# Three essays in Time Series Econometrics

Bo Wang

A dissertation

submitted to the Faculty of

the department of Economics

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

November 2020

© Copyright 2020 Bo Wang

**Three essays in Time Series Econometrics**

Bo Wang

Advisor: Prof. Zhijie Xiao
Prof. Arthur Lewbel
Prof. Shakeeb Khan

The first two chapters study the copula Markov model combined with nonstationarity.

The last chapter proposes a new structural break test with good size and power.

# Contents

[1] Joint work with Zhijie Xiao xiaoz@bc.edu and Zhongjun Qu qu@bu.edu

# List of Tables

# List of Figures

## Part I

# Chapter 1: Semiparametric Sieve Estimation of Dynamic Copula Models with Filtered Nonstationarity


**Abstract**

This paper considers estimation of copula based dynamic semiparametric models coupled with nonstationary filtration. A two-step sieve method is proposed and new theoretical results are obtained regarding the effects of nonstationarity on limiting distributions. Simulation results indicate that the tail dependence brings a finite sample bias in the two-step sieve estimator. For this reason, a joint sieve estimator is proposed that is found to always be superior to all other estimators in a variety of Monte Carlo simulation designs. An empirical estimation for cointegration between weekly stock price and consensus target price highlights the theoretical finding. The results are important for value-at-risk calculation and stock price prediction conditional on consensus target price for potential improvement of stock price forecast accuracy.[2]


## 1 Introduction

Nonstationarity is an important empirical feature in economic and financial time series. Many observed time series seem to display nonstaionary characteristics. In economics and finance, many time series grow in a secular way over long periods of time, some appear to wander around as if they have no fixed mean. Growth characteristics are especially evident in time series that represent aggregate economic behavior like gross domestic product and industrial production. Random wandering behavior is evident in many financial time series like interest rates and asset prices. In addition, nonlinearity is another important characteristic for economic and financial empirical practice. Investors respond to good and bad news in an asymmetric way, asymmetric dependence and other types nonlinearity appear in a wide range of economic and financial series. For example, Granger [2003] points out that the classical linear time series modeling based on Gaussian distribution assumption clearly fails to explain the stylized facts observed in economic and financial data. The above mentioned nonstationarity are usually modeled by deterministic trend and/or unit roots, and a very important and popular tool to model nonlinear dynamics is to use copula, which provides a parsimonious tool for capturing nonlinearity, asymmetry and tail dependence.

In this paper, we consider a time series modeled where $Y_t$ can be decomposed into a nonstationary component and a stationary part:

$$Y_t = X_t'\beta_0 + V_t$$

where $X_t'\beta_0$ is the nonstationary component and $V_t$ is the stationary part that may display nonlinear dynamics.

We use copula to capture nonlinear temporal dependence of $V_t$. If we assume that $V_t$ is a first-order Markov process, by the Sklar [1959] theorem, the joint distribution of $V_t$ and $V_{t-1}$ can be modeled using a parametric bivariate copula that is dependent on some unknown parameter $\alpha_0$, and the marginal distribution of $V_t$, $F_0(\cdot)$. Analysis of the above model entails estimation the coefficients $\beta_0$, the copula parameter $\alpha_0$, and the marginal distribution function of $V_t$, $F_0(\cdot)$. In this paper, we model $F_0(\cdot)$, the marginal of $V_t$, nonparametrically for robustness reason, and consider semiparametric estimation for the model. In particular, our focus is the copula parameter $\alpha_0$.

The parameters $\beta_0$, $\alpha_0$, and $F_0$ may be estimated sequentially or jointly. The first method is a two-step estimation where nonstaionarity is filtered (thus $\beta_0$ is estimated) first, followed by a second


[2]The author is deeply indebted to Zhijie Xiao for his guidance, inspiration and encouragement. He is grateful to Arthur Lewbel, Shakeeb Khan for their endless encouragement and support. He benefited insightful comments from the dissertation workshop in the Department of Economics at Boston College, and BU-BC joint workshop in Econometrics. The usual disclaimer applies.

stage estimation of the copula model, i.e. $(\alpha_0, F_0)$, based on the filtered data. The second method is a joint estimation where $\beta_0$, $\alpha_0$, and $F_0$ are all estimated simultaneously.

This paper shows that the performance and relative effectiveness of these two methods depend on the functional form of the copula and on marginal $F_0$. For some copulas, like the Clayton or Gumbel copula, the limiting distribution of $\hat{\alpha}$ is non-normal and depends on the time series properties of $Y_t$ and $X_t$. For other copulas which are symmetric around $(1/2, 1/2)$, like the Gaussian, Frank and EFGM copula, if $F_0$ is symmetric then the distribution of $\widehat{\alpha}$ is normal and not affected by the nonstationarity of $Y_t$ and $X_t$. In this latter case, the estimation of $\alpha_0$ is, in theory, equally efficient whether $\beta_0$ is estimated first or simultaneously. However, even in this case we find that simultaneous estimation performs better in finite sample simulations. For comparison purpose, we also consider two cases in simulation: the infeasible case when $V_t$ is observed rather than estimated (thus no filtration is needed) and the wrongly-specified parametric case when marginal $F_0(\cdot)$ is endowed with a false parametric structure and we estimate three terms jointly.

Chen and Fan [2006] considered a three-step copula based estimation for time series with filtered nonstationarity. They estimates $\beta$ through OLS to get fitted residuals $\hat{V}_t$ first. Marginal distribution $F_0$ is estimated by the empirical distribution of $\hat{V}_t$. Copula parameter $\alpha_0$ is estimated via MLE during last step. They show that limiting distribution of their three-step empirical estimator is not affected by nonstationary structure of $X_t$. The proposed estimation is simple and convenient for application. But their simulation results show that three-step empirical estimator is biased in finite sample for tail dependence copula, such as Clayton and Gumbel copula.

Chen and Xiao [2016]'s theoretical results cannot be fully extended to our two-step sieve estimator, where $\beta$ is estimated through OLS at first step, then $\alpha_0$ and $F_0$ are jointly estimated through sieve MLE based on fitted residuals $\hat{V}_t$. Theoretical results and corresponding simulation demonstrate that nonstationary structure of $X_t$ will affect limiting distribution when copulas are Clayton or Gumbel, even when marginal is ymmetric. And this effect is positively related with strength of tail dependence.

We apply our methods to study the cointegration model between weekly stock price and consensus target price, the relationship found in Brav and Lehavy [2003], to take into account the non-linear structure in the unobserved residuals. We employ Gaussian copula (without tail dependence) on in-sample estimation and find that our sieve method have better out-of-sample prediction power than empirical method. Clayton copula (with lower tail dependence) is also applied to residuals and our sieve method correctly reject the Clayton structure (no convergence), while empirical method reaches an evidently higher tail dependence estimation. Results show that sieve method is robust and sieve estimator of copula parameter is more convincing and meaningful than empirical method. The estimation results for copula parameter, cointegration coefficient and marginal distribution are attractive for value-at-risk calculation and stock price prediction conditional on consensus target price.

**Related literature.** There are a growing number of papers using copulas to model the temporal dependence of univariate nonlinear time series. Darsow et al. [1992], Victor et al. [2006] and Ibragimov [2009] provide characterizations of a copula-based time series to be a Markov process. Joe [1997] proposes a general structure with parametric stationary Markov models based on parametric copulas and parametric marginal distributions. Sieve application in a semiparametric setting is considered in Chen and Shen [1998] and Ai and Chen [2003]. Chen et al. [2006] apply sieve method on semiparametric copula model under i.i.d. setting. Most related literature to the model considered in this paper are Chen and Fan [2006] and Chen et al. [2009]. Both of them analyze the ideal case where $Y_t$ is directly observed. Chen and Fan [2006] used empirical function based method and Chen et al. [2009] considers sieve approximation. Chen and Xiao [2016] research on same model as us. They not only consider semiparametric problem based on empirical distribution function, as we mention above, but also analyze the parametric marginal case.

**Organization of the paper.** The rest of this article is organized as follows. In **Section** 2 we introduce nonstationary structure, sieve spaces and several commonly used copulas. In **Section** 4 and **Section** 5, we derive consistency and limiting distribution of our sieve estimator, showing whether nonstationary structure will not affect limiting distribution of the estimator. In **Section** 6 we summarize simulation results of the sieve MLE for various nonlinearity structure in Gaussian, Frank, Clayton, EFGM and Gumbel copulas. In **Section** 7, we apply our sieve copula method on cointegration between stock price and consensus target price. Mathematical proof and detailed simulation results are left in remaining sections.

# 2 Copula based model with nonstationary filtering

## 2.1 The Model

We assume that the observed time series $\{Y_t\}_{t=1}^n$ can be modeled as:

$$Y_t = X_t'\beta_0 + V_t$$

where $X_t'\beta_0$ is the non-stationary component and $V_t$ is the stationary component with non-linearity. In particular, we assume that $X_t$ is a $d_X$ dimensional vector of dependent variables that may be nonstationary. The second component, $V_t$, is a stationary process with non-linearity that can be captured by a copula function. For simplicity and without loss of generality, we assume in this paper that $\{V_t\}_{t=1}^n$ is a first-order strictly stationary Markov process. Higher order Markov process can be investigated similarly.

Under the assumption that $\{V_t\}_{t=1}^n$ is a first-order stationary Markov process, its statistic property is fully characterized by the true bivariate joint distribution of $Y_{t-1}$ and $Y_t$, say $H_0(y_{t-1}, y_t)$. Further suppose that $Y_t$ is continuously distributed. Denote marginal distribution function of $Y_t$ be $F_0(\cdot)$. Then by Sklar's theorem, there exists unique copula function $C(\cdot,\cdot)$ satisfying:

$$H_0(a,b) = C(F_0(a), F_0(b))$$

which holds for all $(a,b) \in \mathbb{R}^2$.

Here the copula function $C(\cdot,\cdot)$ is a bivariate probability distribution function with uniform marginals. Denote the corresponding copula density of $C(u_1,u_2)$ by $c(u_1,u_2)$, and the density of the marginal distribution $F_0(\cdot)$ by $f_0(\cdot)$, then the true conditional density of $V_t$ given $V_{t-1}$ is:

$$p(V_t|V_{t-1}) = f_0(V_t)c(F_0(V_{t-1}), F_0(V_t))$$

Thus, given $\{V_t\}_{t=1}^n$, the log likelihood of the sample is:

$$\frac{1}{n}\sum_{t=1}^{n}\log f_0(V_t) + \frac{1}{n}\sum_{t=2}^{n}\log c(F_0(V_{t-1}), F_0(V_t))$$

For convenience of asymptotic analysis, we assume the following assumptions on the dynamics of the process $\{Y_t\}$.

**Assumption 1.** $\{V_t : t = 1, 2, \cdots, n\}$ *is a sample of a stationary first-order Markov process generated from* $(F_0(\cdot), C(\cdot,\cdot;\alpha_0))$*, where* $F_0(\cdot)$ *is the true invariant distribution that is absolutely continuous with respect to Lebesgue measure on the real line;* $C(\cdot,\cdot;\alpha_0)$ *is the true parametric copula for* $(V_{t-1}, V_t)$ *up to unknown value* $\alpha_0$*, is absolutely continuous with respect to Lebesgue measure on* $[0,1]^2$*.*

*Remark* 1. Assumption of absolute continuity of the bi-variate copula $C(\cdot,\cdot;\alpha_0)$ rules out the Fréchet-Hoeffding upper $(C(u_1,u_2) = \min\{u_1,u_2\})$ and the the lower $(C(u_1,u_2) = \max\{u_1+u_2-1, 0\})$ bounds.

*Remark* 2. This assumption also implies the time series $\{V_t\}_{t=1}^n$ is strictly stationary ergodic, see Chen and Fan [2006].

## 2.2 The Nonstationary Component and Filtration

Concerning on the nonstationary component and the related filtration, we make the following assumptions to facilitate asymptotic analysis.

**Assumption 2.** *There exists a scaling matrix $G_n$ such that $G_n^{-1}X_{[nr]} \Rightarrow X(r)$, $r \in [0,1]$. As $n \to \infty$, there exists a random variable $\xi$ such that $\sqrt{n}G_n(\hat{\beta}_n - \beta_0) \Rightarrow \xi$.*

Due to nonstationarity in $X_t$, we introduced appropriate re-standardization via the scaling matrix $G_n$ to facilitate asymptotic analysis. The limit of the standardized nonstationary component, $X(r)$, may be stochastic or deterministic or a mixture of stochastic and deterministic functions. The limiting distribution, $\xi$, of the filtration parameter is a function of $X(\cdot)$ and may not be a normal variate. Leading cases that are widely used in time series application include the following:

**Example 1.** Deterministic trend.

$X_t$ is a vector of deterministic trend function and $G_n^{-1}X_{[nr]} \Rightarrow X(r)$, where $X(r)$ is a continuous limiting trending function. Let the OLS estimator of $\beta$ be $\hat{\beta}_n$,

$$D_n(\hat{\beta}_n - \beta_0) \Rightarrow \xi_1$$

where in general $\xi_1$ a normal variate.

Then the detrended data is given by $\hat{V}_t = Y_t - X_t'\hat{\beta}_n$. For example, if the observed time series $\{Y_t\}_{t=1}^n$ contains a linear trend:

$$Y_t = \beta_{01} + \beta_{02} \cdot t + V_t$$

In practice, we estimate copula model based on:

$$\hat{V}_t = Y_t - \hat{\beta}_{01} - \hat{\beta}_{02} \cdot t$$

The corresponding standardization matrix is $G_n = \text{diag}(1, n)$, $D_n = \sqrt{n}G_n = \text{diag}(n^{\frac{1}{2}}, n^{\frac{3}{2}})$, $X_t = (1, t)'$ and $X(r) = (1, r)'$. Limiting distribution is:

$$\xi_1 = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \lambda W(1) \\ \lambda \int_0^1 s dW(s) \end{pmatrix}$$

here $W(s)$, $s \in [0,1]$ is the standard Brownian motion, $\omega_V^2 := \mathbb{E}\, V_t^2 + 2\sum_{s=1}^{\infty} \mathbb{E}\, V_t V_{t+s}$ is the long run variance of $V_t$.

**Example 2.** Unit Root Time Series.

$X_t = Y_{t-1}$ and $\beta_0 = 1$. In this case,

$$Y_t = \beta_0 Y_{t-1} + V_t$$

In finite sample, we cannot distinguish a unit root process with a near unit process. We estimate $V_t$ by $\hat{V}_t = X_t - \hat{\beta}_n X_{t-1}$ for robustness. Thus, $X_t$ (or, $Y_t$) is a unit root process, $G_n = \sqrt{n}$, and $G_n^{-1}X_{[nr]} \Rightarrow X(r) = \omega_V W(r)$. Here $W(s)$, $s \in [0,1]$ is the standard Brownian motion, $\omega_V^2 := \mathbb{E}\, V_t^2 + 2\sum_{s=1}^{\infty} \mathbb{E}\, V_t V_{t+s}$ is the long run variance of $V_t$.

The estimator $\hat{\beta}_n$ in transformation converges at rate-$n$ to a non-normal limit, $n(\hat{\beta}_n - \beta_0) \Rightarrow \xi_2$, where

$$\xi_2 = \frac{W^2(1) - \mathbb{E}V_t^2/\omega_V^2}{2\int_0^1 W^2(r)\mathrm{d}r}$$

**Example 3.** Cointegration Time Series.

$X_t$ is a vector of nonstationary unit root process independent of $V_t$, $X_t = X_{t-1} + \varepsilon_t$, $G_n = \sqrt{n}$, $G_n^{-1}X_{[nr]} \Rightarrow X(r) = \omega_\varepsilon W_1(r)$. Here $W_1(s)$, $s \in [0,1]$ is a standard Brownian motion, $\omega_\varepsilon^2 := \mathbb{E}\, \varepsilon_t^2 + 2\sum_{s=1}^{\infty} \mathbb{E}\, \varepsilon_t \varepsilon_{t+s}$ is the long run variance of $\varepsilon_t$.

Then the estimator $\hat{\beta}_n$ is still rate-$n$ converging $n(\hat{\beta}_n - \beta_0) \Rightarrow \xi_3$, where

$$\xi_3 = \frac{\omega_V \int_0^1 W_1(r)\mathrm{d}W_2(r)}{\omega_\varepsilon \int_0^1 W_1^2(r)\mathrm{d}r}$$

here $\omega_V^2 := \mathbb{E}\, V_t^2 + 2\sum_{s=1}^{\infty} \mathbb{E}\, V_t V_{t+s}$ is the long run variance of $V_t$, $W_2(s)$, $s \in [0,1]$ is another standard Brownian motion independent with $W_1(\cdot)$.

### 2.3 The Marginal and The Sieve space

If the marginal distribution of $V_t$, $F_0(\cdot)$, were fully known, under **Assumption** 1, we could estimate the copula model based on maximizing:

$$Q_n(\alpha) = \frac{1}{n}\sum_{t=2}^{n} \log c(F_0(\hat{Y}_{t-1}), F_0(\hat{Y}_t); \alpha)$$

here $\hat{V}_t = Y_t - X_t'\hat{\beta}_n$ is the residual process obtained from filtration to remove nonstationarity.

Denote the solution of this maximization problem by $\breve{\alpha}_n$, under some regularity conditions:

$$\sqrt{n}(\breve{\alpha} - \alpha_0) = H_{n\alpha}^{-1} S_{n\alpha} + o_p(1)$$

where

$$H_{n\alpha} = -\frac{1}{n}\sum_{t=2}^{n} \frac{\partial^2 \log c(F_0(\hat{V}_{t-1}), F_0(\hat{V}_t); \alpha_0)}{\partial\alpha\partial\alpha'} \xrightarrow{p} H_\alpha$$

$$S_{n\alpha} = \frac{1}{\sqrt{n}}\sum_{t=2}^{n} \frac{\partial \log c(F_0(\hat{V}_{t-1}), F_0(\hat{V}_t); \alpha_0)}{\partial\alpha}$$

and

$$H_\alpha = -\,\mathbb{E}\,\frac{\partial^2 \log c(F_0(V_{t-1}), F_0(V_t); \alpha_0)}{\partial\alpha\partial\alpha'}$$

$S_{n\alpha}$ will converge in distribution to a random variable and hence is $O_p(1)$. However, the limiting distribution of $S_{n\alpha}$ will be affected by non-stationarity filtration and usually non-normal.

In practice, true distribution function $F_0(\cdot)$ and its density $f(\cdot)$ are unknown and need to be modeled and estimated appropriately. In this article, we model the unknown symmetric marginal density nonparametrically by approximate the true function with various parametric family of densities with increasing complexity. There exists many sieves for approximating a univariate symmetric probability density function. We will focus on using linear sieves to directly approximate either a square root density:

$$\mathcal{F}_n = \left\{ f(y) = \left[\sum_{k=1}^{K_n} a_k A_k(y)\right]^2, \int_{\mathbb{R}^1} f(y)\mathrm{d}y = 1 \right\} \tag{1}$$

or a log density:

$$\mathcal{F}_n = \left\{ f(y) = \exp\left[\sum_{k=1}^{K_n} a_k A_k(y)\right], \int_{\mathbb{R}^1} f(y)\mathrm{d}y = 1 \right\} \tag{2}$$

*Remark* 3. If we concentrate on symmetric marginal, $A_k(\cdot)$ could be selected to be symmetric, then elements in $\mathcal{F}_n$ are automatically symmetric. We apply this technique in **Section** 6.3.

Before presenting some concrete examples of known sieve basis functions $\{A_k(\cdot) : k \geq 1\}$, we first recall a popular smoothness function class used in the non-parametric estimation literature (see e.g. Stone [1982]; Robinson [1988]). Suppose that the support $\mathcal{Y}$ (of the true density $f_0(\cdot)$) is either a compact interval(say, $[0,1]$) or the whole real line $\mathbb{R}^1$. A real-valued function $h$ on $\mathcal{Y}$ is said to be $r$-smooth if it is bounded continuously differentiable on $\mathcal{Y}$ up to order $[r]$ (i.e. there is a positive number $K$ such that $\max_{s=0,1,\cdots,J} |D^s f(y)| \leq K$ for all $y \in \mathcal{Y}$)[3] and its $[r]$th derivative is Hölder continuous with exponent $\{r\} \equiv r - [r] \in (0,1]$ (i.e. there is a positive number $K$ such that $|D^J h(y_1) - D^J h(y_2)| \leq K|y_1 - y_2|^{\{r\}}$ for all $y_1, y_2 \in \mathcal{Y}$). We denote $\Lambda^r(\mathcal{Y})$ as the class of all real-valued functions on $\mathcal{Y}$ that are $r$-smooth. Define Hölder norm of order $r$ to be:

$$\|h\|_{\Lambda^r} := \max_{s=0,1,\cdots,[r]} \sup_{y\in\mathcal{Y}} |D^s f(y)| + \sup_{y_1,y_2\in\mathcal{Y}, y_1\neq y_2} \frac{|D^r f(y_1) - D^r f(y_2)|}{|y_1 - y_2|^{r-[r]}}$$

[3] $[r]$ is the largest integer strictly smaller than $r$. For example, $[2.5] = 2$, $[2] = 1$, $[0.8] = 0$.

then we can define Hölder space as $\{h \in C^{[r]}(\mathcal{Y}) : \|h\|_{\Lambda^r} < +\infty\}$.

Let the true marginal density function $f_0$ satisfy either $\sqrt{f_0} \in \Lambda^r(\mathcal{Y})$ or $\log f_0 \in \Lambda^r(\mathcal{Y})$. Then any function in $\Lambda^r(y)$ can be approximated by some appropriate sieve spaces. For example, if $\mathcal{Y}$ is a bounded interval and $r > \frac{1}{2}$, it can be approximated by the spline sieve $Spl(s, K_n)$ with $s > [r]$, the polynomial sieve , the trigonometric sieve, the cosine series and etc. When the support of $\mathcal{Y}$ is unbounded, thin-tailed density can be approximated by Hermite polynomial sieve, while polynomial fat-tailed density can be approximated by spline wavelet sieve. See Chen [2007] for detailed descriptions of various sieve spaces $\mathcal{G}_n$.

## 2.4 Copulas

Before analyzing large sample property of our two-step sieve estimator, we first introduce some commonly used copulas and their properties.

Suppose $(U_1^{(1)}, U_2^{(1)})$ and $(U_1^{(2)}, U_2^{(2)})$ are two pairs bivariate uniformly distributed random variables, joint distribution following copula $C(\cdot, \cdot)$. Then the Kendall's tau is defined as the probability of concordance minus the probability of discordance, see Nelson [1999] chapter 5:

$$\begin{aligned}\tau &= \mathbb{P}\left[(U_1^{(1)} - U_1^{(2)})(U_2^{(1)} - U_2^{(2)}) > 0\right] - \mathbb{P}\left[(U_1^{(1)} - U_1^{(2)})(U_2^{(1)} - U_2^{(2)}) < 0\right] \\ &= 4\int_0^1\int_0^1 C(u_1, u_2)c(u_1, u_2)\mathrm{d}u_1\mathrm{d}u_2 - 1\end{aligned}$$

Because Kendall's tau is the difference of two probabilities, we have $-1 \le \tau \le 1$. Positive $\tau$ means positive dependence and negative $\tau$ means negative dependence.

Spearman's rho is another commonly used measure of association based on concordance and discordance. Suppose $(U_1^{(1)}, U_2^{(1)})$, $(U_1^{(2)}, U_2^{(2)})$ and $(U_1^{(3)}, U_2^{(3)})$ are three pairs bivariate uniformly distributed random variables, joint distribution following copula $C(\cdot, \cdot)$. Then the Spearman's rho is defined to be proportional to the probability of concordance minus the probability of discordance for the two vectors $(U_1^{(1)}, U_2^{(1)})$ and $(U_1^{(2)}, U_2^{(3)})$, see Nelson [1999] chapter 5:

$$\begin{aligned}\rho &= 3\left(\mathbb{P}\left[(U_1^{(1)} - U_1^{(2)})(U_2^{(1)} - U_2^{(3)}) > 0\right] - \mathbb{P}\left[(U_1^{(1)} - U_1^{(2)})(U_2^{(1)} - U_2^{(3)}) < 0\right]\right) \\ &= 12\int_0^1\int_0^1 C(u_1, u_2)\mathrm{d}u_1\mathrm{d}u_2 - 3\end{aligned}$$

Spearman's rho is also ranged in $[-1, 1]$, like Kendall's tau. Positive $\rho$ means positive dependence and negative $\rho$ means negative dependence.

Tail dependence measures the dependence between the variables in the upper right quadrant and in the lower left quadrant of $[0, 1]^2$. The lower and upper tail dependence coefficients $\lambda_L$ and $\lambda_U$ in terms of copula are defined as:

$$\lambda_L = \lim_{u \to 0^+} \mathbb{P}(U_2 \le u | U_1 \le u) = \lim_{u \to 0^+} \frac{C(u, u)}{u}$$

$$\lambda_U = \lim_{u \to 1^-} \mathbb{P}(U_2 \ge u | U_1 \ge u) = \lim_{u \to 1^-} \frac{1 + C(u, u) - 2u}{1 - u}$$

Tail dependence is a useful structure to model effect of extreme event in empirical research. See **Section** 7 for detail.

We consider five copulas in this paper, each with four choices of copula parameter:

- Gaussian copula

$$C(u_1, u_2; \alpha_0) = \Phi_\alpha(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$$

$$c(u_1, u_2; \alpha_0) = \frac{\phi_\alpha(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi(\Phi^{-1}(u_1)) \cdot \phi(\Phi^{-1}(u_2))}$$

here $\Phi(\cdot)$ and $\phi(\cdot)$ denote the CDF and PDF of standard normal distribution respectively. $\Phi_\alpha(\cdot)$ and $\phi_\alpha(\cdot)$ denote the CDF and PDF of bivariate normal distribution with correlation $\alpha$ respectively. Range for $\alpha$ is $-1 \leq \alpha \leq 1$.
When $\alpha > 0$, dependence is positive:
$\alpha = 0.9 \Rightarrow \tau = 0.713,\ \rho = 0.891;\ \alpha = 0.5 \Rightarrow \tau = 0.333,\ \rho = 0.483$
When $\alpha < 0$, dependence is negative:
$\alpha = -0.9 \Rightarrow \tau = -0.713,\ \rho = -0.891;\ \alpha = -0.5 \Rightarrow \tau = -0.333,\ \rho = -0.483$
There is no tail dependence for Gaussian copula.

- Frank copula

$$C(u_1, u_2; \alpha) = -\frac{1}{\alpha} \cdot \log\left(1 - \frac{(1 - e^{-\alpha u_1})(1 - e^{-\alpha u_2})}{1 - e^{-\alpha}}\right)$$

$$c(u_1, u_2; \alpha) = \alpha \cdot \frac{e^{-\alpha u_1} e^{-\alpha u_2}}{1 - e^{-\alpha}} \cdot \left(1 - \frac{(1 - e^{-\alpha u_1})(1 - e^{-\alpha u_2})}{1 - e^{-\alpha}}\right)^{-2}$$

here $\alpha \in \mathbb{R}^1$.
When $\alpha > 0$, dependence is positive:
$\alpha = 15 \Rightarrow \tau = 0.7626,\ \rho = 0.9294;\ \alpha = 5 \Rightarrow \tau = 0.4567,\ \rho = 0.6435$.
When $\alpha < 0$, dependence is negative:
$\alpha = -15 \Rightarrow \tau = -0.7626,\ \rho = -0.9294;\ \alpha = -5 \Rightarrow \tau = -0.4567,\ \rho = -0.6435$
There is no tail dependence for Frank copula.

- Clayton copula

$$C(u_1, u_2; \alpha) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}}$$

$$c(u_1, u_2; \alpha) = (1 + \alpha) \cdot u_1^{-\alpha-1} \cdot u_2^{-\alpha-1} \cdot (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}-2}$$

here $\alpha$ is positive.
Clayton copula has Kendall's tau $\tau = \frac{\alpha}{2+\alpha}$ and lower tail dependence coefficient $\lambda_L = 2^{-1/\alpha}$ that is increasing in $\alpha$, but no upper tail dependence.
When $\alpha = 2$, $\tau = 0.5$, $\rho = 0.682$, $\lambda_L = 0.7071$.
When $\alpha = 5$, $\tau = 0.7143$, $\rho = 0.885$, $\lambda_L = 0.871$.
When $\alpha = 10$, $\tau = 0.833$, $\rho = 0.958$, $\lambda_L = 0.933$.
When $\alpha = 12$, $\tau = 0.857$, $\rho = 0.969$, $\lambda_L = 0.944$.

- EFGM copula

$$C(u_1, u_2; \alpha) = u_1 u_2 [1 + \alpha(1 - u_1)(1 - u_2)]$$

$$c(u_1, u_2; \alpha) = 1 + \alpha(1 - 2u_1)(1 - 2u_2)$$

here range for $\alpha$ is $-1 \leq \alpha \leq 1$.
When $\alpha > 0$, dependence is positive:
$\alpha = 0.9 \Rightarrow \tau = 0.2,\ \rho = 0.3;\ \alpha = 0.5 \Rightarrow \tau = 0.111,\ \rho = 0.167$
When $\alpha < 0$, dependence is negative:
$\alpha = -0.9 \Rightarrow \tau = -0.2,\ \rho = -0.3;\ \alpha = -0.5 \Rightarrow \tau = -0.111,\ \rho = -0.167$
There is no tail dependence for EFGM copula.

- Gumbel copula

$$C(u_1, u_2; \alpha) = \exp\left[-\left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}}\right]$$

$$\begin{aligned} c(u_1, u_2; \alpha) &= \exp\left[-\left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}}\right] \cdot \left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}-2} \\ &\quad \cdot (-\log u_1)^{\alpha-1} \cdot (-\log u_2)^{\alpha-1} \cdot \frac{1}{u_1 u_2} \cdot \\ &\quad \left[\left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}} + (\alpha - 1)\right] \end{aligned}$$

here $\alpha$ needs to be larger than 1.
Gumbel copula has Kendall's tau $\tau = 1 - \frac{1}{\alpha}$ and upper tail dependence coefficient $\lambda_U = 2 - 2^{1/\alpha}$ that is increasing in $\alpha$, but no lower tail dependence.
When $\alpha = 2$, $\tau = 0.5$, $\rho = 0.682$, $\lambda_U = 0.586$.
When $\alpha = 3.5$, $\tau = 0.7143$, $\rho = 0.887$, $\lambda_U = 0.781$.
When $\alpha = 6$, $\tau = 0.833$, $\rho = 0.96$, $\lambda_U = 0.8775$.
When $\alpha = 7$, $\tau = 0.857$, $\rho = 0.971$, $\lambda_U = 0.896$.

### 2.5 The Semiparametric Estimators

There are two alternative procedures in estimation of copula-based models based on filtering residues using sieve methods: the two-step sieve estimation and the three-step sieve estimation procedures. For three-step estimation procedure, we first get OLS residuals $\hat{V}_t = Y_t - X_t'\hat{\beta}$. During second step, marginal densities is estimated by sieve method. Copula parameter is estimated in last step through MLE.

First step: get OLS residuals

$$\hat{V}_t = Y_t - X_t'\hat{\beta}$$

Second step:get density estimator $\hat{f}$ through

$$\max_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{t=1}^{n} \log f(\hat{V}_t)$$

Second step: get copula estimator $\hat{\alpha}$ through

$$\max_{\alpha \in \mathcal{A}} \frac{1}{n} \sum_{t=2}^{n} \log c \left( \int_{-\infty}^{\hat{V}_{t-1}} \hat{f}(y) \mathrm{d}y, \int_{-\infty}^{\hat{V}_t} \hat{f}(y) \mathrm{d}y; \alpha \right)$$

The two-step sieve method estimates the marginal and copula parameters simultaneously in the second step. Get density estimator $\hat{f}$ and copula estimator $\hat{\alpha}$ together through

$$\max_{\alpha \in \mathcal{A}, f \in \mathcal{F}_n} \frac{1}{n} \sum_{t=1}^{n} \log f(\hat{V}_t) + \frac{1}{n} \sum_{t=2}^{n} \log c \left( \int_{-\infty}^{\hat{V}_{t-1}} f(y) \mathrm{d}y, \int_{-\infty}^{\hat{V}_t} f(y) \mathrm{d}y; \alpha \right)$$

The two-step estimation procedure is usually more efficient as copula parameter and marginal density are estimated at the same time, while the computation complexity is not increasing too much comparing to three-step sieve estimator. For this reason, we focus our discussion on the two-step estimation procedures in this paper.

All the above two estimators are based on filtered residuals. There is also other possible ways not depending on linear filtering. For example, in finite sample simulation part we consider an estimator where we do not use OLS regression to estimate $\hat{\beta}$ at first step but optimize all estimators all together[4]. See **Section** 6.2.1 for detail.

## 3 Summary of Main Results

We summarize in this section the main results of the paper.

By definition, the two-step sieve estimator $(\hat{\alpha}, \hat{f})$ maximize the following criterion:

$$\max_{\alpha \in \mathcal{A}, f \in \mathcal{F}_n} \frac{1}{n} \sum_{t=1}^{n} \log f(\hat{V}_t) + \frac{1}{n} \sum_{t=2}^{n} \log c \left( \int_{-\infty}^{\hat{V}_{t-1}} f(y) \mathrm{d}y, \int_{-\infty}^{\hat{V}_t} f(y) \mathrm{d}y; \alpha \right)$$

[4]This estimator will be referred as joint sieve estimator in the following

Under appropriate regularity conditions (see **Section** 4), we first establish the consistency result by showing that(**Theorem** 1): as $n \to \infty$

$$\|\hat{\alpha} - \alpha_0\|_2 + \|\hat{f} - f_0\|_c \xrightarrow{p} 0$$

for some norm $\|\cdot\|_c$ on the functional space $\mathcal{F}$. Here $\|\cdot\|_2$ refers to Euclidean norm, $\|\alpha\|_2 := \alpha^T \alpha$. Specifically, we will have consistency of the copula parameter: $\hat{\alpha} \xrightarrow{p} \alpha_0$.

Then we derive the limiting distribution of $\hat{\alpha}$ (see **Section** 5). It still has the usual root-n convergence rate but its limiting distribution may be not normally distributed(**Theorem** 2):

$$\sqrt{n}(\hat{\alpha}_n - \alpha_0) \Rightarrow -F(\gamma_0, v^*) \times \int_0^1 X(r)dr \times \xi + N(0, \|v^*\|^2)$$

$X(r)$ is the limiting process of $X_{[nr]}$ and $\xi$ is the limiting distribution. These two terms depend on nonstationary structure.

$v^*$ is an abstract term determined by copula and marginal structure, which is defined in **Section** 5. As we cannot directly make derivative in a function space, like the usual performance in an Euclidean space, a directional derivative based on $v^*$ is developed. The direction $v^*$ is set up in Hilbert space following Riesz representation theorem and does not have analytical expression. The second term is the normal limiting distribution Chen et al. [2009] reaches when $V_t$ is directly observed and there is no filtration.

Expression for $F(\gamma_0, v^*)$ is complicate:

$$\mathbb{E}\frac{\partial^2 l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma\partial y_1}[v^*] + \mathbb{E}\frac{\partial^2 l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma\partial y_2}[v^*]$$

Notice $F(\gamma_0, v^*)$ is a constant determined solely by copula and marginals and is not relevant with the nonstationary structure. If $F(\gamma_0, v^*) \equiv 0$, nonstationary structure of $X_t$ will not affect limiting distribution of $\hat{\alpha}$, a general property holding by three-step empirical estimator proved by Chen and Xiao [2016].

However, for two-step sieve estimator, we show that this property holds only when both copula and marginal are symmetric. Gaussian, Frank and EFGM copulas satisfy the symmetry property. See **Corollary** 1 for formal description.

When copula is asymmetric, such as Clayton copula (with lower tail dependence) and Gumbel copula (with upper tail dependence), we show that $|F(\gamma_0, v^*)|$ is nonzero and strictly increasing with strength of tail dependence. These theoretical results are also corroborated by finite sample simulation.

For finite sample simulation, two-step sieve estimator performs quite well, comparing to three-step empirical method and mis-specified parametric method. However, for tail dependence copula, especially when tail dependence is strong, both two-step sieve method and three-step empirical method do not perform well. Hence joint sieve estimator is analyzed and its finite simulation results are pretty good even under strong tail dependence. We find this phenomenon is due to the bad estimation of first step estimation of filtration for residuals based methods under strong tail dependence.

In empirical application, we analyze the model:

$$Y_t = X_t\beta_0 + V_t$$

where $Y_t$ is weekly stock price and $X_t$ is consensus target price. Residual term $V_t$ is modeled as first-order Markov process with parametric copula structure. Using Gaussian copula(no tail dependence), our sieve methods have better out-of-sample predicting power than three-step empirical methods. Using Clayton copula(lower tail dependence), three-step empirical method results in an evidently higher tail dependence estimation.

# 4 Consistency

In this section, we will establish consistency result for two-step sieve estimator.

In the following, we denote:

$$Q_n(\alpha, f) := \frac{1}{n}\sum_{t=1}^{n}\log f(\hat{V}_t) + \frac{1}{n}\sum_{t=2}^{n}\log c\left(\int_{-\infty}^{\hat{V}_{t-1}} f(y)\mathrm{d}y, \int_{-\infty}^{\hat{V}_t} f(y)\mathrm{d}y; \alpha\right)$$

and

$$Q(\alpha, f) := \mathbb{E}\log f(V_t) + \mathbb{E}\log c\left(\int_{-\infty}^{V_{t-1}} f(y)\mathrm{d}y, \int_{-\infty}^{V_t} f(y)\mathrm{d}y; \alpha\right)$$

here $\mathbb{E}$ is the expectation under the true parameter $(\alpha_0, f_0)$ (i.e. **Assumption** 1).

