{\cdot,l}$ solves the above optimization problem, then $\frac{w_{\mu,jl}^t}{\sigma_{0,j}^t} \equiv c_{\mu,jl}$ and $\frac{w_{\sigma,jl}^t}{\sigma_{0,j}^t} \equiv c_{\sigma,jl}$, where $c_{\mu,jl}$ and $c_{\sigma,jl}$ are constants. The optimization problems (C.10) can be simplified to, for $l = 1,\cdots,d_\theta$ :

$$\inf_{v_{\cdot,l}\in\prod_{j=1}^k \overline{\mathbb{V}_j^r},\ c_{\cdot,l}\in\mathcal{R}^{2k}} E\left(\frac{\partial \log c(U_t^0,\theta_0)}{\partial\theta_l} - \sum_{j=1}^k\left\{\frac{\partial \log c(U_t^0,\theta_0)}{\partial u_j}\int_{-\infty}^{\xi_{j,t}} v_{j,l}(x)dx + \frac{v_{j,l}(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right\}\right. \tag{C.11}$$
$$\left. + \sum_{j=1}^k c_{\sigma,jl} + \sum_{j=1}^k\left[\frac{\partial \log c(U_t^0,\theta_0)}{\partial u_j} f_{0,j}(\xi_{j,t}) + \frac{f_{0,j}'(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right]\left[c_{\mu,jl} + \xi_{j,t}c_{\sigma,jl}\right]\right)^2,$$

where $c_{\cdot,l} = (c_{\mu,1l},\cdots,c_{\mu,kl},c_{\sigma,1l},\cdots,c_{\sigma,kl})^\intercal$.

For $l = 1,\cdots,d_\theta$ and $j = 1,\cdots,k$, any $v_{j,l}\in\overline{\mathbb{V}_j^r}$ and $(c_{\mu,jl},c_{\sigma,jl})\in\mathcal{R}^2$, there exists a unique function

$$b_{j,l}(u_j) = \frac{v_{j,l}\left[F_{0,j}^{-1}(u_j)\right]}{f_{0,j}\left[F_{0,j}^{-1}(u_j)\right]} - \frac{f_{0,j}'\left[F_{0,j}^{-1}(u_j)\right]}{f_{0,j}\left[F_{0,j}^{-1}(u_j)\right]}c_{\mu,jl} - \frac{f_{0,j}'(F_{0,j}^{-1}(u_j))F_{0,j}^{-1}(u_j) + f_{0,j}(F_{0,j}^{-1}(u_j))}{f_{0,j}(F_{0,j}^{-1}(u_j))}c_{\sigma,jl}. \tag{C.12}$$

Under Assumption 7 and Eq.(26), we have

$$E\left[b_{j,l}\left(U_{j,t}^0\right)\right] = E\left(\frac{-f_{0,j}'(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right)c_{\mu,jl} + E\left(\frac{-f_{0,j}'(\xi_{j,t})\xi_{j,t} - f_{0,j}(\xi_{j,t})}{f_{0,j}(\xi_{j,t})}\right)c_{\sigma,jl} = 0,\ E\left(\left[b_{j,l}\left(U_{j,t}^0\right)\right]^2\right) < \infty.$$

Thus $b_{j,l}\in\mathcal{L}_2^0([0,1])$. On the other hand, for $l = 1,\cdots,d_\theta$ and $j = 1,\cdots,k$, any $b_{j,l}\in\mathcal{L}_2^0([0,1])$, Eq.(C.12) uniquely identifies $(c_{\mu,jl}, c_{\sigma,jl}) \in \mathcal{R}^2$ and $v_{j,l} \in \overline{\mathbb{V}_j^r}$ :

$$E\left[\xi_{j,t} b_{j,l}\left(F_{0,j}\left(\xi_{j,t}\right)\right)\right] = c_{\mu,jl} - E\left(\frac{f'_{0,j}(\xi_{j,t})\xi_{j,t}^2}{f_{0,j}(\xi_{j,t})}\right) c_{\sigma,jl} = c_{\mu,jl},$$

$$E\left[\xi_{j,t}^2 b_{j,l}\left(F_{0,j}\left(\xi_{j,t}\right)\right)\right] = -E\left(\frac{f'_{0,j}(\xi_{j,t})\xi_{j,t}^2}{f_{0,j}(\xi_{j,t})}\right) c_{\mu,jl} - E\left(\frac{f'_{0,j}(\xi_{j,t})\xi_{j,t}^3 + f_{0,j}(\xi_{j,t})\xi_{j,t}^2}{f_{0,j}(\xi_{j,t})}\right) c_{\sigma,jl} = 2c_{\sigma,jl},$$

$$v_{j,l}(x) = b_{j,l}\left(F_{0,j}\left(x\right)\right) f_{0,j}(x) + c_{\mu,jl} f'_{0,j}(x) + c_{\sigma,jl}\left(f'_{0,j}(x)x + f_{0,j}(x)\right), \text{ for all } x \in \Xi_j, \tag{C.13}$$

which are obtained using Assumptions 1 (i) and 7, Eq.(26) and the Cauchy–Schwarz inequality. Therefore, Eq.(C.12) defines a one-to-one and onto mapping of $\overline{\mathbb{V}_j^r} \times \mathcal{R}^2$ to $\mathcal{L}_2^0([0,1])$ for $j = 1, \cdots, k$. Thus by change of variables, the optimization problems (C.11) can be rewritten as, for $l = 1, \cdots, d_\theta$:

$$\inf_{\left(b_{1,l},\cdots,b_{k,l}\right) \in \left\{\mathcal{L}_2^0([0,1])\right\}^k} E\left(\frac{\partial \log c(U_t^0, \theta_0)}{\partial \theta_l} - \sum_{j=1}^k \left\{\frac{\partial \log c(U_t^0, \theta_0)}{\partial u_j} \int_0^{U_{j,t}^0} b_{j,l}(u)du + b_{j,l}\left(U_{j,t}^0\right)\right\}\right)^2,$$

which coincides with the optimization problems (22) and can be solved by $B^*$ (defined in Eq.(21)). Thus the semiparametric Fisher information matrix for $\theta_0$ in the full model equals $\mathcal{I}_*(\theta_0) = E\left(S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right) S_{\theta_0}\left(U_t^0, \alpha_0, \kappa_0\right)^\intercal\right)$.