Denote $\gamma = (\alpha, f)$, $\gamma_0 = (\alpha_0, f_0)$, $\hat{\gamma} = (\hat{\alpha}, \hat{f})$, $\Gamma = \mathcal{A}\times\mathcal{F}$, $\Gamma_n = \mathcal{A}\times\mathcal{F}_n$. Let $(\hat{\alpha}, \hat{f})$ be our two-step sieve estimator that maximize the following criterion:

$$\max_{\alpha\in\mathcal{A}, f\in\mathcal{F}_n} Q_n(\alpha, f) := \frac{1}{n}\sum_{t=1}^{n}\log f(\hat{V}_t) + \frac{1}{n}\sum_{t=2}^{n}\log c\left(\int_{-\infty}^{\hat{V}_{t-1}} f(y)\mathrm{d}y, \int_{-\infty}^{\hat{V}_t} f(y)\mathrm{d}y; \alpha\right) \tag{3}$$

For function space $\mathcal{F}$ and $\mathcal{F}_n$, define a norm $\|\cdot\|_c$. We could either take sup norm $\|f\|_\infty$ or a lower order Hölder norm $\|f\|_{\Lambda^{r'}}$ for some $0 < r' < r$. See Chen et al. [2006]. Norm $\|\cdot\|_c$ on $\mathcal{F}$ can induce a natural Cartesian extension on $\Gamma$ and $\Gamma_n$, a new norm defined as: $\|\gamma\|_c := \|\alpha\|_2 + \|f\|_c$. Here $\|\cdot\|_2$ is norm on Euclidean space, $\|\alpha\|_2 := \sqrt{\alpha^\top\alpha}$.

We make the following assumptions to establish consistency.

**Assumption 3.** *$\alpha_0 \in \mathcal{A}$, where $\mathcal{A}$ is a compact subset of $\mathbb{R}^1$ with nonempty interior, $c(u_1, u_2; \alpha) > 0$ for all $(u_1, u_2) \in (0,1)\times(0,1)$, $\alpha \in \mathcal{A}$.*

**Assumption 4.** *$f_0 \in \mathcal{F}$, either $\mathcal{F} = \{f > 0$ on $\mathcal{Y} : \sqrt{f} \in \Lambda^r(\mathcal{Y}), \int_{\mathcal{Y}} f(y)\mathrm{d}y = 1\}$ and $\mathcal{F}_n$ given in equation (1), or $\mathcal{F} = \{f > 0$ on $\mathcal{Y} : \log f \in \Lambda^r(\mathcal{Y}), \int_{\mathcal{Y}} f(y)\mathrm{d}y = 1\}$ and $\mathcal{F}_n$ given in equation (2). $r > \frac{1}{2}$.*

*Remark* 4. When symmetry restriction is added to to the density, we can define the space $\mathcal{F}$ as either either $\mathcal{F} = \{f > 0$ on $\mathcal{Y} : \sqrt{f} \in \Lambda^r(\mathcal{Y}), \int_{\mathcal{Y}} f(y)\mathrm{d}y = 1, f(y) = f(-y)\}$ and $\mathcal{F}_n$ given in equation (1), or $\mathcal{F} = \{f > 0$ on $\mathcal{Y} : \log f \in \Lambda^r(\mathcal{Y}), \int_{\mathcal{Y}} f(y)\mathrm{d}y = 1, f(y) = f(-y)\}$ and $\mathcal{F}_n$ given in equation (2). And sieve basis $A_n(\cdot)$ in $\mathcal{F}_n$ are also selected to be symmetric.

**Assumption 5.** *$Q(\alpha_0, f_0) > -\infty$, there exists a positive measurable function $\eta(\cdot)$ such that $\forall\varepsilon > 0$, $\forall n \geq 1$,*

$$Q(\alpha_0, f_0) - \sup_{\alpha\in\mathcal{A}, f\in\mathcal{F}_n} Q(\alpha, f) \geq \eta(\varepsilon) > 0$$

**Assumption 6.** *For all $n \geq 1$, the sieve space $\mathcal{F}_n$ is compact under norm $\|\cdot\|_c$.*

**Assumption 7.** *For all $n \geq 1$, there exists $\Pi_n f_0 \in \mathcal{F}_n$ such that $Q(\alpha_0, \Pi_n f_0) - Q(\alpha_0, f_0) = o(1)$.*

*Remark* 5. **Assumption** 3 is a standard regularity condition. **Assumption** (4) ensures that the true density functional space could be well approximated by its sieve counterpart. **Assumption** 5 is the standard identification condition. **Assumption** 6 ensures the feasibility of optimization on compact spaces. **Assumption** 7 ensures smoothness around true density $f_0$ and its sieve approximation $\Pi_n f_0$, see Chen et al. [2009] assumptions 3.1.

Uniform law of large numbers is crucial for proof of consistency, which ensures us to replace $Q_n(\cdot)$ with $Q(\cdot)$ and then apply properties of $Q(\cdot)$. Rather than expression in terms of filtering residuals $\hat{V}_t = V_t + X_t'(\beta_0 - \hat{\beta}_n)$, it is often easier to verify the statement in terms of transformed variables, directly in stationary process $V_t$, without explicitly reference of nonstationarity structure of $X_t$. Let:

$$L_n(\gamma, \mathsf{b}) := \frac{1}{n}\sum_{t=1}^{n} \log f(V_t + \frac{\mathsf{b}_t}{\sqrt{n}}) + \frac{1}{n}\sum_{t=2}^{n} \log c\left(\int_{-\infty}^{V_{t-1}+\frac{\mathsf{b}_{t-1}}{\sqrt{n}}} f(y)\mathrm{d}y, \int_{-\infty}^{V_t+\frac{\mathsf{b}_t}{\sqrt{n}}} f(y)\mathrm{d}y; \alpha\right) \quad (4)$$

Define $\|\cdot\|_1$ be sup norm on Euclidean space. For $\mathsf{b} = (\mathsf{b}_1, \mathsf{b}_2, \cdots, \mathsf{b}_n)^\top \in \mathbb{R}^n$, $\|\mathsf{b}\|_1 := \max_{1\le t\le n} |\mathsf{b}_t|$.

**Assumption 8.** *For all $B > 0$,* $\sup_{\substack{\mathsf{b}\in\mathbb{R}^n, \|\mathsf{b}\|_1\le B \\ \gamma\in\Gamma_n}} |L_n(\gamma, \mathsf{b}) - Q(\beta, f)| = o_p(1)$.

**Lemma 1.** *Under **Assumption** 2 and **Assumption** 8, we can derive uniform law of large numbers:* $\sup_{\alpha\in\mathcal{A}, f\in\mathcal{F}_n} |Q_n(\alpha, f) - Q(\alpha, f)| = o_p(1)$.

*Remark* 6. Verification of **Assumption** 8 needs two steps. Notice:

$$\sup_{\substack{\mathsf{b}\in\mathbb{R}^n, \|\mathsf{b}\|_1\le B \\ \gamma\in\Gamma_n}} |L_n(\gamma, \mathsf{b}) - Q(\alpha, f)| \le \sup_{\substack{\mathsf{b}\in\mathbb{R}^n, \|\mathsf{b}\|_1\le B \\ \gamma\in\Gamma_n}} |L_n(\gamma, \mathsf{b}) - L_n(\gamma, 0)| + \sup_{\gamma\in\Gamma_n} |L_n(\gamma, 0) - Q(\alpha, f)|$$

The first term $\sup_{\substack{\mathsf{b}\in\mathbb{R}^n, \|\mathsf{b}\|_1\le B \\ \gamma\in\Gamma_n}} |L_n(\gamma, \mathsf{b}) - L_n(\gamma, 0)|$ is $o_p(1)$ if $\log f$ and $\log c$ are continuous uniformly over $\gamma \in \Gamma_n$, which is a mild condition. For the second term $\sup_{\gamma\in\Gamma_n} |L_n(\gamma, 0) - Q(\alpha, f)|$ to be $o_p(1)$, strictly stationary ergodicity derived from **Assumption** 1 implies that Glivenko Cantelli theorem for stationary ergodic processes is applicable, see Chen et al. [2009] proof of proposition 3.1.

**Theorem 1.** *Under **Assumption** 1-8, $\|\hat{\gamma}_n - \gamma_0\|_c = o_p(1)$. In praticular, we have consistency of copula parameter $\hat{\alpha}_n \xrightarrow{p} \alpha_0$.*

# 5 Limiting distribution

In this section, we first establish root-$n$ convergence rate for sieve copula estimator $\hat{\alpha}_n$.

**Assumption 9.** $\alpha_0 \in \mathrm{int}(\mathcal{A})$

**Assumption 10.** *There exists a neighborhood $\mathcal{N}_0$ of $\gamma_0 = (\alpha_0, f_0)$ such that the following second-order partial derivatives are all well-defined and continuous in $\mathcal{N}_0$:* $\frac{\partial^2 \log c(u_1,u_2;\alpha)}{\partial\alpha\partial\alpha'}$, $\frac{\partial^2 \log c(u_1,u_2;\alpha)}{\partial u_j\partial\alpha}$, $\frac{\partial^2 \log c(u_1,u_2;\alpha)}{\partial u_j\partial u_k}$ *for $j, k = 1, 2$.*

Let $l(\gamma, y_1, y_2) = \log f(y_2) + \log c(\int_{-\infty}^{y_1} f(y)\mathrm{d}y, \int_{-\infty}^{y_2} f(y)\mathrm{d}y; \alpha)$. Denote $\mathbf{V}$ as the linear span of $\Gamma - \{\gamma_0\}$. Under **Assumption** 10, for any $v = (v_\alpha, v_f)' \in \mathbf{V}$, we have that $l(\gamma_0 + \eta v, y_1, y_2)$ is continuously differential in $\eta \in [0, 1]$. For any $\gamma \in \mathcal{N}_0$, define the first-order directional derivative of $l(\gamma, y_1, y_2)$ at the direction $v \in \mathbf{V}$ as:

$$\frac{\partial l(\gamma, y_1, y_2)}{\partial\gamma'}[v] := \left.\frac{\mathrm{d}l(\gamma + \eta v, y_1, y_2)}{\mathrm{d}\eta}\right|_{\eta=0} \quad (5)$$

and the second-order directional derivative as:

$$\begin{aligned}\frac{\partial^2 l(\gamma, y_1, y_2)}{\partial\gamma\partial\gamma'}[v, \tilde{v}] \quad &:= \quad \frac{\mathrm{d}}{\mathrm{d}\tilde{\eta}}\left.\left\{\frac{\partial l(\gamma + \tilde{\eta}\tilde{v}, y_1, y_2)}{\partial\gamma'}[v]\right\}\right|_{\tilde{\eta}=0} \\ &= \quad \left.\left.\frac{\mathrm{d}^2 l(\gamma + \eta v + \tilde{\eta}\tilde{v}, y_1, y_2)}{\mathrm{d}\tilde{\eta}\mathrm{d}\eta}\right|_{\eta=0}\right|_{\tilde{\eta}=0}\end{aligned}$$

We also define cross second-order directional derivative as:

$$\frac{\partial^2 l(\gamma, y_1, y_2)}{\partial\gamma\partial y_1}[v] := \frac{\mathrm{d}}{\mathrm{d}y_1}\left\{\frac{\partial l(\gamma, y_1, y_2)}{\partial\gamma'}[v]\right\}$$
$$\frac{\partial^2 l(\gamma, y_1, y_2)}{\partial\gamma\partial y_2}[v] := \frac{\mathrm{d}}{\mathrm{d}y_2}\left\{\frac{\partial l(\gamma, y_1, y_2)}{\partial\gamma'}[v]\right\}$$

**Assumption 11.** $0 < \mathbb{E}[(\frac{\partial l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma'}[v])^2] < \infty$ *for all* $v \neq 0$, $v \in \mathbf{V}$

**Assumption 12.** $\int \sup_{\eta\in S_v} \left|\frac{dh(y|V_{t-1};\gamma_0+\eta v)}{d\eta}\right| \mathrm{d}y < \infty$ *and* $\int \sup_{\eta\in S_v} \left|\frac{d^2h(y|V_{t-1};\gamma_0+\eta v)}{d\eta^2}\right| \mathrm{d}y < \infty$ *almost surely, for* $S_v := \{\eta \in [0,1] : \gamma_0 + \eta v \in \mathcal{N}_0\}$, $v \neq 0$, $v \in \mathbf{V}$.

Here $h(\cdot|y_{t-1};\gamma)$ is the density of $V_t$ conditional on $V_{t-1}$.

*Remark* 7. Following Chen et al. [2009], **Assumption** 9, 10, 11 and 12 are sufficient to establish the Fisher inner product on the space $\mathbf{V}$ as:

$$< v, \tilde{v} >:= \mathbb{E}\left[\left(\frac{\partial l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma'}[v]\right)\left(\frac{\partial l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma'}[\tilde{v}]\right)\right]$$

and the Fisher norm for $v \in \mathbf{V}$ as $\|v\|^2 :=< v, v >$. Let $\overline{\mathbf{V}}$ be the closed span of $\mathbf{V}$ under the Fisher norm. Then $(\overline{\mathbf{V}}, \|\cdot\|)$ is a Hilbert space.

Define functional $\rho : \Gamma \to \mathbb{R}^1$ as $\rho(\gamma) := \lambda^T\alpha$. We want to analyze the limiting distribution of $\sqrt{n}\left(\rho(\hat{\gamma}_n) - \rho(\gamma_0)\right) = \sqrt{n}\left(\lambda^T\hat{\alpha}_n - \lambda^T\alpha_0\right)$.

*Remark* 8. For simplicity of expression, we mainly consider one dimensional parameter $\alpha_0 \in \mathbb{R}^1$. Hence $\lambda = 1$ and $\rho(\gamma) := \alpha$, natural projection for $\gamma = (\alpha, f)$ to its first component $\alpha$. For copula parameter $\alpha_0$ of multiple dimension, the analysis is similar. All five copulas we consider in this paper (see **Section** 2.4) belong to one parameter families copula.

**Assumption 13.** $\int_{-\infty}^{+\infty} \frac{\partial c(u_1,u_2;\alpha)}{\partial u_1}\mathrm{d}u_2 = \frac{\partial}{\partial u_1}\int_{-\infty}^{+\infty} c(u_1, u_2; \alpha_0)\mathrm{d}u_2 = 0$ *and* $\int_{-\infty}^{+\infty} \frac{\partial c(u_1,u_2;\alpha)}{\partial u_2}\mathrm{d}u_1 = \frac{\partial}{\partial u_2}\int_{-\infty}^{+\infty} c(u_1, u_2; \alpha)\mathrm{d}u_1 = 0.$

**Assumption 14.** $\mathbb{E}\left(\frac{\partial \log c(U_{t-1},U_t;\alpha_0)}{\partial\beta} \cdot \frac{\partial \log c(U_{t-1},U_t;\alpha_0)}{\partial\beta'}\right)$ *is finite and positive definite. Here* $U_{t-1} = F_0(V_{t-1}) = \int_{-\infty}^{V_{t-1}} f_0(y)\mathrm{d}y$, $U_t = F_0(V_t) = \int_{-\infty}^{V_t} f_0(y)\mathrm{d}y$.

**Assumption 15.** $\int_{-\infty}^{+\infty} \frac{\partial^2 c(u_1,u_2;\alpha_0)}{\partial u_1\partial\alpha}\mathrm{d}u_2 = \frac{\partial^2}{\partial u_1\partial\alpha}\int_{-\infty}^{+\infty} c(u_1, u_2; \alpha_0)\mathrm{d}u_2 = 0$ *and* $\int_{-\infty}^{+\infty} \frac{\partial^2 c(u_1,u_2;\alpha_0)}{\partial u_2\partial\alpha}\mathrm{d}u_1 = \frac{\partial^2}{\partial u_2\partial\alpha}\int_{-\infty}^{+\infty} c(u_1, u_2; \alpha_0)\mathrm{d}u_1 = 0.$

**Assumption 16.** *There exists a positive constant* $K$ *such that*

$$\max_{j=1,2} \sup_{0<u_j<1} \mathbb{E}\left[\left(u_j(1-u_j)\frac{\partial \log c(U_1, U_2; \alpha_0)}{\partial u_j}\right)^2 \middle| U_j = u_j\right] \leq K$$

Following Chen et al. [2009], from **Assumption** 13, 14, 15 and 16, we can apply Riesz representation theorem on operator $\rho$: there exists a $v^* \in \overline{\mathbf{V}}$ such that

$$\frac{\partial\rho(\gamma_0)}{\partial\gamma'}[v] = v_\alpha =< v^*, v >, \ \forall v = (v_\alpha, v_f) \in \overline{\mathbf{V}}$$

$$\|v^*\|^2 = \left\|\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\right\|^2 = \sup_{v\in\overline{\mathbf{V}}:\|v\|>0} \frac{\left|\frac{\partial\rho(\gamma_0)}{\partial\gamma'}[v]\right|^2}{\|v\|^2} = \sup_{v\in\overline{\mathbf{V}}:\|v\|>0} \frac{{v_\alpha}^2}{\|v\|^2} < +\infty$$

**Assumption 17.** $\|\hat{\gamma}_n - \gamma_0\| = O_p(\delta_n)$ *for a decreasing sequence* $\delta_n$ *satisfying* $\delta_n \to 0$ *when* $n \to \infty$.

**Assumption 18.** *There exists* $\Pi_n v^* \in \Gamma_n - \{\gamma_0\}$ *such that* $\delta_n \times \|\Pi_n v^* - v^*\| = o(1/\sqrt{n})$

**Condition 1.** For all $\tilde{\gamma}_n \in \mathcal{N}_0 \cap \Gamma_n$ with $\|\tilde{\gamma}_n - \gamma_0\| = O(\delta_n)$ and all $v_n = (v_\alpha, v_f) \in \overline{\mathbf{V}}$ with $\|v_n\| = O(\delta_n)$, we have:

$$\mathbb{E}\left(\frac{\partial^2 l(\tilde{\gamma}, V_{t-1}, V_t)}{\partial\gamma\partial\gamma'}[v,v] - \frac{\partial^2 l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma\partial\gamma'}[v,v]\right) = o(1/n)$$

**Condition 2.** $\left\{\frac{\partial l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma'}[\Pi_n v^*] : \gamma \in \mathcal{N}_0, \|\gamma - \gamma_0\| = O(\delta_n)\right\}$ is a Donsker class.

**Assumption** 17, 18, together with **Condition** 1 and 2, are conditions assumed in Chen et al. [2009] for root-$n$ convergence rate of copula parameter when there is no filtering process. To deal with the non-linear filtering, we need stronger versions of **Condition** 1 and **Condition** 2.

**Assumption 19.** *There exists a positive sequence* $\tilde{\varepsilon}_n = o(1)$ *such that, for all sequence* $\tilde{\gamma}_n$ *with* $\tilde{\gamma}_n \in \mathcal{N}_0 \cap \Gamma_n$, $\|\tilde{\gamma}_n - \gamma_0\| = O(\delta_n)$ *and all* $v_n = (v_\alpha, v_f) \in \overline{\mathbf{V}}$ *with* $\|v_n\| = O(\delta_n)$, *we have:*

$$\mathbb{E}\left(\frac{\partial^2 l(\tilde{\gamma}_n, V_{t-1}, V_t)}{\partial\gamma\partial\gamma'}[v,v] - \frac{\partial^2 l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma\partial\gamma'}[v,v]\right) = \tilde{\varepsilon}_n \cdot o(n^{-1})$$

*Remark* 9. **Assumption** 19 is stronger than **Assumption** 1. We introduce $\tilde{\varepsilon}_n$ here to serve as a gap between this term and $o_p(n^{-1})$, which will be used as a directional step size for proof of convergence rate of copula estimator, see **Section** (9.3).

**Assumption 20.**

$$\sup_{\|\mathfrak{b}\|=O(1/\sqrt{n})} \sup_{\|\gamma-\gamma_0\|=O(\delta_n)} \mu_n\left[\frac{1}{n}\sum_{t=2}^{n}\frac{\partial l(\gamma, V_{t-1}+\mathfrak{b}_{t-1}, V_t+\mathfrak{b}_t)}{\partial\gamma'}[\Pi_n v^*]\right] - \mu_n\left[\frac{\partial l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma'}[\Pi_n v^*]\right] = o_p(1/\sqrt{n})$$

*Remark* 10. **Assumption** 20 is stronger than **Condition** 2. Setting $\mathfrak{b}$ to be the zero vector, **Assumption** 20 is simplified to the stochastic equicontinuity property implied by Donsker class in **Condition** 2. We allow $V_t$ to move around in an order $\frac{1}{\sqrt{n}}$ neighborhood. Here $\mu_n$ denote the corresponding empirical process. The expectation is based on randomness of $V_{t-1}$ and $V_t$. For example:

$$\mu_n\left[\frac{1}{n}\sum_{t=2}^{n}\frac{\partial l(\gamma, V_{t-1}+\mathfrak{b}_{t-1}, V_t+\mathfrak{b}_t)}{\partial\gamma'}[\Pi_n v^*]\right] = \frac{1}{n}\sum_{t=2}^{n}\frac{\partial l(\gamma, V_{t-1}+\mathfrak{b}_{t-1}, V_t+\mathfrak{b}_t)}{\partial\gamma'}[\Pi_n v^*] - \frac{1}{n}\sum_{t=2}^{n}\mathbb{E}\frac{\partial l(\gamma, V_{t-1}+\mathfrak{b}_{t-1}, V_t+\mathfrak{b}_t)}{\partial\gamma'}[\Pi_n v^*]$$

$$\mu_n\left[\frac{1}{n}\sum_{t=2}^{n}\frac{\partial l(\gamma, V_{t-1}, V_t)}{\partial\gamma'}[\Pi_n v^*]\right] = \frac{1}{n}\sum_{t=2}^{n}\frac{\partial l(\gamma, V_{t-1}, V_t)}{\partial\gamma'}[\Pi_n v^*] - \frac{1}{n}\sum_{t=2}^{n}\mathbb{E}\frac{\partial l(\gamma, V_{t-1}, V_t)}{\partial\gamma'}[\Pi_n v^*]$$

**Assumption 21.** *For all sequence* $\tilde{\gamma}_n$ *with* $\tilde{\gamma}_n \in \mathcal{N}_0 \cap \Gamma_n$, $\|\tilde{\gamma}_n - \gamma_0\| = O(\delta_n)$ *and all* $\mathfrak{b} \in \mathbb{R}^n$, $\|\mathfrak{b}\|_1 = O(1/\sqrt{n})$, *we have*

$$\left|\mathbb{E}\left(\frac{\partial^2 l(\gamma, V_{t-1}+\mathfrak{b}_{t-1}, V_t+\mathfrak{b}_t)}{\partial\gamma'\partial y_1}[\Pi_n v^*] - \frac{\partial^2 l(\gamma, V_{t-1}, V_t)}{\partial\gamma'\partial y_1}[\Pi_n v^*]\right)\right| = o(1)$$

$$\left|\mathbb{E}\left(\frac{\partial^2 l(\gamma, V_{t-1}+\mathfrak{b}_{t-1}, V_t+\mathfrak{b}_t)}{\partial\gamma'\partial y_2}[\Pi_n v^*] - \frac{\partial^2 l(\gamma, V_{t-1}, V_t)}{\partial\gamma'\partial y_2}[\Pi_n v^*]\right)\right| = o(1)$$

*Remark* 11. **Assumption** 21 is analogous to **Condition** 1. We need assumption to demonstrate continuity for second order derivative. Notice we only need order $o(1)$ here, unlike $o(n^{-1})$ in **Condition** 1.

**Theorem 2.** *Under **Assumption** 1-21, we have asymptotic distribution for $\hat{\alpha}_n$:*

$$\sqrt{n}(\hat{\alpha}_n - \alpha_0) \Rightarrow -F(\gamma_0, v^*) \times \int_0^1 X(r)dr \times \xi + N(0, \|v^*\|^2)$$

*here $X(r)$ and $\xi$ are limiting distribution of $G_n^{-1}X_{[nr]}$ and $\sqrt{n}G_n(\hat{\beta}_n - \beta_0)$, respectively. See **Assumption** 2.*

*$F(\gamma_0, v^*)$ is $\mathbb{E}\frac{\partial^2 l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma\partial y_1}[v^*] + \mathbb{E}\frac{\partial^2 l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma\partial y_2}[v^*]$. The detailed expression refers to proof in **Section** 9.3.*

*Remark* 12. **Theorem** 2 shows that our copula estimator achieves root-$n$ convergence rate. We can split the limiting distribution into two parts. The second normal distribution term $N(0, \|v^*\|^2)$ is the limiting distribution when there is no filtering. The first term is due to nonlinear filtering. $\int_0^1 X(r)dr$ and $\xi$ compress all information from the nonstationarity. $F(\gamma_0, v^*)$ is a constant irrelevant with filtering process, fully characterized by the structure of Markov process $V_t$.

Unless $F(\gamma_0, v^*)$ is exactly zero, the limiting distribution will be generally non-normal due to non-normality of $\int_0^1 X(r)dr \times \xi$. We will simulate constant $F(\gamma_0, v^*)$ and asymptotic variance $\|v^*\|^2$ in **Section** (6.3). In the following derivation, we will show that $F(\gamma_0, v^*)$ is exactly zero for Gaussian, Frank and EFGM copula.

We call a copula being symmetric if it is symmetric around $(1/2, 1/2)$:

$$c(u_1, u_2; \alpha) \equiv c(1 - u_1, 1 - u_2; \alpha)$$

Notice if a copula only has lower tail dependence or only has upper tail dependence(e.g. Clayton or Gumbel), then it is not symmetric around $(1/2, 1/2)$ as symmetric copula must have the same tail dependence on both sides:

$$\begin{aligned}
\lambda_L &= \lim_{u \to 0^+} \mathbb{P}(U_2 \leq u | U_1 \leq u) \\
&= \lim_{u \to 0^+} \mathbb{P}(1 - U_2 \geq 1 - u | 1 - U_1 \geq 1 - u) \\
&= \lim_{u \to 1^-} \mathbb{P}(U_2 \geq u | U_1 \geq u) = \lambda_U
\end{aligned}$$

However, for Gaussian, Frank and EFGM copula, where $\lambda_L = \lambda_U = 0$, symmetric property is satisfied.

For Gaussian copula:

$$\begin{aligned}
& c(1 - u_1, 1 - u_2; \alpha) \\
&= \frac{\phi_\alpha(\Phi^{-1}(1 - u_1), \Phi^{-1}(1 - u_2))}{\phi(\Phi^{-1}(1 - u_1)) \cdot \phi(\Phi^{-1}(1 - u_2))} \\
&= \frac{\phi_\alpha(-\Phi^{-1}(u_1), -\Phi^{-1}(u_2))}{\phi(-\Phi^{-1}(u_1)) \cdot \phi(-\Phi^{-1}(u_2))} \\
&= \frac{\phi_\alpha(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi(\Phi^{-1}(u_1)) \cdot \phi(\Phi^{-1}(u_2))} \\
&= c(u_1, u_2; \alpha)
\end{aligned}$$

For Frank copula:

$$\begin{aligned}
&c(1-u_1,1-u_2;\alpha)\\
&=\alpha\cdot\frac{e^{\alpha u_1}e^{\alpha u_2}\cdot e^{-2\alpha}}{1-e^{-\alpha}}\cdot\left(1-\frac{(1-e^{-\alpha}\cdot e^{\alpha u_1})(1-e^{-\alpha}\cdot e^{\alpha u_2})}{1-e^{-\alpha}}\right)^{-2}\\
&=\alpha\cdot\frac{e^{-\alpha u_1}e^{-\alpha u_2}}{1-e^{-\alpha}}\cdot\left(\frac{e^{-\alpha}\cdot e^{\alpha u_1}+e^{-\alpha}\cdot e^{\alpha u_2}-e^{-\alpha}-e^{-2\alpha}e^{\alpha u_1}e^{\alpha u_2}}{1-e^{-\alpha}}\right)^{-2}\cdot(e^{-\alpha u_1}e^{-\alpha u_2}\cdot e^{\alpha})^{-2}\\
&=\alpha\cdot\frac{e^{-\alpha u_1}e^{-\alpha u_2}}{1-e^{-\alpha}}\cdot\left(\frac{e^{-\alpha u_2}+e^{-\alpha u_1}-e^{-\alpha u_1}e^{-\alpha u_2}-e^{-\alpha}}{1-e^{\alpha}}\right)^{-2}\\
&=\alpha\cdot\frac{e^{-\alpha u_1}e^{-\alpha u_2}}{1-e^{-\alpha}}\cdot\left(1-\frac{(1-e^{-\alpha u_1})(1-e^{-\alpha u_2})}{1-e^{-\alpha}}\right)^{-2}\\
&=c(u_1,u_2;\alpha)
\end{aligned}$$

For EFGM copula:

$$\begin{aligned}
&c(1-u_1,1-u_2;\alpha)\\
&=1+\alpha(2u_1-1)(2u_2-1)\\
&=1+\alpha(1-2u_1)(1-2u_2)\\
&=c(u_1,u_2;\alpha)
\end{aligned}$$

Thus we also have:

$$c_1(u_1,u_2;\alpha)=-c_1(1-u_1,1-u_2;\alpha)$$

$$c_{1\alpha}(u_1,u_2;\alpha)=-c_{1\alpha}(1-u_1,1-u_2;\alpha)$$

$$c_2(u_1,u_2;\alpha)=-c_2(1-u_1,1-u_2;\alpha)$$

$$c_{2\alpha}(u_1,u_2;\alpha)=-c_{2\alpha}(1-u_1,1-u_2;\alpha)$$

$$c_{11}(u_1,u_2;\alpha)=c_{11}(1-u_1,1-u_2;\alpha)$$

$$c_{22}(u_1,u_2;\alpha)=c_{22}(1-u_1,1-u_2;\alpha)$$

$$c_{12}(u_1,u_2;\alpha)=c_{12}(1-u_1,1-u_2;\alpha)$$

When the marginal distribution is symmetric around zero, like student t distribution, symmetry restriction on sieve space $\mathcal{F}$ and $\mathcal{F}_n$ could be applied, see **Remark** 4. Hence we have $f_0(y)\equiv f_0(-y)$ and $v_f^*(y)\equiv v_f^*(-y)$.

Refer to expression in **Section** 9.3. The first term is:

$$\mathbb{E}\frac{\dot{v}_f^*(V_t)\cdot f_0(V_t)-\dot{f}_0(V_t)\cdot v_f^*(V_t)}{[f_0(V_t)]^2}$$

As both $f_0$ and $v_f^*$ are symmetric around zero, this expectation is zero.

The second term is:.

$$\mathbb{E}c_{11}(F_0(V_{t-1}),F_0(V_t);\alpha_0)\cdot v_F^*(V_{t-1})\cdot f_0(V_{t-1})$$

As $v_f^*(y) \equiv v_f^*(-y)$, we have $v_F^*(y) = -v_F^*(y)$. Following symmetric property of $c$, $c_{11}$ and $c_1$, this item will equal to zero:

$$\begin{aligned}
&\mathbb{E}c_{11}(F_0(V_{t-1}), F_0(V_t); \alpha_0) \cdot v_F^*(V_{t-1}) \cdot f_0(V_{t-1}) \\
=& \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} c_{11}(F_0(y_1), F_0(y_2); \alpha_0) \cdot v_F^*(y_1) \cdot c(F_0(y_1), F_0(y_2); \alpha_0) \cdot [f_0(y_1)]^2 \cdot f_0(y_2)\mathrm{d}y_1\mathrm{d}y_2 \\
=& \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} c_{11}(F_0(-y_1), F_0(-y_2); \alpha_0) \cdot v_F^*(-y_1) \cdot c(F_0(-y_1), F_0(-y_2); \alpha_0) \cdot \\
& [f_0(-y_1)]^2 \cdot f_0(-y_2)\mathrm{d}y_1\mathrm{d}y_2 \\
=& \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} c_{11}(1 - F_0(y_1), 1 - F_0(y_2); \alpha_0) \cdot [-v_F^*(y_1)] \cdot c(1 - F_0(y_1), 1 - F_0(y_2); \alpha_0) \cdot \\
& [f_0(y_1)]^2 \cdot f_0(y_2)\mathrm{d}y_1\mathrm{d}y_2 \\
=& -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} c_{11}(F_0(y_1), F_0(y_2); \alpha_0) \cdot v_F^*(y_1) \cdot c(F_0(y_1), F_0(y_2); \alpha_0) \cdot [f_0(y_1)]^2 \cdot f_0(y_2)\mathrm{d}y_1\mathrm{d}y_2 \\
=& -\mathbb{E}c_{11}(F_0(V_{t-1}), F_0(V_t); \alpha_0) \cdot v_F^*(V_{t-1}) \cdot f_0(V_{t-1})
\end{aligned}$$

Based on similar logic, we have the remaining terms all being zeros. Hence $F(\gamma_0, v^*) = 0$.

**Corollary 1.** *For Gaussian, Frank and EFGM copula, nonstationary structure will not affect limiting distribution for two-step sieve estimator of these three copulas, when marginal is symmetric.*

Notice $c(u_1, u_2; \alpha) = c(1-u_1, 1-u_2; \alpha)$ is not satisfied for Clayton and Gumbel copula, $F(\gamma_0, v^*)$ is not zero in general for these two copulas, we need to simulate $v^*$ to get $F(\gamma_0, v^*)$ in **Section** 6.3 for further analysis.

*Remark* 13. Chen et al. [2009] demonstrates that ideal sieve estimator is semiparametric efficient when $V_t$ is directly observed and there is no linear filtering. We find that two-step sieve estimator has the same limiting distribution, as if there is no nonstationary filtering, thus also being semiparametric efficient, when both marginal and copula are symmetric. Also, when $F(\gamma_0, v^*)$ is exactly zero, our limiting distribution will be normally distributed same as ideal estimator in Chen et al. [2009]. Hence procedures for estimation of variance and corresponding inference could be applied as proposition 4.2 in Chen et al. [2009] and theorem 5.1 in Ai and Chen [2003].

*Remark* 14. All the above derivation focus on OLS regression $\hat{\beta}_n$. However, the only information we utilize for proof is the convergence rate and the limiting distribution. If we consider another estimator of nonstationary filtering coefficient with the assumption that it is of the usual root-n convergence rate after our $G_n$ standardization, then the limiting distribution of estimator for copula parameter will follow the same pattern:

$$-F(\gamma_0, v^*) \times \int_0^1 X(r)dr \times \tilde{\xi} + N(0, \|v^*\|^2)$$

constant term $F(\gamma_0, v^*)$ and nonstationarlity term $\int_0^1 X(r)\mathrm{d}r$ remain the same. The only difference is $\tilde{\xi}$, limiting distribution of $\sqrt{n}G_n(\tilde{\beta}_n - \beta_0)$. When both marginal and copula are symmetric, then $F(\gamma_0, v^*)$ is still zero and thus the new estimator $\hat{\alpha}$ is also semiparametric efficient following results of Chen et al. [2009]. See **Section** 6.2.6 for analysis of joint sieve estimator.

# 6 Simulation

## 6.1 Copula and marginal choice

We consider five copulas, each with four choices of copula parameter introduced in **Section** 2.4. Marginal distribution we experiment are student t distribution with degree of freedom 3 and 5.

If $(V_1, V_2) \sim C(F_0(\cdot), F_0(\cdot))$ and we know $V_1$ in advance, we can apply the conditional approach described in Nelson [1999] and Chen and Fan [2006] to generate uniform distributed time series satisfying the specific copula, then apply inverse distribution function $F_0^{-1}(\cdot)$ to get $V_2$:

1. Let $U_1 = F_0(V_1)$.
2. Generate a uniformly distributed random variable $\varepsilon$. Solve $U_2$ by $C_1(U_1, U_2) = \varepsilon$.
   Here $C_1 := \frac{\partial C}{\partial u_1}$ is the conditional distribution of $U_2$ given $U_1$.
3. $V_2 = F_0^{-1}(U_2)$, here $F_0(\cdot)$ is true marginal distribution function of $V_1$ and $V_2$.

To generate a first order Markov process specified by a copula $C(\cdot, \cdot)$ and a marginal $F_0(\cdot)$, we can repeat this algorithm sequentially.

For five copulas we consider in **Section** 2.4, expressions of conditional distribution $C_1$ are:

- Gaussian copula

$$\begin{aligned} C_1(u_1, u_2, \alpha) &= \frac{1}{2\pi\sqrt{1-\alpha^2}} \cdot \frac{1}{\phi(\Phi^{-1}(u_1))} \cdot \\ &\int_{-\infty}^{\Phi^{-1}(u_2)} \exp\left[-\frac{[\Phi^{-1}(u_1)]^2 + x^2 - 2\alpha x \cdot \Phi^{-1}(u_1)}{2(1-\alpha^2)}\right] \mathrm{d}x \end{aligned} \tag{6}$$

- Frank copula

$$C_1(u_1, u_2; \alpha) = \left(1 - \frac{(1-e^{-\alpha u_1})(1-e^{-\alpha u_2})}{1-e^{-\alpha}}\right)^{-1} \cdot \frac{1-e^{-\alpha u_2}}{1-e^{-\alpha}} \cdot e^{-\alpha u_1}$$

- Clayton copula

$$C_1(u_1, u_2; \alpha) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}-1} \cdot u_1^{-\alpha-1}$$

- EFGM copula

$$C_1(u_1, u_2; \alpha) = u_2[1 + \alpha(1-u_2)(1-2u_1)]$$

- Gumbel copula

$$\begin{aligned} C_1(u_1, u_2; \alpha) &= \exp\left[-\left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}}\right] \cdot \\ &\left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}-1} \cdot (-\log u_1)^{\alpha-1} \cdot \frac{1}{u_1} \end{aligned}$$

## 6.2 Finite sample performance

### 6.2.1 Estimator

In this section we address the finite sample performance of sieve estimator $\hat{\beta}_S$ by comparing it to the empirical estimator proposed in Chen and Xiao [2016] and the infeasible (or ideal) estimator proposed in Chen et al. [2009] where we observe $V_t$ directly and there is no filtering. In all simulation, sample size is $T = 500$, number of repetition is $N = 2000$.

For empirical approach, it is in fact a three step method. During the first step, we do OLS to get $\hat{V}_t = Y_t - X_t'\hat{\beta}_n$. In second step, we compute empirical distribution according to $\hat{V}_t$ to approximate the true distribution function $F(V_t)$:

$$F_0(V_t) \to \hat{F}_n(\hat{V}_t) := \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{I}(\hat{V}_t \leq \hat{V}_i)$$

On the last step, we do optimization to get $\hat{\alpha}$ based on $\hat{F}_n(\hat{Y}_t)$:

$$\max_{\alpha \in \mathcal{A}} \sum_{t=2}^{n} \log c(\hat{F}_n(\hat{V}_{t-1}), \hat{F}_n(\hat{V}_t); \alpha)$$

For infeasible estimator $\hat{\alpha}$, assume there is no filtering and $Y_t$ is already observed. We do optimization based on OLS residuals $\hat{Y}_t$ and true distribution function $F_0$ to get $\hat{\alpha}$:

$$\max_{\alpha \in \mathcal{A}, f \in \mathcal{F}_n} \sum_{t=1}^{n} \log f(V_t) + \sum_{t=2}^{n} \log c \left( \int_{-\infty}^{V_{t-1}} f(y)\mathrm{d}y, \int_{-\infty}^{V_t} f(y)\mathrm{d}y; \alpha \right)$$

Later our finite sample simulation results will show that both two-step sieve estimator and three-step empirical estimator perform relatively bad when tail dependence of copula is relatively strong. To solve this problem, we also consider joint estimator, which is not based on filtering residuals, a natural extension when we optimize filtering coefficient $\beta$, copula parameter $\alpha$ and marginal $f$ all together:

$$\max_{\alpha \in \mathcal{A}, \beta \in \mathcal{B}, f \in \mathcal{F}_n} \sum_{t=1}^{n} \log f(Y_t - X_t'\beta) + \sum_{t=2}^{n} \log c \left( \int_{-\infty}^{Y_{t-1} - X_{t-1}'\beta} f(y)\mathrm{d}y, \int_{-\infty}^{Y_t - X_t'\beta} f(y)\mathrm{d}y; \alpha \right)$$

To illustrate robustness of nonparametric method, we also simulate the joint parametric estimator where marginal parametric structure is mis-specified. Here we use normal distribution $N(0, \sigma^2)$ with unknown variance to approximate student t distribution with degree 3 or 5. We optimize filtering coefficient $\beta$, copula parameter $\alpha$ and normal distribution variance $\sigma$ all together:

$$\max_{\alpha \in \mathcal{A}, \beta \in \mathcal{B}, \sigma > 0} \sum_{t=1}^{n} \log \left[ \frac{1}{\sigma} \phi \left( \frac{Y_t - X_t'\beta}{\sigma} \right) \right] + \sum_{t=2}^{n} \log c \left( \Phi \left( \frac{Y_{t-1} - X_{t-1}'\beta}{\sigma} \right), \Phi \left( \frac{Y_t - X_t'\beta}{\sigma} \right); \alpha \right)$$

here $\phi(\cdot)$ and $\Phi(\cdot)$ are density and distribution function for standard normal distribution, respectively.

### 6.2.2 Non-stationarity structure of $X_t$

We consider several choices of nonstationarity structure of $X_t$:

- $X_t$ is deterministic trend
  $X_t = t$, $Y_t = X_t\beta_0 + V_t$, $\beta_0 = 1$
- Unit root process
  $Y_t = X_t\beta_0 + V_t$, $\beta_0 = 1$, $X_t = Y_{t-1}$
- Cointegration process
  $X_t = X_{t-1} + \eta_t$, $\eta_t \sim N(0, 1)$ independent of $\{V_t\}$. $Y_t = X_t\beta_0 + V_t$, $\beta_0 = 1$

### 6.2.3 Sieve choice

We use Laguerre polynomial sieve to approximate student $t$ distribution with degree of freedom 3 and 5:

$$\mathcal{F}_n = \left\{ f(y) = \left[ \sum_{k=0}^{K_n} a_k \cdot \frac{L_k(|y|)}{\sqrt{2}} \cdot e^{-\frac{x}{2}} \right]^2, \int_{-\infty}^{+\infty} f(y)dy = 1 \right\} \tag{7}$$

The first several Laguerre polynomials to order 5 are:

$$\begin{aligned}
L_0(x) &= 1 \\
L_1(x) &= 1 - x \\
L_2(x) &= \frac{1}{2}\left(x^2 - 4x + 2\right) \\
L_3(x) &= \frac{1}{6}\left(-x^3 + 9x^2 - 18x + 6\right) \\
L_4(x) &= \frac{1}{24}\left(x^4 - 16x^3 + 72x^2 - 96x + 24\right) \\
L_5(x) &= \frac{1}{120}\left(-x^5 + 25x^4 - 200x^3 + 600x^2 - 600x + 120\right)
\end{aligned}$$

Laguerre polynomials are orthogonal based on kernel $e^{-x}$:

$$\int_0^{+\infty} L_i(y)L_j(y)e^{-x}\mathrm{d}y = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Because we consider symmetric marginal $t(3)$ and $t(5)$ defined in whole real line, we extend Laguerre polynomial $L_k(y)$ to $\mathbb{R}^1$ as $\frac{L_k(|y|)}{\sqrt{2}}$ by dividing $\sqrt{2}$ in equation (7) to keep orthogonality.

For the finite sample simulation reports in this paper, we select $K_n = 5$. For real application, the selection of number of sieve terms $\hat{K}_n$ could be based on small sample AIC of Burnham and Anderson [2003]: $\hat{K}_n = \arg\max_K \{L_n(\hat{\gamma}_n(K)) - \frac{K}{n-K-1}\}$, where $\hat{\gamma}_n(K)$ is the sieve MLE of $\gamma_0 = (\alpha_0, f_0)$ using $K$ as the sieve number of terms. We apply this criterion for sieve selection in empirical application, see **Section** 7. Other criterion is also available like out-of-sample validation.

We can further simplify the constraint $\int_{\mathbb{R}^1} f(y)dy = 1$ in $\mathcal{F}_n$. Notice orthogonality:

$$\int_{-\infty}^{+\infty} \frac{L_i(|y|)}{\sqrt{2}} \cdot \frac{L_j(|y|)}{\sqrt{2}}\mathrm{d}y = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

hence condition $\int_{\mathbb{R}^1} f(y)dy = 1$ can be simplified as:

$$a_0^2 + a_1^2 + a_2^2 + a_3^2 + a_4^2 + a_5^2 = 1$$

through which we can solve $a_0$ analytically as:

$$a_0 = \sqrt{1 - \sum_{k=1}^{5} a_k^2}$$

and then we can solve the gradient of both objective and constraint analytically rather than numerically centered finite difference. Finite sample simulation shows that passing analytical gradient will save us half computation time.

Comparing to power series log sieve applied in Chen et al. [2009], Laguerre polynomial square root sieve does not need numerical integration and there is only one tuning parameter $K_n$ (for log sieve method, grid length is also need to predetermined). We can also write out analytical expression of gradient for acceleration of optimization, whereas analytical gradient is nearly impossible for log sieve. Furthermore, we utilize the prior information that the marginal distribution is symmetric so we can extend Laguerre polynomial from $[0, +\infty)$ to $\mathbb{R}^1$.[5]

#### 6.2.4 Discussion of Results

Selective tables for Clayton copula are presented in **Section** 11[6]. $S1$ stands for joint sieve estimator, $S2$ stands for two-step sieve estimator, $S$ stands for ideal sieve estimator when $V_t$ is directly observed, $E$ stands for three-step empirical estimator, $P$ stands for joint parametric estimator but with misspecified parametric structure. See **Section** 6.2.1 for detail description of these estimators.

We concentrate on comparison between performance of copula parameter. At the same time, we also list simulation results for linear filtering coefficient $\beta$ bases on joint sieve method, joint parametric(wrong) method and ordinary least square.

For Gaussian copula and Frank copula, joint sieve estimator, two-step sieve estimator and three-step empirical estimator behave roughly the same. When positive dependence is strong($\alpha_0 = 0.9$ for Gaussian, $\alpha_0 = 15$ for Frank), three-step empirical estimator is a bit worse for all three nonstationary structure(time trend, unit root and cointegration).

[5] Chen et al. [2009] also incorporate this prior knowledge, they use sieve basis $\{1, |y|^{3/2}, y^2, y^4\}$ for $t(5)$ and $\{1, |y|^{5/4}, |y|^{3/2}, y^2, y^4\}$ for $t(3)$, all are symmetric. We use these two sieves for robustness test for Clayton copula with $\alpha_0 = 10, 12$ in **Section** 6.2.5.

[6] For full simulation results: please click link here.

Two step estimator behave quite well for Gaussian copula and Frank copula. Exceptions are unit root case with strong negative dependence($\alpha_0 = -0.9$ for Gaussian, $\alpha_0 = -15$ for Frank).

Simulation results below show that poor performance of OLS estimation of linear filtering $\beta$ leads to catastrophic estimation of copula coefficient $\alpha$. Whereas joint sieve method achieves an accurate estimator of $\beta$, hence $\alpha$.

Table 1: Gaussian copula, $\alpha_0 = -0.9$; $X_t$ unit root, $t(3)$

| $\beta_0 = 1$ | bias$\times 10^3$ | std$\times 10^3$ | MSE$\times 10^6$ | 2.5% | 97.5% |
|---|---|---|---|---|---|
| sieve | -1.999171 | 4.538162 | 24.591603 | 0.986184 | 1.003092 |
| OLS | -75.277011 | 84.950852 | 12883.275687 | 0.706989 | 0.995729 |

Table 2: Gaussian copula, $\alpha_0 = -0.9$; $X_t$ unit root, $t(5)$

| $\beta_0 = 1$ | bias$\times 10^3$ | std$\times 10^3$ | MSE$\times 10^6$ | 2.5% | 97.5% |
|---|---|---|---|---|---|
| sieve | -2.990440 | 5.728011 | 41.752843 | 0.981622 | 1.003138 |
| OLS | -88.074103 | 90.051871 | 15866.387154 | 0.667253 | 0.994240 |

Table 3: Frank copula, $\alpha_0 = -15$; $X_t$ unit root, $t(3)$

| $\beta_0 = 1$ | bias$\times 10^3$ | std$\times 10^3$ | MSE$\times 10^6$ | 2.5% | 97.5% |
|---|---|---|---|---|---|
| sieve | -0.956898 | 2.909754 | 9.382321 | 0.991497 | 1.002767 |
| OLS | -11.429428 | 15.176432 | 360.955899 | 0.946577 | 1.001403 |

Table 4: Frank copula, $\alpha_0 = -15$; $X_t$ unit root, $t(5)$

| $\beta_0 = 1$ | bias$\times 10^3$ | std$\times 10^3$ | MSE$\times 10^6$ | 2.5% | 97.5% |
|---|---|---|---|---|---|
| sieve | -0.804994 | 2.633968 | 7.585801 | 0.992068 | 1.002814 |
| OLS | -42.106453 | 48.386098 | 4114.167866 | 0.821676 | 0.998329 |

*Remark* 15. We consider a simplified model to show why OLS estimator performs so bad for unit root model when dependence is strong negative.

Let $x_t$ be a unit root process, $x_t = x_{t-1}\beta_0 + \varepsilon_t$, $\beta_0 = 1$, $\varepsilon_t$ is an $AR(1)$ process with strong negative dependence ($\rho$ is negative and closer to $-1$): $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t$, the residual term $\eta_t$ is i.i.d mean zero $\mathbb{E}\eta_t = 0$ and finite variance $\mathbb{E}\eta_t^2 < +\infty$. Then the variance for $\varepsilon_t$ is $\sigma_\varepsilon^2 := \mathbb{E}\varepsilon_t^2 = \frac{\mathbb{E}\eta_t^2}{1-\rho^2}$, the long run variance for $\varepsilon_t$ is $\omega_\varepsilon^2 := \mathbb{E}\varepsilon_t^2 + 2\sum_{k=1}^{\infty}\mathbb{E}\varepsilon_t\varepsilon_{t+k} = \frac{1+\rho}{1-\rho}\sigma_\varepsilon^2$. Notice the limiting distribution of OLS estimator $\hat{\beta}_n$ is:

$$n(\hat{\beta}_n - 1) \xrightarrow{d} \frac{W^2(1) - \sigma_\varepsilon^2/\omega_\varepsilon^2}{2\int_0^1 W^2(r)\mathrm{d}r} = \frac{W^2(1) - \frac{1-\rho}{1+\rho}}{2\int_0^1 W^2(r)\mathrm{d}r}$$

when $\rho$ is closer to $-1$, $\frac{1-\rho}{1+\rho}$ tends to infinity. Thus we will have strong negative asymptotic bias and this will be more obvious for finite sample simulation. This simple example illustrate why OLS estimator has strong downward bias in unit root setting with strong negative dependence.

For EFGM copula, joint sieve estimator, two-step sieve estimator and three-step empirical estimator behave quite similar, for all three nonstationary structure and four copula parameter choice.

Theoretical explanation is shown in Chen et al. [2009] as EFGM copula is very close to the independent copula ($C(u_1, u_2) := u_1u_2$, $c(u_1, u_2) \equiv 1$), because the distance between EFGM copula function to the independent copula function is $\alpha u_1 u_2(1-u_1)(1-u_2) \leq 0.0625\alpha$ for $\alpha \in [-1, 1]$.[7]

In summary, for Gaussian copula, Frank copula and EFGM copula, we can keep using two-step sieve estimator and three-step empirical estimator, unless for extreme case (unit root with strong negative dependence, Gaussian copula $\alpha_0 = -0.9$; Frank copula $\alpha_0 = -15$). When dependence is strongly positive (Gaussian $\alpha_0 = 0.9$, Frank $\alpha_0 = 15$), joint > two-step > three-step. However, the efficiency loss is not huge and we can still believe in the estimation from two-step sieve method and three-step empirical method.

For copula with tail dependence, both two-step sieve estimator and three-step empirical estimator do not converge well. Two-step sieve estimator has an acceptable bias but the variance is exploding, whereas three-step empirical estimator is strongly downward biased. The stronger of tail dependence, the worse performance for these two estimators. Meanwhile, joint sieve estimator is always stable even for extreme tail dependence, such as Clayton copula ($\alpha_0 = 10, 12$) and Gumbel copula ($\alpha_0 = 6, 7$). Hence we recommend using joint sieve method when dealing with copula with tail dependence, especially when data shows strong tail dependence.

In most cases, joint parametric estimator with mis-specified parametric marginal diverges from the true value and is not comparable from other estimators. If we do not have strong belief about parametric family structure of marginal, semi-parametric sieve method (joint or two-step) will definitely be a better choice.

#### 6.2.5 Robustness and prior information

Variance of two-step sieve estimator explodes for Clayton copula, under Laguerre polynomial with square sieve (**Equation** 1), especially when tail dependence is large ($\alpha_0 = 10, 12$), although the bias of copula estimator is quite mild for two-step sieve estimator. In this section, we consider power series log sieve (**Equation** 2) to check whether this phenomenon is general when we select different sieve. Also we want to see if joint sieve estimator will still dominant other estimators for this log sieve.

Same as Chen et al. [2009], we use sieve basis $\{1, |y|^{3/2}, y^2, y^4\}$ to approximate $t(5)$ and sieve basis $\{1, |y|^{5/4}, |y|^{3/2}, y^2, y^4\}$ to approximate $t(3)$. Then when marginal is $t(5)$, $K_N = 3$, $A_0(y) \equiv 1$, $A_1(y) = |y|^{3/2}$, $A_2(y) = y^2$ and $A_3(y) = y^4$. When marginal is $t(3)$, $K_N = 4$, $A_0(y) \equiv 1$, $A_1(y) = |y|^{5/4}$, $A_2(y) = |y|^{3/2}$, $A_3(y) = y^2$, $A_4(y) = y^4$. Then the marginal density function $f_0(y)$ can be approximated as:

$$\frac{\exp\left(\sum_{k=1}^{K_N} a_k A_k(y)\right)}{\int \exp\left(\sum_{k=1}^{K_N} a_k A_k(y)\right) \mathrm{d}y}$$

We approximate the density $f_0$ the support $[\min(V_t) - s_V, \max(V_t) + s_V]$, where $s_Y$ is the sample standard deviation of $\{V_t\}$. To evaluate the integral that appears in above density approximation, we use a grid of equidistant points on $[\min(V_t) - s_V, \max(V_t) + s_V]$. The grid size in our estimation report was chosen to be 0.005. In all simulation, sample size is $T = 2000$, number of repetition is $N = 500$, marginal is student t distribution with degree of freedom 3 or 5, nonstationary structure is as **Section** 6.2.2. Detail results are listed in **Section** 12.

Our results show that variance of two-step sieve estimator still diverges under power series log sieve, although its bias is even a little smaller than joint estimator. Meanwhile joint sieve estimator is quite stable, dominating both two-step sieve estimator and three-step empirical estimator proposed in Chen and Xiao [2016]. For example, Clayton copula $\alpha_0 = 12$, $X_t$ time trend together with marginal $t(3)$, the bias of two-step sieve estimator is even a little better than ideal estimator ($0.848 < 0.890$), whereas the variance of two-step sieve estimator explodes. Joint estimator (MSE 4.942) dominates two-step sieve estimator (MSE 29.732) and three-step empirical estimator (MSE 38.277) in this case.

[7] This could also be illustrated by Kendall's tau. $\alpha = 0.9 \Rightarrow \tau = 0.2$, $\rho = 0.3$; $\alpha = 0.5 \Rightarrow \tau = 0.111$, $\rho = 0.167$; $\alpha = -0.9 \Rightarrow \tau = -0.2$, $\rho = -0.3$; $\alpha = -0.5 \Rightarrow \tau = -0.111$, $\rho = -0.167$. They are quite similar and close to 0, which is Kendall's tau and Spearman's rho for independent copula.

Above results indicate that our finite sample simulation results that two-step sieve estimator explodes when tail dependence is strong is not coincident. To further research why two-step sieve estimator perform very bad for Clayton copula and Gumbel copula and how its performance are affected by non-stationary linear filtering, we make simulation on $F(\gamma_0, v^*)$ and $\|v^*\|^2$ in next section. Before then, we will make a little modification for three-step estimator.

Notice that for all sieve based estimator (joint sieve estimator S1, two-step sieve estimator S2 and ideal sieve estimator S), we incorporate the prior information that the true marginal density is symmetric around zero, whereas for three-step empirical estimator we still utilize the natural empirical distribution function to approximate the true distribution function, not considering the addition information from symmetry. Could three-step empirical estimator improve a lot if we add this prior information appropriately?

For data $V_1, V_2, \cdots, V_n$, if we know in advance that the marginal density is symmetric, then a better approximation for true distribution $F_0(\cdot)$ in empirical form is:

$$\tilde{F}(x) = \frac{1}{2}\left(1 + \text{sign}(x) \cdot \frac{1}{n+1}\sum_{t=1}^{n} \mathbb{I}(|V_t| \leq x)\right) \tag{8}$$

We denote this estimator as E1, reported in the last column in **Section** 12. With symmetric information incorporated, bias is smaller and the resulting MSE is nearly halved. For example, Clayton copula, $\alpha_0 = 12$, $X_t$ time trend, marginal $t(3)$, MSE decreased from 38.277 to 20.807, when we use $\tilde{F}(\cdot)$ instead of natural empirical distribution function. However, MSE of joint sieve estimator is only 4.942, a much better performance. Hence, symmetric version of empirical distribution function does help a lot for estimator performance, it will still be strictly dominated by joint sieve estimator, when tail dependence is relatively strong, like Clayton copula and Gumbel copula, if copula parameter is large.

### 6.2.6 Semiparametric efficiency

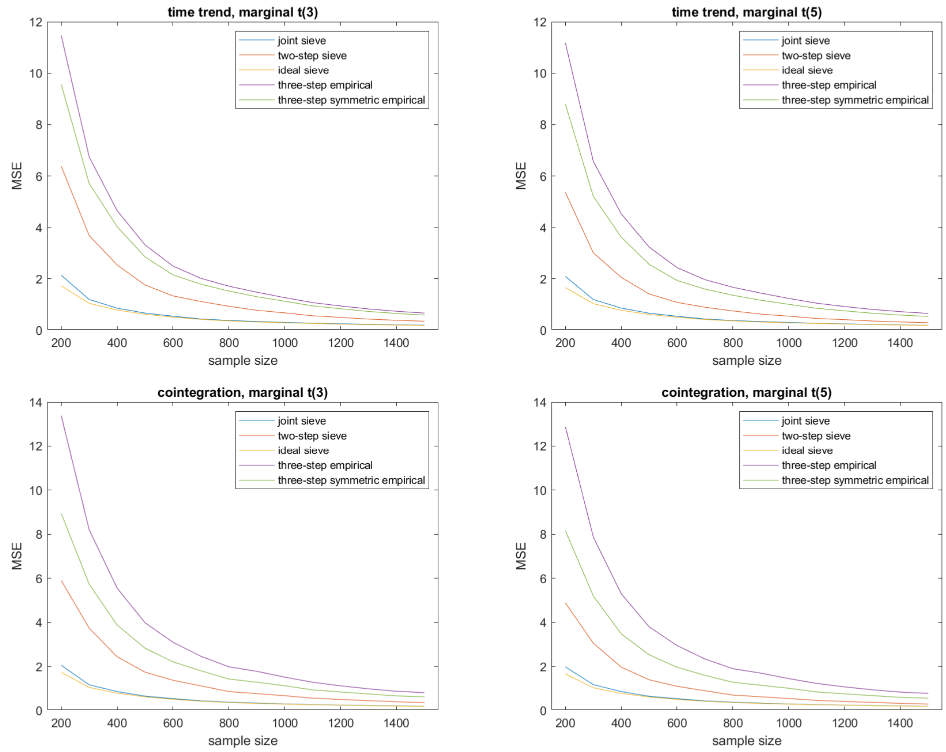
If joint sieve estimator of linear filtering coefficient $\hat{\beta}_n^{joint}$ has the usual convergence rate: $G_n(\hat{\beta}_n^{joint} - \beta_0) = O_p(1/\sqrt{n})$, then joint sieve estimator of copula parameter $\hat{\alpha}_n^{joint}$ is semiparametric efficient, when both marginal and copula are symmetric. This fact follows from **Remark** 14 and Chen et al. [2009].

Analyze the cases when marginal is symmetric and our focus is copula parameter. For Frank copula and EFGM copula, both joint-sieve estimator and two-step sieve estimator are theoretically equivalent to ideal sieve estimator hence seimparametric efficient. For Gaussian copula, joint sieve estimator, two-step sieve estimator and three-step empirical estimator has the exactly the same limiting distribution[8] as ideal sieve estimator, hence all three estimators being semiparametric efficient. However, finite sample results above show that joint sieve estimator is still better than its competitors, especially when the dependence is extreme.

To further illustrate the convergence speed, for Frank copula with time trend and cointegration, sample size ranging from $T = 200$ to $T = 1500$, we plot the finite sample MSE of five estimators: joint sieve estimator, two-step sieve estimator, ideal sieve estimator(theoretically best), three-step empirical estimator and three-step modified empirical estimator(see above section). Simulation repetition is $N = 2000$. I attach results for $\alpha_0 = 15$ here. For full simulation results are in: please click link here.

[8]Chen et al. [2006] show that for stationary model where $Y_t$ is directly observed, empirical distribution method reaches semiparameteric efficiency bound for Gaussian copula. While Chen and Xiao [2016] demonstrates that non-stationary structure will not affect limiting distribution of three-step empirical estimator. Thus three-step empirical estimator is semiparametric efficient when copula is Gaussian.

Figure 1: Frank copula, $\alpha_0 = 15$




When dependence is low ($\alpha_0 = \pm 0.5$), all estimators behave similar, the efficiency loss is quite trifling when we take different estimation methods. When dependence is large ($\alpha_0 = \pm 15$), all estimators are asymptotically equivalent. However, only joint sieve estimator keeps the same path of ideal sieve estimator. Other estimators display some efficiency loss when sample size is small. These figures not only demonstrate that joint sieve estimator has superb finite sample performance but also show us an intuition that the assumption $G_n(\hat{\beta}_n^{joint} - \beta_0) = O_p(1/\sqrt{n})$ should be satisfied and thus $\hat{\alpha}_n^{joint}$ is root-n normally distributed with same limiting variance as ideal sieve estimator.

## 6.3 Analysis of $F(\gamma_0, v^*)$ and $\|v^*\|^2$

In this section, we use simulation in a new sieve space $\mathbf{B}_n$ to approximate true $v^*$ and finally get $\|v^*\|^2$ and $F(\gamma_0, v^*)$ for Clayton copula and Gumbel copula.

### 6.3.1 Simulation scheme

Denote space $\tilde{\mathcal{L}}_2^0[0,1]$ as function space from unit interval to real line satisfying zero integration, finite square integration and symmetry around $\frac{1}{2}$:

$$\tilde{\mathcal{L}}_2^0[0,1] := \left\{ e : [0,1] \to \mathbb{R}^1 \,\middle|\, \int_0^1 e(u)\mathrm{d}u = 0, \int_0^1 [e(u)]^2 \,\mathrm{d}u < +\infty, e(u) = e(1-u) \right\}$$

Let $e^*$ solves the following infinite dimensional optimization problems:

$$\inf_{e \in \tilde{\mathcal{L}}_2^0[0,1]} \quad \mathbb{E}\Bigg( \tfrac{\partial \log c(U_1,U_2;\alpha_0)}{\partial \alpha} - e(U_2) - \tfrac{\partial \log c(U_1,U_2;\alpha_0)}{\partial u_1} \int_0^{U_1} e(u)\mathrm{d}u - \tfrac{\partial \log c(U_1,U_2;\alpha_0)}{\partial u_2} \int_0^{U_2} e(u)\mathrm{d}u \Bigg)^2$$

and $\mathcal{I}_*(\alpha_0)$ be the minimum objective function value. Here both $U_1$ and $U_2$ are uniformly distributed random variables, joint distribution following copula $c(\cdot,\cdot;\alpha_0)$.

Then from Chen et al. [2009], we have $\|v^*\|^2 = \frac{1}{\mathcal{I}_*(\alpha_0)}$ and $v^* = [1, -e^*(F_0(\cdot))f_0(\cdot)] \cdot \|v^*\|^2$.

In general, there is no closed form solution of $e^*$ and $\mathcal{I}_*(\alpha_0)$. Nevertheless we use a sieve for $\mathcal{L}_2^0[0,1]$. Sieve space $\tilde{\mathbf{B}}_n$ is recommended from Chen et al. [2006] due to its simple structure[9]:

$$\tilde{\mathbf{B}}_n = \left\{ e(u) = \sum_{k=1}^{K_n} a_k \sqrt{2}\cos(2k\pi u), u \in [0,1], \sum_{k=1}^{K_n} a_k^2 < +\infty \right\}$$

In experiment, we let $N = 1,000,000$ and $K_N = 20$ to solve optimization problems for $a_1, a_2, \cdots, a_{20}$.

Simulate bivariate uniformly distributed random variables $(U_{1i}, U_{2i})$, joint distribution following copula function $c(\cdot,\cdot;\alpha_0)$, independent across $i = 1, 2, \cdots, n$. Then we need to solve optimization problem:

$$\min_{a_1,\cdots,a_{20}} \sum_{i=1}^{1,000,000} \left( \frac{\partial \log c(U_{1i},U_{2i};\alpha_0)}{\partial \alpha} - \sum_{k=1}^{K_N} a_k\sqrt{2}\cos(2k\pi U_{2i}) - \frac{\partial \log c(U_{1i},U_{2i};\alpha_0)}{\partial u_1} \cdot \right.$$
$$\left. \sum_{k=1}^{K_N} \frac{a_k\sqrt{2}\sin(2k\pi U_{1i})}{2k\pi} - \frac{\partial \log c(U_1,U_2;\alpha_0)}{\partial u_2} \cdot \sum_{k=1}^{K_N} \frac{a_k\sqrt{2}\sin(2k\pi U_{2i})}{2k\pi} \right)^2$$

to get $a_1, a_2, \cdots, a_{20}$ and denote the minimum objective function value as $\mathcal{I}_*(\alpha_0)$.

Notice $v_f^*(y) = -\frac{\sqrt{2}f_0(y)}{\mathcal{I}_*} \sum_{k=1}^{K_N} a_k \cos(2k\pi F_0(y))$, then:

$$v_F^*(y) = -\frac{\sqrt{2}}{2k\pi\mathcal{I}_*} \sum_{k=1}^{K_N} a_k \sin(2k\pi F_0(y))$$

$$\dot{v}_f^*(y) = \frac{\sqrt{2}}{\mathcal{I}_*} \sum_{k=1}^{K_N} a_k \left\{ 2k\pi \sin\left[2k\pi F_0(y)\right] \cdot \left[f_0(y)\right]^2 - \cos\left[2k\pi F_0(y)\right] \cdot \dot{f}_0(y) \right\}$$

With the help of detail equation in **Section** 9.3, we can simulate to get $F(\gamma_0, v^*)$.

$$\begin{aligned} F(\gamma_0, v^*) &= \mathbb{E}\frac{\dot{v}_f^*(V_t)\cdot f(V_t) - f'(V_t)\cdot v_f^*(V_t)}{[f(V_t)]^2} \\ &+ c_1(U_{t-1},U_t;\alpha_0)\cdot v_f^*(V_{t-1}) + c_{11}(U_{t-1},U_t;\alpha_0)\cdot v_F^*(V_{t-1})\cdot f(V_{t-1}) \\ &+ c_{12}(U_{t-1},U_t;\alpha_0)\cdot v_F^*(V_t)\cdot f(V_{t-1}) + c_{12}(U_{t-1},U_t;\alpha_0)\cdot v_F^*(V_{t-1})\cdot f(V_t) \\ &+ c_{1\alpha}(U_{t-1},U_t;\alpha_0)\cdot v_\alpha^* + c_{2\alpha}(U_{t-1},U_t;\alpha_0)\cdot v_\alpha^* \\ &+ c_2(U_{t-1},U_t;\alpha_0)\cdot v_f^*(Y_t) + c_{22}(U_{t-1},U_t;\alpha_0)\cdot v_F^*(V_t)\cdot f(V_t) \end{aligned}$$

Notice $c(u_1, u_2; \alpha) \equiv c(u_2, u_1; \alpha)$ holds for all five copulas introduced in **Section** 2.4, we have:

$$\mathbb{E}c_1(U_{t-1},U_t;\alpha_0)\cdot v_f^*(V_{t-1}) = \mathbb{E}c_2(U_{t-1},U_t;\alpha_0)\cdot v_f^*(V_t)$$

$$\mathbb{E}c_{11}(U_{t-1},U_t;\alpha_0)\cdot v_F^*(V_{t-1})\cdot f(V_{t-1}) = \mathbb{E}c_{22}(U_{t-1},U_t;\alpha_0)\cdot v_F^*(V_t)\cdot f(V_t)$$

$$\mathbb{E}c_{12}(U_{t-1},U_t;\alpha_0)\cdot v_F^*(V_t)\cdot f(V_{t-1}) = \mathbb{E}c_{12}(U_{t-1},U_t;\alpha_0)\cdot v_F^*(V_{t-1})\cdot f(V_t)$$

$$\mathbb{E}c_{1\alpha}(U_{t-1},U_t;\alpha_0) = \mathbb{E}c_{2\alpha}(U_{t-1},U_t;\alpha_0)$$

[9] As we restrict our density to be symmetric around zero, function $e(\cdot)$ should be symmetric around $\frac{1}{2}$, thus we consider sieve space $\tilde{\mathbf{B}}_n$ only including even terms $\cos(2k\pi u)$, but without odd terms $\cos((2k-1)\pi u)$ as they are not symmetric around $\frac{1}{2}$.

take last equation for example. From $c(u_1,u_2;\alpha)\equiv c(u_2,u_1;\alpha)$ we have $\log c(u_1,u_2;\alpha)\equiv\log c(u_2,u_1;\alpha)$, thus $c_{1\alpha}(u_1,u_2;\alpha)\equiv c_{2\alpha}(u_1,u_2;\alpha)$, hence

$$\begin{aligned}\mathbb{E}c_{1\alpha}(U_{t-1},U_t;\alpha_0) &= \int_0^1\int_0^1 c_{1\alpha}(u_1,u_2;\alpha_0)\cdot c(u_1,u_2;\alpha_0)\mathrm{d}u_1\mathrm{d}u_2\\ &= \int_0^1\int_0^1 c_{2\alpha}(u_2,u_1;\alpha_0)\cdot c(u_2,u_1;\alpha_0)\mathrm{d}u_1\mathrm{d}u_2\\ &= \int_0^1\int_0^1 c_{2\alpha}(u_1,u_2;\alpha_0)\cdot c(u_1,u_2;\alpha_0)\mathrm{d}u_1\mathrm{d}u_2\\ &= \mathbb{E}c_{2\alpha}(U_{t-1},U_t;\alpha_0)\end{aligned}$$

Notice both $v_f^*$ and $f_0$ are symmetric around zero, the first term will be zero:

$$\mathbb{E}\frac{\dot{v}_f^*(V_t)\cdot f_0(V_t)-\dot{f}_0(V_t)\cdot v_f^*(V_t)}{[f_0(V_t)]^2}=0$$

Hence we only need to simulate: $\mathbb{E}c_1(U_{t-1},U_t;\alpha_0)\cdot v_f^*(Y_{t-1})$, $\mathbb{E}c_{11}(U_{t-1},U_t;\alpha_0)\cdot v_F^*(V_{t-1})\cdot f(V_{t-1})$, $\mathbb{E}c_{12}(U_{t-1},U_t;\alpha_0)\cdot v_F^*(V_t)\cdot f(V_{t-1})$ and $\mathbb{E}c_{1\alpha}(U_{t-1},U_t;\alpha_0)$. We generate $M=1,000,000$ random pairs of $(Y_{1i},Y_{2i})\sim C(F_0(\cdot),F_0(\cdot);\alpha_0)$, independent each other, to simulate these constants: $U_{1i}=F_0^{-1}(V_{1i})$, $U_{2i}=F_0^{-1}(V_{2i})$.[10]

Detail results are listed in **Section** 10.

#### 6.3.2 Results discussion

For Clayton copula, $\mathcal{I}_*(0.05857\to0.01265\to0.00431\to0.00328)$ decreases with $\alpha_0(2\to5\to10\to12)$. For Gumbel copula, $\mathcal{I}_*(0.15178\to0.02298\to0.00529\to0.00369)$ decreases with $\alpha_0(2\to3.5\to6\to7)$. Both results are consistent with our finite sample results that the larger the copula parameter $\alpha_0$(hence larger tail dependence), the larger MSE of the ideal estimator.

For Clayton copula, $|F(\gamma_0,v^*)|$ increase with $\alpha_0$ for both $t(3)$ and $t(5)$. This indicates that filtering term may have a huge difference to our two step sieve estimator comparing with ideal estimator. Consistent with bad performance of two step estimator relative to ideal estimator when tail dependence is large following our finite sample results.

For Gumbel copula, $|F(\gamma_0,v^*)|$ also increase with $\alpha_0$ for both $t(3)$ and $t(5)$. However, the magnitude is smaller than Clayton copula. Hence although two step estimator is also affected by linear filtering for Gumbel copula, its difference with ideal estimator is smaller than Clayton copula.

Notice all estimators are significant in a 99% confidence interval[11], combining with finite sample simulation results, we can conclude that nonstationary structure will have a significant influence on two-step sieve estimator for Clayton copula and Gumbel copula.

#### 6.3.3 Robustness

All the above simulation results are based on sieve choice $K_N=20$ for space $\mathbf{B}_n$. In this section, we set instead $K_N=22$ to check whether the simulation results are robust when we approximate $v_f^*$ with different sieve complexity.

Results for $a_1,a_2,\cdots,a_{22}$ and $\mathcal{I}(\alpha_*)$ are attached referred to full simulation results. $\mathcal{I}(\alpha^*)$ for $K_N=20$ implies the limiting variance for ideal sieve estimator $\|v^*\|^2=1/\mathcal{I}(\alpha_*)$ will increase with tail dependence. Results for $K_N=22$ illustrates same pattern and the implied standard deviation are quite similar:

[10] For example, we simulate $\mathbb{E}c_1(U_{t-1},U_t;\alpha_0)\cdot v_f^*(V_{t-1})$ as:

$$\frac{1}{1,000,000}\sum_{i=1}^{1,000,000}c_1(U_{1i},U_{2i};\alpha_0)\cdot v_f^*(V_{1i})$$

[11] 99% confidence interval is computed as: mean value $\pm$ 2.58$\times$ standard deviation $/\sqrt{1,000,000}$

Table 5: Simulation results for $\|v^*\|$ under $K_N = 20$ and $K_N = 22$

| Clayton copula | $\alpha_0 = 2$ | $\alpha_0 = 5$ | $\alpha_0 = 10$ | $\alpha_0 = 12$ |
|---|---|---|---|---|
| $K_N = 20$ | 4.131889428 | 8.892009093 | 15.22619114 | 17.46279624 |
| $K_N = 22$ | 4.138018088 | 8.905668555 | 15.2466739 | 17.48473315 |

| Gumbel copula | $\alpha_0 = 2$ | $\alpha_0 = 3.5$ | $\alpha_0 = 6$ | $\alpha_0 = 7$ |
|---|---|---|---|---|
| $K_N = 20$ | 2.566823748 | 6.597379413 | 13.75524445 | 16.45891828 |
| $K_N = 22$ | 2.572639121 | 6.643081106 | 13.92829938 | 16.68856978 |

Results for $F(\gamma_0, v^*)$ are attached in **Section** 10.2. Estimation for $F(\gamma_0, v^*)$ is quite similar for $K_N = 20$ and $K_N = 22$. For example, with Clayton copula $\alpha_0 = 5$ and marginal $t(3)$, $F(\gamma_0, v^*) \approx 3.795$ for $K_N = 20$ and $F(\gamma_0, v^*) \approx 3.807$ for $K_N = 22$. And the results are all significant for both Clayton copula and Gumbel copula and for both $t(3)$ and $t(5)$.

All above results indicate robustness of our simulation scheme.

### 6.4 Variance of $\xi$

From Remark 14, the effect of nonstaionary structure to limiting distribution of estimator has three effects: $F(\gamma_0, v^*)$, $\int_0^1 X(r)\mathrm{d}r$ and limiting distribution of linear filtering coefficient. In above section, there is clear evidence to reveal that $F(\gamma_0, v^*)$ is not zero for Clayton copula and Gumbel copula. The magnitude $|F(\gamma_0, v^*)|$ will increase with tail dependence. $\int_0^1 X(r)\mathrm{d}r$ is irrelevant with copula structure. In this section, through a simplified model, we will demonstrate that limiting variance of linear filtering coefficient estimator through OLS method explosed as tail dependence gets larger, for Clayton copula and Gumbel copula. And this provides an intuition why two-step sieve estimator gets diverged but joint sieve estimator is stable, when tail dependence becomes larger.

Suppose now both marginal distribution ($F_0$ and $f_0$) and copula parameter ($\alpha_0$) are observed. Only time trend and cointegration are considered here such that $X_t$ is independent of the $Y_t$ system. We need to incorporate the copula structure into the objective function for MLE estimator $\hat{\beta}_n^{MLE}$ by maximizing:

$$\max \sum_{t=1}^{n} \log f_0(Y_t - X_t\beta) + \sum_{t=2}^{n} \log c(F_0(Y_{t-1} - X_{t-1}\beta), F_0(Y_t - X_t\beta), \alpha_0)$$

to get $\hat{\beta}_n^{MLE}$.