$(ii)$ It is a direct conclusion of part $(i)$ and Theorem 2. □

## C.2 Proofs for Theoretical Results in Section 4

### Proof of Theorem 4

$$\begin{aligned}
&\sqrt{n}\left(\widetilde{F}_{j,2s}(x) - F_{0,j}(x)\right) = \sqrt{n}\left(\widetilde{F}_{j,2s}(x) - \widehat{F}_{j,2s}(x)\right) + \sqrt{n}\left(\widehat{F}_{j,2s}(x) - F_{0,j}(x)\right) \\
=&\sqrt{n}\left\{\frac{1}{n+1}\sum_{t=1}^n \left(\mathbb{I}\left\{\widehat{\xi}_{j,t} \le x\right\} - \mathbb{I}\left\{\xi_{j,t} \le x\right\}\right)\right\} + \sqrt{n}\left\{\frac{1}{n+1}\sum_{t=1}^n \left(\mathbb{I}\left\{\xi_{j,t} \le x\right\} - F_{0,j}(x)\right)\right\} + O_p\left(n^{-1/2}\right) \\
=&\sqrt{n}\mu_n\left(\mathbb{I}\left\{\widehat{\xi}_{j,t} \le x\right\} - \mathbb{I}\left\{\xi_{j,t} \le x\right\}\right) + \sqrt{n}\mathbb{E}_X\left(\mathbb{I}\left\{\widehat{\xi}_{j,t} \le x\right\} - \mathbb{I}\left\{\xi_{j,t} \le x\right\}\right) \\
&+ n^{-1/2}\sum_{t=1}^n \left(\mathbb{I}\left\{\xi_{j,t} \le x\right\} - F_{0,j}(x)\right) + O_p\left(n^{-1/2}\right).
\end{aligned}$$

First, $\mu_n\left(\mathbb{I}\left\{\widehat{\xi}_{j,t} \le x\right\} - \mathbb{I}\left\{\xi_{j,t} \le x\right\}\right) = \mu_n\left(\mathbb{I}\left\{\xi_{j,t} \le \frac{\widehat{\mu}_j^t - \mu_{0,j}^t}{\sigma_{0,j}^t} + \frac{\widehat{\sigma}_j^t}{\sigma_{0,j}^t}x\right\} - \mathbb{I}\left\{\xi_{j,t} \le x\right\}\right) = o_p\left(n^{-1/2}\right)$ can be shown similar to Example 1 of Chen et al. (2003) and is also a much weaker version of

Eq.(C.20). Second,

$$
\begin{aligned}
&\sqrt{n}\mathbb{E}_X\left(\mathbb{I}\left\{\widehat{\xi}_{j,t}\le x\right\}-\mathbb{I}\left\{\xi_{j,t}\le x\right\}\right)=\sqrt{n}\mathbb{E}_X\left(\mathbb{I}\left\{\xi_{j,t}\le\frac{\widehat{\mu}_j^t-\mu_{0,j}^t}{\sigma_{0,j}^t}+\frac{\widehat{\sigma}_j^t}{\sigma_{0,j}^t}x\right\}-\mathbb{I}\left\{\xi_{j,t}\le x\right\}\right)\\
&=\sqrt{n}\mathbb{E}_X\left[F_{0,j}\left(\frac{\widehat{\mu}_j^t-\mu_{0,j}^t}{\sigma_{0,j}^t}+\frac{\widehat{\sigma}_j^t}{\sigma_{0,j}^t}x\right)-F_{0,j}\left(x\right)\right]\\
&=\sqrt{n}\mathbb{E}_X\left[f_{0,j}\left(x\right)\left(\frac{\widehat{\mu}_j^t-\mu_{0,j}^t}{\sigma_{0,j}^t}+\frac{\widehat{\sigma}_j^t-\sigma_{0,j}^t}{\sigma_{0,j}^t}x\right)+\frac{1}{2}f_{0,j}'\left(\widetilde{x}\right)\left(\frac{\widehat{\mu}_j^t-\mu_{0,j}^t}{\sigma_{0,j}^t}+\frac{\widehat{\sigma}_j^t-\sigma_{0,j}^t}{\sigma_{0,j}^t}x\right)^2\right]\\
&=f_{0,j}\left(x\right)\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}_j^t-\mu_{0,j}^t}{\sigma_{0,j}^t}\right)+xf_{0,j}\left(x\right)\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}_j^t-\sigma_{0,j}^t}{\sigma_{0,j}^t}\right)+o_p\left(1\right),
\end{aligned}
$$

where the second equality uses the fact that $\xi_{j,t}$ being independent of $\mathcal{F}^{t-1}$ ; the third equality is due to Taylor expansion around $x$, and $\widetilde{x}$ lies between $x$ and $\frac{\widehat{\mu}_j^t-\mu_{0,j}^t}{\sigma_{0,j}^t}+\frac{\widehat{\sigma}_j^t}{\sigma_{0,j}^t}x$. Since $f_{0,j}\left(\cdot\right)$ is continuously differentiable (Assumption 2(ii)), $\left|f_{0,j}'\left(\widetilde{x}\right)\right|<\infty$ and $\left|f_{0,j}'\left(\widetilde{x}\right)x\right|<\infty$. Along with the fact that $\|\widehat{\kappa}_j-\kappa_{0,j}\|_j^2=o_p\left(n^{-1/2}\right)$, we obtain the fourth equality. Thus, $\sqrt{n}\left(\widetilde{F}_{j,2s}(x)-F_{0,j}\left(x\right)\right)=f_{0,j}\left(x\right)\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\mu}_j^t-\mu_{0,j}^t}{\sigma_{0,j}^t}\right)+xf_{0,j}\left(x\right)\sqrt{n}\mathbb{E}_X\left(\frac{\widehat{\sigma}_j^t-\sigma_{0,j}^t}{\sigma_{0,j}^t}\right)+n^{-1/2}\sum_{t=1}^n\left(\mathbb{I}\left\{\xi_{j,t}\le x\right\}-F_{0,j}\left(x\right)\right)+o_p\left(1\right)$. Then Theorem 4 holds under Assumption 8. □

We will establish Lemma 3 in the theoretical framework of Liu and Yang (2016) (see Appendix B.2).