Suppose we still have consistency condition $\hat{\beta}_n^{MLE} \xrightarrow{p} \beta_0$, make Taylor expansion, we have:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} h(V_{t-1}, V_t) \approx A_3 \cdot G_n^{-1} X_t \cdot \sqrt{n} G_n(\hat{\beta}_n^{MLE} - \beta_0)$$

for $h(V_{t-1}, V_t) := \dot{g}(V_t) + l_1(V_{t-1}, V_t; \alpha_0) + l_2(V_{t-1}, V_t; \alpha_0)$.

Here $A_3$ is defined as the expectation of:

$$\ddot{g}(V_t) + c_{11}(V_{t-1}, V_t; \alpha_0) + 2c_{12}(V_{t-1}, V_t; \alpha_0) + c_{22}(V_{t-1}, V_t; \alpha_0)$$

Then the limiting distribution of $\hat{\beta}_n^{MLE}$ is:

$$\sqrt{n} G_n(\hat{\beta}_n^{MLE} - \beta_0) \Rightarrow \frac{\omega_1}{A_3} \cdot \frac{\int_0^1 X(r)\mathrm{d}W(r)}{\int_0^1 [X(r)]^2 \mathrm{d}r}$$

here $W(\cdot)$ is a standard Wiener process independent of $X(\cdot)$ and $\omega_1^2$ is the long run variance of $h(V_{t-1}, V_t) := \dot{g}(V_t) + l_1(V_{t-1}, V_t; \alpha_0) + l_2(V_{t-1}, V_t; \alpha_0)$. However, we can see that $h(V_{t-1}, V_t)$ is in fact a martingale difference sequence $\mathbb{E}\left(h(V_{t-1}, V_t)|\mathcal{F}_{t-1}\right) = 0$, where $\mathcal{F}_t$ is defined as the $\sigma-$algebra generated by $V_t, V_{t-1}, V_{t-2}, \cdots$. See remark below for detail. Hence $\omega_1^2$ is indeed the variance of $h(V_{t-1}, V_t)$, $\omega_1^2 = \mathbb{E}\left[\dot{g}(V_t) + l_1(V_{t-1}, V_t; \alpha_0) + l_2(V_{t-1}, V_t; \alpha_0)\right]^2$.

*Remark* 16. Notice $\mathbb{E}\left(h(V_{t-1},V_t)|\mathcal{F}_{t-1}\right) = \mathbb{E}\left(h(V_{t-1},V_t)|V_{t-1}\right)$ due to Markov property. We first show that $\mathbb{E}l_1(V_{t-1},V_t;\alpha_0|Y_{t-1}) = 0$:

$$\begin{aligned}
\mathbb{E}l_1(V_{t-1},V_t;\alpha_0|V_{t-1}) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} c_1(F_0(y_1),F_0(y_2);\alpha_0)\cdot[f_0(y_1)]^2\cdot f_0(y_2)\mathrm{d}y_1\mathrm{d}y_2 \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} c_1(F_0(y_1),F_0(y_2);\alpha_0)\cdot[f_0(y_1)]^2\cdot f_0(y_2)\mathrm{d}y_2\mathrm{d}y_1 \\
&= \int_{-\infty}^{+\infty} [f_0(y_1)]^2\cdot\left(\int_0^1 c_1(F_0(y_1),u;\alpha_0)\mathrm{d}u\right)\mathrm{d}y_1
\end{aligned}$$

$\int_0^1 c(F_0(y_1),u;\alpha_0)\mathrm{d}u$ will always be 1 no matter what value $y_1$ takes, as this is the conditional density of second element of a bivariate copula system given the first element. If we assume differentiation and integration can exchange, then we have $\int_0^1 c(F_0(y_1),u;\alpha_0)\mathrm{d}u \equiv 0$. Hence $\mathbb{E}l_1(V_{t-1},V_t;\alpha_0|V_{t-1})$ is always zero.

We next show that $\mathbb{E}\left(\dot{g}(V_t)|V_{t-1}\right) + \mathbb{E}\left(l_2(V_{t-1},V_t;\alpha_0)|V_{t-1}\right) \equiv 0$:

$$\begin{aligned}
\mathbb{E}\left(l_2(V_{t-1},V_t;\alpha_0)|V_{t-1}\right) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} c_2(F_0(y_1),F_0(y_2);\alpha_0)\cdot f_0(y_1)\cdot[f_0(y_2)]^2\,\mathrm{d}y_1\mathrm{d}y_2 \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} c_2(F_0(y_1),F_0(y_2);\alpha_0)\cdot f_0(y_1)\cdot[f_0(y_2)]^2\,\mathrm{d}y_2\mathrm{d}y_1 \\
&= \int_{-\infty}^{+\infty} f_0(y_1)\cdot\left(\int_{-\infty}^{+\infty} f_0(y_2)\mathrm{d}c(F_0(y_1),F_0(y_2);\alpha_0)\right)\mathrm{d}y_1 \\
&= -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_0'(y_2)\cdot c(F_0(y_1),F_0(y_2);\alpha_0)\cdot f_0(y_2)\mathrm{d}y_1\mathrm{d}y_2 \\
&= -\mathbb{E}\left(\dot{g}(V_t)|V_{t-1}\right)
\end{aligned}$$

the integration will hold once the regulation condition is satisfied:

$$\lim_{y_2\to+\infty} f_0(y_2)c(F_0(y_1),F_0(y_2);\alpha_0) = \lim_{y_2\to-\infty} f_0(y_2)c(F_0(y_1),F_0(y_2);\alpha_0) = 0$$

Limiting distribution for OLS estimator $\hat{\beta}_n^{OLS}$ follows similar pattern as we just replace the variance of $V_t$ with its long run variance:

$$\sqrt{n}G_n(\hat{\beta}_n^{OLS}-\beta_0) \Rightarrow \omega_V\cdot\frac{\int_0^1 X(r)\mathrm{d}W(r)}{\int_0^1[X(r)]^2\mathrm{d}r}$$

We need to compare $\omega_1/|A_3|$ with $\omega_V$ to measure performance of $\hat{\beta}_n^{MLE}$ and $\hat{\beta}_n^{OLS}$. We generate a 10 million length time series to simulate the long run variance of $Y_t$ through method in Newey and West [1987] with Bartlett kernel[12], $\omega_V^2 \approx \hat{\gamma}_0 + 2\sum_{k=1}^{\infty}\left(1-\frac{k}{L+1}\right)\hat{\gamma}_k$.

Simulation repetition is 80 times and results are reported as an simple average. Detail simulation results are attached below:

Table 6: Clayton copula

| marginal $t(3)$ | $\alpha_0 = 2$ | $\alpha_0 = 5$ | $\alpha_0 = 10$ | $\alpha_0 = 12$ |
|---|---|---|---|---|
| $\sqrt{lrvar(Y_t)}$ | 5.18401319 | 11.42061148 | 22.41561542 | 26.73142608 |
| $\sqrt{var\ h(Y_{t-1},Y_t)}/\vert A_3\vert$ | 1.72819808 | 1.42744164 | 1.20674291 | 1.16647114 |

| marginal $t(5)$ | $\alpha_0 = 2$ | $\alpha_0 = 5$ | $\alpha_0 = 10$ | $\alpha_0 = 12$ |
|---|---|---|---|---|
| $\sqrt{lrvar(Y_t)}$ | 3.71094687 | 8.04308102 | 15.64510863 | 18.61165040 |
| $\sqrt{var\ h(Y_{t-1},Y_t)}/\vert A_3\vert$ | 1.66953871 | 1.39415455 | 1.18403436 | 1.14351576 |

[12]Kernel window size is set to be $L = \max\{L \le 2000|\hat{\gamma}(L+1) \le 0.005\}$.

Table 7: Gumbel copula

| marginal $t(3)$ | $\alpha_0 = 2$ | $\alpha_0 = 3.5$ | $\alpha_0 = 6$ | $\alpha_0 = 7$ |
|---|---|---|---|---|
| $\sqrt{lrvar(Y_t)}$ | 4.17103442 | 7.83155913 | 13.94437015 | 16.61133628 |
| $\sqrt{var\ h(Y_{t-1}, Y_t)}/\lvert A_3\rvert$ | 2.13638789 | 2.39868612 | 2.39310372 | 2.38050671 |

| marginal $t(5)$ | $\alpha_0 = 2$ | $\alpha_0 = 3.5$ | $\alpha_0 = 6$ | $\alpha_0 = 7$ |
|---|---|---|---|---|
| $\sqrt{lrvar(Y_t)}$ | 3.08007116 | 5.76228169 | 10.33584921 | 12.37445128 |
| $\sqrt{var\ h(Y_{t-1}, Y_t)}/\lvert A_3\rvert$ | 2.21904427 | 2.65616174 | 2.69937590 | 2.68964755 |

Simulation results show that MLE estimator $\hat{\beta}_n^{MLE}$ will outperform OLS estimator $\hat{\beta}_n^{OLS}$, especially when $Y_t$ has large tail dependence. Take Clayton copula marginal $t(5)$ for example, the long run variance of $Y_t$ is strictly increasing with the copula parameter($\alpha_0 : 2 \to 5 \to 10 \to 12$, $\omega_Y : 3.71 \to 8.08 \to 15.65 \to 18.61$), indicating bad performance of OLS estimator $\hat{\beta}_n^{OLS}$ when copula parameter is extreme. However, $\omega_1/|A_3|$ is roughly constant($\alpha_0 : 2 \to 5 \to 10 \to 12$, $\omega_1/|A_3| : 1.67 \to 1.39 \to 1.18 \to 1.14$). These results further illustrates another excuse for explosion of two-step sieve estimator and the advantage of joint sieve estimator under strong tail dependence, due to different precision of estimation for nonstationary filtering coefficient.

# 7 Empirical Application

## 7.1 Model

Academics, practitioners and individual investors have long been interested in understanding the value and usefulness of sell-side analysts' equity reports. In recent years, security analysts have been increasingly disclosing target prices in these reports, along with their stock recommendations and earnings forecasts. Theses target prices provide market participants with analysts' most concise and explicit statement on the magnitude of the firm's expected value. Brav and Lehavy [2003] analyzes the long-term behavior of market and target prices by establish a cointegration relationship between them:

$$Y_t = X_t\beta_0 + V_t$$

here $Y_t$ is stock price and $X_t$ is consensus stock price. Both time series follow a unit root pattern and there exists a cointegration relation $\beta_0$ such that the residuals $V_t = Y_t - X_t\beta_0$ are stationary.

When making prediction of $Y_t$ based on $X_t$, we also have past information $Y_{t-1}$ and $X_{t-1}$ available. Although $V_{t-1}$ is still not available as $\beta_0$ is not observed, we can get residuals $\hat{V}_{t-1}$ from our estimator $\hat{\beta}$. If we can research on the residual structure of $Y_t$ and get some useful inference of $V_t$ based on $\hat{V}_{t-1}$, a more efficient and precise prediction of $V_t$ (or its quantile) would be desirable. However, the prediction cannot be done until we specify the serial relationship of $V_t$. In this part, we furthermore utilize our copula technique to analyze dependence structure in the residuals derived from this cointegration to see whether the suggested dependency is consistent and robust.

Given copula structure $c(u_1, u_2; \alpha)$ and past information $\mathcal{F}_{t-1} = \sigma\{X_t, Y_{t-1}, X_{t-1}, Y_{t-2}, X_{t-2} \cdots\}$, the 5% quantile of stock price $Y_t$ is predicted as $X_t\hat{\beta} + \hat{V}$, here $\hat{V}$ is the solution of:

$$\int_{-\infty}^{\hat{V}} c(\hat{F}(Y_{t-1} - X_{t-1}\hat{\beta}), \hat{F}(y); \hat{\alpha})\hat{f}(y)\mathrm{d}y = 0.05 \quad (9)$$

and the conditional expectation of stock price $Y_t$ is predicted as:

$$X_t\hat{\beta} + \int_{-\infty}^{+\infty} c(\hat{F}(Y_{t-1} - X_{t-1}\hat{\beta}), \hat{F}(y); \hat{\alpha})\hat{f}(y)y\mathrm{d}y$$

$\hat{\alpha}$ is estimator for copula parameter, $\hat{\beta}$ is estimator for cointegration coefficient, $\hat{F}$ and $\hat{f}$ are nonparametric estimator for marginal distribution and density respectively. Better estimation results of these variables are attractive for value-at-risk calculation and stock price prediction conditional on consensus target price.

Stock price data is from CRSP[13], $Z_t$ is weekly stock close price. Target price data is from IBES[14], $X_t$ is consensus target price issued over the proceeding past range from 1 day to 90 days[15]. Both data are split adjusted. Intuitively, left side is closing price in one week, which we want to predict. On one day before, we had collected all target price available in 90 days then take an average. As each target price prediction is for 12 months, this consensus target price could be viewed as a proxy of analysts' consensus prediction of stock's one year profitability.

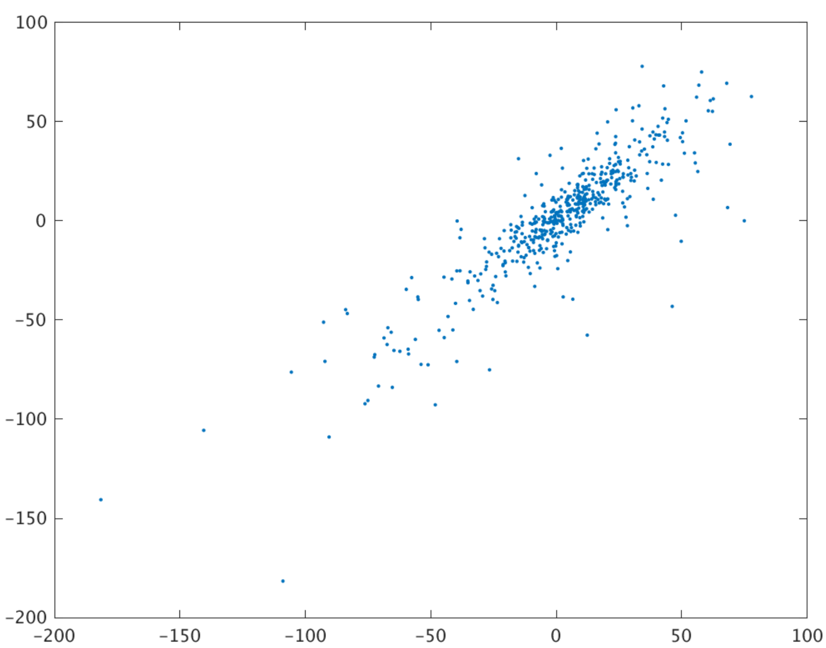
We consider stock price of Amazon incorporation(Nasdaq AMZN) in recent 10 years, from 2008 January to 2018 April (Apr. 20, 2018). Length of time series is 537. We first conduct augmented Dickey Fuller test for both stock weekly price and consensus target price. Both fail to reject the null hypothesis $H_0$ : existing unit root, with p-value larger than 0.99. Then we conduct Johansen cointegration test. Result shows that we should reject the null hypothesis $H_0$ : no integration, at 1% level significance level.[16]

Table 8: Cointegration test

| $H_0$ : no cointegration | test statistic | 10% | 5% | 1% |
|---|---|---|---|---|
| AMZN | 80.20 | 12.91 | 14.90 | 19.19 |

First, we run ordinary least square to get residuals $\hat{V}_t$ and make a plot of $(\hat{V}_{t-1}, \hat{V}_t)$, from which we can observe positive dependence clearly:

Figure 2: Plot of OLS residuals $(\hat{Y}_{t-1}, \hat{Y}_t)$, NASDAQ AMZN




[13]The Center for Research in Security Prices. Thanks to Boston College for purchasing the data.

[14]Institutional Brokers Estimate System. Thanks to Boston College for purchasing the data.

[15]We will select other days for robustness test in **Section** 7.4

[16]10%, 5%, 1% means critical value at corresponding significance level. Out test statistic 102.38 is larger than 1% significance level critical value 19.19. Hence we can reject the null hypothesis at 1% significance level.

We first try Gaussian copula to compare joint sieve method, two-step sieve method and three-step empirical method. Our choice of sieve is consistent with Section 6.2.3, Laguerre polynomials to order $K = 3$, 4 or 5. Selection of number of sieve terms $\hat{K}_n$ are based on small sample AIC of Burnham and Anderson [2003]: $\hat{K}_n = \arg\max_K \{L_n(\hat{\gamma}_n(K)) - \frac{K}{n-K-1}\}$. We also add a variance term $\sigma$ to control the scale:

$$f(y) = \frac{1}{\sigma}\left[\sum_{k=0}^{K_n} a_k \cdot \frac{L_k(|y|/\sigma)}{\sqrt{2}} \cdot e^{-\frac{x}{2}}\right]^2$$

hence we will estimate the standard variance $\sigma$ at the same time with $a_i$, $i = 0, 1, \cdots, 5$.

Results for Gaussian copula are based on data to September 2018. Length of time series is 507. The remaining 30 samples are left for out-of-sample prediction in next section.

Table 9: Gaussian copula, in-sample estimation

| | joint | two step | empirical |
|---|---|---|---|
| filter $\beta$ | 0.86143812 | 0.87177696 | 0.87177696 |
| copula $\alpha$ | 0.88745260 | 0.88753331 | 0.87059250 |

Estimation of copula parameter is similar between between four methods. Whereas two-step sieve method shows a little larger positive dependence. Meanwhile, our estimation of linear filtering coefficient from joint sieve method is also robust, which is quite close to OLS result(0.86143812 $\rightarrow$ 0.87177696).

## 7.2 Prediction

### 7.2.1 Conditional quantile

The estimated marginal distribution function $\hat{F}$ is automatically monotonic for both sieve method and empirical method. We can easily pass **Equation** 9 to step 2 in Section 6.1 for solving the conditional quantile in analytic expression.

First step, we compute the estimated residuals in past period as: $\hat{V}_t = Y_t - X_t\hat{\beta}$. Then we transfer $\hat{V}_{t-1}$ to uniform: $\hat{U}_{t-1} = \hat{F}(\hat{V}_{t-1})$. Let $C_1(\hat{U}_{t-1}, \hat{U}_t) = 0.05$ to solve $\hat{U}_t$. For Gaussian copula, we have an analytic expression from **Equation** 6 as $\hat{U}_t = \Phi(V)$, then the conditional quantile $\hat{Q}_\tau(Y_t|\mathcal{F}_{t-1})$ is computed as $X_t\hat{\beta} + \hat{F}^{-1}(V)$:

$$V = \Phi^{-1}(0.05 \times \sqrt{2\pi} \cdot \phi\left(\Phi^{-1}(\hat{U}_{t-1})\right) \cdot \exp\left(\left[\Phi^{-1}(\hat{U}_{t-1})\right]^2/2\right); \mu = \hat{\alpha}\Phi^{-1}(\hat{U}_{t-1}), \sigma^2 = 1 - \hat{\alpha}^2)$$

here $\phi(\cdot)$ and $\Phi(\cdot)$ are density and distribution functions for standard normal distribution, respectively. $\Phi(\cdot; \mu, \sigma^2)$ is the distribution function for normal distribution with mean $\mu$ and variance $\sigma^2$, $\Phi^{-1}(\cdot; \mu, \sigma^2)$ is the corresponding inverse.

As the conditional quantile is unobserved from the data, we conduct a simulation with similar copula parameter to test whether sieve methods have better prediction precision for conditional quantile comparing to empirical method.

We generate time series $Y_t = X_t\beta_0 + V_t$ for $t = 1, 2, \cdots, T = 600$. $(V_{t-1}, V_t)$ satisfies Gaussian copula for $\alpha_0 = 0.9$ together with marginal $t(3)$. $X_t$ follows unit root $X_t = X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, 1)$ independent of $V_t$. The first 500 data are used for estimation of $\hat{\beta}$, $\hat{\alpha}$ and $\hat{F}(\cdot)$. Remaining 100 data are for out-of-sample test of conditional quantile. We consider two fitness measures: mean square

$$\sqrt{\frac{1}{100}\sum_{t=501}^{600}\left(\hat{Q}(\hat{Y}_t|\mathcal{F}_{t-1}) - Q(Y_t|\mathcal{F}_{t-1})\right)^2}$$

and mean absolute value

$$\frac{1}{100}\sum_{t=501}^{600}\left|\hat{Q}(\hat{Y}_t|\mathcal{F}_{t-1}) - Q(Y_t|\mathcal{F}_{t-1})\right|$$

here $Q(Y_t|\mathcal{F}_{t-1})$ is the conditional quantile obtained with same procedure above but based on true parameter $\beta_0 = 1$, $\alpha_0 = 0.9$ and $F_0 \sim t(3)$.

For both criterions, joint sieve method provides the best prediction of conditional quantile, whereas three-step empirical method is the worst among three.

Table 10: Gaussian copula, conditional quantile

| | joint | two step | empirical |
|---|---|---|---|
| mean square | 0.06049159 | 0.10284027 | 0.25545852 |
| mean absolute value | 0.04910866 | 0.08871507 | 0.10042358 |

#### 7.2.2 Conditional expectation

Unlike conditional quantile where we have analytic expression, theoretically calculation of:

$$\int_{-\infty}^{+\infty} c(\hat{F}(Y_{t-1} - X_{t-1}\hat{\beta}), \hat{F}(y); \hat{\alpha})\hat{f}(y)y\mathrm{d}y$$

is time consuming. In this section, we consider an simulation approximation utilizing the special structure of Gaussian copula to check whether our sieve methods have a better predicting power.

Notice the fact that if $(V_{t-1}, V_t) \sim C(F(V_{t-1}), F(V_t); \alpha)$ for Gaussian copula, then the conditional distribution of $\Phi^{-1}(F(V_t))$ given $V_{t-1}$ is:

$$\Phi^{-1}(F(V_t))|V_{t-1} \sim N\left(\alpha \cdot \Phi^{-1}(F(V_{t-1})), 1 - \alpha^2\right)$$

Generate i.i.d. random variables $\xi_i \sim N(0,1)$, $i = 1, 2, \cdots, M = 20000$. Given estimation of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{F}$ (no matter based on sieve method or empirical method), an approximation of $V_t$ given past information could be:

$$\frac{1}{M}\sum_{i=1}^{M}\Phi\left(\hat{F}^{-1}(\hat{\alpha} \cdot \Phi^{-1}(F(Y_{t-1} - X_{t-1}\hat{\beta})) + \sqrt{1-\alpha^2}\xi_i)\right)$$

thus conditional expectation of $Z_t$ given $X_t$ is approximated as:

$$X_t\hat{\beta} + \frac{1}{M}\sum_{i=1}^{M}\Phi\left(\hat{F}^{-1}(\hat{\alpha} \cdot \Phi^{-1}(F(Y_{t-1} - X_{t-1}\hat{\beta})) + \sqrt{1-\alpha^2}\xi_i)\right)$$

For out-of-sample prediction, we consider two criterion functions: mean square and mean absolute value. Results based on 30 observations are attached below:

Table 11: Gaussian copula, out-of-sample prediction

| | joint | two step | empirical |
|---|---|---|---|
| mean square | 47.49788177 | 45.99579732 | 59.91085578 |
| mean absolute value | 40.65527030 | 38.83713746 | 46.77323927 |

Both joint sieve method and two-step sieve method have a better prediction power than three-step empirical method. It seems surprising at first glance that two-step sieve method is a little better than joint sieve method. However, this result is indeed consistent with our finite sample simulation rsults. For Gaussian copula, $\alpha_0 = 0.5$, cointegration, time length $T = 500$, when marginal is $t(3)$, relative efficiency is 1.072 for joint sieve estimator and 1.155 for two-step sieve estimator. When marginal is $t(5)$, relative efficiency is 1.021 for joint sieve estimator and 1.018 for two-step sieve estimator. Their performance are quite close and it is not amazing that two-step sieve method will outperform joint sieve method in a single experiment.

### 7.3 Tail dependence

However, Gaussian copula has zero tail dependence and hence cannot reflect effect of extreme event. Financial markets tend to exhibit tail dependence, especially lower tail dependence. For example, major stock returns in normal times have a correlation of, roughly 0.5, but in September/October 2008, some pairs had correlation of over 0.9. They were both falling massively. Copulas without tail dependence, such as Gaussian copula, will under estimate potential loss when market disaster happens. Wen et al. [2012], Jondeau and Rockinger [2006] and Nguyen and Bhatti [2012] are examples when people considers copulas rather than Gaussian copula for modeling tail dependence. Here we apply Clayton copula to model the lower tail dependence of residuals $V_t$.

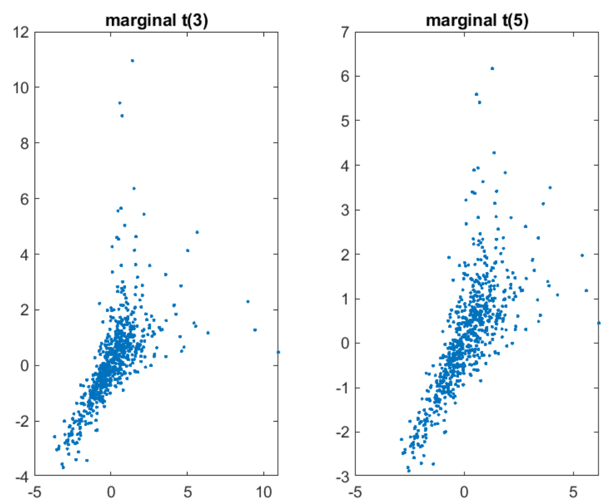
Result for Clayton copula:

Table 12: Clayton copula, in-sample estimation

| | joint | two step | empirical |
|---|---|---|---|
| filter $\beta$ | 1.11467093 | 0.87177696 | 0.87177696 |
| copula $\alpha$ | no convergence | no convergence | 3.52417476 |

Estimation of copula parameter will not converge for both joint sieve method and two-step sieve method. Whereas three-step empirical method and its modification show a larger tail dependence($\alpha = 3.5 \rightarrow \lambda_L = 82.03\%$). We make plot of $(V_{t-1}, V_t)$ from simulation of Clayton copula $\alpha_0 = 3.5$ together with marginal $t(3)$ and $t(5)$:

Figure 3: Clayton copula, $\alpha_0 = 3.5$, $(Y_{t-1}, Y_t)$




When $\alpha_0 = 3.5$, cluster of lower tail is quite obvious. Comparing to plot of OLS residuals $(\hat{V}_{t-1}, \hat{V}_t)$ where lower tail is still disperse, we believe that no evidence showing strong lower tail dependence. Hence result from three-step empirical method (and its modification) is misleading for this application and our sieve methods illustrates correctly that Clayton copula may not be a good choice to model serial dependence for the residual terms and whether we should utilize tail dependence in this setting is questionable.

We conduct a simulation to illustrate the above results. Suppose the true model is generated through Gaussian copula but we use Clayton copula to fit the data. We want to show that the above empirical results are not coincident but with some certainty, that empirical method will lead to a misleading result for sure and sieve method will diverge with large probability.

$Y_t = X_t\beta_0 + V_t$ for $t = 1, 2, \cdots, T = 500$. $(V_{t-1}, V_t)$ satisfies Gaussian copula for $\alpha_0 = 0.9$ together with marginal $t(3)$. $X_t$ follows unit root $X_t = X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, 1)$ independent of

$V_t$. Results based on Clayton copula are attached below, showing that joint sieve estimator and two-step sieve estimator tend to explode while empirical estimator is quite stable around 3, which is a misleading result as we state above.

Table 13: Clayton copula, true data from Gaussian copula

| | joint | two step | empirical |
|---|---|---|---|
| mean | 55.48412474 | 70.36815328 | 2.74811295 |
| median | 35.43051811 | 44.86515709 | 2.71396259 |
| std | 75.62500201 | 79.59951726 | 0.42425493 |
| 2.5% | 2.99817449 | 3.19520767 | 2.00109238 |
| 97.5% | 289.89230810 | 315.23311081 | 3.65729449 |

### 7.4 Robustness

The above results are based on consensus target price average in 90 days. We also consider other choices for robustness, like 75 days and 105 days, difference in around two weeks. Both new target price series pass the augmented Dickey Fuller test with p-value greater than 0.99. Conclusion are similar and we attach detail output in **Section** 13.

All results here indicate that if people want to apply our model to characterize the unobserved stationary residuals as parametric copula together with nonparametric marginal, sieve method could be a better choice than empirical method.

## 8 Conclusion and extensions

This paper considers estimation of copula based dynamic semiparametric models coupled with nonstationary filtration. A two-step sieve method is proposed and asymptotic properties of the proposed estimators are developed. We show that its limiting distribution is not affected by nonstationary structure if both the marginal and copula are symmetric. In the absence of symmetry, the limiting distributions are usually non-normal due to nonstationarity, and the impact of the preliminary filtration is increasing with the strength of asymmetric tail dependence. Simulation results indicate that the tail dependence brings a finite sample bias in the two-step sieve estimator. For this reason, a joint sieve estimator is also proposed and studied. Monte Carlo simulation demonstrates that the joint estimation is superior in all cases than two-step estimators. An empirical estimation for cointegration between weekly stock price and consensus target price to highlight the theoretical finding. The results are important for value-at-risk calculation and stock price prediction conditional on consensus target price.

**EXTENSION.** When nonstationarity exists in objective function, identification is quite hard without some higher order assumption (see e.g. De Jong [2002]). Moreover, the sieve structure will make the problem quite complex. Hence theoretical property of joint sieve method is still unknown. For example, when both marginal and copula are symmetric, how can we derive the convergence rate of linear coefficient estimator; when symmetric is violated (e.g. Clayton copula), what is the limiting distribution of joint sieve estimator?

Another extension would be inference. Our model fully characterizes the joint probability distribution, making sieve likelihood ratio test available as in Chen et al. [2009]. Simulations of bootstrap likelihood ratio test on joint sieve estimation have been done and the size is great. Theoretical validation of this inference procedure is of practical value for further research.

## References

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

Donald WK Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856, 1993.

Donald WK Andrews and Hong-Yuan Chen. Approximately median-unbiased estimation of autoregressive models. *Journal of Business & Economic Statistics*, 12(2):187–204, 1994.

Donald WK Andrews and Werner Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, pages 1383–1414, 1994.

Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78, 1998.

Brendan K Beare. Copulas and temporal dependence. *Econometrica*, 78(1):395–410, 2010.

Brendan K Beare. Archimedean copulas and temporal dependence. *Econometric Theory*, 28(6): 1165–1185, 2012.

Alon Brav and Reuven Lehavy. An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics. *The Journal of Finance*, 58(5):1933–1968, 2003.

Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

Xiaohong Chen. Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics*, 6:5549–5632, 2007.

Xiaohong Chen and Yanqin Fan. Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335, 2006.

Xiaohong Chen and Xiaotong Shen. Sieve extremum estimates for weakly dependent data. *Econometrica*, pages 289–314, 1998.

Xiaohong Chen and Zhijie Xiao. Copula-based time series with filtered nonstationarity. 2016.

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

Xiaohong Chen, Wei Biao Wu, and Yanping Yi. Efficient estimation of copula-based semiparametric markov models. *The Annals of Statistics*, pages 4214–4253, 2009.

Ciprian M Crainiceanu and Timothy J Vogelsang. Nonmonotonic power for tests of a mean shift in a time series. *Journal of Statistical Computation and Simulation*, 77(6):457–476, 2007.

William F Darsow, Bao Nguyen, Elwood T Olsen, et al. Copulas and markov processes. *Illinois Journal of Mathematics*, 36(4):600–642, 1992.

Robert M De Jong. Nonlinear minimization estimators in the presence of cointegrating relations. *Journal of Econometrics*, 110(2):241–259, 2002.

Ai Deng and Pierre Perron. The limit distribution of the cusum of squares test under general mixing conditions. *Econometric Theory*, 24(3):809–822, 2008.

Francis X Diebold and Celia Chen. Testing structural stability with endogenous breakpoint a size comparison of analytic and bootstrap procedures. *Journal of Econometrics*, 70(1):221–241, 1996.

Paul Doukhan, Jean-David Fermanian, Gabriel Lang, et al. *Copulas of a vector-valued stationary weakly dependent process*. INSEE, 2005.

Ray C Fair. Computing median unbiased estimates in macroeconometric models. *Journal of Applied Econometrics*, 11(4):431–435, 1996.

Evarist Giné and Joel Zinn. Bootstrapping general empirical measures. *The Annals of Probability*, pages 851–869, 1990.

Clive WJ Granger. Time series concepts for conditional distributions. *Oxford Bulletin of Economics and Statistics*, 65:689–701, 2003.

Bruce E Hansen. Testing for structural change in conditional models. *Journal of Econometrics*, 97 (1):93–115, 2000.

Bruce E Hansen. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(03):726–748, 2008.

Rustam Ibragimov. Copula-based characterizations for higher order markov processes. *Econometric Theory*, 25(03):819–846, 2009.

Harry Joe. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

Eric Jondeau and Michael Rockinger. The copula-garch model of conditional dependencies: An international stock market application. *Journal of international money and finance*, 25(5):827–853, 2006.

Ted Juhl and Zhijie Xiao. Tests for changing mean with monotonic power. *Journal of Econometrics*, 148(1):14–24, 2009.

Lutz Kilian. Accounting for lag order uncertainty in autoregressions: the endogenous lag order bootstrap algorithm. *Journal of Time Series Analysis*, 19(5):531–548, 1998.

Martial Longla and Magda Peligrad. Some aspects of modeling dependence in copula-based markov chains. *Journal of Multivariate Analysis*, 111:234–240, 2012.

RB Nelson. An introduction to copulas, lectures notes in statistics vol. 39, 1999.

Whitney K Newey and Kenneth D West. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, pages 777–787, 1987.

Cuong C Nguyen and M Ishaq Bhatti. Copula model dependency between oil prices and stock markets: Evidence from china and vietnam. *Journal of International Financial Markets, Institutions and Money*, 22(4):758–773, 2012.

Joon Y Park. An invariance principle for sieve bootstrap in time series. *Econometric Theory*, 18 (02):469–490, 2002.

Andrew J Patton. A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18, 2012.

Pierre Perron. *A test for changes in a polynomial trend function for a dynamic time series*. Université de Montréal, Centre de recherche et développement en économique, 1991.

Pierre Perron et al. Dealing with structural breaks. *Palgrave handbook of econometrics*, 1(2):278–352, 2006.

Werner Ploberger and Walter Krämer. The cusum test with ols residuals. *Econometrica: Journal of the Econometric Society*, pages 271–285, 1992.

Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954, 1988.

M Sklar. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8: 229–231, 1959.

Charles J Stone. Optimal global rates of convergence for nonparametric regression. *The annals of statistics*, pages 1040–1053, 1982.

H Victor, Rustam Ibragimov, Shaturgun Sharakhmetov, et al. Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series. In *Optimality*, pages 183–209. Institute of Mathematical Statistics, 2006.

Timothy J Vogelsang. Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics*, 88(2):283–299, 1999.

Xiaoqian Wen, Yu Wei, and Dengshi Huang. Measuring contagion between energy market and stock market during financial crisis: A copula approach. *Energy Economics*, 34(5):1435–1446, 2012.

# 9 Technical Appendix

## 9.1 Proof of Lemma 1

*Proof.* For any positive $\varepsilon > 0$, we consider $\mathbb{P}\left(\sup_{\alpha\in\mathcal{A},f\in\mathcal{F}_n} |Q_n(\alpha,f) - Q(\alpha,f)| > \varepsilon\right)$.

Let $\mathsf{b}_n^* = \sqrt{n}(X_1'(\beta_0 - \hat{\beta}), X_2'(\beta_0 - \hat{\beta}), \cdots, X_n'(\beta_0 - \hat{\beta}))$. Here $\hat{\beta}$ is the estimator satisfying Assumption 2.

Then $Q_n(\alpha, f) = L_n(\gamma, \mathsf{b}_n^*)$. Definition of $L_n$ is referred to equation (4).

Now we analyze $\|\mathsf{b}_n^*\|_1$:

$$\begin{aligned}
\|\mathsf{b}_n^*\|_1 &= \max_{1\le t\le n} \sqrt{n}|X_t'(\beta_0 - \hat{\beta})| \\
&= \max_{1\le t\le n} |X_t' G_n^{-1} \cdot \sqrt{n} G_n(\beta_0 - \hat{\beta})| \\
&\le \max_{1\le t\le n} \|G_n^{-1} X_t\|_2 \cdot \|\sqrt{n} G_n(\hat{\beta} - \beta_0)\|_2
\end{aligned}$$

The last step is because for two vectors $\alpha$ and $\beta$, $|\alpha^T\beta| \le \sqrt{\alpha^T\alpha} \cdot \sqrt{\beta^T\beta}$ by Cauchy Schwarz inequality.

Following **Assumption** 2, $G_n^{-1}X_{[nr]} \Rightarrow X(r)$, then by continuous mapping theorem, $\max_{1\le t\le n} \|G_n^{-1}X_t\|_2 \Rightarrow \sup_{0<r<1} \|X(r)\|_2$. Also $\sqrt{n}G_n(\hat{\beta} - \beta_0) = O_p(1)$, then $\|\sqrt{n}G_n(\hat{\beta} - \beta_0)\|_2 = O_p(1)$. Thus $\|\mathsf{b}_n^*\|_1 = O_p(1)$.

By definition of $O_p(1)$, for any positive $\delta > 0$, there exists $B > 0$ and $N_0$ such that $\mathbb{P}(\|\mathsf{b}_n^*\|_1 > B) \le \delta$ when $n \ge N_0$.

Then when $n \ge N_0$, the event $\sup_{\alpha\in\mathcal{A},f\in\mathcal{F}_n} |L_n(\gamma, \mathsf{b}_n^*) - Q(\alpha, f)| > \varepsilon$ implies either $\|\mathsf{b}_n^*\|_1 > B$ or $\|\mathsf{b}_n^*\|_1 \le B$, $\sup_{\beta\in\mathcal{B},f\in\mathcal{F}_n} |L_n(\gamma, \mathsf{b}_n^*) - Q(\beta, f)| > \varepsilon$.