**Assumption C.5.** *For $j=1,\cdots,k$, the first-order partial derivative $\frac{\partial C_0(u)}{\partial u_j}$ exists and is continuous on $\{u\in[0,1]^k;0<u_j<1\}$.*

**Assumption C.6.** *For $i,j=1,\cdots,k$, (i) the second-order partial derivatives $\frac{\partial^2F_0(z)}{\partial z_i\partial z_j}$ exist and satisfy $\sup_{z\in R^k}\left|\frac{\partial^2F_0(z)}{\partial z_i\partial z_j}z_iz_j\right|<\infty$; (ii) Assumption 7 (i); (iii) $\inf_{u\in[0,1]}h_{j,\gamma_{0,j}}(u)>0$; (iv) $\sup_{u\in[0,1]}\left|\frac{\partial h_{j,\gamma_{0,j}}(u)}{\partial\gamma_j^\intercal}\right|<\infty$, $\sup_{u\in[0,1]}\left|\frac{\partial h_{j,\gamma_{0,j}}(u)}{\partial u}\right|<\infty$, $\sup_{u\in[0,1]}\left|\frac{\partial^2h_{j,\gamma_{0,j}}(u)}{\partial\gamma_j^\intercal\partial u}\right|<\infty$, $\sup_{u\in[0,1]}\left|\frac{\partial^2h_{j,\gamma_{0,j}}(u)}{\partial u^2}\right|<\infty$, $\sup_{u\in[0,1]}\left|\frac{\partial^3h_{j,\gamma_{0,j}}(u)}{\partial\gamma_j^\intercal\partial u^2}\right|<\infty$.*

**Assumption C.7.** *$\{Y_t,X_{1,\gamma_1,t},\cdots,X_{k,\gamma_k,t}\ 1\le t\le n\}$ is a strictly stationary $\beta$-mixing sequence with $\beta(t)\le\beta_0t^{-\zeta}$ for some $\beta_0>0$ and $\zeta=\gamma-2>2$, $\forall\gamma_j\in\Gamma_j$ with $||\gamma_j-\gamma_{0,j}||\le O\left(\log(\log n)n^{-1/2}\right)$ for $j=1,\cdots,k$.*

## Proof of Lemma 3

For simplicity, we prove Lemma 3 for $k=2$. When $k>2$, the proof is similar. Denote the rescaled empirical copula functions of $\xi_t = (\xi_{1,t}, \xi_{2,t})^\intercal$ and $\widehat{\xi}_t = \left(\widehat{\xi}_{1,t}, \widehat{\xi}_{2,t}\right)^\intercal$ as

$$\widehat{C}_n(u_1,u_2) = \frac{1}{n+1}\sum_{t=1}^{n}\mathbb{I}\left(\xi_{1,t} \le \widehat{F}_{1,2s}^{-1}(u_1), \xi_{2,t} \le \widehat{F}_{2,2s}^{-1}(u_2)\right) = \frac{1}{n+1}\sum_{t=1}^{n}\mathbb{I}\left(\widehat{U}_{1,t} \le u_1, \widehat{U}_{2,t} \le u_2\right),$$
$$\widetilde{C}_n(u_1,u_2) = \frac{1}{n+1}\sum_{t=1}^{n}\mathbb{I}\left(\widehat{\xi}_{1,t} \le \widetilde{F}_{1,2s}^{-1}(u_1), \widehat{\xi}_{2,t} \le \widetilde{F}_{2,2s}^{-1}(u_2)\right) = \frac{1}{n+1}\sum_{t=1}^{n}\mathbb{I}\left(\widetilde{U}_{1,t} \le u_1, \widetilde{U}_{2,t} \le u_2\right),$$

where $\widehat{F}_{j,2s}$ and $\widetilde{F}_{j,2s}$ are the rescaled empirical distribution functions of $\xi_{j,t}$ and $\widehat{\xi}_{j,t}$ for $j=1,2$, respectively. Therefore, $\widehat{\mathbb{C}}_n(u_1,u_2) = \sqrt{n}\left(\widehat{C}_n(u_1,u_2) - C_0(u_1,u_2)\right) + O_p\left(\frac{1}{n}\right)$ and $\widetilde{\mathbb{C}}_n(u_1,u_2) = \sqrt{n}\left(\widetilde{C}_n(u_1,u_2) - C_0(u_1,u_2)\right) + O_p\left(n^{-1}\right)$. Because $\widehat{\mathbb{C}}_n \rightsquigarrow \widehat{\mathbb{C}}$ is a well-known result, in this proof, essentially we show

$$\widetilde{\mathbb{C}}_n \rightsquigarrow \widehat{\mathbb{C}}. \tag{C.14}$$

To prove Eq.(C.14), we define two distribution functions on $[0,1]^2$ [13] : $\widehat{G}_n(u_1,u_2) = \frac{1}{n+1}\sum_{t=1}^{n}\mathbb{I}(F_{0,1}(\xi_{1,t}) \le u_1, F_{0,2}(\xi_{2,t}) \le u_2)$ and $\widetilde{G}_n(u_1,u_2) = \frac{1}{n+1}\sum_{t=1}^{n}\mathbb{I}\left(F_{0,1}\left(\widehat{\xi}_{1,t}\right) \le u_1, F_{0,2}\left(\widehat{\xi}_{2,t}\right) \le u_2\right)$. Note that :

$$\widetilde{C}_n(u_1,u_2) = \widetilde{G}_n(\widetilde{G}_{1,n}^{-1}(u_1), \widetilde{G}_{2,n}^{-1}(u_2)), \tag{C.15}$$