The probability of first event is smaller than $\delta$. The probability of second event is smaller than:

$$\mathbb{P}\left(\sup_{\substack{\mathsf{b}\in\mathbb{R}^n, \|\mathsf{b}\|_1\le B \\ \gamma\in\Gamma_n}} |L_n(\gamma, \mathsf{b}) - Q(\beta, f)| > \varepsilon\right)$$

Notice this term tends to zero when $n \to \infty$ following **Assumption** 8. We have:

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{\alpha\in\mathcal{A},f\in\mathcal{F}_n} |Q_n(\alpha,f) - Q(\alpha,f)| > \varepsilon\right) \le \delta$$

Choice of $\delta > 0$ is arbitrary. Hence we have

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{\alpha\in\mathcal{A},f\in\mathcal{F}_n} |Q_n(\alpha,f) - Q(\alpha,f)| > \varepsilon\right) = 0$$

Thus uniform continuity $\sup_{\alpha\in\mathcal{A},f\in\mathcal{F}_n} |Q_n(\alpha,f) - Q(\alpha,f)| = o_p(1)$ is verified. □

## 9.2 Proof of Theorem 1

*Proof.* First notice $\hat{\gamma}_n = (\hat{\alpha}_n, \hat{f}_n)$ will maximize $Q_n(\alpha, f)$ in $\Gamma_n$, for any $\varepsilon > 0$, we have:

$$\mathbb{P}\left(\|\hat{\gamma}_n - \gamma_0\|_c > \varepsilon\right) \leq \mathbb{P}\left(\sup_{\substack{\gamma\in\Gamma_n \\ \|\gamma-\gamma_0\|_c>\varepsilon}} Q_n(\alpha, f) \geq Q_n(\alpha_0, \pi_n f_0)\right)$$

Following triangle inequality:

$$\begin{aligned}\sup_{\substack{\gamma\in\Gamma_n \\ \|\gamma-\gamma_0\|_c>\varepsilon}} Q_n(\alpha, f) &\leq \sup_{\substack{\gamma\in\Gamma_n \\ \|\gamma-\gamma_0\|_c>\varepsilon}} [Q(\alpha, f) + |Q_n(\alpha, f) - Q(\alpha, f)|] \\ &\leq \sup_{\substack{\gamma\in\Gamma_n \\ \|\gamma-\gamma_0\|_c>\varepsilon}} Q(\alpha, f) + \sup_{\gamma\in\Gamma_n} |Q_n(\alpha, f) - Q(\alpha, f)|\end{aligned}$$

$$\begin{aligned}Q_n(\alpha_0, \Pi_n f_0) &\geq Q(\alpha_0, \Pi_n f_0) - |Q_n(\alpha_0, \Pi_n f_0) - Q(\alpha_0, \Pi_n f_0)| \\ &\geq Q(\alpha_0, \Pi_n f_0) - \sup_{\gamma\in\Gamma_n} |Q_n(\alpha, f) - Q(\alpha, f)|\end{aligned}$$

Combine above equations, we have:

$$\sup_{\substack{\gamma\in\Gamma_n \\ \|\gamma-\gamma_0\|_c>\varepsilon}} Q_n(\alpha, f) \geq Q_n(\alpha_0, \pi_n f_0)$$

implies

$$\sup_{\substack{\gamma\in\Gamma_n \\ \|\gamma-\gamma_0\|_c>\varepsilon}} Q(\alpha, f) \geq Q(\alpha_0, \Pi_n f_0) - 2\sup_{\gamma\in\Gamma_n} |Q_n(\alpha, f) - Q(\alpha, f)|$$

further implies

$$\begin{aligned}2\sup_{\gamma\in\Gamma_n} |Q_n(\alpha, f) - Q(\alpha, f)| + [Q(\alpha_0, f_0) - Q(\alpha_0, \Pi_n f_0)]& \\ \geq Q(\alpha_0, f_0) - \sup_{\alpha\in\mathcal{A}, f\in\mathcal{F}_n} Q(\alpha, f)&\end{aligned} \tag{10}$$

Following **Assumption** 5, right hand side: $Q(\alpha_0, f_0) - \sup_{\alpha\in\mathcal{A}, f\in\mathcal{F}_n} Q(\alpha, f)$ is larger than $\eta(\varepsilon) > 0$. Following **Assumption** 7 and uniform continuity $\sup_{\alpha\in\mathcal{A}, f\in\mathcal{F}_n} |Q_n(\alpha, f) - Q(\alpha, f)| = o_p(1)$, left hand side: $2\sup_{\gamma\in\Gamma_n} |Q_n(\alpha, f) - Q(\alpha, f)| + [Q(\alpha_0, f_0) - Q(\alpha_0, \Pi_n f_0)]$ is $o_p(1) + o(1)$, which is still $o_p(1)$.

By definition of convergence in probability, the probability of this event (i.e. equation (10)) will tend to zero when $n \to \infty$.

Hence $\mathbb{P}\left(\|\hat{\gamma} - \gamma_0\|_c > \varepsilon\right)$ tends to zero when $n$ towards to infinity. Thus $\|\hat{\gamma} - \gamma_0\| = o_p(1)$.

Furthermore, we have $\|\hat{\gamma} - \gamma_0\|_c = \|\hat{\alpha} - \alpha_0\|_2 + \|\hat{f} - f_0\|_c = o_p(1)$, this implies $\hat{\alpha} \xrightarrow{p} \alpha_0$. □

## 9.3 Proof of Theorem 2

*Proof.* Let $r(\gamma, \gamma_0) = l(\gamma, \hat{V}_{t-1}, \hat{V}_t) - l(\gamma_0, V_{t-1}, V_t) - \frac{\partial l(\gamma_0, V_t)}{\partial \gamma'}[\gamma - \gamma_0]$.

Denote $r^*(\gamma, \gamma_0) = l(\gamma, V_{t-1}, V_t) - l(\gamma_0, V_{t-1}, V_t) - \frac{\partial l(\gamma_0, V_{t-1}, V_t)}{\partial \gamma'}[\gamma - \gamma_0]$.

Then for a sequence $\varepsilon_n$ satisfying $\tilde{\varepsilon}_n < \sqrt{n}\varepsilon_n = o(1)$, where $\tilde{\varepsilon}_n$ is defined in **Assumption** 19, we have:

$$
\begin{aligned}
0 &\leq Q_n(\hat{\gamma}_n) - Q_n(\hat{\gamma}_n \pm \varepsilon_n \Pi_n v^*) \\
&= \frac{1}{n}\sum_{t=2}^{n}[l(\hat{\gamma}_n, \hat{V}_{t-1}, \hat{V}_t) - l(\hat{\gamma}_n \pm \varepsilon_n \Pi_n v^*, \hat{V}_{t-1}, \hat{V}_t) \\
&\quad + \frac{1}{n}[\log \hat{f}_n(V_1) - \log(\hat{f}_n(V_1) \pm \varepsilon_n \Pi_n v_f^*(V_1))] \\
&= \mp\varepsilon_n \cdot \frac{\partial l(\gamma_0, V_{t-1}, V_t)}{\partial \gamma'}[\Pi_n v^*] + \mu_n(r(\hat{\gamma}_n, \gamma_0) - r(\hat{\gamma}_n \pm \varepsilon_n \Pi_n v^*, \gamma_0)) \\
&\quad + \mathbb{E}_0 \frac{1}{n}\sum_{t=2}^{n}[r^*(\hat{\gamma}_n, \gamma_0) - r^*(\hat{\gamma} \pm \varepsilon_n \Pi_n v^*, \gamma_0)] \\
&\quad + \mathbb{E}_0 \frac{1}{n}\sum_{t=2}^{n}\Big[l(\hat{\gamma}_n, \hat{V}_{t-1}, \hat{V}_t) - l(\hat{\gamma}_n \pm \varepsilon_n \Pi_n v^*, \hat{V}_{t-1}, \hat{V}_t) \\
&\quad - l(\hat{\gamma}_n, V_{t-1}, V_t) + l(\hat{\gamma}_n \pm \varepsilon_n \Pi_n v^*, V_{t-1}, V_t)\Big] \\
&\quad + \frac{1}{n}[\log \hat{f}_n(V_1) - \log(\hat{f}_n(V_1) \pm \varepsilon_n \Pi_n v_f^*(V_1))] \\
&:= A_1 + A_2 + A_3 + A_4 + \varepsilon_n \cdot o_p(n^{-\frac{1}{2}})
\end{aligned}
$$

The first term $A_1 := \mp\varepsilon_n \cdot \frac{\partial l(\gamma_0, Y_{t-1}, Y_t)}{\partial \gamma'}[\Pi_n v^*] = \mp\varepsilon_n \cdot \frac{\partial l(\gamma_0, V_{t-1}, V_t)}{\partial \gamma'}[v^*] + \varepsilon_n \cdot o_p(n^{-\frac{1}{2}})$ from Chen et al. [2009] proof **Theorem** 4.1 Claim 1.

The second term $A_2 := \mu_n(r(\hat{\gamma}_n, \gamma_0) - r(\hat{\gamma} \pm \varepsilon_n \Pi_n v^*, \gamma_0)) = \varepsilon_n \cdot o_p(n^{-\frac{1}{2}})$ from **Assumption** 20.

The third term $A_3 := \mathbb{E}\frac{1}{n}\sum_{t=2}^{n}[r^*(\hat{\gamma}_n, \gamma_0) - r^*(\hat{\gamma}_n \pm \varepsilon_n \Pi_n v^*, \gamma_0)] = \pm\varepsilon_n \times < \hat{\gamma}_n - \gamma_0, v^* > + \varepsilon_n \times o_p(n^{-1/2})$ from proof of Chen et al. [2009] **Theorem** 4.1 Claim 3.

The last term $A_4 := \mathbb{E}_0 \frac{1}{n}\sum_{t=2}^{n}[l(\hat{\gamma}_n, \hat{V}_{t-1}, \hat{V}_t) - l(\hat{\gamma}_n \pm \varepsilon_n \Pi_n v^*, \hat{V}_{t-1}, \hat{V}_t) - l(\hat{\gamma}_n, V_{t-1}, V_t) + l(\hat{\gamma}_n \pm \varepsilon_n \Pi_n v^*, V_{t-1}, V_t)]$ could be simplified to $\mp\varepsilon_n \cdot F(\gamma_0, v^*) \cdot \int_0^1 X(r)\mathrm{d}r \cdot \xi + \varepsilon_n \cdot o_p(1/\sqrt{n})$ due to **Assumption** 21, 21, 18 and 2.

Combine the above terms, we get:

$$
\sqrt{n}(\hat{\alpha}_n - \alpha_0) \Rightarrow -F(\gamma_0, v^*) \times \int_0^1 X(r)dr \times \xi + N(0, \|v^*\|^2)
$$

Here $N(0, \|v^*\|^2)$ is the limiting distribution of $\sqrt{n} \cdot \frac{\partial l(\gamma_0, V_{t-1}, V_t)}{\partial \gamma'}[v^*]$, see Chen et al. [2009] proof of **Theorem** 4.1.

$F(\gamma_0, v^*)$ is defined as $\mathbb{E}\frac{\partial^2 l(\gamma_0, V_{t-1}, V_t)}{\partial \gamma \partial y_1}[v^*] + \mathbb{E}\frac{\partial^2 l(\gamma_0, V_{t-1}, V_t)}{\partial \gamma \partial y_2}[v^*]$. The detail expression is:

$$
\begin{aligned}
& \mathbb{E}\frac{\dot{v}_f^*(V_t) \cdot f_0(V_t) - \dot{f}_0(V_t) \cdot v_f^*(V_t)}{[f_0(V_t)]^2} \\
+ \quad & \mathbb{E}c_{11}(U_{t-1}, U_t; \alpha_0) \cdot v_F^*(V_{t-1}) \cdot f_0(V_{t-1}) + \mathbb{E}c_{12}(U_{t-1}, U_t; \alpha_0) \cdot v_F^*(V_{t-1}) \cdot f_0(V_t) \\
+ \quad & \mathbb{E}c_1(U_{t-1}, U_t; \alpha_0) \cdot v_f^*(V_{t-1}) + \mathbb{E}c_2(U_{t-1}, U_t; \alpha_0) \cdot v_f^*(V_t) \\
+ \quad & \mathbb{E}c_{12}(U_{t-1}, U_t; \alpha_0) \cdot v_F^*(V_t) \cdot f_0(Y_{t-1}) + \mathbb{E}c_{22}(U_{t-1}, U_t; \alpha_0) \cdot v_F^*(V_t) \cdot f_0(V_t) \\
+ \quad & \mathbb{E}c_{1\alpha}(U_{t-1}, U_t; \alpha_0) \cdot f_0(V_{t-1}) \cdot v_\alpha^* + \mathbb{E}c_{2\alpha}(U_{t-1}, U_t; \alpha_0) \cdot f_0(V_t) \cdot v_\alpha^*
\end{aligned}
$$

Here $v_F^*(x) = \int_{-\infty}^{x} v_f^*(y)\mathrm{d}y$, $U_{t-1} = F_0(V_{t-1})$, $U_t = F_0(V_t)$.

$c_1(u_1, u_2, \alpha) = \frac{\partial \log c(u_1, u_2, \alpha)}{\partial u_1}$, $c_2(u_1, u_2, \alpha) = \frac{\partial \log c(u_1, u_2, \alpha)}{\partial u_2}$.

$c_\alpha(u_1, u_2, \alpha) = \frac{\partial \log c(u_1, u_2, \alpha)}{\partial \beta}$, $c_{11}(u_1, u_2, \alpha) = \frac{\partial^2 \log c(u_1, u_2, \alpha)}{\partial u_1^2}$.

$c_{12}(u_1, u_2, \alpha) = \frac{\partial^2 \log c(u_1, u_2, \alpha)}{\partial u_1 \partial u_2}$, $c_{22}(u_1, u_2, \alpha) = \frac{\partial^2 \log c(u_1, u_2, \alpha)}{\partial u_2^2}$.

$c_{1\alpha}(u_1, u_2, \alpha) = \frac{\partial^2 \log c(u_1, u_2, \alpha)}{\partial u_1 \partial \alpha}$, $c_{2\alpha}(u_1, u_2, \alpha) = \frac{\partial^2 \log c(u_1, u_2, \alpha)}{\partial u_2 \partial \alpha}$. □

*Remark* 17. Definition of $l(\gamma, y_1, y_2)$ is:

$$l(\gamma, y_1, y_2) = \log f(y_2) + \log c\left(\int_{-\infty}^{y_1} f(y)dy, \int_{-\infty}^{y_2} f(y)dy; \alpha\right)$$

Notice that $v^*$ has two parts: $v^* = (v^*_\alpha, v^*_f)$, then:

$$\begin{aligned} l(\gamma_0 + \eta v^*, y_1, y_2) &= \log\left[f_0(y_2) + \eta v^*_f(y_2)\right] + \\ &\log c\left(\int_{-\infty}^{y_1} f_0(y)\mathrm{d}y + \eta\int_{-\infty}^{y_1} v^*_f(y)\mathrm{d}y, \int_{-\infty}^{y_2} f_0(y)\mathrm{d}y + \eta\int_{-\infty}^{y_2} v^*_f(y)\mathrm{d}y; \alpha_0 + \eta v^*_\alpha\right) \end{aligned}$$

From equation 5, define $F_0(x) := \int_{-\infty}^{x} f_0(y)\mathrm{d}y$ and $v^*_F(x) = \int_{-\infty}^{x} v^*_f(y)\mathrm{d}y$, we have:

$$\begin{aligned} \frac{\partial l(\gamma, y_1, y_2)}{\partial \gamma'}[v^*] &= \frac{v^*_f(y_2)}{f(y_2)} + c_1\left(F_0(y_1), F_0(y_2); \alpha_0\right) \cdot v^*_F(y_1) + \\ &c_2\left(F_0(y_1), F_0(y_2); \alpha_0\right) \cdot v^*_F(y_2) + c_\alpha\left(F_0(y_1), F_0(y_2); \alpha_0\right) \cdot v^*_\alpha \end{aligned}$$

Take derivative to $y_1$:

$$\begin{aligned} \frac{\partial^2 l(\gamma_0, y_1, y_2)}{\partial\gamma\partial y_1}[v^*] &= c_{11}\left(F_0(y_1), F_0(y_2); \alpha_0\right) \cdot f_0(y_1) \cdot v^*_F(y_1) + \\ &c_1\left(F_0(y_1), F_0(y_2); \alpha_0\right) \cdot v^*_f(y_1) + c_{12}\left(F_0(y_1), F_0(y_2); \alpha_0\right) \cdot f_0(y_1) \cdot v^*_F(y_2) + \\ &c_{1\alpha}\left(F_0(y_1), F_0(y_2); \alpha_0\right) \cdot f_0(y_1) \cdot v^*_\alpha \end{aligned}$$

Take derivative to $y_2$:

$$\begin{aligned} \frac{\partial^2 l(\gamma_0, y_1, y_2)}{\partial\gamma\partial y_2}[v^*] &= \frac{\dot{v}^*_f(y_2) \cdot f_0(y_2) - \dot{f}_0(y_2) \cdot v^*_f(y_2)}{\left[f(y_2)\right]^2} + c_{11}\left(F_0(y_1), F_0(y_2); \alpha_0\right) \cdot f_0(y_1) \cdot v^*_F(y_1) + \\ &c_1\left(F_0(y_1), F_0(y_2); \alpha_0\right) \cdot v^*_f(y_1) + c_{12}\left(F_0(y_1), F_0(y_2); \alpha_0\right) \cdot f_0(y_1) \cdot v^*_F(y_2) + \\ &c_{1\alpha}\left(F_0(y_1), F_0(y_2); \alpha_0\right) \cdot f_0(y_1) \cdot v^*_\alpha \end{aligned}$$

Replace $y_1$ with $V_{t-1}$, $y_2$ with $Y_t$. Also denote $U_{t-1} = F_0(V_{t-1})$, $U_t = F_0(V_t)$, we have:

$$\begin{aligned} & \mathbb{E}\frac{\partial^2 l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma\partial y_1}[v^*] + \mathbb{E}\frac{\partial^2 l(\gamma_0, V_{t-1}, V_t)}{\partial\gamma\partial y_2}[v^*] \\ = \; & \mathbb{E}\frac{\dot{v}^*_f(V_t) \cdot f_0(V_t) - \dot{f}_0(V_t) \cdot v^*_f(V_t)}{[f_0(V_t)]^2} \\ + \; & \mathbb{E}c_{11}(U_{t-1}, U_t; \alpha_0) \cdot v^*_F(V_{t-1}) \cdot f_0(V_{t-1}) + \mathbb{E}c_{12}(U_{t-1}, U_t; \alpha_0) \cdot v^*_F(V_{t-1}) \cdot f_0(V_t) \\ + \; & \mathbb{E}c_1(U_{t-1}, U_t; \alpha_0) \cdot v^*_f(V_{t-1}) + \mathbb{E}c_2(U_{t-1}, U_t; \alpha_0) \cdot v^*_f(V_t) \\ + \; & \mathbb{E}c_{12}(U_{t-1}, U_t; \alpha_0) \cdot v^*_F(V_t) \cdot f_0(Y_{t-1}) + \mathbb{E}c_{22}(U_{t-1}, U_t; \alpha_0) \cdot v^*_F(V_t) \cdot f_0(V_t) \\ + \; & \mathbb{E}c_{1\alpha}(U_{t-1}, U_t; \alpha_0) \cdot f_0(V_{t-1}) \cdot v^*_\alpha + \mathbb{E}c_{2\alpha}(U_{t-1}, U_t; \alpha_0) \cdot f_0(V_t) \cdot v^*_\alpha \end{aligned}$$

A more detailed proof is omitted due to the length of the paper but is available upon request.

# 10 Simulation for $F(\gamma_0, v^*)$ for Clayton and Gumbel copula

## 10.1 $K_N = 20$

Table 14: Clayton copula

| Clayton | $\alpha_0 = 2$ | $\alpha_0 = 5$ | $\alpha_0 = 10$ | $\alpha_0 = 12$ |
|---|---|---|---|---|
| $t(3)$ | 1.10991432 | 3.79487138 | 9.41862191 | 11.63949125 |
| left | 1.08583010 | 3.71753452 | 9.20862531 | 11.35345182 |
| right | 1.13399854 | 3.87220824 | 9.62861852 | 11.92553068 |
| $t(5)$ | 1.49694402 | 4.39367369 | 10.04265303 | 12.29797097 |
| left | 1.46836395 | 4.29856560 | 9.79114444 | 11.96080819 |
| right | 1.52552410 | 4.48878179 | 10.29416161 | 12.63513376 |

Table 15: Gumbel copula

| Gumbel | $\alpha_0 = 2$ | $\alpha_0 = 3.5$ | $\alpha_0 = 6$ | $\alpha_0 = 7$ |
|---|---|---|---|---|
| $t(3)$ | -0.35962340 | -1.21945840 | -3.27915940 | -4.14579532 |
| left | -0.37250963 | -1.26587638 | -3.43938546 | -4.37154749 |
| right | -0.34673717 | -1.17304041 | -3.11893333 | -3.92004315 |
| $t(5)$ | -0.45954655 | -1.26717679 | -2.97835988 | -3.69481887 |
| left | -0.47427085 | -1.32361721 | -3.18271507 | -3.98469622 |
| right | -0.44482225 | -1.21073637 | -2.77400469 | -3.40494152 |

## 10.2 $K_N = 22$

Table 16: Clayton copula

| Clayton | $\alpha_0 = 2$ | $\alpha_0 = 5$ | $\alpha_0 = 10$ | $\alpha_0 = 12$ |
|---|---|---|---|---|
| $t(3)$ | 1.11351056 | 3.80699264 | 9.44613433 | 11.67302713 |
| left | 1.08939602 | 3.72974093 | 9.23669040 | 11.38772307 |
| right | 1.13762510 | 3.88424434 | 9.65557826 | 11.95833119 |
| $t(5)$ | 1.50173782 | 4.40824070 | 10.07445166 | 12.33684502 |
| left | 1.47318417 | 4.31365127 | 9.82503578 | 12.00244720 |
| right | 1.53029147 | 4.50283014 | 10.32386755 | 12.67124284 |

Table 17: Gumbel copula

| Gumbel | $\alpha_0 = 2$ | $\alpha_0 = 3.5$ | $\alpha_0 = 6$ | $\alpha_0 = 7$ |
|---|---|---|---|---|
| $t(3)$ | -0.36122994 | -1.23548582 | -3.36153808 | -4.26244901 |
| left | -0.37415915 | -1.28229997 | -3.52443241 | -4.49254466 |
| right | -0.34830072 | -1.18867166 | -3.19864375 | -4.03235336 |
| $t(5)$ | -0.46140088 | -1.28162959 | -3.04303470 | -3.78464738 |
| left | -0.47615549 | -1.33832683 | -3.24977658 | -4.07869208 |
| right | -0.44664627 | -1.22493234 | -2.83629282 | -3.49060268 |

# 11 Square root sieve

## 11.1 Time trend

Table 18: Clayton copula, $\alpha_0 = 12$; $X_t$ time trend; marginal $t(3)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | P |
|---|---|---|---|---|---|
| mean | 12.390020 | 12.091684 | 12.340644 | 6.104206 | 16.029835 |
| bias | 0.390020 | 0.091684 | 0.340644 | -5.895794 | 4.029835 |
| std | 2.424539 | 6.170698 | 2.177290 | 1.875286 | 4.300961 |
| MSE | 6.030506 | 38.085919 | 4.856632 | 38.277081 | 34.737834 |
| relative | 1.241705 | 7.842044 | 1 | 7.881405 | 7.152659 |
| 2.5% | 9.455326 | 5.316279 | 10.296216 | 3.285070 | 8.409757 |
| 97.5% | 19.884956 | 25.000000 | 20.022432 | 10.576679 | 25.000000 |

| $\beta_0 = 1$ | bias$\times 10^3$ | std$\times 10^3$ | MSE$\times 10^6$ | 2.5% | 97.5% |
|---|---|---|---|---|---|
| sieve | 0.078258 | 2.195015 | 4.824215 | 0.999039 | 1.000786 |
| OLS | 0.065571 | 2.921240 | 8.537945 | 0.992728 | 1.003194 |
| parametric | 19.901275 | 47.339936 | 2637.130306 | 0.996754 | 1.156430 |

Table 19: Clayton copula, $\alpha_0 = 12$; $X_t$ time trend; marginal $t(5)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | P |
|---|---|---|---|---|---|
| mean | 12.502974 | 12.145863 | 12.374938 | 6.128115 | 14.022040 |
| bias | 0.502974 | 0.145863 | 0.374938 | -5.871885 | 2.022040 |
| std | 2.580538 | 5.789021 | 2.246420 | 1.912847 | 3.551555 |
| MSE | 6.912157 | 33.534035 | 5.186981 | 38.138012 | 16.702190 |
| relative | 1.332597 | 6.465039 | 1 | 7.352641 | 3.220021 |
| 2.5% | 9.449404 | 5.546247 | 10.321597 | 3.268238 | 8.024558 |
| 97.5% | 23.026350 | 25.000000 | 21.763029 | 10.699746 | 25.000000 |

| $\beta_0 = 1$ | bias$\times 10^3$ | std$\times 10^3$ | MSE$\times 10^6$ | 2.5% | 97.5% |
|---|---|---|---|---|---|
| sieve | -0.055905 | 0.610807 | 0.376211 | 0.998851 | 1.000753 |
| OLS | 0.046979 | 2.242401 | 5.030571 | 0.994035 | 1.002667 |
| parametric | 20.324849 | 41.614826 | 2144.893247 | 0.998357 | 1.131316 |

## 11.2 Unit root

Table 20: Clayton copula, $\alpha_0 = 12$; $X_t$ unit root; marginal $t(3)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | P |
|---|---|---|---|---|---|
| mean | 12.399042 | 11.937420 | 12.340731 | 6.219080 | 15.834364 |
| bias | 0.399042 | -0.062580 | 0.340731 | -5.780920 | 3.834364 |
| std | 2.298290 | 5.510775 | 2.177099 | 1.950947 | 3.651948 |
| MSE | 5.441371 | 30.372562 | 4.855858 | 37.225234 | 28.039068 |
| relative | 1.120579 | 6.254829 | 1 | 7.666047 | 5.774277 |
| 2.5% | 9.721399 | 5.640541 | 10.296216 | 3.291824 | 9.020098 |
| 97.5% | 18.612457 | 25.000000 | 20.022435 | 10.811720 | 25.000000 |

| $\beta_0 = 1$ | bias$\times 10^3$ | std$\times 10^3$ | MSE$\times 10^6$ | 2.5% | 97.5% |
|---|---|---|---|---|---|
| sieve | -0.144993 | 1.015039 | 1.051327 | 0.997780 | 1.001457 |
| OLS | 2.408317 | 1.892584 | 9.381863 | 0.999844 | 1.006944 |
| parametric | -1.052325 | 7.451765 | 56.636187 | 0.994238 | 1.004058 |

Table 21: Clayton copula, $\alpha_0 = 12$; $X_t$ unit root; marginal $t(5)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | P |
|---|---|---|---|---|---|
| mean | 12.510076 | 12.069806 | 12.374938 | 6.264702 | 13.895997 |
| bias | 0.510076 | 0.069806 | 0.374938 | -5.735298 | 1.895997 |
| std | 2.458474 | 5.252980 | 2.246420 | 2.011569 | 3.044865 |
| MSE | 6.304270 | 27.598676 | 5.186981 | 36.940052 | 12.866010 |
| relative | 1.215403 | 5.320759 | 1 | 7.121686 | 2.480443 |
| 2.5% | 9.606274 | 5.881196 | 10.321597 | 3.276511 | 8.486591 |
| 97.5% | 21.130121 | 25.000000 | 21.763029 | 11.086462 | 25.000000 |

| $\beta_0 = 1$ | bias$\times 10^3$ | std$\times 10^3$ | MSE$\times 10^6$ | 2.5% | 97.5% |
|---|---|---|---|---|---|
| sieve | -0.126832 | 1.020734 | 1.057985 | 0.997611 | 1.001754 |
| OLS | 2.381391 | 1.850620 | 9.095819 | 0.999873 | 1.006666 |
| parametric | -0.294973 | 12.419458 | 154.329937 | 0.996393 | 1.002898 |

### 11.3 Cointegration

Table 22: Clayton copula, $\alpha_0 = 12$; $X_t$ cointegration; marginal $t(3)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | P |
|---|---|---|---|---|---|
| mean | 12.327888 | 11.817449 | 12.340682 | 5.874031 | 15.756159 |
| bias | 0.327888 | -0.182551 | 0.340682 | -6.125969 | 3.756159 |
| std | 2.277177 | 5.223738 | 2.177206 | 1.814410 | 3.358357 |
| MSE | 5.293045 | 27.320762 | 4.856291 | 40.819577 | 25.387293 |
| relative | 1.089936 | 5.625850 | 1 | 8.405505 | 5.227713 |
| 2.5% | 9.733464 | 5.445915 | 10.296216 | 3.150459 | 11.564163 |
| 97.5% | 18.842197 | 25.000000 | 20.022436 | 10.238318 | 25.000000 |

| $\beta_0 = 1$ | bias$\times 10^3$ | std$\times 10^3$ | MSE$\times 10^6$ | 2.5% | 97.5% |
|---|---|---|---|---|---|
| sieve | -0.118581 | 5.503727 | 30.305069 | 0.988646 | 1.010854 |
| OLS | 1.672495 | 65.560251 | 4300.943799 | 0.893515 | 1.113421 |
| parametric | -0.077138 | 12.432901 | 154.582985 | 0.974382 | 1.024501 |

Table 23: Clayton copula, $\alpha_0 = 12$; $X_t$ cointegration; marginal $t(5)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | P |
|---|---|---|---|---|---|
| mean | 12.378719 | 11.969450 | 12.374938 | 5.893641 | 13.974475 |
| bias | 0.378719 | -0.030550 | 0.374938 | -6.106359 | 1.974475 |
| std | 2.356531 | 5.022992 | 2.246420 | 1.822452 | 2.796164 |
| MSE | 5.696666 | 25.231382 | 5.186981 | 40.608949 | 11.717088 |
| relative | 1.098262 | 4.864367 | 1 | 7.829014 | 2.258942 |
| 2.5% | 9.760541 | 5.716948 | 10.321597 | 3.161708 | 10.886874 |
| 97.5% | 21.213324 | 25.000000 | 21.763029 | 10.262889 | 25.000000 |

| $\beta_0 = 1$ | bias$\times 10^3$ | std$\times 10^3$ | MSE$\times 10^6$ | 2.5% | 97.5% |
|---|---|---|---|---|---|
| sieve | -0.049948 | 5.165665 | 26.686589 | 0.990082 | 1.010590 |
| OLS | 1.144755 | 50.229566 | 2524.319730 | 0.911823 | 1.095445 |
| parametric | -0.075015 | 6.572480 | 43.203117 | 0.987086 | 1.013605 |

## 12 Log sieve

### 12.1 Time trend

Table 24: Clayton copula, $\alpha_0 = 12$; $X_t$ time trend; marginal $t(3)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | E1 |
|---|---|---|---|---|---|
| mean | 10.820670 | 11.152059 | 11.109732 | 6.104206 | 7.983544 |
| bias | -1.179330 | -0.847941 | -0.890268 | -5.895794 | -4.016456 |
| std | 1.884414 | 5.386411 | 1.180313 | 1.875286 | 2.162303 |
| MSE | 4.941834 | 29.732428 | 2.185714 | 38.277081 | 20.807470 |
| relative | 2.260970 | 13.603070 | 1 | 17.512388 | 9.519757 |
| 2.5% | 8.063555 | 5.123517 | 9.007488 | 3.285070 | 4.489908 |
| 97.5% | 14.438227 | 25.000000 | 13.108848 | 10.576679 | 12.687488 |

Table 25: Clayton copula, $\alpha_0 = 12$; $X_t$ time trend; marginal $t(5)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | E1 |
|---|---|---|---|---|---|
| mean | 11.156425 | 11.449140 | 11.301299 | 6.128115 | 8.190875 |
| bias | -0.843575 | -0.550860 | -0.698701 | -5.871885 | -3.809125 |
| std | 1.710430 | 5.289210 | 1.057431 | 1.912847 | 2.342150 |
| MSE | 3.637191 | 28.279192 | 1.606343 | 38.138012 | 19.995100 |
| relative | 2.264268 | 17.604705 | 1 | 23.742136 | 12.447591 |
| 2.5% | 8.460203 | 5.258930 | 9.316576 | 3.268238 | 4.564534 |
| 97.5% | 14.929536 | 25.000000 | 13.332056 | 10.699746 | 13.467270 |

## 12.2 Unit root

Table 26: Clayton copula, $\alpha_0 = 12$; $X_t$ unit root; marginal $t(3)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | E1 |
|---|---|---|---|---|---|
| mean | 10.969984 | 11.175585 | 11.114346 | 6.219080 | 8.031966 |
| bias | -1.030016 | -0.824415 | -0.885654 | -5.780920 | -3.968034 |
| std | 1.605914 | 5.062578 | 1.164483 | 1.950947 | 1.965568 |
| MSE | 3.639892 | 26.309357 | 2.140403 | 37.225234 | 19.608748 |
| relative | 1.700564 | 12.291777 | 1 | 17.391693 | 9.161241 |
| 2.5% | 8.272588 | 5.343988 | 9.041769 | 3.291824 | 4.626597 |
| 97.5% | 14.223595 | 25.000000 | 13.108848 | 10.811720 | 11.918983 |

Table 27: Clayton copula, $\alpha_0 = 12$; $X_t$ unit root; marginal $t(5)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | E1 |
|---|---|---|---|---|---|
| mean | 11.269823 | 11.524558 | 11.301302 | 6.264702 | 8.199972 |
| bias | -0.730177 | -0.475442 | -0.698698 | -5.735298 | -3.800028 |
| std | 1.696507 | 4.997589 | 1.057432 | 2.011569 | 2.051522 |
| MSE | 3.411296 | 25.201937 | 1.606343 | 36.940052 | 18.648955 |
| relative | 2.123642 | 15.689018 | 1 | 22.996373 | 11.609576 |
| 2.5% | 8.590813 | 5.570762 | 9.316576 | 3.276511 | 4.665593 |
| 97.5% | 14.842655 | 25.000000 | 13.332055 | 11.086462 | 12.382381 |

## 12.3 Cointegration

Table 28: Clayton copula, $\alpha_0 = 12$; $X_t$ cointegration; marginal $t(3)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | E1 |
|---|---|---|---|---|---|
| mean | 10.924942 | 10.880250 | 11.107974 | 5.874031 | 7.966668 |
| bias | -1.075058 | -1.119750 | -0.892026 | -6.125969 | -4.033332 |
| std | 1.354789 | 4.568501 | 1.191008 | 1.814410 | 1.945689 |
| MSE | 2.991203 | 22.125040 | 2.214210 | 40.819577 | 20.053475 |
| relative | 1.350912 | 9.992294 | 1 | 18.435277 | 9.056717 |
| 2.5% | 8.340442 | 5.180986 | 8.999724 | 3.150459 | 4.467643 |
| 97.5% | 13.402914 | 25.000000 | 13.108848 | 10.238318 | 11.872324 |

Table 29: Clayton copula, $\alpha_0 = 12$; $X_t$ cointegration; marginal $t(5)$

| $\alpha_0 = 12$ | S1 | S2 | S | E | E1 |
|---|---|---|---|---|---|
| mean | 11.158440 | 11.151072 | 11.301302 | 5.893641 | 8.105867 |
| bias | -0.841560 | -0.848928 | -0.698698 | -6.106359 | -3.894133 |
| std | 1.260010 | 4.457627 | 1.057432 | 1.822452 | 2.015088 |
| MSE | 2.295848 | 20.591118 | 1.606343 | 40.608949 | 19.224851 |
| relative | 1.429239 | 12.818635 | 1 | 25.280380 | 11.968089 |
| 2.5% | 8.666289 | 5.360964 | 9.316576 | 3.161708 | 4.500327 |
| 97.5% | 13.710618 | 25.000000 | 13.332055 | 10.262889 | 12.289780 |

# 13 Empirical results

## 13.1 Target price: 75 days

Table 30: Cointegration test

| $H_0$ : no cointegration | test statistic | 10% | 5% | 1% |
|---|---|---|---|---|
| AMZN | 94.58 | 12.91 | 14.90 | 19.19 |

Table 31: Gaussian copula, in-sample estimation

| | joint | two step | empirical |
|---|---|---|---|
| filter $\beta$ | 0.85885274 | 0.86628940 | 0.86628940 |
| copula $\alpha$ | 0.88111838 | 0.87691301 | 0.85902297 |

Table 32: Gaussian copula, out-of-sample prediction

| | joint | two step | empirical |
|---|---|---|---|
| mean square | 49.22776111 | 46.27214580 | 62.11960223 |
| mean absolute value | 42.09943212 | 39.48921785 | 49.27176517 |

Table 33: Clayton copula, in-sample estimation

| | joint | two step | empirical |
|---|---|---|---|
| filter $\beta$ | 1.01641083 | 0.86628940 | 0.86628940 |
| copula $\alpha$ | no convergence | no convergence | 3.14160691 |

## 13.2 Target price: 105 days

Table 34: Cointegration test

| $H_0$ : no cointegration | test statistic | 10% | 5% | 1% |
|---|---|---|---|---|
| AMZN | 102.38 | 12.91 | 14.90 | 19.19 |

Table 35: Gaussian copula, in-sample estimation

| | joint | two step | empirical |
|---|---|---|---|
| filter $\beta$ | 0.86604591 | 0.87677088 | 0.87677088 |
| copula $\alpha$ | 0.87837231 | 0.89761183 | 0.88144247 |

Table 36: Gaussian copula, out-of-sample prediction

| | joint | two step | empirical |
|---|---|---|---|
| mean square | 49.68143941 | 45.53424167 | 57.73661726 |
| mean absolute value | 41.98012533 | 38.14044441 | 44.67737604 |

Table 37: Clayton copula, in-sample estimation

| | joint | two step | empirical |
|---|---|---|---|
| filter $\beta$ | 1.12172763 | 0.87677088 | 0.87677088 |
| copula $\alpha$ | no convergence | no convergence | 3.58999512 |

# Part II

# Chapter 2: Estimation of Parametric Dynamic Copula Models with Filtered Nonstationarity

**Abstract**

This paper considers estimation of copula based dynamic parametric models coupled with nonstationary filtration. Two new methods are proposed: joint estimator and two-step estimator. New theoretical results are obtained regarding: (1) conditions under which these estimator are equivalent asymptotically; (2) the effects of nonstationarity on limiting distributions and its relationship to tail dependence of copula. Monte Carlo simulation compares the performance between literature three-step estimator and our two new estimators. Three-step estimator is in general inferior to our two estimators. Joint estimator is found to always be superior to all other estimators in a variety of Monte Carlo simulation designs, especially in the presence of strong tail dependence. Hence joint method is what we suggest in practical use.