where $\widetilde{G}_{j,n}, j=1,2$ are the marginals of $\widetilde{G}_n$. Define the empirical process $\widetilde{\mathbb{G}}_n(u_1,u_2) \equiv \sqrt{n}(\widetilde{G}_n(u_1,u_2) - C_0(u_1,u_2)) = \sqrt{n}\left(\widetilde{G}_n(u_1,u_2) - \widehat{G}_n(u_1,u_2)\right) + \sqrt{n}\left(\widehat{G}_n(u_1,u_2) - C_0(u_1,u_2)\right)$, where $\sqrt{n}(\widehat{G}_n(u_1,u_2) - C_0(u_1,u_2)) = \mathbb{C}_n(u_1,u_2) + O_p\left(n^{-1}\right) \rightsquigarrow \mathbb{C}(u_1,u_2)$, $\mathbb{C}_n$ and $\mathbb{C}$ are defined in Eq.(31). In the following (*T.1.*) and (*T.2.*), we show weak convergence of $\widetilde{\mathbb{G}}_n$, and then use Eq.(C.15) to establish the weak convergence of $\widetilde{\mathbb{C}}_n$.

[13] Neumeyer et al. (2019) uses the same transformation to establish the invariance of the empirical copula processes to the first-stage estimation error. They consider nonparametric ARCH models and use the local polynomial estimation, while we consider semi-nonparametric GARCH models and use the B spline estimation.

**(*T.1.*) We first prove**

$$
\begin{aligned}
&\mathbb{E}_X\left[\sqrt{n}\left(\widetilde{G}_n(u_1,u_2)-\widehat{G}_n(u_1,u_2)\right)\right] \\
=&\frac{1}{2}\frac{\partial C_0(u_1,u_2)}{\partial u_1}f_{0,1}\left(F_{0,1}^{-1}(u_1)\right)F_{0,1}^{-1}(u_1)\left[E\left(\frac{\partial \log h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}{\partial \gamma_1^{\intercal}}\right)\Upsilon_1+\mathbb{H}_1\right] \\
+&\frac{1}{2}\frac{\partial C_0(u_1,u_2)}{\partial u_2}f_{0,2}\left(F_{0,2}^{-1}(u_2)\right)F_{0,2}^{-1}(u_2)\left[E\left(\frac{\partial \log h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}{\partial \gamma_2^{\intercal}}\right)\Upsilon_2+\mathbb{H}_2\right]+o_p(1),
\end{aligned}
\tag{C.16}
$$

where the expectation takes over the data, and the first-stage estimators are considered to be deterministic. $\Upsilon_j$ and $\mathbb{H}_j$, for $j=1,2$, are defined in Lemmas B.2 and B.3. Note that

$$
\begin{aligned}
&\mathbb{E}_X\left[\sqrt{n}\left(\widetilde{G}_n(u_1,u_2)-\widehat{G}_n(u_1,u_2)\right)\right] \\
=&\sqrt{n}\mathbb{E}_X\left\{\mathbb{I}\left[\xi_{1,t}\le F_{0,1}^{-1}(u_1)\sqrt{\frac{\widehat{h}_{1,\widehat{\gamma}_1}(U_{1,\widehat{\gamma}_1,t})}{h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}},\ \xi_{2,t}\le F_{0,2}^{-1}(u_2)\sqrt{\frac{\widehat{h}_{2,\widehat{\gamma}_2}(U_{2,\widehat{\gamma}_2,t})}{h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}}\right]\right\} \\
&-\sqrt{n}E\left\{\mathbb{I}\left[\xi_{1,t}\le F_{0,1}^{-1}(u_1),\ \xi_{2,t}\le F_{0,2}^{-1}(u_2)\right]\right\}+O_p\left(n^{-1}\right) \\
=&\sqrt{n}\left\{\mathbb{E}_X\left[F_0\left(F_{0,1}^{-1}(u_1)\sqrt{\frac{\widehat{h}_{1,\widehat{\gamma}_1}(U_{1,\widehat{\gamma}_1,t})}{h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}},\ F_{0,2}^{-1}(u_2)\sqrt{\frac{\widehat{h}_{2,\widehat{\gamma}_2}(U_{2,\widehat{\gamma}_2,t})}{h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}}\right)\right]\right. \\
&\left.-F_0\left(F_{0,1}^{-1}(u_1),\ F_{0,2}^{-1}(u_2)\right)\right\}+O_p\left(n^{-1}\right) \\
=&\mathbb{E}_X\left[\frac{\partial F_0\left(F_{0,1}^{-1}(u_1),F_{0,2}^{-1}(u_2)\right)}{\partial z_1}F_{0,1}^{-1}(u_1)\sqrt{n}\left(\sqrt{\frac{\widehat{h}_{1,\widehat{\gamma}_1}(U_{1,\widehat{\gamma}_1,t})}{h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}}-1\right)\right] \\
&+\mathbb{E}_X\left[\frac{\partial F_0\left(F_{0,1}^{-1}(u_1),F_{0,2}^{-1}(u_2)\right)}{\partial z_2}F_{0,2}^{-1}(u_2)\sqrt{n}\left(\sqrt{\frac{\widehat{h}_{2,\widehat{\gamma}_2}(U_{2,\widehat{\gamma}_2,t})}{h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}}-1\right)\right]+o_p(1) \\
=&\frac{\partial C_0(u_1,u_2)}{\partial u_1}f_{0,1}\left(F_{0,1}^{-1}(u_1)\right)F_{0,1}^{-1}(u_1)\mathbb{E}_X\left[\sqrt{n}\left(\sqrt{\frac{\widehat{h}_{1,\widehat{\gamma}_1}(U_{1,\widehat{\gamma}_1,t})}{h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}}-1\right)\right] \\
&+\frac{\partial C_0(u_1,u_2)}{\partial u_2}f_{0,2}\left(F_{0,2}^{-1}(u_2)\right)F_{0,2}^{-1}(u_2)\mathbb{E}_X\left[\sqrt{n}\left(\sqrt{\frac{\widehat{h}_{2,\widehat{\gamma}_2}(U_{2,\widehat{\gamma}_2,t})}{h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}}-1\right)\right]+o_p(1),
\end{aligned}
\tag{C.17}
$$

where the second equality is due to $\xi_t$ being independent of $\mathcal{F}^{t-1}$[14], the third equality is implied by Taylor expansion, Assumption C.6 (i)(iii) and Eq.(B.8). Note that, for $j=1,2$,

$$\sqrt{n}\left(\sqrt{\frac{\widehat{h}_{j,\widehat{\gamma}_j}(U_{j,\widehat{\gamma}_j,t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}}-1\right)=\frac{1}{2}\sqrt{n}\left(\frac{\widehat{h}_{j,\widehat{\gamma}_j}(U_{j,\widehat{\gamma}_j,t})-h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}\right)+o_p(1)$$

$$=\frac{1}{2}\sqrt{n}\left(\frac{\widehat{h}_{j,\widehat{\gamma}_j}(U_{j,\widehat{\gamma}_j,t})-\widehat{h}_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})+\widehat{h}_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})-h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}\right)+o_p(1)$$

$$=\frac{1}{2}\sqrt{n}\left(\frac{\partial\log h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{\partial\gamma_j^{\intercal}}(\widehat{\gamma}_j-\gamma_{0,j})+\frac{\widehat{h}_{j,\gamma_{0,j}}\left(U_{j,\gamma_{0,j},t}\right)-h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}\right)+o_p(1), \quad \text{(C.18)}$$

where the first equality is by Taylor expansion of $\sqrt{\widehat{h}_{j,\widehat{\gamma}_j}(U_{j,\widehat{\gamma}_j,t})}$ around $\sqrt{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}$ and Eq.(B.8), the second equality is again by Taylor expansion and Eq.(B.8), and $\widehat{\gamma}_j-\gamma_{0,j}=O_p\left(n^{-1/2}\right)$. Taking the expectation of Eq. (C.18), Lemmas B.2 and B.3 imply, for $j=1,2$

$$\mathbb{E}_X\left[\sqrt{n}\left(\sqrt{\frac{\widehat{h}_{j,\widehat{\gamma}_j}(U_{j,\widehat{\gamma}_j,t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}}-1\right)\right]=\frac{1}{2}E\left(\frac{\partial\log h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{\partial\gamma_j^{\intercal}}\right)\Upsilon_j+\frac{1}{2}\mathbb{H}_j+o_p(1). \quad \text{(C.19)}$$

Then Eqs.(C.17) and (C.19) lead to Eq.(C.16).

***(T.2.)* We then show**

$$\sup_{u\in[0,1]^2}\left|\sqrt{n}\left(\widetilde{G}_n(u_1,u_2)-\widehat{G}_n(u_1,u_2)\right)-\mathbb{E}_X\left[\sqrt{n}\left(\widetilde{G}_n(u_1,u_2)-\widehat{G}_n(u_1,u_2)\right)\right]\right|=o_p(1). \quad \text{(C.20)}$$

Following Akritas and Keilegom (2001), Dette et al. (2009) and Neumeyer et al. (2019), we define the class of functions:

$$\mathcal{F}=\left\{(u_1,u_2,\xi_1,\xi_2)\mapsto\mathbb{I}\left[\xi_1\le y_1\psi_1\left(u_1\right),\ \xi_2\le y_2\psi_2\left(u_2\right)\right],\ y_1,y_2\in\mathcal{R},\psi_1,\psi_2\in\widetilde{C}_2^{1+\delta}\left([0,1]\right)\right\},$$

where $\delta\in(0,1]$, $\widetilde{C}_2^{1+\delta}([0,1])$ is the class of all differential functions $\psi$ defined on $[0,1]$ such that $||\psi||_{1+\delta}\le 2$, $\inf_{u\in[0,1]}\psi(u)\ge\frac{1}{2}$, and $||\psi||_{1+\delta}=\max\left\{\sup_{u\in[0,1]}|\psi(u)|,\sup_{u\in[0,1]}|\psi'(u)|\right\}+\sup_{u\neq u'\in[0,1]}\frac{|\psi'(u)-\psi'(u')|}{|u-u'|^{\delta}}$. Note that

$$||\psi_1+\psi_2||_{1+\delta}\le||\psi_1||_{1+\delta}+||\psi_2||_{1+\delta},\quad\text{and}\quad||\lambda\psi||_{1+\delta}=|\lambda|||\psi||_{1+\delta},\quad\forall\lambda\in\mathcal{R}. \quad \text{(C.21)}$$

[14] If random variable $W$ is independent of Z, let $F$ denote the CDF of $W$, then $E\left\{\mathbb{I}\left(W\le Z\right)\right\}=E\left\{E\left[\mathbb{I}\left(W\le Z\right)|Z\right]\right\}=E\left\{F(Z)\right\}$.

Denote the centered process

$$Z_n(g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[ g\left(\xi_t, \mathcal{F}^{t-1}\right) - \mathbb{E}_X\left(g\left(\xi_t, \mathcal{F}^{t-1}\right)\right)\right], \ \forall g \in \mathcal{F}. \tag{C.22}$$

Note that $\sqrt{n}\left(\widetilde{G}_n(u_1,u_2) - \widehat{G}_n(u_1,u_2)\right) - \mathbb{E}_X\left[\sqrt{n}\left(\widetilde{G}_n(u_1,u_2) - \widehat{G}_n(u_1,u_2)\right)\right]$

$$\begin{aligned}
=&Z_n\left(\mathbb{I}\left[\xi_{1,t} \le F_{0,1}^{-1}(u_1)\sqrt{\frac{\widehat{h}_{1,\widehat{\gamma}_1}(U_{1,\widehat{\gamma}_1,t})}{h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}},\ \xi_{2,t} \le F_{0,2}^{-1}(u_2)\sqrt{\frac{\widehat{h}_{2,\widehat{\gamma}_2}(U_{2,\widehat{\gamma}_2,t})}{h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}}\right]\right)\\
&-Z_n\left(\mathbb{I}\left[\xi_{1,t} \le F_{0,1}^{-1}(u_1),\ \xi_{2,t} \le F_{0,2}^{-1}(u_2)\right]\right) + O_p\left(n^{-1}\right).
\end{aligned} \tag{C.23}$$