## 14 Introduction

Chen and Xiao [2016] propose the following model:

$$Z_t = X_t'\beta + Y_t$$

$$(Y_{t-1}, Y_t) \sim C(F(Y_{t-1}, \theta), F(Y_t, \theta); \alpha)$$

where $X_t$ is a non-stationary trend and distribution of $Y_t$ is characterized by a parametric copula and marginal. Chen and Xiao [2016] research on this model based on three step estimator. After first step linear filter $\beta$ to get residues, they estimate the marginal $\theta$ at the second step and estimate the copula parameter $\alpha$ at last. In this paper, we consider two alternative estimators: joint estimator and two-step estimator. For two-step estimator, after first step linear filtration $\beta$, we estimate marginal $\theta$ and copula $\alpha$ together. For joint estimator, we estimate three terms $\beta$, $\theta$ and $\alpha$ at the same time.

Chen and Xiao [2016] show that the nonstationarity will affect the limiting distribution of three-step estimator in general case. We derive the same results for joint estimator and two-step estimator. However, in the special case when both copula and marginal are symmetric, all three estimators will not be affected by the structure of nonstationary regressor $X_t$. Both joint estimator and two-step estimator are asymptotically equivalent to the infeasible estimator. And we show that all of them are better than three-step estimator asymptotically.

When either copula or marginal is asymmetric, the above equivalence relationship does not hold and the nonstationarity will affect the limiting distribution of above estimators. Especially for copulas with tail dependence, such as Clayton and Gumbel, we show that this effect is positively related to the strength of tail dependence. The theoretical results are consistent with finite simulation findings that relative MSE of our estimators comparing to the infeasible benchmark is increasing with tail dependence.

Three step estimator is found to suffer finite sample bias when tail dependence is strong. Our simulation results demonstrate that our joint method is very stable, even under extreme tail dependence. Furthermore, for copulas without tail dependence, three-step estimator is still inferior to our joint estimator and two-step estimator in finite sample performance.

### 14.1 Related literature

While a large number of previous work using copulas has focused on modeling the contemporaneous dependence between multiple univariate series, there are also a growing number of papers using copulas to model the temporal dependence of univariate nonlinear time series. Darsow et al. [1992], Victor et al. [2006] and Ibragimov [2009] provide characterizations of a copula-based time series to be a Markov process. See Patton [2012] for a review of copula models in economic time series.

There is fairly extensive literature concentrating only on nonlinearity by assuming that $Y_t$ is directly observed in the copula Markov process. Joe [1997] proposes a general structure with parametric stationary Markov models based on parametric copulas and parametric marginal distributions. Chen and Fan [2006] characterizes copula function by parametric family, but deals with marginal distribution nonparametrically with empirical distribution. Chen et al. [2009] use sieve maximum likelihood estimation method to solve the model with parametric copula function and nonparametric marginal distribution function. Doukhan et al. [2005] research on the case for both copula function and univariate marginal distribution function estimated nonparametrically.

All the above models are based on assumption of stationarity. However, nonstationarity is an important empirical features in economic and financial time series. Many observed time series seem to display non-stationary characteristics. To the best of the author's knowledge, Chen and Xiao [2016] is the only literature combining nonstationarity and nonlinearity together based on copula Markov model in a fully parametric way. Chen and Xiao [2016] also analyze the semiparametric model by treating marginal nonparametrically. In this paper, we only research on fully parametric model. Results for semiparametric model will be presented in another project.

The rest of this article is organized as follows. In **Section** 15 and 16, we introduce our estimators, the nonstationary structure and several commonly used copulas. In **Section** 17, we analyze the relationship among estimators when both copula and marginal are symmetric. In **Section** 18 and 19, we concentrate on how asymmetry of copula and marginal will affect our estimators. In **Section** 20 we summarize simulation results on finite sample performance. **Section** 21 concludes.

# 15 Background

## 15.1 The Model

We assume that the observed time series $\{Z_t\}_{t=1}^n$ can be modeled as:

$$Z_t = X_t'\beta_0 + Y_t$$

where $X_t'\beta_0$ is the non-stationary component and $Y_t$ is the stationary component with non-linearity. In particular, we assume that $X_t$ is a $d$ dimensional vector of dependent variables that may be non-stationary. The second component, $Y_t$, is a stationary process with non-linearity that can be captured by a copula function. For simplicity and without loss of generality, we assume in this paper that $\{Y_t\}_{t=1}^n$ is a first-order strictly stationary Markov process. Higher order Markov process can be investigated similarly.

Under the assumption that $\{Y_t\}_{t=1}^n$ is a first-order stationary Markov process, its statistic property is fully characterized by the true bi-variate joint distribution of $Y_{t-1}$ and $Y_t$, say $H_0(y_{t-1}, y_t)$. Further suppose that $Y_t$ is continuously distributed. Denote marginal distribution function of $Y_t$ be $F_0(\cdot)$, respectively. Then by Sklar's theorem, there exists unique copula function $C_0(\cdot,\cdot)$ satisfying:

$$H_0(a,b) = C_0(F_0(a), F_0(b))$$

which holds for all $(a,b) \in \mathbb{R}^2$.

Here the copula function $C_0(\cdot,\cdot)$ is a bi-variate probability distribution function with uniform marginals. Denote the corresponding copula density of $C_0(u_1,u_2)$ by $c_0(u_1,u_2)$, and the density of the marginal distribution $F_0(\cdot)$ by $f_0(\cdot)$, the true conditional density of $Y_t$ given $Y_{t-1}$ is:

$$p(Y_t|Y_{t-1}) = f_0(Y_t)c(F_0(Y_{t-1}), F_0(Y_t))$$

Thus, given $\{Y_t\}_{t=1}^n$, the log likelihood of the sample is

$$\frac{1}{n}\sum_{t=1}^{n}\log f_0(Y_t) + \frac{1}{n}\sum_{t=2}^{n}\log c(F_0(Y_{t-1}), F_0(Y_t)) \tag{11}$$

For convenience of asymptotic analysis, we make following assumptions on the dynamics of the latent process $\{Y_t\}$.

**Assumption 22.** *$\{Y_t : t = 1, 2, \cdots, n\}$ is a sample of a stationary first-order Markov process generated from $(F(\cdot;\theta_0), C(\cdot,\cdot;\alpha_0))$, where $F(\cdot;\theta_0)$ is the true invariant distribution up to unknown value $\theta_0$, is absolutely continuous with respect to Lebesgue measure on the real line; $C(\cdot,\cdot;\alpha_0)$ is the true parametric copula for $(Y_{t-1}, Y_t)$ up to unknown value $\alpha_0$, is absolutely continuous with respect to Lebesgue measure on $[0,1]^2$, and is neither the Fréchet-Hoeffding upper ($C(u_1,u_2) = \min\{u_1,u_2\}$) nor the the lower ($C(u_1,u_2) = \max\{u_1+u_2-1, 0\}$) bound.*

**Assumption 23.** *The process $\{Y_t\}$ is absolutely regular with mixing coefficient $\beta(\tau) = O(\tau^{-\delta})$, for a constant $\delta > 0$.*

*Remark* 18. See Chen and Fan [2006], Chen et al. [2009], Beare [2010], Beare [2012], Longla and Peligrad [2012] and others about sufficient conditions that most commonly used copula-based Markov process is geometric ergodic and hence absolutely regular with polynomial decay mixing coefficinets.
.

Concerning on the non-stationary component and the related filtration, we make the following assumptions to facilitate asymptotic analysis.

**Assumption 24.** *There exists a scaling matrix $G_n$ such that:*

$$\begin{pmatrix} X_n(r) \\ Y_n(r) \end{pmatrix} \Rightarrow \begin{pmatrix} X(r) \\ \lambda_Y W(r) \end{pmatrix}$$

*here $X_n(r) := G_n^{-1} X_{[nr]}$ and $Y_n(r) := \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} Y_t$. $\lambda_Y^2 := \mathbb{E}Y_t^2 + 2\sum_{s=1}^{\infty} \mathbb{E}Y_t Y_{t+s}$ is the long run variance of $Y_t$. $W(r)$ is a standard Wiener process.*

*Remark* 19. Due to non-stationarity in $X_t$, we introduced appropriate re-standardization via the scaling matrix $G_n$ to facilitate asymptotic analysis. The limit of the standardized non-stationary component, $X(r)$, may be stochastic or deterministic or a mixture of stochastic and deterministic functions. In the case when $X(r)$ contains stochastic functions, $W(r)$ and $X(r)$ may be correlated.

The limiting distribution, $\xi$, of the filtration parameter is a function of $X(\cdot)$ and may not be a normal variate. Leading cases that are widely used in time series application includes the following:

**Example 4.** Deterministic trend.

$X_t$ is a vector of deterministic trend function and $G_n^{-1} X_{[nr]} \Rightarrow X(r)$, where $X(r)$ is a continuous limiting trending function. Let the estimator of $\beta$ be $\hat{\beta}_n$,

$$\sqrt{n} G_n(\hat{\beta}_n - \beta_0) \Rightarrow \xi_1$$

where in general $\xi_1$ is a normal variate.

Then the detrend data is given by $\hat{Y}_t = Z_t - X_t'\hat{\beta}$. For example, if the observed time series $\{Z_t\}_{t=1}^n$ contains a linear trend:

$$Z_t = \beta_{01} + \beta_{02} \cdot t + Y_t$$

In practice, we estimate copula model based on OLS:

$$\hat{Y}_t = Z_t - \hat{\beta}_{01} - \hat{\beta}_{02} \cdot t$$

The corresponding standardization matrix is $G_n = \operatorname{diag}(1, n)$, $\sqrt{n}G_n = \operatorname{diag}(n^{\frac{1}{2}}, n^{\frac{3}{2}})$, $X_t = (1, t)'$ and $X(r) = (1, r)'$. Limiting distribution is $\xi_1 = \lambda\xi$, where:

$$\xi = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}^{-1} \cdot \begin{pmatrix} W(1) \\ \int_0^1 s dW(s) \end{pmatrix}$$

here $W(s)$, $s \in [0,1]$ is the standard Brownian motion, If we estimate copula model through LAD to get $\hat{\beta}$, then the limiting distribution is $\xi_1 = \frac{\lambda_Y^*}{2f_0(0)}\xi$, here $\lambda_Y^{*2} := \mathbb{E}\left[Y_t^*\right]^2 + 2\sum_{s=1}^{\infty}\mathbb{E}Y_t^*Y_{t+s}^*$ is the long run variance of $Y_t^* = 2\mathbb{I}\left(Y_t \geq 0\right) - 1$, the sign of $Y_t$. $f_0(\cdot)$ is the true marginal density of $Y_t$ and we need to assume that it is positive at zero.

**Example 5.** Cointegration Time Series.

$X_t$ is a vector of non-stationary process with unit roots, $X_t = X_{t-1} + \varepsilon_t$, $G_n = \sqrt{n}$, $G_n^{-1}X_{[nr]} \Rightarrow X(r) = \lambda_\varepsilon W_1(r)$. Here $W_1(s)$, $s \in [0,1]$ is a standard Brownian motion, $\lambda_\varepsilon^2 := \mathbb{E}\,\varepsilon_t^2 + 2\sum_{s=1}^{\infty}\mathbb{E}\,\varepsilon_t\varepsilon_{t+s}$ is the long run variance of $\varepsilon_t$. Assume $\varepsilon_t$ is independent of $Y_t$.

Then the OLS estimator $\hat{\beta}$ is still rate-$n$ converging $n(\hat{\beta} - \beta_0) \Rightarrow \xi_2 = \frac{\lambda_Y}{\lambda_\varepsilon}\xi$, where

$$\xi = \left[\int_0^1 W_1^2(r)dr\right]^{-1}\left[\int_0^1 W_1(r)dW_2(r)\right]$$

here $\lambda_Y^2 := \mathbb{E}\,Y_t^2 + 2\sum_{s=1}^{\infty}\mathbb{E}\,Y_tY_{t+s}$ is the long run variance of $Y_t$, $W_2(s)$, $s \in [0,1]$ is another standard Brownian motion independent with $W_1(\cdot)$.

Similar as previous example, if we estimate copula model through LAD to get $\hat{\beta}$, then the limiting distribution is $\xi_2 = \frac{\lambda_Y^*}{2\lambda_\varepsilon f_0(0)}\xi$, here $\lambda_Y^{*2} := \mathbb{E}\left[Y_t^*\right]^2 + 2\sum_{s=1}^{\infty}\mathbb{E}Y_t^*Y_{t+s}^*$ is the long run variance of $Y_t^* = 2\mathbb{I}\left(Y_t \geq 0\right) - 1$, the sign of $Y_t$. $f_0(\cdot)$ is the true marginal density of $Y_t$ and we need to assume that it is positive at zero.

## 15.2 Copulas

Before analyzing large sample property of our two step estimator, we first introduce some commonly used copulas and their properties.

Suppose $(U_1, U_2)$ and $(V_1, V_2)$ are two pairs bivariate uniformly distributed random variables, joint distribution following copula $C(\cdot,\cdot)$. Then the Kendall's tau is defined as the probability of concordance minus the probability of discordance, see Nelson [1999] chapter 5:

$$\begin{aligned}\tau &= \mathbb{P}\left[(U_1 - V_1)(U_2 - V_2) > 0\right] - \mathbb{P}\left[(U_1 - V_1)(U_2 - V_2) < 0\right] \\ &= 4\int_0^1\int_0^1 C(u_1,u_2)c(u_1,u_2)\mathrm{d}u_1\mathrm{d}u_2 - 1\end{aligned}$$

Because Kendall's tau is the difference of two probabilities, we have $-1 \leq \tau \leq 1$. Positive $\tau$ means positive dependence and negative $\tau$ means negative dependence.

Tail dependence measures the dependence between the variables in the upper right quadrant and in the lower left quadrant of $[0,1]^2$. The lower and upper tail dependence coefficients $\lambda_L$ and $\lambda_U$ in terms of copula are defined as:

$$\lambda_L = \lim_{u\to 0^+}\mathbb{P}(U_2 \leq u | U_1 \leq u) = \lim_{u\to 0^+}\frac{C(u,u)}{u}$$

$$\lambda_U = \lim_{u\to 1^-}\mathbb{P}(U_2 \geq u | U_1 \geq u) = \lim_{u\to 1^-}\frac{1 + C(u,u) - 2u}{1-u}$$

We consider five copulas in this paper, each with four choices of copula parameter:

- Gaussian copula
$$C(u_1,u_2;\alpha_0) = \Phi_\alpha(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$$
$$c(u_1,u_2;\alpha_0) = \frac{\phi_\alpha(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi(\Phi^{-1}(u_1))\cdot\phi(\Phi^{-1}(u_2))}$$
here $\Phi(\cdot)$ and $\phi(\cdot)$ denote the CDF and PDF of standard normal distribution respectively. $\Phi_\alpha(\cdot)$ and $\phi_\alpha(\cdot)$ denote the CDF and PDF of bivariate normal distribution with correlation $\alpha$

respectively. Range for $\alpha$ is $-1 \le \alpha \le 1$.
When $\alpha > 0$, dependence is positive:
$\alpha = 0.9 \Rightarrow \tau = 0.713$, $\rho = 0.891$; $\alpha = 0.5 \Rightarrow \tau = 0.333$, $\rho = 0.483$
When $\alpha < 0$, dependence is negative:
$\alpha = -0.9 \Rightarrow \tau = -0.713$, $\rho = -0.891$; $\alpha = -0.5 \Rightarrow \tau = -0.333$, $\rho = -0.483$
There is no tail dependence for Gaussian copula.

- Frank copula

$$C(u_1, u_2; \alpha) = -\frac{1}{\alpha} \cdot \log\left(1 - \frac{(1-e^{-\alpha u_1})(1-e^{-\alpha u_2})}{1-e^{-\alpha}}\right)$$

$$c(u_1, u_2; \alpha) = \alpha \cdot \frac{e^{-\alpha u_1} e^{-\alpha u_2}}{1-e^{-\alpha}} \cdot \left(1 - \frac{(1-e^{-\alpha u_1})(1-e^{-\alpha u_2})}{1-e^{-\alpha}}\right)^{-2}$$

here $\alpha \in \mathbb{R}^1$.
When $\alpha > 0$, dependence is positive:
$\alpha = 15 \Rightarrow \tau = 0.7626$, $\rho = 0.9294$; $\alpha = 5 \Rightarrow \tau = 0.4567$, $\rho = 0.6435$.
When $\alpha < 0$, dependence is negative:
$\alpha = -15 \Rightarrow \tau = -0.7626$, $\rho = -0.9294$; $\alpha = -5 \Rightarrow \tau = -0.4567$, $\rho = -0.6435$
There is no tail dependence for Frank copula.

- Clayton copula

$$C(u_1, u_2; \alpha) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}}$$

$$c(u_1, u_2; \alpha) = (1+\alpha) \cdot u_1^{-\alpha-1} \cdot u_2^{-\alpha-1} \cdot (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}-2}$$

here $\alpha$ is positive.
Clayton copula has Kendall's tau $\tau = \frac{\alpha}{2+\alpha}$ and lower tail dependence coefficient $\lambda_L = 2^{-1/\alpha}$ that is increasing in $\alpha$, but no upper tail dependence.
When $\alpha = 2$, $\tau = 0.5$, $\rho = 0.682$, $\lambda_L = 0.7071$.
When $\alpha = 5$, $\tau = 0.7143$, $\rho = 0.885$, $\lambda_L = 0.871$.
When $\alpha = 10$, $\tau = 0.833$, $\rho = 0.958$, $\lambda_L = 0.933$.
When $\alpha = 12$, $\tau = 0.857$, $\rho = 0.969$, $\lambda_L = 0.944$.

- EFGM copula

$$C(u_1, u_2; \alpha) = u_1 u_2 [1 + \alpha(1-u_1)(1-u_2)]$$

$$c(u_1, u_2; \alpha) = 1 + \alpha(1-2u_1)(1-2u_2)$$

here range for $\alpha$ is $-1 \le \alpha \le 1$.
When $\alpha > 0$, dependence is positive:
$\alpha = 0.9 \Rightarrow \tau = 0.2$, $\rho = 0.3$; $\alpha = 0.5 \Rightarrow \tau = 0.111$, $\rho = 0.167$
When $\alpha < 0$, dependence is negative:
$\alpha = -0.9 \Rightarrow \tau = -0.2$, $\rho = -0.3$; $\alpha = -0.5 \Rightarrow \tau = -0.111$, $\rho = -0.167$
There is no tail dependence for EFGM copula.

- Gumbel copula

$$C(u_1, u_2; \alpha) = \exp\left[-\left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}}\right]$$

$$\begin{aligned} c(u_1, u_2; \alpha) &= \exp\left[-\left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}}\right] \cdot \left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}-2} \\ &\cdot (-\log u_1)^{\alpha-1} \cdot (-\log u_2)^{\alpha-1} \cdot \frac{1}{u_1 u_2} \cdot \\ &\left[\left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}} + (\alpha - 1)\right] \end{aligned}$$

here $\alpha$ needs to be larger than 1.
Gumbel copula has Kendall's tau $\tau = 1 - \frac{1}{\alpha}$ and upper tail dependence coefficient $\lambda_U = 2 - 2^{1/\alpha}$

that is increasing in $\alpha$, but no lower tail dependence.
When $\alpha = 2$, $\tau = 0.5$, $\rho = 0.682$, $\lambda_U = 0.586$.
When $\alpha = 3.5$, $\tau = 0.7143$, $\rho = 0.887$, $\lambda_U = 0.781$.
When $\alpha = 6$, $\tau = 0.833$, $\rho = 0.96$, $\lambda_U = 0.8775$.
When $\alpha = 7$, $\tau = 0.857$, $\rho = 0.971$, $\lambda_U = 0.896$.

# 16 Estimators

Now we have three variables to estimate: linear filtering coefficient $\beta$, copula parameter $\alpha$ and parametric marginal $\theta$. The copula $\alpha$ and marginal $\theta$ are our main focuses.

## 16.1 Three step method

Chen and Xiao [2016] propose the three-step method. During first step, they estimate $\beta$ through some moment condition or quantile condition. Then $\theta$ is estimated at the second step. $\alpha$ is solved during last stage.

If residual $Y_t$ satisfies the moment condition $\mathbb{E}Y_t = 0$, then $\tilde{\beta}$ can be estimated through OLS:

$$\min \sum_{t=1}^{n} (Z_t - X_t'\beta)^2$$

If median of $Y_t$ is zero, $\operatorname{Med} Y_t = 0$. Then during first step, we run absolute mean deviation to get $\tilde{\beta}$:

$$\min \sum_{t=1}^{n} |Z_t - X_t'\beta|$$

More generally, if $Y_t$ has known quantile at $\tau$, $Q_\tau(Y_t) = a$, then use quantile regression to solve $\tilde{\beta}$:

$$\min \sum_{t=1}^{n} \rho_\tau \left(Z_t - X_t'\beta\right)$$

here $\rho_\tau(u) = u\left(\tau - \mathbb{I}(u < 0)\right)$. Absolute mean deviation is the specific case when $\tau = 0.5$.

Once $\tilde{\beta}$ is reached from first step estimation, residuals $\hat{Y}_t$ could be computed as:

$$\hat{Y}_t = Z_t - X_t'\tilde{\beta}$$

During second step, we estimate marginal parameter $\theta$ through filtering residuals $\hat{Y}_t$:

$$\max \sum_{t=1}^{n} \log f(\hat{Y}_t, \theta)$$

At last step, we estimate $\alpha$ by:

$$\max_{\alpha} \sum_{t=2}^{n} \log c(F(\hat{Y}_{t-1}, \hat{\theta}), F(\hat{Y}_t, \hat{\theta}), \alpha)$$

Denote the estimators as $\hat{\beta}^{3step} = \tilde{\beta}$, $\hat{\alpha}^{3step}$ and $\hat{\theta}^{3step}$.

## 16.2 Two step method

With similar first step estimation to get $\tilde{\beta}$ and $\hat{Y}_t$, two-step estimation differs with three-step method from the second step, as $\alpha$ and $\theta$ are estimated jointly rather than separately:

$$\max_{\alpha,\theta} \sum_{t=1}^{n} \log f(\hat{Y}_t, \theta) + \sum_{t=2}^{n} \log c(F(\hat{Y}_{t-1}, \theta), F(\hat{Y}_t, \theta), \alpha)$$

Denote the estimators as $\hat{\beta}^{2step} = \tilde{\beta}$, $\hat{\alpha}^{2step}$ and $\hat{\theta}^{2step}$.

### 16.3 Joint method

For joint estimator, $\alpha$, $\beta$ and $\theta$ are estimated at the same time:

$$\max_{\alpha,\beta,\theta} \sum_{t=1}^{n} \log f(Z_t - X_t\beta, \theta) + \sum_{t=2}^{n} \log c(F(Z_{t-1} - X_{t-1}\beta, \theta), F(Z_t - X_t\beta, \theta), \alpha)$$

Denote the estimators as $\hat{\beta}^{joint}$, $\hat{\alpha}^{joint}$ and $\hat{\theta}^{joint}$.

### 16.4 Infeasible(ideal) estimator

For comparison purpose, we also research on infeasible(ideal) estimator where $Y_t$ is assumed to be directly observed. $\alpha$ and $\theta$ are jointly estimated:

$$\max_{\alpha,\theta} \sum_{t=1}^{n} \log f(Y_t, \theta) + \sum_{t=2}^{n} \log c(F(Y_{t-1}, \theta), F(Y_t, \theta), \alpha)$$

Denote the estimators as $\hat{\alpha}^{ideal}$ and $\hat{\theta}^{ideal}$.

### 16.5 Comparisons among above estimators

In later sections, we will demonstrate the following relationship for limiting distribution of both copula $\alpha$ and marginal $\theta$:

- **When both copula and marginal are symmetric**
  Linear filtering method has no effects on all the estimators: joint, two-step and three-step. Furthermore, asymptotically:
  joint $\approx$ two-step(OLS) $\approx$ two-step(LAD) $\approx$ ideal $\succeq$ three-step(OLS) $\approx$ three-step(LAD)

- **When either copula or marginal is asymmetric**
  The above equivalence relationship broke up. However, we show that joint estimator is the best among several candidates:
  ideal $\succeq$ joint $\succeq$ two-step $\succeq$ three-step

## 17 Effect of nonstationarity

### 17.1 Limiting distribution

For $\hat{\beta}^{OLS}$ when $\mathbb{E}Y_t = 0$ and $\hat{\beta}^{MED}$ when $\operatorname{Med} Y_t = 0$, we have $G_n(\hat{\beta} - \beta_0) = O_p(1/\sqrt{n})$. Now suppose $\hat{\beta}^{joint}$ also satisfies this property:

$$G_n(\hat{\beta}^{joint} - \beta_0) = O_p(1/\sqrt{n})$$

Then for both joint and two-step estimators, $\hat{\alpha}$ and $\hat{\theta}$ will maximize:

$$\max \sum_{t=1}^{n} \log f(\hat{Y}_t, \theta) + \sum_{t=2}^{n} \log c(F(\hat{Y}_{t-1}, \theta), F(\hat{Y}_t, \theta), \alpha)$$

*Remark* 20. Denote $g(y,\theta) := \log f(y,\theta)$, $g_i(y,\theta) := \frac{\partial g(y,\theta)}{\partial i}$, $i = y, \theta$,
$g_{ij}(y,\theta) := \frac{\partial^2 g(y,\theta)}{\partial i \partial j}$, $i, j$ could be $y, \theta$,
$l(y_1, y_2, \alpha, \theta) := \log c(F(y_1,\theta), F(y_2,\theta); \alpha)$, $l_i(y_1, y_2, \alpha) := \frac{\partial l(y_1,y_2,\alpha)}{\partial y_i}$, $i = 1, 2, \alpha, \theta$,[17]
$l_{ij}(y_1, y_2, \alpha) := \frac{\partial^2 l(y_1,y_2,\alpha)}{\partial y_i \partial y_j'}$, $i, j$ could be $1, 2, \alpha, \theta$.

[17] For symbolic simplification, $y_\alpha$ is viewed as $\alpha$ and $y_\beta$ is viewed as $\beta$.

Consider the first order condition for $\hat{\alpha}$ and $\hat{\theta}$:

$$\sum_{t=2}^{n} l_\alpha(\hat{Y}_{t-1}, \hat{Y}_t, \hat{\alpha}, \hat{\theta}) = 0$$

$$\sum_{t=1}^{n} g_\theta(\hat{Y}_t, \hat{\theta}) + \sum_{t=2}^{n} l_\theta(\hat{Y}_{t-1}, \hat{Y}_t, \hat{\alpha}, \hat{\theta}) = 0$$

Notice:

$$\hat{Y}_t = Z_t - X_t'\hat{\beta}$$

We make first order Taylor expansion around $(Y_{t-1}, Y_t, \alpha_0, \theta_0)$, ignoring the higher order terms:

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=2}^{n} l_\alpha(Y_{t-1}, Y_t, \alpha_0, \theta_0) + \frac{1}{n}\sum_{t=2}^{n} l_{1\alpha}(Y_{t-1}, Y_t, \alpha_0, \theta_0) \cdot G_n^{-1}X_{t-1}' \cdot \sqrt{n}G_n(\beta_0 - \hat{\beta}) \\
&+\sum_{t=2}^{n} l_{2\alpha}(Y_{t-1}, Y_t, \alpha_0, \theta_0) \cdot G_n^{-1}X_t' \cdot \sqrt{n}G_n(\beta_0 - \hat{\beta}) + \frac{1}{n}\sum_{t=2}^{n} l_{\alpha\alpha}(Y_{t-1}, Y_t, \alpha_0, \theta_0) \cdot \sqrt{n}(\hat{\alpha} - \alpha_0) \\
&+\frac{1}{n}\sum_{t=2}^{n} l_{\alpha\theta}(Y_{t-1}, Y_t, \alpha_0, \theta_0) \cdot \sqrt{n}(\hat{\theta} - \theta_0) \\
&\approx 0
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_\theta(Y_t, \theta_0) + \frac{1}{n}\sum_{t=1}^{n} g_{\theta\theta}(Y_t, \theta_0) \cdot \sqrt{n}(\hat{\theta} - \theta_0) + \frac{1}{n}\sum_{t=1}^{n} \dot{g}_\theta(Y_t, \theta_0) \cdot G_n^{-1}X_t' \cdot \sqrt{n}G_n(\beta_0 - \hat{\beta}) \\
&+\frac{1}{\sqrt{n}}\sum_{t=2}^{n} l_\alpha(Y_{t-1}, Y_t, \alpha_0, \theta_0) + \frac{1}{n}\sum_{t=2}^{n} l_{1\theta}(Y_{t-1}, Y_t, \alpha_0, \theta_0) \cdot G_n^{-1}X_{t-1}' \cdot \sqrt{n}G_n(\beta_0 - \hat{\beta}) \\
&+\sum_{t=2}^{n} l_{2\theta}(Y_{t-1}, Y_t, \alpha_0, \theta_0) \cdot G_n^{-1}X_t' \cdot G_n(\beta_0 - \hat{\beta}) + \frac{1}{n}\sum_{t=2}^{n} l_{\alpha\theta}(Y_{t-1}, Y_t, \alpha_0, \theta_0) \cdot \sqrt{n}(\hat{\alpha} - \alpha_0) \\
&+\frac{1}{n}\sum_{t=2}^{n} l_{\theta\theta}(Y_{t-1}, Y_t, \alpha_0, \theta_0) \cdot \sqrt{n}(\hat{\theta} - \theta_0) \\
&\approx 0
\end{aligned}$$

*Remark* 21. From **Assumption** 23, we have following weak laws of large numbers once $g_{ij}$ and $l_{ij}$ are assumed to be smooth enough:

$$\frac{1}{n}\sum_{t=2}^{n} l_{ij}(Y_{t-1}, Y_t, \alpha_0, \theta_0) \xrightarrow{p} \mathbb{E}l_{ij}(Y_{t-1}, Y_t, \alpha_0, \theta_0) \quad i, j = 1, 2, \alpha, \theta$$

$$\frac{1}{n}\sum_{t=1}^{n} g_{ij}(Y_t) \xrightarrow{p} \mathbb{E}g_{ij}(Y_t) \quad i, j = y, \theta$$

For symbol simplification, denote: $\mathbb{E}g_{ij} := \mathbb{E}g_{ij}(Y_t, \theta_0)$ and $\mathbb{E}l_{ij} := \mathbb{E}l_{ij}(Y_{t-1}, Y_t, \alpha_0, \theta_0)$.

From Assumption 24, $\sum_{t=1}^{n} G_n^{-1}X_t \Rightarrow \int_0^1 X(r)\mathrm{d}r$. Hence we have:

$$\begin{pmatrix} \mathbb{E}l_{\alpha\alpha} & \mathbb{E}l_{\alpha\theta} \\ \mathbb{E}l_{\alpha\theta} & \mathbb{E}l_{\theta\theta} + \mathbb{E}g_{\theta\theta} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{n}(\hat{\alpha} - \alpha_0) \\ \sqrt{n}(\hat{\theta} - \theta_0) \end{pmatrix} \approx -\frac{1}{\sqrt{n}} \begin{pmatrix} \sum l_\alpha \\ \sum g_\theta + \sum l_\theta \end{pmatrix} + \begin{pmatrix} \mathbb{E}l_{1\alpha} + \mathbb{E}l_{2\alpha} \\ \mathbb{E}g_{y\theta} + \mathbb{E}l_{1\theta} + \mathbb{E}l_{2\theta} \end{pmatrix} \cdot \left[\int_0^1 X(r)\mathrm{d}r\right]^\top \cdot \sqrt{n}G_n(\hat{\beta} - \beta_0)$$

Assume the left matrix is revertible, split copula $\alpha$ and marginal $\theta$ separately, we have:

$$\sqrt{n}(\hat{\alpha}-\alpha) \approx A \cdot \left[\int_0^1 X(r)\mathrm{d}r\right]^\top \cdot \sqrt{n}G_n(\hat{\beta}-\beta_0)+\xi_\alpha$$

$$\sqrt{n}(\hat{\theta}-\theta_0) \approx B \cdot \left[\int_0^1 X(r)\mathrm{d}r\right]^\top \cdot \sqrt{n}G_n(\hat{\beta}-\beta_0)+\xi_\theta$$

here $\xi = (\xi_\alpha, \xi_\theta)$ is a normally distributed random variable, which is the limiting distribution of infeasible estimator when $Y_t$ is directly observed (Section 16.4). $A$ and $B$ are two constant terms only depending on copula and marginal:

$$A = \frac{[\mathbb{E}l_{1\alpha}+\mathbb{E}l_{2\alpha}]\cdot[\mathbb{E}l_{\theta\theta}+\mathbb{E}g_{\theta\theta}] - [\mathbb{E}g_{y\theta}+\mathbb{E}l_{1\theta}+\mathbb{E}l_{2\theta}]\cdot \mathbb{E}l_{\alpha\theta}}{\mathbb{E}l_{\alpha\alpha}\cdot[\mathbb{E}l_{\theta\theta}+\mathbb{E}g_{\theta\theta}] - (\mathbb{E}l_{\alpha\theta})^2}$$

$$B = \frac{-\mathbb{E}l_{\alpha\theta}\cdot[\mathbb{E}l_{1\alpha}+\mathbb{E}l_{2\alpha}] + [\mathbb{E}g_{y\theta}+\mathbb{E}l_{1\theta}+\mathbb{E}l_{2\theta}]\cdot \mathbb{E}l_{\alpha\alpha}}{\mathbb{E}l_{\alpha\alpha}\cdot[\mathbb{E}l_{\theta\theta}+\mathbb{E}g_{\theta\theta}] - (\mathbb{E}l_{\alpha\theta})^2}$$

If we can show that $A = B = 0$, then estimation of $\hat{\beta}$ will not affect the limiting distribution of $\hat{\alpha}$ and $\hat{\theta}$. Thus both joint estimator and two step estimator will have the same limiting distribution with ideal estimator where $Y_t$ is directly observed (see **Section** 16.4).

*Remark* 22. This also means that linear filtering method (through OLS or LAD) will not affect the limiting distribution of two step estimator asymptotically (and both are equivalent to ideal estimator).

The detail derivations of expression for $A$ and $B$ are listed in Appendix.

## 17.2 Symmetry

Suppose the copula has symmetric property:

$$c(u_1, u_2, \alpha) = c(1-u_1, 1-u_2, \alpha)$$

and the marginal distribution is also symmetric:

$$f(y, \theta) = f(-y, \theta)$$

In this section, we will show that under the above two symmetry restrictions, $A = B = 0$.
From symmetry property we have:

$$g(Y_t, \theta) \equiv g(-Y_t, \theta)$$
$$F(Y_t, \theta) \equiv 1 - F(-Y_t, \theta)$$
$$l(Y_{t-1}, Y_t, \alpha, \theta) = l(-Y_{t-1}, -Y_t, \alpha, \theta)$$

Hence:

$$\mathbb{E}l_{1\alpha}(Y_{t-1}, Y_t, \alpha_0, \theta_0) = 0$$
$$\mathbb{E}l_{2\alpha}(Y_{t-1}, Y_t, \alpha_0, \theta_0) = 0$$
$$\mathbb{E}l_{1\theta}(Y_{t-1}, Y_t, \alpha_0, \theta_0) = 0$$

$$\mathbb{E}l_{2\theta}(Y_{t-1}, Y_t, \alpha_0, \theta_0) = 0$$

$$\mathbb{E}g_{\theta y}(Y_t) = 0$$

For the first term:

$$l_1(y_1, y_2, \alpha, \theta) \equiv -l_1(-y_1, -y_2, \alpha, \theta)$$

$$l_{1\alpha}(y_1, y_2, \alpha, \theta) \equiv -l_{1\alpha}(-y_1, -y_2, \alpha, \theta)$$

$$\begin{aligned}
\mathbb{E}l_{1\alpha}(y_1, y_2, \alpha, \theta) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} l_{1\alpha}(y_1, y_2, \alpha, \theta) f(y_1, \theta_0) f(y_2, \theta_0) c(F(y_1, \theta_0), F(y_2, \theta_0); \alpha_0) \mathrm{d}y_1 \mathrm{d}y_2 \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} l_{1\alpha}(-y_1, -y_2, \alpha, \theta) f(-y_1, \theta_0) f(-y_2, \theta_0) c(F(-y_1, \theta_0), F(-y_2, \theta_0); \alpha_0) \mathrm{d}y_1 \mathrm{d}y_2 \\
&= -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} l_{1\alpha}(y_1, y_2, \alpha, \theta) f(y_1, \theta_0) f(y_2, \theta_0) c(F(y_1, \theta_0), F(y_2, \theta_0); \alpha_0) \mathrm{d}y_1 \mathrm{d}y_2 \\
&= -\mathbb{E}l_{1\alpha}(y_1, y_2, \alpha, \theta)
\end{aligned}$$

For the last term:

$$g_\theta(y, \theta) \equiv g_\theta(-y, \theta)$$

$$g_{\theta y}(y, \theta) \equiv -g_{\theta y}(-y, \theta)$$

$$\begin{aligned}
\mathbb{E}g_{\theta y}(Y_t, \theta) &= \int_{-\infty}^{+\infty} g_{\theta y}(y, \theta) f(y, \theta_0) \mathrm{d}y \\
&= \int_{-\infty}^{+\infty} g_{\theta y}(-y, \theta) f(-y, \theta_0) \mathrm{d}y \\
&= -\int_{-\infty}^{+\infty} g_{\theta y}(y, \theta) f(y, \theta_0) \mathrm{d}y \\
&= -\mathbb{E}g_{\theta y}(Y_t, \theta)
\end{aligned}$$

The remaining terms can be analyzed similarly.

From simple algebra in the appendix, we show that Gaussian, Frank and EFGM copulas are symmetric. Hence, when the marginal distribution is also symmetric around zero, like student t distribution, joint estimator and two-step estimator (both OLS and LAD) of $\hat{\alpha}$ and $\hat{\theta}$ are equivalent to the ideal estimator for these three copulas.

### 17.2.1 Three step estimator

In this section, we analyze three-step estimator when both copula and marginal are symmetric.

Assume $Y_t$ is observed, rather than the infeasible estimator where copula $\alpha$ and marginal $\theta$ are jointly estimated, consider the 'inference functions for margins' (IFM) as in Joe [1997], chapter 10.

Estimate $\theta$ first through true $Y_t$:

$$\max_\theta \sum_{t=1}^{n} \log f(Y_t, \theta)$$

Then estimate $\alpha$ by:

$$\max_\alpha \sum_{t=2}^{n} \log c(F(Y_{t-1}, \hat{\theta}), F(Y_t, \hat{\theta}), \alpha)$$

Following Chen and Xiao [2016], the three-step estimator will be equivalent to IFM asymptotically if:

$$\mathbb{E}l_{1\alpha} = \mathbb{E}l_{2\alpha} = \mathbb{E}l_{1\theta} = \mathbb{E}l_{2\theta} = \mathbb{E}l_{\alpha\theta} = \mathbb{E}g_{y\theta} = 0$$

This will be ensured by the symmetry properties. Hence, either through OLS or LAD in the first step filtration, three-step estimator behave asymptotically the same as if the residual $Y_t$ is directly observed. Hence the last part of equivalence equation in **Section** 16.5 is established: three-step(OLS) $\approx$ three-step(LAD).

However, above argument cannot apply to the equivalence relationship between three-step and our previous class (joint, two-step and ideal), because IFM is usually less efficient than our infeasible(ideal) estimator. Although Joe [1997] points out that the efficiency loss may be quite low under some cases, we use a brief simulation to illustrate that three-step estimator is indeed inferior in an asymptotic sense.

### 17.2.2 Simulation

To further illustrate the asymptotic properties of these estimators, we run simulation using Gaussian copula with time trend, copula parameter $\alpha_0 = 0.9$, sample size ranging from $T = 1500$ to $T = 5000$. We plot the finite sample MSE of five estimators: joint, two-step(OLS), two-step(LAD), three-step(OLS), three-step(LAD) and ideal estimators(theoretically the best). Simulation repetition is $M = 2000$. Marginal distribution is student t with degree of freedom 3 or 5. X axis is the sample size and Y axis is the **square root** of finite sample MSE. I attach results for $\alpha_0 = 0.9$ here for both copula $\alpha$ and marginal $\theta$.

Figure 4: Gaussian copula, time trend, $\alpha_0 = 0.9$, marginal $t(3)$

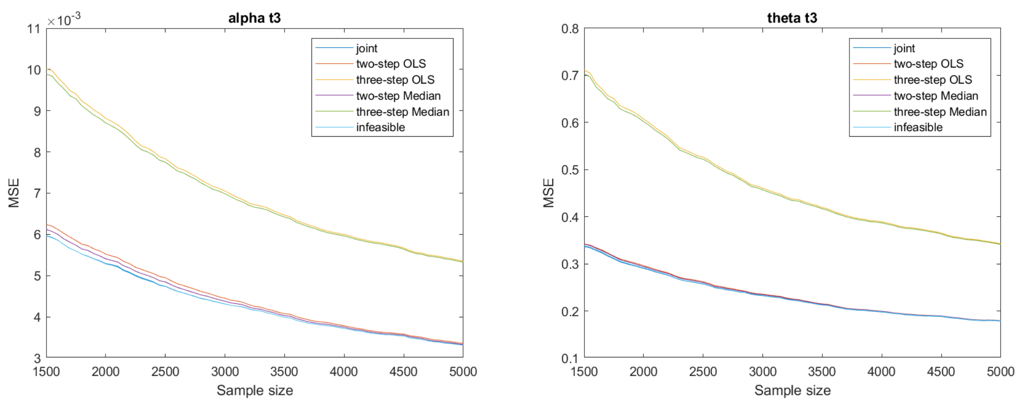
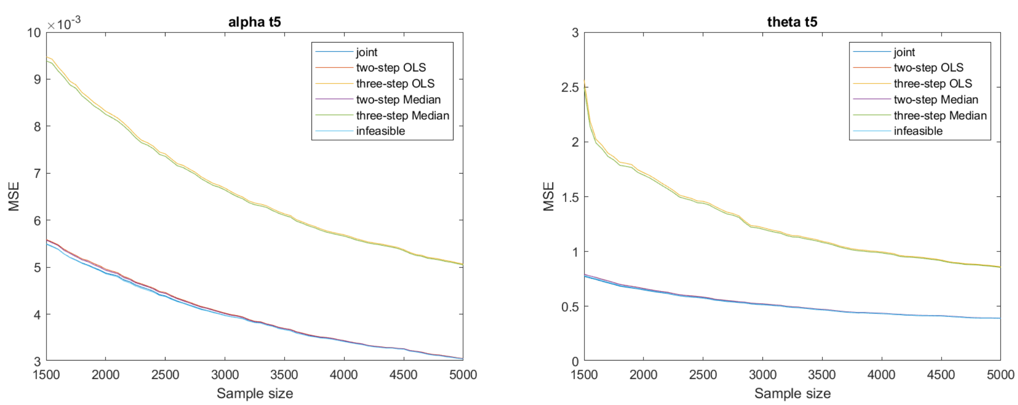



Figure 5: Gaussian copula, time trend, $\alpha_0 = 0.9$, marginal $t(5)$




When sample size is large, joint, two-step(OLS) and two-step(LAD) estimators behave roughly the same as ideal estimator. Three-step(OLS) and three-step(LAD) also coincides, but there is a clear gap between two classes. Above two figures are strong evidences for the following asymptotic relationship when both copula and marginal are symmetric:

- joint $\approx$ two-step(OLS) $\approx$ two-step(LAD) $\approx$ ideal $\succeq$ three-step(OLS) $\approx$ three-step(LAD)

# 18 Tail Dependence

For copulas with one-side tail dependence such as Clayton copula and Gumbel copula, symmetric property is not holding. The constant terms $A$ and $B$ are not zeros and the nonstationary structure will affect the limiting distribution of copula $\alpha$ and marginal $\theta$. In this section, we use simulation to illustrate the relationship between this effect and strength of tail dependence.

If the marginal distribution is symmetric, we still have: $\mathbb{E}g_{y\theta} = 0$. Also notice that both two copulas satisfy:

$$c(u_1, u_2, \alpha) \equiv c(u_2, u_1, \alpha)$$

hence

$$\mathbb{E}l_{1\alpha}(Y_{t-1}, Y_t, \alpha_0, \theta_0) = \mathbb{E}l_{2\alpha}(Y_{t-1}, Y_t, \alpha_0, \theta_0)$$

$$\mathbb{E}l_{1\theta}(Y_{t-1}, Y_t, \alpha_0, \theta_0) = \mathbb{E}l_{2\theta}(Y_{t-1}, Y_t, \alpha_0, \theta_0)$$

Thus we only need to simulate: $\mathbb{E}l_{1\alpha}$, $\mathbb{E}l_{1\theta}$, $\mathbb{E}l_{\alpha\alpha}$, $\mathbb{E}l_{\alpha\theta}$, $\mathbb{E}l_{\theta\theta}$ and $\mathbb{E}g_{\theta\theta}$. We generate $M = 2 \times 10^6$ random pairs $(Y_{1i}, Y_{2i}) \sim C(F_0(\cdot), F_0(\cdot); \alpha_0)$ independent each other, to simulate these constants. For example, $\mathbb{E}l_{1\alpha}$ is approximated as:

$$\frac{1}{M}\sum_{i=1}^{M} l_{1\alpha}(Y_{1i}, Y_{2i}, \alpha_0, \theta_0)$$

Copula parameter $\alpha$ ranges from 2 to 15 for both Clayton and Gumbel. X axis is $\alpha$ and $Y$ axis is constant $A$ or $B$. Simulation results are graphed below:

Figure 6: Clayton copula

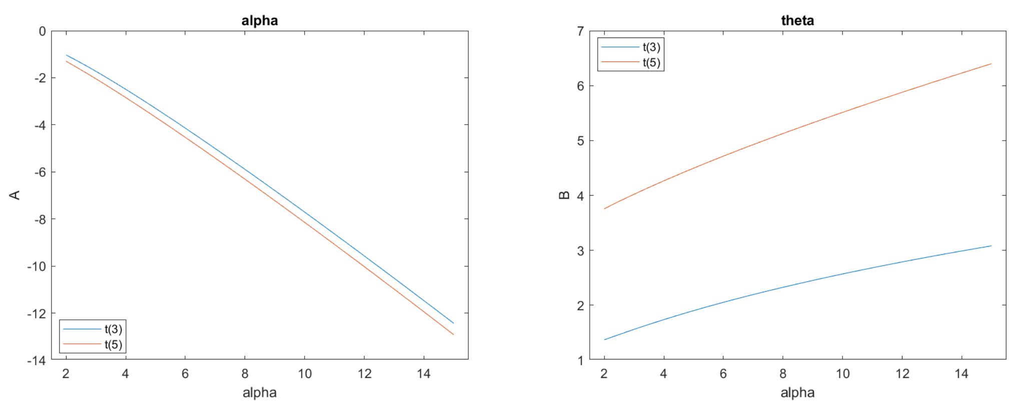
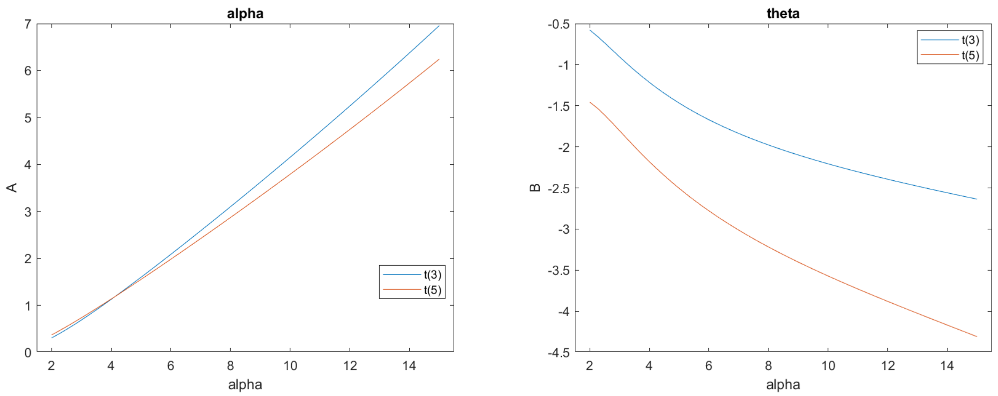


Figure 7: Gumbel copula




For marginal being both $t(3)$ and $t(5)$, absolute value of $A$ and $B$ is strictly increasing with tail dependence. The theoretical result here is consistent with finite simulation findings that relative MSE of our estimators comparing to the infeasible bechmark is increasingly with tail dependence. This fact indicates that nonstationary filtering term may have a huge difference to our estimators comparing with ideal estimator when tail dependence is very strong.

For Clayton copula, we also make a similar experiment as in **Section** 17.2.2, where the X axis the sample size and Y axis the square root of finite sample MSE.

Figure 8: Clayton copula, time trend, $\alpha_0 = 2$, marginal $t(3)$

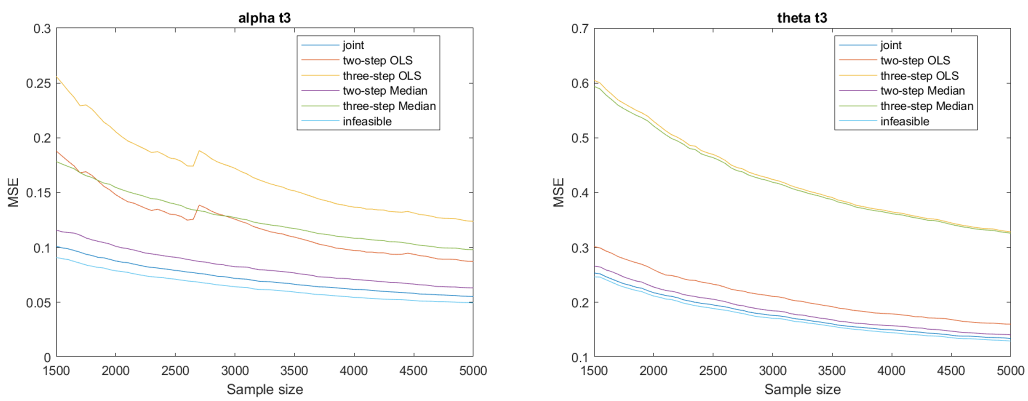
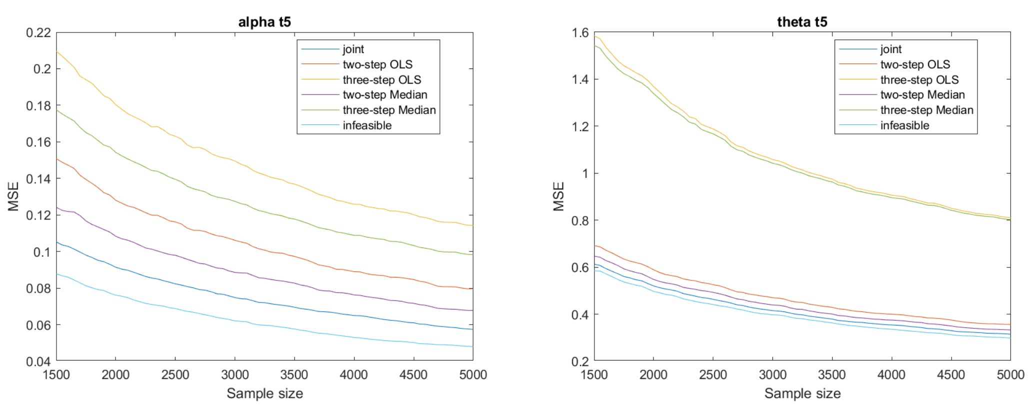



Figure 9: Clayton copula, time trend, $\alpha_0 = 2$, marginal $t(5)$



As the constant term $A \neq 0$ and $B \neq 0$, linear filtering method does affect the limiting distribution:

- two-step (OLS) $\not\approx$ two-step (LAD)
- three-step (OLS) $\not\approx$ three-step (LAD)

Joint estimator is still the best estimator available closed to ideal estimator. And the following general relationship holds:

- ideal $\succeq$ joint $\succeq$ two-step $\succeq$ three-step

Notice for both figures, especially square root of marginal $\theta$, three step estimator is far behind other competitiors.

# 19 Asymmetric marginal

When both copula and marginal are symmetric, nonstationarity has no effect on copula parameter estimation. We have shown that, when copula is asymmetric, despite the fact that marginal is still symmetric, $\sqrt{n}G_n(\hat{\beta} - \beta_0)$ does involve into the limiting distribution of $\sqrt{n}(\hat{\alpha} - \alpha_0)$ and $\sqrt{n}(\hat{\theta} - \theta_0)$. Also, this effect is positively dependent on degree of asymmetry (tail dependence). Then a natural question arises: what if the copula is symmetric but the marginal is asymmetric? Could we get a similar figure as above?

Before moving forward to the technical detail, we first introduce an asymmetric marginal distribution: linear combination of student t distribution and centered chi-square distribution.

Denote density of t distribution $f_t$:

$$f_t(y, \theta) = \frac{\Gamma\left(\frac{1+\theta}{2}\right)}{\sqrt{\pi\theta}\Gamma\left(\frac{\theta}{2}\right)} \left(1 + \frac{y^2}{\theta}\right)^{-\frac{1+\theta}{2}}$$

Denote density of chi-square distribution $f_\chi$:

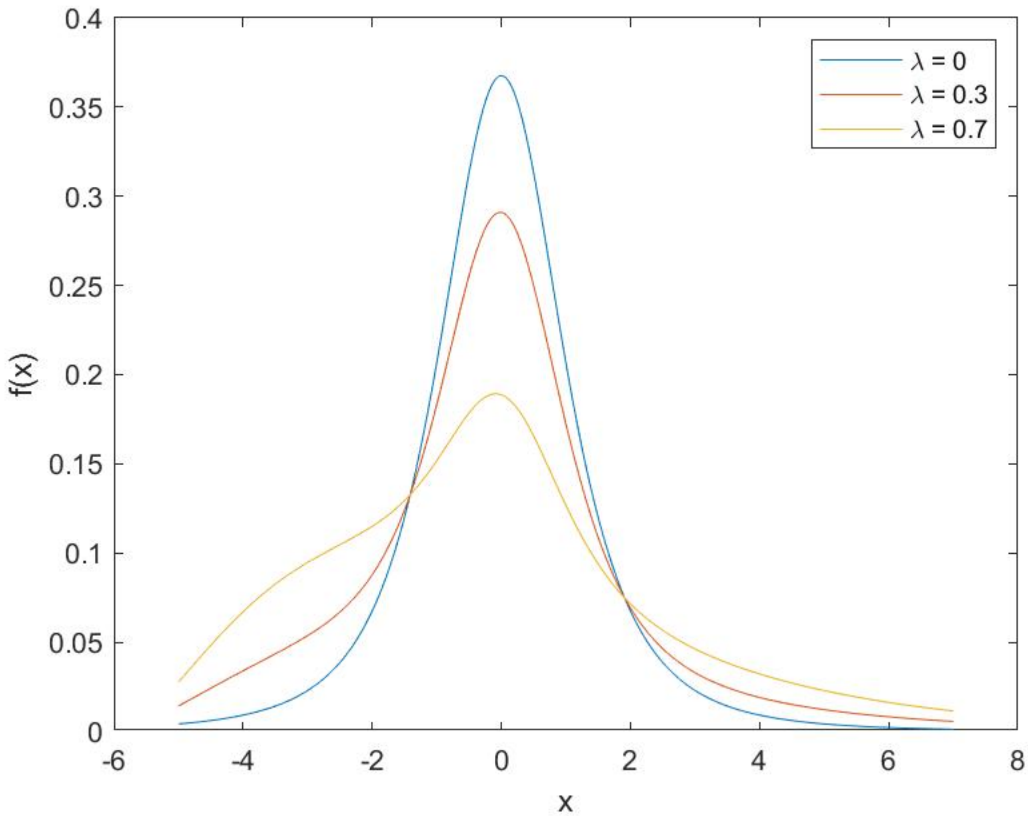
$$f_\chi(y, \theta) = \frac{1}{2^{\frac{\theta}{2}}\Gamma\left(\frac{\theta}{2}\right)} y^{\frac{\theta}{2}-1} e^{-\frac{\theta}{2}} \mathbb{I}(y > 0)$$

Then the density of our asymmetric marginal could be expressed as:

$$f(y, \theta) := (1 - \lambda) f_t(y, \theta) + \lambda f_\chi(y + 2\theta, 2\theta) \tag{12}$$

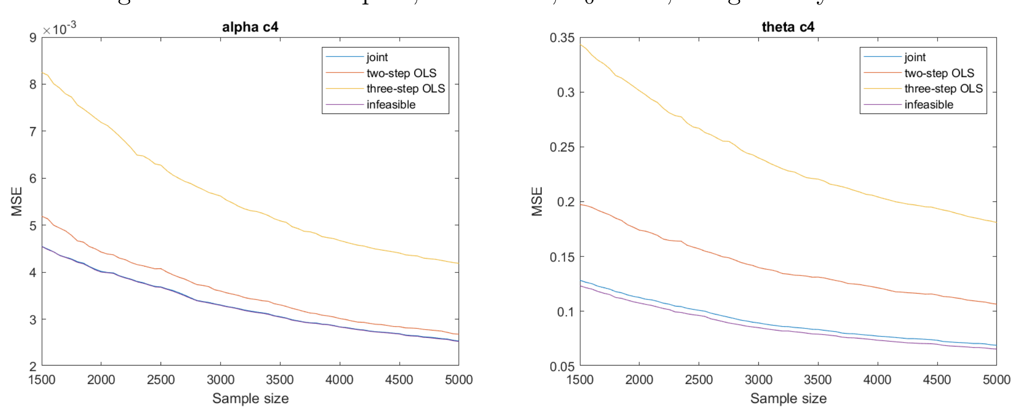
The following figure demonstrates the effect of different weight $\lambda$ on asymmetry. The more weight (larger $\lambda$) we put on the mixed normal, the more asymmetry the density has.

Figure 10: Plot of density for different weight $\lambda$




Simulation design is similar as in **Section** 17.2.2, except that the marginal is asymmetric with the weight $\lambda = 0.7$ and true parameter value $\theta_0 = 4$. Also, as the median of the marginal is not zero any more, both two-step and three-stem estimator are generated only through OLS in the first step linear filtering.

Figure 11: Gaussian copula, time trend, $\alpha_0 = 0.9$, marginal asymmetric




Gaps are obvious among several estimators when sample size is large. Joint estimator is unbelievely well for copula parameter $\alpha$ as it could catch up with the ideal estimator, although for marginal $\theta$, joint estimator is still a little behind the ideal estimator. Above figure is a strong evidence for the following asymptotic relationship when symmetry relationship is broken up:

- ideal $\succeq$ joint $\succeq$ two-step $\succeq$ three-step (where the efficiency loss of joint estimator comparing to ideal case is tiny)

Notice for both figures, three step estimator is far behind other competitiors.

# 20 Finite sample performance

We exam the finite sample performance of the joint, two-step and three-step estimators. For comparison purpose, we also simulate the ideal estimator (I) serving as a benchmark. Simulation repetition times is $M = 2000$ and time series length is $n = 500$. Copula and its parameter choice is as in Section 15.2. Marginal distribution is set as student t distribution with degree of freedom 3 or 5, together with the asymmetric marginal as in equation 12.

We consider two types of nonstationarity:

- Deterministic time trend as in Example 4: $X_t = t$, $\beta_0 = 1$.
- Cointegration time series as in Example 5: $X_t = X_{t-1} + v_t$, $\beta_0 = 1$.

Under different combinations of nonstationary filtering (2), copula functions (5), copula parameters (4) and marginal distributions (3), 120 tables are generated. Due to lack of space, we only include several typical results in **Section** 23. Full simulation results are available through request from the author.

**Section** 23.1 summarizes the results for Gaussian copula with copula parameter $\alpha_0 = 0.9$. **Section** 23.2 summarizes the results for Frank copula with copula parameter $\alpha_0 = 15$. **Section** 23.3 summarizes the results for Clayton copula with copula parameter $\alpha_0 = 5$. Joint estimator is denoted as (P1). Two step estimator through OLS is denoted as (P2). Two step estimator through LAD is denoted as (M2). Three step estimator through OLS is denoted as (P3). Three step estimator through LAD is denoted as (M3). Ideal estimator is denoted as (I). The Monte Carlo bias, standard deviation (std), Mean Square Error (MSE), and the ratio of MSE over the MSE of infeasible estimator (RMSE), for both copula $\alpha$ and marginal $\theta$ are reported in each table. All the simulations reveal the following clear patterns[18]:

**First**, the joint estimator performs very well in terms of bias, variance, MSE compared to the other estimators (two-step and three-step). The RMSE of joint estimator is the smallest in almost all the situations.

**Second**, when the positive dependence is quite strong, three-step estimator may diverge in finite sample. For example, in table 38 and 39, the RMSE for $\theta$(P3) and $\theta$(M3) are larger than 10,000.

**Third**, even when both copula and marginal are symmetric, hence joint estimator and two-step estimator are equivalent asymptotically, joint estimator will strictly dominate in finite sample. For example, in table 44, the RMSE of $\alpha$(P1) is 1.041, very close to the ideal estimator, while the RMSE of $\alpha$(P2) is 3.57 and the the RMSE of $\alpha$(M2) is 2.83.

**Last**, for copulas with strong tail dependence, both two-step estimator and three-step estimator do not perform very well. For example, in table 50, the RMSE of $\alpha$(P2) and $\alpha$(M2) are larger than 70 and the RMSE of $\theta$(M2) and $\theta$(M3) are larger than 16. Joint estimator performs very well in both copula $\alpha$ and marginal $\theta$ even under strong tail dependence (both lower tail for Clayton and upper tail for Gumbel). For example, in above case (table 50), the RMSE of $\alpha$(P1) is 1.588 and the RMSE of $\theta$(P1) is 1.08.

[18] An exception is EFGM copula, where all estimators behave quite similar. Theoretical explanation is shown in Chen et al. [2009] as EFGM copula is very close to the independent copula ($C(u_1, u_2) := u_1u_2$, $c(u_1, u_2) \equiv 1$), because the distance between EFGM copula function to the independent copula function is $\alpha u_1u_2(1-u_1)(1-u_2) \leq 0.0625\alpha$ for $\alpha \in [-1, 1]$. This could also be illustrated by Kendall's tau and Spearman's rho. $\alpha = 0.9 \Rightarrow \tau = 0.2$, $\rho = 0.3$; $\alpha = 0.5 \Rightarrow \tau = 0.111$, $\rho = 0.167$; $\alpha = -0.9 \Rightarrow \tau = -0.2$, $\rho = -0.3$; $\alpha = -0.5 \Rightarrow \tau = -0.111$, $\rho = -0.167$. The value is very close to 0, indicating nearly independence.

## 21 Conclusion

This paper considers estimation of copula based dynamic parametric models coupled with nonstationary filtration. Two new methods are proposed: joint estimator and two-step estimator. New theoretical results are obtained regarding:

- Conditions under which these estimator are equivalent asymptotically:
  **when both copula and marginal are symmetric:**
  joint $\approx$ two-step(OLS) $\approx$ two-step(LAD) $\approx$ ideal $\succeq$ three-step(OLS) $\approx$ three-step(LAD)
  **when either copula or marginal is asymmetric:**
  ideal $\succeq$ joint $\succeq$ two-step $\succeq$ three-step

- Tail dependence $\Rightarrow$ effect of nonstationarity on limiting distributions:
  The stronger the tail dependence, the larger effect of nonstationarity (both the nonstationary structure and the nonstationary estimation method) to the limiting distribution.

Monte Carlo simulation compares the performance between literature three-step estimator and our two new estimators. Three-step estimator is in general inferior to joint estimator and two-step estimator. Joint estimator is found to always be superior to all other estimators in a variety of Monte Carlo simulation designs, especially in the presence of strong tail dependence. Hence joint method is what we suggest in practical use.

**Extension.** We have also done some preliminary simulation for the dependent cointegration case $Z_t = X_t\beta + Y_t$ when $X_t$ is not independent with $Y_t$. Joint estimator is still doing very well. However, we do not explicitly measure the dependent structure into the objective function of the joint method. There should be a more efficient way to estimate the dependent cointegration model.

Another extension would be inference. Our model fully characterizes the joint probability distribution, making likelihood ratio test available as in Chen and Xiao [2016]. Simulations of bootstrap likelihood ration test on joint estimation have been done and the size is great. Theoretical validation of this inference procedure is of practical value for further research.

## References


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

Donald WK Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856, 1993.

Donald WK Andrews and Hong-Yuan Chen. Approximately median-unbiased estimation of autoregressive models. *Journal of Business & Economic Statistics*, 12(2):187–204, 1994.

Donald WK Andrews and Werner Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, pages 1383–1414, 1994.

Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78, 1998.

Brendan K Beare. Copulas and temporal dependence. *Econometrica*, 78(1):395–410, 2010.

Brendan K Beare. Archimedean copulas and temporal dependence. *Econometric Theory*, 28(6): 1165–1185, 2012.

Alon Brav and Reuven Lehavy. An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics. *The Journal of Finance*, 58(5):1933–1968, 2003.

Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach.* Springer Science & Business Media, 2003.

Xiaohong Chen. Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics*, 6:5549–5632, 2007.

Xiaohong Chen and Yanqin Fan. Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335, 2006.

Xiaohong Chen and Xiaotong Shen. Sieve extremum estimates for weakly dependent data. *Econometrica*, pages 289–314, 1998.

Xiaohong Chen and Zhijie Xiao. Copula-based time series with filtered nonstationarity. 2016.

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

Xiaohong Chen, Wei Biao Wu, and Yanping Yi. Efficient estimation of copula-based semiparametric markov models. *The Annals of Statistics*, pages 4214–4253, 2009.

Ciprian M Crainiceanu and Timothy J Vogelsang. Nonmonotonic power for tests of a mean shift in a time series. *Journal of Statistical Computation and Simulation*, 77(6):457–476, 2007.

William F Darsow, Bao Nguyen, Elwood T Olsen, et al. Copulas and markov processes. *Illinois Journal of Mathematics*, 36(4):600–642, 1992.

Robert M De Jong. Nonlinear minimization estimators in the presence of cointegrating relations. *Journal of Econometrics*, 110(2):241–259, 2002.

Ai Deng and Pierre Perron. The limit distribution of the cusum of squares test under general mixing conditions. *Econometric Theory*, 24(3):809–822, 2008.

Francis X Diebold and Celia Chen. Testing structural stability with endogenous breakpoint a size comparison of analytic and bootstrap procedures. *Journal of Econometrics*, 70(1):221–241, 1996.

Paul Doukhan, Jean-David Fermanian, Gabriel Lang, et al. *Copulas of a vector-valued stationary weakly dependent process*. INSEE, 2005.

Ray C Fair. Computing median unbiased estimates in macroeconometric models. *Journal of Applied Econometrics*, 11(4):431–435, 1996.

Evarist Giné and Joel Zinn. Bootstrapping general empirical measures. *The Annals of Probability*, pages 851–869, 1990.

Clive WJ Granger. Time series concepts for conditional distributions. *Oxford Bulletin of Economics and Statistics*, 65:689–701, 2003.

Bruce E Hansen. Testing for structural change in conditional models. *Journal of Econometrics*, 97 (1):93–115, 2000.

Bruce E Hansen. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(03):726–748, 2008.

Rustam Ibragimov. Copula-based characterizations for higher order markov processes. *Econometric Theory*, 25(03):819–846, 2009.

Harry Joe. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

Eric Jondeau and Michael Rockinger. The copula-garch model of conditional dependencies: An international stock market application. *Journal of international money and finance*, 25(5):827–853, 2006.

Ted Juhl and Zhijie Xiao. Tests for changing mean with monotonic power. *Journal of Econometrics*, 148(1):14–24, 2009.

Lutz Kilian. Accounting for lag order uncertainty in autoregressions: the endogenous lag order bootstrap algorithm. *Journal of Time Series Analysis*, 19(5):531–548, 1998.

Martial Longla and Magda Peligrad. Some aspects of modeling dependence in copula-based markov chains. *Journal of Multivariate Analysis*, 111:234–240, 2012.

RB Nelson. An introduction to copulas, lectures notes in statistics vol. 39, 1999.

Whitney K Newey and Kenneth D West. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, pages 777–787, 1987.

Cuong C Nguyen and M Ishaq Bhatti. Copula model dependency between oil prices and stock markets: Evidence from china and vietnam. *Journal of International Financial Markets, Institutions and Money*, 22(4):758–773, 2012.

Joon Y Park. An invariance principle for sieve bootstrap in time series. *Econometric Theory*, 18 (02):469–490, 2002.

Andrew J Patton. A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18, 2012.

Pierre Perron. *A test for changes in a polynomial trend function for a dynamic time series*. Université de Montréal, Centre de recherche et développement en économique, 1991.

Pierre Perron et al. Dealing with structural breaks. *Palgrave handbook of econometrics*, 1(2):278–352, 2006.

Werner Ploberger and Walter Krämer. The cusum test with ols residuals. *Econometrica: Journal of the Econometric Society*, pages 271–285, 1992.

Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954, 1988.

M Sklar. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8: 229–231, 1959.

Charles J Stone. Optimal global rates of convergence for nonparametric regression. *The annals of statistics*, pages 1040–1053, 1982.

H Victor, Rustam Ibragimov, Shaturgun Sharakhmetov, et al. Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series. In *Optimality*, pages 183–209. Institute of Mathematical Statistics, 2006.

Timothy J Vogelsang. Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics*, 88(2):283–299, 1999.

Xiaoqian Wen, Yu Wei, and Dengshi Huang. Measuring contagion between energy market and stock market during financial crisis: A copula approach. *Energy Economics*, 34(5):1435–1446, 2012.

# 22 Technical Appendix

## 22.1 Copula simulation

If $(Y_1, Y_2) \sim C(F_0(\cdot), F_0(\cdot))$ and we know $Y_1$ in advance, we can apply the conditional approach described in Nelson [1999] to generate uniform distributed time series satisfying the specific copula, then apply inverse distribution function $F_0^{-1}(\cdot)$ to get $Y_2$:

1. Let $U_1 = F_0(Y_1)$.

2. Generate a uniformly distributed random variable $\varepsilon$. Solve $U_2$ by $C_1(U_1, U_2) = \varepsilon$. Here $C_1 := \frac{\partial C}{\partial u_1}$ is the conditional distribution of $U_2$ given $U_1$.

3. $Y_2 = F_0^{-1}(U_2)$, here $F_0(\cdot)$ is true marginal distribution function of $Y_1$ and $Y_2$.

To generate a first order Markov process specified by a copula $C(\cdot, \cdot)$ and a marginal $F_0(\cdot)$, we can repeat this algorithm sequentially.

For five copulas we consider in 15.2, expressions of conditional distribution $C_1$ are:

- Gaussian copula

$$\begin{aligned} C_1(u_1, u_2, \alpha) &= \frac{1}{2\pi\sqrt{1-\alpha^2}} \cdot \frac{1}{\phi(\Phi^{-1}(u_1))} \cdot \\ &\int_{-\infty}^{\Phi^{-1}(u_2)} \exp\left[-\frac{[\Phi^{-1}(u_1)]^2 + x^2 - 2\alpha x \cdot \Phi^{-1}(u_1)}{2(1-\alpha^2)}\right] \mathrm{d}x \end{aligned}$$

- Frank copula

$$C_1(u_1, u_2; \alpha) = \left(1 - \frac{(1-e^{-\alpha u_1})(1-e^{-\alpha u_2})}{1-e^{-\alpha}}\right)^{-1} \cdot \frac{1-e^{-\alpha u_2}}{1-e^{-\alpha}} \cdot e^{-\alpha u_1}$$

- Clayton copula

$$C_1(u_1, u_2; \alpha) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}-1} \cdot u_1^{-\alpha-1}$$

- EFGM copula

$$C_1(u_1, u_2; \alpha) = u_2[1 + \alpha(1-u_2)(1-2u_1)]$$

- Gumbel copula

$$\begin{aligned} C_1(u_1, u_2; \alpha) &= \exp\left[-\left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}}\right] \cdot \\ &\left((-\log u_1)^\alpha + (-\log u_2)^\alpha\right)^{\frac{1}{\alpha}-1} \cdot (-\log u_1)^{\alpha-1} \cdot \frac{1}{u_1} \end{aligned}$$

## 22.2 Symmetry of copulas

For Gaussian, Frank and EFGM copula, the following symmetric property is satisfied:

$$c(u_1, u_2; \alpha) = c(1-u_1, 1-u_2; \alpha)$$

For Gaussian copula:

$$\begin{aligned} & c(1-u_1, 1-u_2; \alpha) \\ &= \frac{\phi_\alpha(\Phi^{-1}(1-u_1), \Phi^{-1}(1-u_2))}{\phi(\Phi^{-1}(1-u_1)) \cdot \phi(\Phi^{-1}(1-u_2))} \\ &= \frac{\phi_\alpha(-\Phi^{-1}(u_1), -\Phi^{-1}(u_2))}{\phi(-\Phi^{-1}(u_1)) \cdot \phi(-\Phi^{-1}(u_2))} \\ &= \frac{\phi_\alpha(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi(\Phi^{-1}(u_1)) \cdot \phi(\Phi^{-1}(u_2))} \\ &= c(u_1, u_2; \alpha) \end{aligned}$$

For Frank copula:

$$
\begin{aligned}
& c(1-u_1, 1-u_2; \alpha) \\
&= \alpha \cdot \frac{e^{\alpha u_1} e^{\alpha u_2} \cdot e^{-2\alpha}}{1-e^{-\alpha}} \cdot \left(1 - \frac{(1-e^{-\alpha} \cdot e^{\alpha u_1})(1-e^{-\alpha} \cdot e^{\alpha u_2})}{1-e^{-\alpha}}\right)^{-2} \\
&= \alpha \cdot \frac{e^{-\alpha u_1} e^{-\alpha u_2}}{1-e^{-\alpha}} \cdot \left(\frac{e^{-\alpha} \cdot e^{\alpha u_1} + e^{-\alpha} \cdot e^{\alpha u_2} - e^{-\alpha} - e^{-2\alpha} e^{\alpha u_1} e^{\alpha u_2}}{1-e^{-\alpha}}\right)^{-2} \cdot (e^{-\alpha u_1} e^{-\alpha u_2} \cdot e^{\alpha})^{-2} \\
&= \alpha \cdot \frac{e^{-\alpha u_1} e^{-\alpha u_2}}{1-e^{-\alpha}} \cdot \left(\frac{e^{-\alpha u_2} + e^{-\alpha u_1} - e^{-\alpha u_1} e^{-\alpha u_2} - e^{-\alpha}}{1-e^{\alpha}}\right)^{-2} \\
&= \alpha \cdot \frac{e^{-\alpha u_1} e^{-\alpha u_2}}{1-e^{-\alpha}} \cdot \left(1 - \frac{(1-e^{-\alpha u_1})(1-e^{-\alpha u_2})}{1-e^{-\alpha}}\right)^{-2} \\
&= c(u_1, u_2; \alpha)
\end{aligned}
$$

For EFGM copula:

$$
\begin{aligned}
& c(1-u_1, 1-u_2; \alpha) \\
&= 1 + \alpha(2u_1 - 1)(2u_2 - 1) \\
&= 1 + \alpha(1 - 2u_1)(1 - 2u_2) \\
&= c(u_1, u_2; \alpha)
\end{aligned}
$$

## 22.3 Constants $A$ and $B$

The expression for $A$ and $B$ is:

$$
\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \mathbb{E}l_{\alpha\alpha} & \mathbb{E}l_{\alpha\theta} \\ \mathbb{E}l_{\alpha\theta} & \mathbb{E}l_{\theta\theta} + \mathbb{E}g_{\theta\theta} \end{pmatrix}^{-1} \times \begin{pmatrix} \mathbb{E}l_{1\alpha} + \mathbb{E}l_{2\alpha} \\ \mathbb{E}g_{y\theta} + \mathbb{E}l_{1\theta} + \mathbb{E}l_{2\theta} \end{pmatrix}
$$

The inverse of matrix is:

$$
\begin{pmatrix} \mathbb{E}l_{\alpha\alpha} & \mathbb{E}l_{\alpha\theta} \\ \mathbb{E}l_{\alpha\theta} & \mathbb{E}l_{\theta\theta} + \mathbb{E}g_{\theta\theta} \end{pmatrix}^{-1} = \frac{1}{\mathbb{E}l_{\alpha\alpha} \cdot [\mathbb{E}l_{\theta\theta} + \mathbb{E}g_{\theta\theta}] - (\mathbb{E}l_{\alpha\theta})^2} \cdot \begin{pmatrix} \mathbb{E}l_{\theta\theta} + \mathbb{E}g_{\theta\theta} & -\mathbb{E}l_{\alpha\theta} \\ -\mathbb{E}l_{\alpha\theta} & \mathbb{E}l_{\alpha\alpha} \end{pmatrix}
$$