1. We first show $\psi(u) \equiv \sqrt{\frac{\widehat{h}_{j,\widehat{\gamma}_j}(U_{j,\widehat{\gamma}_j,t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}}\Bigg| U_{j,\gamma_{0,j},t} = u \in \widetilde{C}_2^{1+\delta}([0,1])$ almost surely. Equivalently we show $||\phi||_{1+\delta} = o_p(1)$, where $\phi(u) \equiv \sqrt{\frac{\widehat{h}_{j,\widehat{\gamma}_j}(U_{j,\widehat{\gamma}_j,t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}} - 1\Bigg| U_{j,\gamma_{0,j},t} = u$. As shown in Eq. (C.18),

$$\begin{aligned}
\sqrt{\frac{\widehat{h}_{j,\widehat{\gamma}_j}(U_{j,\widehat{\gamma}_j,t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}} - 1 =& \frac{1}{2}\frac{\partial \log h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{\partial \gamma_j^\intercal}(\widehat{\gamma}_j - \gamma_{0,j}) + \frac{1}{2}\frac{\widehat{h}_{j,\gamma_{0,j}}\left(U_{j,\gamma_{0,j},t}\right) - \widetilde{h}_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}\\
&+ \frac{1}{2}\frac{\widetilde{h}_{j,\gamma_{0,j}}\left(U_{j,\gamma_{0,j},t}\right) - h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})} + R\left(U_{j,\gamma_{0,j},t}\right),
\end{aligned}$$

where $\widetilde{h}_{j,\gamma_{0,j}}(\cdot)$ is the projection of $h_{j,\gamma_{0,j}}(\cdot)$ onto the cubic spline function space and $\sup_{u\in[0,1]}|R(u)| = o_p(n^{-1/2})$. $\left\|\frac{\partial \log h_{j,\gamma_{0,j}}(u)}{\partial \gamma_j^\intercal}(\widehat{\gamma}_j - \gamma_{0,j})\right\|_{1+\delta} = o_p(1)$ is valid under Assumption C.6 (iii) (iv) and Lemma B.2. Also, $h_{j,\gamma_{0,j}}(u)$ is estimated using cubic spline functions, $\sum_{J=1-k}^{N} \lambda_J B_{J,4}(u)$. The support $[0,1]$ is divided into $N+1$ equally-spaced subintervals $J_j = [\frac{j}{N+1}, \frac{j+1}{N+1})$, for $j = 0,1,\cdots,N-1$, $J_N = [\frac{N}{N+1}, 1]$. As shown in Lemma A.9 of Liu and Yang (2016), $\sup_{u\in[0,1]}\left|\widetilde{h}_{j,\gamma_{0,j}}(u) - h_{j,\gamma_{0,j}}(u)\right| = O(N^{-4}) = o(n^{-1/2})$. Thus the two terms $\frac{1}{2}\frac{\widetilde{h}_{j,\gamma_{0,j}}\left(U_{j,\gamma_{0,j},t}\right) - h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}$ and $R\left(U_{j,\gamma_{0,j},t}\right)$ can be handled analogously as in the proof of Theorem 1 in Neumeyer et al. (2019).

As shown in Proposition 2 and Lemma A.14 of Liu and Yang (2016), $\widehat{h}_{j,\gamma_{0,j}}(u) - \widetilde{h}_{j,\gamma_{0,j}}(u) = \sum_{J=1-k}^{N} \widehat{\lambda}_J B_{J,4}(u)$, $\sup_{u\in[0,1]}\left|\widehat{h}_{j,\gamma_{0,j}}(u) - \widetilde{h}_{j,\gamma_{0,j}}(u)\right| = O_p\left(\frac{\log(n)\sqrt{N}}{\sqrt{n}}\right)$ and $\max_{J=1-k}^{N}\left\{\left|\widehat{\lambda}_J\right|\right\} = O_p\left(\frac{\log(n)\sqrt{N}}{\sqrt{n}}\right)$. For any given value $\tilde{u}$, if $\tilde{u} \in J_j = [\frac{j}{N+1}, \frac{j+1}{N+1})$, then $B_{J,4}(\tilde{u}) \neq 0$ only for

$J = j-3, j-2, j-1, j$, thus $\widehat{h}_{j,\gamma_{0,j}}(\tilde{u}) - \widetilde{h}_{j,\gamma_{0,j}}(\tilde{u}) = \sum_{J=j-3}^{j} \widehat{\lambda}_J B_{J,4}(\tilde{u})$. Furthermore, according to Lemma A.3 (ii) of Liu and Yang (2016), $\sup_{u\in[0,1]} \left|\frac{d}{du} B_{J,4}(u)\right| = O(N)$ and $\sup_{u\in[0,1]} \left|\frac{d^2}{du^2} B_{J,4}(u)\right| = O(N^2)$ (see Liu and Yang (2016) pages 5, 17 for detailed descriptions). Therefore, under Assumption C.6 (iii)(iv), for some $M_1, M_2 < \infty$

$$\left\|\frac{\widehat{h}_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t}) - \widetilde{h}_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}\right\|_{1+\delta} \leq M_1 \max_{J=1-k}^{N} \left\{\left|\widehat{\lambda}_J\right|\right\} \sup_{u\in[0,1]} \left|\frac{d}{du} B_{J,4}(u)\right|$$
$$+ M_2 \max_{J=1-k}^{N} \left\{\left|\widehat{\lambda}_J\right|\right\} \sup_{u\in[0,1]} \left|\frac{d^2}{du^2} B_{J,4}(u)\right| = O_p\left(\frac{\log(n) N^{5/2}}{\sqrt{n}}\right) = o_p(1),$$

where the last equality is satisfied under Assumption A6 of Liu and Yang (2016), $n^{1/6} << N << n^{1/5}\log(n)^{-2/5}$, and $N = o\left(n^{1/5}\log(n)^{-2/5}\right)$ is needed. Due to Eq. (C.21), we have $||\phi||_{1+\delta} = o_p(1)$, consequently $\psi(u) \equiv \sqrt{\frac{\widehat{h}_{j,\widehat{\gamma}_j}(U_{j,\widehat{\gamma}_j,t})}{h_{j,\gamma_{0,j}}(U_{j,\gamma_{0,j},t})}}\Bigg| U_{j,\gamma_{0,j},t} = u \in \widetilde{C}_2^{1+\delta}([0,1])$ almost surely. Therefore,

$$\mathbb{I}\left[\xi_{1,t} \leq F_{0,1}^{-1}(u_1)\sqrt{\frac{\widehat{h}_{1,\widehat{\gamma}_1}(U_{1,\widehat{\gamma}_1,t})}{h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}},\ \xi_{2,t} \leq F_{0,2}^{-1}(u_2)\sqrt{\frac{\widehat{h}_{2,\widehat{\gamma}_2}(U_{2,\widehat{\gamma}_2,t})}{h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}}\right] \in \mathcal{F}, \quad \text{a.s.}$$

2. We have $\log N\left(\epsilon, \widetilde{C}_2^{1+\delta}([0,1]), ||\cdot||_\infty\right) = O\left(\epsilon^{-\frac{1}{1+\delta}}\right)$ and $\log N_{[]}\left(\epsilon, \widetilde{C}_2^{1+\delta}([0,1]), ||\cdot||_2\right) = O\left(\epsilon^{-\frac{1}{1+\delta}}\right)$, for every $\epsilon > 0$, due to van der Vaart and Wellner (1996) Theorem 2.7.1 (page 155) and Corollary 2.7.2 (page 157), where $||\cdot||_\infty$ denotes the sup norm and $||\cdot||_2$ denotes the $L_2$ norm. Define the semi-norm (see same definitions in Doukhan et al. (1995), Dedecker and Louhichi (2002) and Dette et al. (2009)): $||g||_{2,\beta}^2 = \int_0^1 \beta^{-1}(u) Q_g^2(u) du$, $\beta^{-1}(u) = \inf\left\{x > 0 : \beta_{[x]} \leq u\right\}$, $Q_g(u) = \inf\left\{x > 0 : \Pr\{|g| > x\} \leq u\right\}$, where $[x]$ is the integer part of $x$ and $\beta_t$ is the mixing coefficient. Following an analogous argument as in the proofs of Lemma 1 of Akritas and Keilegom (2001) and Lemma 1 of Dette et al. (2009), we can show that $\log N_{[]}\left(\epsilon, \mathcal{F}, ||\cdot||_{2,\beta}\right) = O\left(\log\left(\epsilon^{-1}\right)\right) + O\left(\epsilon^{-\frac{2}{1+\delta}}\right)$, for every $\epsilon > 0$. Along with one bracket being sufficient for $\epsilon \geq 1$, we have $\int_0^\infty \sqrt{\log N_{[]}\left(\epsilon, \mathcal{F}, ||\cdot||_{2,\beta}\right)} d\epsilon < \infty$. So the centered process $Z_n(\cdot)$ defined in Eq.(C.22) is asymptotically $||\cdot||_{2,\beta}$-equicontinuous, due to Section 4.3 of Dedecker and Louhichi (2002).

3. We are left to show

$$\sup_{u\in[0,1]^2}\left\|\mathbb{I}\left[\xi_{1,t}\le F_{0,1}^{-1}(u_1)\sqrt{\frac{\widehat{h}_{1,\widehat{\gamma}_1}(U_{1,\widehat{\gamma}_1,t})}{h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}},\ \xi_{2,t}\le F_{0,2}^{-1}(u_2)\sqrt{\frac{\widehat{h}_{2,\widehat{\gamma}_2}(U_{2,\widehat{\gamma}_2,t})}{h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}}\right]\right.$$
$$\left.-\ \mathbb{I}\left[\xi_{1,t}\le F_{0,1}^{-1}(u_1),\ \xi_{2,t}\le F_{0,2}^{-1}(u_2)\right]\right\|_{2,\beta}\longrightarrow 0. \tag{C.24}$$

Denote

$$\lambda(u_1,u_2)=\Pr\left\{\left|\mathbb{I}\left[\xi_{1,t}\le F_{0,1}^{-1}(u_1)\sqrt{\frac{\widehat{h}_{1,\widehat{\gamma}_1}(U_{1,\widehat{\gamma}_1,t})}{h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}},\ \xi_{2,t}\le F_{0,2}^{-1}(u_2)\sqrt{\frac{\widehat{h}_{2,\widehat{\gamma}_2}(U_{2,\widehat{\gamma}_2,t})}{h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}}\right]\right.\right.$$
$$\left.\left.-\ \mathbb{I}\left[\xi_{1,t}\le F_{0,1}^{-1}(u_1),\ \xi_{2,t}\le F_{0,2}^{-1}(u_2)\right]\right|>0\right\}$$
$$\le\left|\Pr\left\{\xi_{1,t}\le F_{0,1}^{-1}(u_1)\sqrt{\frac{\widehat{h}_{1,\widehat{\gamma}_1}(U_{1,\widehat{\gamma}_1,t})}{h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}}\right\}-\Pr\left\{\xi_{1,t}\le F_{0,1}^{-1}(u_1)\right\}\right|$$
$$+\left|\Pr\left\{\xi_{2,t}\le F_{0,2}^{-1}(u_2)\sqrt{\frac{\widehat{h}_{2,\widehat{\gamma}_2}(U_{2,\widehat{\gamma}_2,t})}{h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}}\right\}-\Pr\left\{\xi_{2,t}\le F_{0,2}^{-1}(u_2)\right\}\right|$$
$$=\left|\mathbb{E}_X\left\{F_{0,1}\left[F_{0,1}^{-1}(u_1)\sqrt{\frac{\widehat{h}_{1,\widehat{\gamma}_1}(U_{1,\widehat{\gamma}_1,t})}{h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}}\right]\right\}-u_1\right|$$
$$+\left|\mathbb{E}_X\left\{F_{0,2}\left[F_{0,2}^{-1}(u_2)\sqrt{\frac{\widehat{h}_{2,\widehat{\gamma}_2}(U_{2,\widehat{\gamma}_2,t})}{h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}}\right]\right\}-u_2\right|=O_p\left(n^{-1/2}\right),$$

where the last equality can be shown using the similar argument as Eq.(C.17). By the definition of $\|\cdot\|_{2,\beta}$ and Assumption C.7, we have

$$\left\|\mathbb{I}\left[\xi_{1,t}\le F_{0,1}^{-1}(u_1)\sqrt{\frac{\widehat{h}_{1,\widehat{\gamma}_1}(U_{1,\widehat{\gamma}_1,t})}{h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}},\ \xi_{2,t}\le F_{0,2}^{-1}(u_2)\sqrt{\frac{\widehat{h}_{2,\widehat{\gamma}_2}(U_{2,\widehat{\gamma}_2,t})}{h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}}\right]\right.$$
$$\left.-\ \mathbb{I}\left[\xi_{1,t}\le F_{0,1}^{-1}(u_1),\ \xi_{2,t}\le F_{0,2}^{-1}(u_2)\right]\right\|_{2,\beta}^2$$
$$=\int_0^{\lambda(u_1,u_2)}\beta^{-1}(v)dv\le\int_0^{\lambda(u_1,u_2)}\frac{v^{-1/\zeta}}{\beta_0^{-1/\zeta}}dv=\frac{\lambda(u_1,u_2)^{(\zeta-1)/\zeta}}{\beta_0^{-1/\zeta}(\zeta-1)/\zeta}\longrightarrow 0,$$

uniformly on $(u_1,u_2)\in[0,1]^2$. $\beta_0>0$ and $\zeta>2$ are defined in Assumption C.7. Because the centered process $Z_n(\cdot)$ is asymptotically $||\cdot||_{2,\beta}$-equicontinuous, we establish Eq.(C.20) using Eqs.(C.23) and (C.24).

$(T.1.)$ and $(T.2.)$ imply

$$\sqrt{n}\left(\widetilde{G}_n(u_1,u_2)-\widehat{G}_n(u_1,u_2)\right) \rightsquigarrow \frac{1}{2}\frac{\partial C_0(u_1,u_2)}{\partial u_1} f_{0,1}\left(F_{0,1}^{-1}(u_1)\right) F_{0,1}^{-1}(u_1)\left[E\left(\frac{\partial \log h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}{\partial \gamma_1^{\intercal}}\right)\Upsilon_1+\mathbb{H}_1\right]$$
$$+\frac{1}{2}\frac{\partial C_0(u_1,u_2)}{\partial u_2} f_{0,2}\left(F_{0,2}^{-1}(u_2)\right) F_{0,2}^{-1}(u_2)\left[E\left(\frac{\partial \log h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}{\partial \gamma_2^{\intercal}}\right)\Upsilon_2+\mathbb{H}_2\right],$$

therefore, $\widetilde{\mathbb{G}}_n(u_1,u_2) \equiv \sqrt{n}\left(\widetilde{G}_n(u_1,u_2)-C_0(u_1,u_2)\right) \rightsquigarrow \mathbb{C}(u_1,u_2)$

$$+\frac{1}{2}\frac{\partial C_0(u_1,u_2)}{\partial u_1} f_{0,1}\left(F_{0,1}^{-1}(u_1)\right) F_{0,1}^{-1}(u_1)\left[E\left(\frac{\partial \log h_{1,\gamma_{0,1}}(U_{1,\gamma_{0,1},t})}{\partial \gamma_1^{\intercal}}\right)\Upsilon_1+\mathbb{H}_1\right]$$
$$+\frac{1}{2}\frac{\partial C_0(u_1,u_2)}{\partial u_2} f_{0,2}\left(F_{0,2}^{-1}(u_2)\right) F_{0,2}^{-1}(u_2)\left[E\left(\frac{\partial \log h_{2,\gamma_{0,2}}(U_{2,\gamma_{0,2},t})}{\partial \gamma_2^{\intercal}}\right)\Upsilon_2+\mathbb{H}_2\right].$$

Under Assumption C.5, Theorem 3.9.4 of van der Vaart and Wellner (1996) implies (also see Proposition A.1. of Genest et al. (2007)),

$$\begin{aligned}\widetilde{\mathbb{C}}_n(u_1,u_2) &= \sqrt{n}\left(\widetilde{C}_n(u_1,u_2)-C_0(u_1,u_2)\right)+O_p(n^{-1}) = \sqrt{n}\left(\widetilde{G}_n(\widetilde{G}_{1,n}^{-1}(u_1),\widetilde{G}_{2,n}^{-1}(u_2))-C_0(u_1,u_2)\right)+O_p(n^{-1})\\ &= \widetilde{\mathbb{G}}_n(u_1,u_2)-\frac{\partial C_0(u_1,u_2)}{\partial u_1}\widetilde{\mathbb{G}}_n(u_1,1)-\frac{\partial C_0(u_1,u_2)}{\partial u_2}\widetilde{\mathbb{G}}_n(1,u_2)+o_p(1)\\ &\rightsquigarrow \mathbb{C}(u_1,u_2)-\frac{\partial C_0(u_1,u_2)}{\partial u_1}\mathbb{C}(u_1,1)-\frac{\partial C_0(u_1,u_2)}{\partial u_2}\mathbb{C}(1,u_2)\equiv \widehat{\mathbb{C}}. \quad \square\end{aligned}$$

## Proof of Theorem 5

Similar to the proof of Theorem 2 in Neumeyer et al. (2019). □