Combining the above two equations, we have:

$$
A = \frac{[\mathbb{E}l_{1\alpha} + \mathbb{E}l_{2\alpha}] \cdot [\mathbb{E}l_{\theta\theta} + \mathbb{E}g_{\theta\theta}] - [\mathbb{E}g_{y\theta} + \mathbb{E}l_{1\theta} + \mathbb{E}l_{2\theta}] \cdot \mathbb{E}l_{\alpha\theta}}{\mathbb{E}l_{\alpha\alpha} \cdot [\mathbb{E}l_{\theta\theta} + \mathbb{E}g_{\theta\theta}] - (\mathbb{E}l_{\alpha\theta})^2}
$$

$$
B = \frac{-\mathbb{E}l_{\alpha\theta} \cdot [\mathbb{E}l_{1\alpha} + \mathbb{E}l_{2\alpha}] + [\mathbb{E}g_{y\theta} + \mathbb{E}l_{1\theta} + \mathbb{E}l_{2\theta}] \cdot \mathbb{E}l_{\alpha\alpha}}{\mathbb{E}l_{\alpha\alpha} \cdot [\mathbb{E}l_{\theta\theta} + \mathbb{E}g_{\theta\theta}] - (\mathbb{E}l_{\alpha\theta})^2}
$$

# 23 Tables

## 23.1 Gaussian copula

### 23.1.1 Time trend

Table 38: Normal copula, $\alpha_0 = 0.9$; $X_t$ time trend; marginal $t(3)$

| $\alpha_0 = 0.9$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 0.899586 | 0.895923 | 0.897059 | 0.891702 | 0.892529 | 0.899323 |
| bias | -0.000414 | -0.004077 | -0.002941 | -0.008298 | -0.007471 | -0.000677 |
| std | 0.010914 | 0.011578 | 0.011258 | 0.018205 | 0.017829 | 0.010821 |
| MSE | 0.000119 | 0.000151 | 0.000135 | 0.000400 | 0.000374 | 0.000118 |
| relative | 1.014634 | 1.281681 | 1.151695 | 3.404846 | 3.178599 | 1 |
| 2.5% | 0.877465 | 0.872078 | 0.873887 | 0.854232 | 0.854905 | 0.877386 |
| 97.5% | 0.920068 | 0.917044 | 0.917807 | 0.924029 | 0.924064 | 0.919437 |

| $\theta_0 = 3$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 3.108162 | 3.211068 | 3.159608 | 5.245809 | 5.162396 | 3.133271 |
| bias | 0.108162 | 0.211068 | 0.159608 | 2.245809 | 2.162396 | 0.133271 |
| std | 0.660914 | 0.678009 | 0.683229 | 60.088270 | 59.325965 | 0.665609 |
| MSE | 0.448507 | 0.504246 | 0.492277 | 3615 | 3524 | 0.460797 |
| relative | 0.973329 | 1.094291 | 1.068317 | 7846 | 7648 | 1 |
| 2.5% | 2.159470 | 2.239992 | 2.186189 | 1.887734 | 1.892296 | 2.184898 |
| 97.5% | 4.669615 | 4.815974 | 4.781117 | 10.046583 | 9.744579 | 4.692251 |

Table 39: Normal copula, $\alpha_0 = 0.9$; $X_t$ time trend; marginal $t(5)$

| $\alpha_0 = 0.9$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 0.899402 | 0.897445 | 0.897687 | 0.892387 | 0.893077 | 0.899231 |
| bias | -0.000598 | -0.002555 | -0.002313 | -0.007613 | -0.006923 | -0.000769 |
| std | 0.010171 | 0.010353 | 0.010274 | 0.016714 | 0.016588 | 0.010017 |
| MSE | 0.000104 | 0.000114 | 0.000111 | 0.000337 | 0.000323 | 0.000101 |
| relative | 1.028364 | 1.126465 | 1.098695 | 3.341862 | 3.200943 | 1 |
| 2.5% | 0.878585 | 0.876059 | 0.876449 | 0.858950 | 0.859400 | 0.878924 |
| 97.5% | 0.918945 | 0.916544 | 0.916934 | 0.922681 | 0.923143 | 0.918046 |

| $\theta_0 = 5$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 5.441528 | 5.613763 | 5.593364 | 128.951013 | 115.432356 | 5.516748 |
| bias | 0.441528 | 0.613763 | 0.593364 | 123.951013 | 110.432356 | 0.516748 |
| std | 2.312500 | 2.469017 | 2.544495 | 734.161241 | 690.961193 | 2.549045 |
| MSE | 5.542604 | 6.472750 | 6.826538 | 554356 | 489622 | 6.764658 |
| relative | 0.819347 | 0.956848 | 1.009148 | 81948 | 72379 | 1 |
| 2.5% | 3.314031 | 3.426234 | 3.380807 | 2.835703 | 2.801019 | 3.416062 |
| 97.5% | 9.671101 | 10.334626 | 10.016158 | 1954.050667 | 502.746450 | 9.871510 |

Table 40: Normal copula, $\alpha_0 = 0.9$; $X_t$ time trend; marginal $t(4) + c(8)$

| $\alpha_0 = 0.9$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 0.89926630 | 0.89589313 | 0.89195210 | 0.89928966 |
| bias | -0.00073370 | -0.00410687 | -0.00804790 | -0.00071034 |
| std | 0.00876768 | 0.01053527 | 0.01556571 | 0.00860546 |
| MSE | 0.00007741 | 0.00012786 | 0.00030706 | 0.00007456 |
| relative | 1.03825395 | 1.71487288 | 4.11838185 | 1 |
| 2.5% | 0.88172902 | 0.87263578 | 0.85927207 | 0.88203373 |
| 97.5% | 0.91593527 | 0.91378343 | 0.91854498 | 0.91538371 |

| $\theta_0 = 4$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 3.98303536 | 3.91928173 | 3.84988800 | 3.98217360 |
| bias | -0.01696464 | -0.08071827 | -0.15011200 | -0.01782640 |
| std | 0.25753093 | 0.34476266 | 0.61756440 | 0.24568923 |
| MSE | 0.06660998 | 0.12537673 | 0.40391940 | 0.06068098 |
| relative | 1.09770765 | 2.06616190 | 6.65644150 | 1 |
| 2.5% | 3.46231455 | 3.22082383 | 2.79917365 | 3.48185908 |
| 97.5% | 4.48501683 | 4.60080147 | 5.23210550 | 4.45860162 |

### 23.1.2 Cointegration

Table 41: Normal copula, $\alpha_0 = 0.9$; $X_t$ cointegration; marginal $t(3)$

| $\alpha_0 = 0.9$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 0.899618 | 0.895688 | 0.896885 | 0.891601 | 0.892495 | 0.899323 |
| bias | -0.000382 | -0.004312 | -0.003115 | -0.008399 | -0.007505 | -0.000677 |
| std | 0.010907 | 0.011696 | 0.011368 | 0.018211 | 0.017889 | 0.010821 |
| MSE | 0.000119 | 0.000155 | 0.000139 | 0.000402 | 0.000376 | 0.000118 |
| relative | 1.013234 | 1.321711 | 1.181816 | 3.421089 | 3.201377 | 1 |
| 2.5% | 0.877585 | 0.870970 | 0.874165 | 0.853676 | 0.854366 | 0.877386 |
| 97.5% | 0.920092 | 0.916969 | 0.918120 | 0.925649 | 0.925465 | 0.919437 |

| $\theta_0 = 3$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 3.111106 | 3.216810 | 3.166404 | 6.185583 | 6.123448 | 3.133271 |
| bias | 0.111106 | 0.216810 | 0.166404 | 3.185583 | 3.123448 | 0.133271 |
| std | 0.661508 | 0.686007 | 0.687187 | 102.839123 | 102.853688 | 0.665609 |
| MSE | 0.449937 | 0.517613 | 0.499917 | 10586 | 10588 | 0.460797 |
| relative | 0.976432 | 1.123299 | 1.084897 | 22973 | 22978 | 1 |
| 2.5% | 2.155947 | 2.233299 | 2.178310 | 1.884454 | 1.882621 | 2.184898 |
| 97.5% | 4.692468 | 4.827157 | 4.775221 | 10.019529 | 9.553549 | 4.692250 |

Table 42: Normal copula, $\alpha_0 = 0.9$; $X_t$ cointegration; marginal $t(5)$

| $\alpha_0 = 0.9$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 0.899450 | 0.897219 | 0.897506 | 0.892329 | 0.893027 | 0.899231 |
| bias | -0.000550 | -0.002781 | -0.002494 | -0.007671 | -0.006973 | -0.000769 |
| std | 0.010146 | 0.010450 | 0.010360 | 0.016727 | 0.016645 | 0.010017 |
| MSE | 0.000103 | 0.000117 | 0.000114 | 0.000339 | 0.000326 | 0.000101 |
| relative | 1.022790 | 1.158505 | 1.124997 | 3.354979 | 3.226355 | 1 |
| 2.5% | 0.878972 | 0.875126 | 0.876368 | 0.858020 | 0.858371 | 0.878924 |
| 97.5% | 0.918828 | 0.916528 | 0.916940 | 0.924178 | 0.924550 | 0.918046 |

| $\theta_0 = 5$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 5.445645 | 5.644230 | 5.608832 | 126.384921 | 109.907787 | 5.516748 |
| bias | 0.445645 | 0.644230 | 0.608832 | 121.384921 | 104.907787 | 0.516748 |
| std | 2.102694 | 2.726667 | 2.524513 | 730.030226 | 674.096598 | 2.549043 |
| MSE | 4.619921 | 7.849745 | 6.743841 | 547678 | 465411 | 6.764649 |
| relative | 0.682951 | 1.160407 | 0.996924 | 80961 | 68800 | 1 |
| 2.5% | 3.328483 | 3.382052 | 3.342897 | 2.792522 | 2.769337 | 3.416062 |
| 97.5% | 9.775784 | 10.088137 | 10.020469 | 1512.932685 | 279.039347 | 9.871510 |

Table 43: Normal copula, $\alpha_0 = 0.9$; $X_t$ cointegration; marginal $t(4) + c(8)$

| $\alpha_0 = 0.9$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 0.89924269 | 0.89599635 | 0.89197074 | 0.89928966 |
| bias | -0.00075731 | -0.00400365 | -0.00802926 | -0.00071034 |
| std | 0.00881583 | 0.01051900 | 0.01541873 | 0.00860546 |
| MSE | 0.00007829 | 0.00012668 | 0.00030221 | 0.00007456 |
| relative | 1.05008071 | 1.69905026 | 4.05327818 | 1 |
| 2.5% | 0.88116720 | 0.87307901 | 0.85893169 | 0.88203373 |
| 97.5% | 0.91581359 | 0.91420738 | 0.91924271 | 0.91538371 |

| $\theta_0 = 4$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 3.97992926 | 3.93659601 | 3.85902482 | 3.98217360 |
| bias | -0.02007074 | -0.06340399 | -0.14097518 | -0.01782640 |
| std | 0.25817535 | 0.32945915 | 0.59404548 | 0.24568923 |
| MSE | 0.06705734 | 0.11256340 | 0.37276403 | 0.06068098 |
| relative | 1.10508011 | 1.85500293 | 6.14301268 | 1 |
| 2.5% | 3.47094721 | 3.28299957 | 2.85856438 | 3.48185908 |
| 97.5% | 4.47861583 | 4.58205128 | 5.15340742 | 4.45860162 |

## 23.2 Frank copula

### 23.2.1 Time trend

Table 44: Frank copula, $\alpha_0 = 15$; $X_t$ time trend; marginal $t(3)$

| $\alpha_0 = 15$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 14.975880 | 14.234989 | 14.353372 | 14.204292 | 14.311321 | 14.967962 |
| bias | -0.024120 | -0.765011 | -0.646628 | -0.795708 | -0.688679 | -0.032038 |
| std | 0.714247 | 1.080195 | 0.983946 | 1.089068 | 0.995717 | 0.699688 |
| MSE | 0.510731 | 1.752063 | 1.386278 | 1.819220 | 1.465731 | 0.490589 |
| relative | 1.041056 | 3.571343 | 2.825739 | 3.708233 | 2.987694 | 1 |
| 2.5% | 13.696303 | 11.828572 | 12.242359 | 11.801502 | 12.184435 | 13.688136 |
| 97.5% | 16.446371 | 16.090483 | 16.082049 | 16.088059 | 16.072224 | 16.425091 |

| $\theta_0 = 3$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 3.135608 | 3.199695 | 3.147308 | 3.320591 | 3.286403 | 3.137857 |
| bias | 0.135608 | 0.199695 | 0.147308 | 0.320591 | 0.286403 | 0.137857 |
| std | 0.532003 | 0.567349 | 0.550899 | 0.796558 | 0.774648 | 0.532244 |
| MSE | 0.301417 | 0.361763 | 0.325189 | 0.737284 | 0.682106 | 0.302288 |
| relative | 0.997119 | 1.196748 | 1.075759 | 2.439011 | 2.256476 | 1 |
| 2.5% | 2.360041 | 2.376653 | 2.347515 | 2.281861 | 2.267837 | 2.363160 |
| 97.5% | 4.349428 | 4.510561 | 4.390622 | 5.287423 | 5.182795 | 4.349278 |

Table 45: Frank copula, $\alpha_0 = 15$; $X_t$ time trend; marginal $t(5)$

| $\alpha_0 = 15$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 14.978688 | 14.383173 | 14.415088 | 14.345261 | 14.375190 | 14.972209 |
| bias | -0.021312 | -0.616827 | -0.584912 | -0.654739 | -0.624810 | -0.027791 |
| std | 0.704850 | 0.952712 | 0.929078 | 0.967364 | 0.941058 | 0.688958 |
| MSE | 0.497268 | 1.288136 | 1.205308 | 1.364476 | 1.275978 | 0.475435 |
| relative | 1.045921 | 2.709384 | 2.535170 | 2.869954 | 2.683812 | 1 |
| 2.5% | 13.703685 | 12.296102 | 12.429349 | 12.249768 | 12.412091 | 13.716361 |
| 97.5% | 16.434110 | 16.123656 | 16.118490 | 16.095313 | 16.092844 | 16.396803 |

| $\theta_0 = 5$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 5.427665 | 5.572849 | 5.482046 | 8.527833 | 8.424681 | 5.427466 |
| bias | 0.427665 | 0.572849 | 0.482046 | 3.527833 | 3.424681 | 0.427466 |
| std | 1.473976 | 1.746573 | 1.692623 | 111.697258 | 111.696841 | 1.466280 |
| MSE | 2.355503 | 3.378673 | 3.097342 | 12488.723163 | 12487.912708 | 2.332705 |
| relative | 1.009773 | 1.448392 | 1.327790 | 5353.750531 | 5353.403099 | 1 |
| 2.5% | 3.678427 | 3.680425 | 3.633327 | 3.539867 | 3.497319 | 3.679464 |
| 97.5% | 8.893497 | 9.255556 | 8.929565 | 12.060738 | 11.615047 | 8.859804 |

Table 46: Frank copula, $\alpha_0 = 15$; $X_t$ time trend; marginal $t(4) + c(8)$

| $\alpha_0 = 15$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 14.95123620 | 14.30776466 | 14.22129885 | 14.95503530 |
| bias | -0.04876380 | -0.69223534 | -0.77870115 | -0.04496470 |
| std | 0.70195759 | 1.05389251 | 1.08743938 | 0.69067359 |
| MSE | 0.49512236 | 1.58987919 | 1.78889988 | 0.47905183 |
| relative | 1.03354653 | 3.31880410 | 3.73425117 | 1 |
| 2.5% | 13.70918320 | 11.93998538 | 11.81915939 | 13.69693652 |
| 97.5% | 16.37324070 | 16.10425599 | 16.03290645 | 16.38130845 |

| $\theta_0 = 4$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 3.94812795 | 3.93748755 | 3.88713798 | 3.95112650 |
| bias | -0.05187205 | -0.06251245 | -0.11286202 | -0.04887350 |
| std | 0.24502390 | 0.37274686 | 0.51629464 | 0.24342824 |
| MSE | 0.06272742 | 0.14284803 | 0.27929799 | 0.06164593 |
| relative | 1.01754359 | 2.31723373 | 4.53068020 | 1 |
| 2.5% | 3.43126609 | 3.20945303 | 2.86512379 | 3.43689742 |
| 97.5% | 4.39935160 | 4.68992269 | 4.96361424 | 4.40169095 |

#### 23.2.2 Cointegration

Table 47: Frank copula, $\alpha_0 = 15$; $X_t$ cointegration; marginal $t(3)$

| $\alpha_0 = 15$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 14.981861 | 14.224796 | 14.346228 | 14.196673 | 14.307627 | 14.967962 |
| bias | -0.018139 | -0.775204 | -0.653772 | -0.803327 | -0.692373 | -0.032038 |
| std | 0.710132 | 1.075943 | 0.980935 | 1.085849 | 0.993013 | 0.699688 |
| MSE | 0.504616 | 1.758594 | 1.389652 | 1.824401 | 1.465456 | 0.490589 |
| relative | 1.028592 | 3.584657 | 2.832618 | 3.718795 | 2.987133 | 1 |
| 2.5% | 13.702979 | 11.744206 | 12.112664 | 11.723884 | 12.093277 | 13.688144 |
| 97.5% | 16.481840 | 16.111950 | 16.157846 | 16.071548 | 16.103368 | 16.425103 |

| $\theta_0 = 3$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 3.135664 | 3.198107 | 3.148140 | 3.311231 | 3.277886 | 3.137858 |
| bias | 0.135664 | 0.198107 | 0.148140 | 0.311231 | 0.277886 | 0.137858 |
| std | 0.532424 | 0.567431 | 0.551989 | 0.781120 | 0.760275 | 0.532244 |
| MSE | 0.301880 | 0.361224 | 0.326637 | 0.707013 | 0.655239 | 0.302288 |
| relative | 0.998650 | 1.194968 | 1.080549 | 2.338872 | 2.167599 | 1 |
| 2.5% | 2.362109 | 2.384196 | 2.357822 | 2.274032 | 2.257115 | 2.363166 |
| 97.5% | 4.352706 | 4.541677 | 4.400361 | 5.316165 | 5.206149 | 4.349280 |

Table 48: Frank copula, $\alpha_0 = 15$; $X_t$ cointegration; marginal $t(5)$

| $\alpha_0 = 15$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 14.984991 | 14.372991 | 14.404355 | 14.338307 | 14.368163 | 14.972208 |
| bias | -0.015009 | -0.627009 | -0.595645 | -0.661693 | -0.631837 | -0.027792 |
| std | 0.700414 | 0.951904 | 0.930451 | 0.966884 | 0.942239 | 0.688957 |
| MSE | 0.490805 | 1.299263 | 1.220531 | 1.372702 | 1.287032 | 0.475434 |
| relative | 1.032330 | 2.732795 | 2.567196 | 2.887264 | 2.707071 | 1 |
| 2.5% | 13.717547 | 12.209744 | 12.310220 | 12.167040 | 12.310457 | 13.716359 |
| 97.5% | 16.456527 | 16.139807 | 16.168742 | 16.078081 | 16.112626 | 16.396771 |

| $\theta_0 = 5$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 5.425945 | 5.570076 | 5.483073 | 6.036676 | 5.934112 | 5.427467 |
| bias | 0.425945 | 0.570076 | 0.483073 | 1.036676 | 0.934112 | 0.427467 |
| std | 1.471667 | 1.635090 | 1.581719 | 2.752677 | 2.621064 | 1.466277 |
| MSE | 2.347234 | 2.998507 | 2.735193 | 8.651927 | 7.742542 | 2.332697 |
| relative | 1.006232 | 1.285425 | 1.172545 | 3.708980 | 3.319137 | 1 |
| 2.5% | 3.679114 | 3.693157 | 3.652556 | 3.522234 | 3.476637 | 3.679459 |
| 97.5% | 8.852422 | 9.248426 | 9.043495 | 12.211053 | 11.653345 | 8.859796 |

Table 49: Frank copula, $\alpha_0 = 15$; $X_t$ cointegration; marginal $t(4) + c(8)$

| $\alpha_0 = 15$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 14.95458399 | 14.31933663 | 14.24088918 | 14.95503565 |
| bias | -0.04541601 | -0.68066337 | -0.75911082 | -0.04496435 |
| std | 0.70035035 | 1.05078606 | 1.07988563 | 0.69067396 |
| MSE | 0.49255323 | 1.56745396 | 1.74240220 | 0.47905231 |
| relative | 1.02818257 | 3.27198916 | 3.63718569 | 1 |
| 2.5% | 13.68603371 | 11.83766346 | 11.69052361 | 13.69693116 |
| 97.5% | 16.39749940 | 16.16419693 | 16.14412387 | 16.38127765 |

| $\theta_0 = 4$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 3.94648959 | 3.95418829 | 3.90042707 | 3.95112643 |
| bias | -0.05351041 | -0.04581171 | -0.09957293 | -0.04887357 |
| std | 0.24758449 | 0.36969669 | 0.47773050 | 0.24342805 |
| MSE | 0.06416144 | 0.13877435 | 0.23814120 | 0.06164584 |
| relative | 1.04080733 | 2.25115509 | 3.86305363 | 1 |
| 2.5% | 3.41116653 | 3.23130785 | 2.94809193 | 3.43689603 |
| 97.5% | 4.39536611 | 4.70760663 | 4.84631928 | 4.40169048 |

## 23.3 Clayton copula

### 23.3.1 Time trend

Table 50: Clayton copula, $\alpha_0 = 5$; $X_t$ time trend; marginal $t(3)$

| $\alpha_0 = 5$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 5.004177 | 5.180811 | 5.087672 | 4.695532 | 4.856148 | 4.993004 |
| bias | 0.004177 | 0.180811 | 0.087672 | -0.304468 | -0.143852 | -0.006996 |
| std | 0.382395 | 7.241352 | 2.589089 | 1.420804 | 1.216885 | 0.303379 |
| MSE | 0.146243 | 52.469864 | 6.711067 | 2.111385 | 1.501503 | 0.092088 |
| relative | 1.588085 | 569.781215 | 72.876875 | 22.927970 | 16.305134 | 1 |
| 2.5% | 4.323025 | 3.366441 | 3.750594 | 3.225135 | 3.584222 | 4.445209 |
| 97.5% | 5.765372 | 7.740841 | 7.079887 | 8.107023 | 7.573691 | 5.614109 |

| $\theta_0 = 3$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 3.098873 | 3.154934 | 3.046188 | 4.302119 | 4.147187 | 3.102354 |
| bias | 0.098873 | 0.154934 | 0.046188 | 1.302119 | 1.147187 | 0.102354 |
| std | 0.500921 | 0.785852 | 0.653033 | 1.780595 | 1.630571 | 0.480340 |
| MSE | 0.260697 | 0.641568 | 0.428585 | 4.866034 | 3.974801 | 0.241203 |
| relative | 1.080823 | 2.659873 | 1.776868 | 20.174052 | 16.479096 | 1 |
| 2.5% | 2.333418 | 1.668549 | 1.835400 | 1.461138 | 1.444718 | 2.383341 |
| 97.5% | 4.319237 | 4.775363 | 4.481057 | 8.546290 | 7.964598 | 4.292948 |

Table 51: Clayton copula, $\alpha_0 = 5$; $X_t$ time trend; marginal $t(5)$

| $\alpha_0 = 5$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 5.007355 | 4.972956 | 5.092554 | 4.752493 | 4.903096 | 4.991784 |
| bias | 0.007355 | -0.027044 | 0.092554 | -0.247507 | -0.096904 | -0.008216 |
| std | 0.405413 | 2.079547 | 1.580341 | 1.233135 | 1.107888 | 0.293124 |
| MSE | 0.164414 | 4.325245 | 2.506044 | 1.581881 | 1.236805 | 0.085989 |
| relative | 1.912026 | 50.299881 | 29.143719 | 18.396276 | 14.383269 | 1 |
| 2.5% | 4.297535 | 3.466686 | 3.710439 | 3.341005 | 3.563765 | 4.449929 |
| 97.5% | 5.829868 | 7.684039 | 7.485375 | 7.777593 | 7.419336 | 5.590978 |

| $\theta_0 = 5$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 5.356028 | 5.573711 | 5.342138 | 49.565179 | 34.680431 | 5.350891 |
| bias | 0.356028 | 0.573711 | 0.342138 | 44.565179 | 29.680431 | 0.350891 |
| std | 1.417924 | 2.152214 | 1.840395 | 416.915655 | 325.848714 | 1.332860 |
| MSE | 2.137263 | 4.961168 | 3.504113 | 175804 | 107058 | 1.899640 |
| relative | 1.125089 | 2.611636 | 1.844619 | 92546 | 56357 | 1 |
| 2.5% | 3.612740 | 2.610428 | 2.769917 | 2.176628 | 2.124112 | 3.671294 |
| 97.5% | 8.835572 | 10.648695 | 9.782142 | 46.491965 | 31.197025 | 8.708020 |

Table 52: Clayton copula, $\alpha_0 = 5$; $X_t$ time trend; marginal $t(4) + c(8)$

| $\alpha_0 = 5$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 5.00040533 | 4.92492217 | 4.19226542 | 4.99346500 |
| bias | 0.00040533 | -0.07507783 | -0.80773458 | -0.00653500 |
| std | 0.38432473 | 1.19804838 | 1.34371591 | 0.32287738 |
| MSE | 0.14770566 | 1.44095660 | 2.45800760 | 0.10429251 |
| relative | 1.41626340 | 13.81649185 | 23.56840040 | 1 |
| 2.5% | 4.28247098 | 3.11712872 | 2.36204297 | 4.36875859 |
| 97.5% | 5.78940646 | 7.80589064 | 7.57466813 | 5.66008408 |

| $\theta_0 = 4$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 3.98781894 | 3.86226228 | 3.24216553 | 3.98560651 |
| bias | -0.01218106 | -0.13773772 | -0.75783447 | -0.01439349 |
| std | 0.28139238 | 0.69068341 | 0.73807542 | 0.27231290 |
| MSE | 0.07933005 | 0.49601526 | 1.11906842 | 0.07436149 |
| relative | 1.06681626 | 6.67032429 | 15.04903153 | 1 |
| 2.5% | 3.42672389 | 3.02912450 | 2.03076418 | 3.45200574 |
| 97.5% | 4.57925787 | 4.48766034 | 5.08011945 | 4.58028729 |

### 23.3.2 Cointegration

Table 53: Clayton copula, $\alpha_0 = 5$; $X_t$ cointegration; marginal $t(3)$

| $\alpha_0 = 5$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 4.999695 | 4.913607 | 4.996693 | 4.655947 | 4.807326 | 4.993004 |
| bias | -0.000305 | -0.086393 | -0.003307 | -0.344053 | -0.192674 | -0.006996 |
| std | 0.355793 | 4.409542 | 2.123469 | 1.172701 | 1.052907 | 0.303379 |
| MSE | 0.126588 | 19.451520 | 4.509131 | 1.493601 | 1.145736 | 0.092088 |
| relative | 1.374650 | 211.228122 | 48.965596 | 16.219328 | 12.441784 | 1 |
| 2.5% | 4.329355 | 3.425135 | 3.751143 | 3.292677 | 3.635771 | 4.445209 |
| 97.5% | 5.715926 | 6.731062 | 6.515857 | 7.690381 | 7.341182 | 5.614109 |

| $\theta_0 = 3$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 3.102791 | 3.178882 | 3.091936 | 4.281357 | 4.152847 | 3.102354 |
| bias | 0.102791 | 0.178882 | 0.091936 | 1.281357 | 1.152847 | 0.102354 |
| std | 0.498549 | 0.693241 | 0.615315 | 2.026734 | 1.798961 | 0.480340 |
| MSE | 0.259117 | 0.512582 | 0.387064 | 5.749528 | 4.565318 | 0.241203 |
| relative | 1.074271 | 2.125109 | 1.604727 | 23.836925 | 18.927317 | 1 |
| 2.5% | 2.358046 | 1.912758 | 2.051686 | 1.402880 | 1.391340 | 2.383341 |
| 97.5% | 4.354536 | 4.772549 | 4.546638 | 8.499422 | 8.038827 | 4.292948 |

Table 54: Clayton copula, $\alpha_0 = 5$; $X_t$ cointegration; marginal $t(5)$

| $\alpha_0 = 5$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 4.998967 | 4.892269 | 4.998314 | 4.715263 | 4.839748 | 4.991784 |
| bias | -0.001033 | -0.107731 | -0.001686 | -0.284737 | -0.160252 | -0.008216 |
| std | 0.364725 | 2.026218 | 1.519511 | 1.031309 | 0.957447 | 0.293124 |
| MSE | 0.133026 | 4.117164 | 2.308915 | 1.144673 | 0.942386 | 0.085989 |
| relative | 1.547003 | 47.880023 | 26.851232 | 13.311824 | 10.959352 | 1 |
| 2.5% | 4.309386 | 3.552950 | 3.726283 | 3.403930 | 3.620481 | 4.449929 |
| 97.5% | 5.746006 | 6.860581 | 6.745947 | 7.364923 | 7.204350 | 5.590978 |

| $\theta_0 = 5$ | P1 | P2 | M2 | P3 | M3 | I |
|---|---|---|---|---|---|---|
| mean | 5.368794 | 5.564606 | 5.396362 | 49.450230 | 39.429079 | 5.350891 |
| bias | 0.368794 | 0.564606 | 0.396362 | 44.450230 | 34.429079 | 0.350891 |
| std | 1.424991 | 1.902027 | 1.733268 | 425.870993 | 368.746049 | 1.332860 |
| MSE | 2.166609 | 3.936487 | 3.161322 | 183341 | 137159 | 1.899640 |
| relative | 1.140537 | 2.072228 | 1.664169 | 96514 | 72202 | 1 |
| 2.5% | 3.616808 | 2.992781 | 3.079820 | 2.084198 | 2.044659 | 3.671294 |
| 97.5% | 9.016202 | 10.436735 | 9.662374 | 46.549061 | 36.267775 | 8.708020 |

Table 55: Clayton copula, $\alpha_0 = 5$; $X_t$ cointegration; marginal $t(4) + c(8)$

| $\alpha_0 = 5$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 4.99514978 | 4.86945055 | 4.11373254 | 4.99346500 |
| bias | -0.00485022 | -0.13054945 | -0.88626746 | -0.00653500 |
| std | 0.35956211 | 0.95801696 | 1.17511658 | 0.32287738 |
| MSE | 0.12930843 | 0.93483965 | 2.16636899 | 0.10429251 |
| relative | 1.23986312 | 8.96363188 | 20.77204801 | 1 |
| 2.5% | 4.32780709 | 3.25541648 | 2.45197685 | 4.36875859 |
| 97.5% | 5.73282528 | 7.12774365 | 7.07648049 | 5.66008408 |

| $\theta_0 = 4$ | P1 | P2 | P3 | I |
|---|---|---|---|---|
| mean | 3.98391982 | 3.91307617 | 3.24458305 | 3.98560650 |
| bias | -0.01608018 | -0.08692383 | -0.75541695 | -0.01439350 |
| std | 0.28064693 | 0.36563003 | 0.79121477 | 0.27231291 |
| MSE | 0.07902127 | 0.14124107 | 1.19667558 | 0.07436149 |
| relative | 1.06266390 | 1.89938457 | 16.09267865 | 1 |
| 2.5% | 3.42907370 | 3.21806552 | 2.02015083 | 3.45200571 |
| 97.5% | 4.57105847 | 4.61500087 | 5.13306225 | 4.58028739 |

# Part III

# Chapter 3: Testing for Structural Change with Good Size and Power[19]

**Abstract**

This paper studies procedures for testing structural changes with good size and power properties. We focus on dynamic models and the analysis covers a wide range of important inference problems. A leading case is testing for changing trends in dynamic models. In this case, existing tests either suffer from substantial size distortions or exhibit non-monotonic power. Size and power problems also surface in other dynamic models. We propose to address these two issues simultaneously by constructing estimates for nuisance parameters using nonparametrically detrended residuals to achieve good power and an appropriate bootstrap procedure to improve the size. The core of the construction is a modified bootstrap procedure. It is of sieve type and it differs from the conventional bootstrap procedure in two aspects: (1) it uses estimates from the nonparametric regression to generate bootstrap samples, and (2) it uses simulations to correct for the bias associated with the estimates for the largest auto-regressive root. We show that the procedure yields tests with adequate size and good power against a broad class of structural changes, including one time discrete change, smooth change and multiple structural changes. It is hoped that the results obtained in this paper will be of interests not only from the perspective of testing for structural changes, but also from the broader perspective of understanding the size and power properties of bootstrap testing procedures applied to dynamic models.

## 24 Introduction

Testing for structural changes in dynamic models is a common practice in empirical time series analysis. Most of the commonly used tests are asymptotic tests, relying on asymptotic approximations for relevant critical values. As a direct consequence, the resulting tests may suffer from size distortions when the sample size is small, or when series is persistent. The problem is particularly acute when nuisance parameters (long run variance) are estimated under the alternative hypothesis. For example, let's consider using the Sup-Wald (Sup-W) test as in Andrews [1993] to test for structural change in a linear regression with a constant and linear trend and an AR(1) error process driven by independently and identically distributed normal innovations. Then, the rejection frequency at 5% nominal level can be as large as 45% (29%) when the auto-regressive coefficient is 0.9, the sample size is $T = 100$ ($T = 200$) and the trimming proportion is 10%. The rejection frequencies increase to 56% and 44% respectively when the auto regressive coefficient is 0.95. To fix the problem, Diebold and Chen [1996] suggested using a bootstrap procedure. However, they do not provide a theory and, more importantly, the power property of the procedure is not examined. Hansen [2000] proposed a "fixed regressor" bootstrap and established its asymptotic validity. His focus was different in that the goal was to provide asymptotic valid inference when marginal distributions of the regressors change overtime. His results greatly facilitate the analysis in this paper.

Another important issue is the power properties of structural change tests. Perron [1991] and Vogelsang [1999] documented a rather disturbing phenomenon, namely the issue of non-monotonic power. Specifically, Vogelsang [1999] considered the issue of testing for a shift in the mean of a dynamic time series. He showed that if the variance is estimated under the null hypothesis, then the power of many commonly used tests eventually decreases as the magnitude of the structural change increases. Such power problem is also observed in other situations when nuisance parameters are estimated under the null hypothesis. Juhl and Xiao [2009] investigated the problem of non-monotonic power in tests for a changing mean. They provided a theoretical explanation for the non-monotonic power problem and proposed a modification using a non-parametric estimator for the mean function to obtain residuals for variance estimation. The source of non-monotonic power is eliminated and the resulting procedure has good power properties. However, the empirical size of such tests is

[19] Joint work with Zhijie Xiao xiaoz@bc.edu and Zhongjun Qu qu@bu.edu

affected by the bandwidth choice. A procedure with good size and power performance awaits to be developed.

In this paper, we address the size and power issues simultaneously and develop testing procedures with improved size and power properties. This goal is achieved by coupling bootstrap based tests with non-parametric methods.

We first investigate the performance of a conventional bootstrap procedure that involves estimating all parameters under the null hypothesis and subsequently using them to generate bootstrap samples. We show that this procedure significantly improves the size; however it also exhibits non-monotonic power. What is interesting is that it delivers better power than the corresponding asymptotic tests, in the sense that the power decreases at a latter stage, due to the fact that when a break presents, the estimate of the auto-correlation coefficient is biased upward and accordingly bootstrap critical values are smaller than the asymptotic ones. Unfortunately, the power still diminishes and this occurs for parameter values than are of particular importance in practice.

We then propose a new procedure, in which we use nonparametrically detrended residuals to construct tests to achieve good power and a modified bootstrap procedure to improve the size. First, we follow Juhl and Xiao [2009] and use nonparametrically estimated residuals to construct the long run variance estimate. This ensures that the long run variance estimate will be consistent even under the alternative hypothesis, including the cases of multiple structural changes and smooth changes. Second, we introduce a modified bootstrap procedure to account for uncertainty associated with parameter estimation, with special attention paid to the effect of the non-parametric procedure with a particular bandwidth. The modified bootstrap procedure is of sieve type and it differs from the conventional ones in two aspects: (1) it uses nonparametrically estimated residuals to generate the bootstrap sample, and (2) it uses and additional layer of bootstrap to reduce the bias associated with estimating the dynamics of the model. We prove that the procedure is asymptotically valid. We also use simulations to show that it yields tests with adequate size and significantly, sometimes drastically, improved power over the asymptotic tests and the conventional bootstrap tests.

The importance and application of bias correction in dynamic models has been studied in the literature. Kilian [1998] proposes a bootstrap after bootstrap procedure to construct confidence intervals for bias reduced estimate of the impulse response. More remotely, Andrews and Chen [1994] and Fair [1996] also studied the median unbiased procedures which aimed to eliminate the median biased procedures which aimed to eliminate the median bias associated with the largest auto-regressive root in finite order autoregressions.

Another advantage of our procedure is that it allows a rather broad class of regressors, including stationary regressors as well as deterministic trends. This property is also shared by Hansen [2000]. And it is different from most of the existing literature, in which models with trends and with stationary regressors are considered separately.

The paper is structured as follows. Section 2 presents the model of interest and discusses the null and alternative hypothesis. Leading examples are also given. It also reports the result of a small simulation to illustrate the size and power issue associated with a model with a linear trend. Section 3 examines the power property of the conventional bootstrap tests. Section 4 proposes modified testing procedures and establishes its asymptotic validity. It also conducts simulations to evaluate its finite sample performance. Section 5 concludes.

# 25 The model and assumptions

Consider the following time series model:

$$y_t = x_t'\gamma + z_t'\beta_t + u_t, \quad t = 1, 2, \cdots, T \tag{13}$$

where $u_t$ is an error term that may be serially correlated, and the regressor $x_t$ and $z_t$ can be deterministic or stochastic. More specific assumptions about $u_t$, $x_t$ and $z_t$ will be given later. The issue of interest is to test whether $\beta_t$ is constant, while $\gamma$ is restricted to be time invariant, i.e.:

$$H_0 : \beta_t \equiv \beta \text{ for all } t \geq 1$$

Under the alternative hypothesis, $\beta_t$ has one or more structural changes, i.e.:

$$H_1 : \begin{cases} \beta_t = \beta_1 & \text{if } 1 \le t \le k_1 \\ \beta_t = \beta_2 & \text{if } k_1 < t \le k_2 \\ \cdots & \cdots \\ \beta_t = \beta_m & \text{if } k_{m-1} < t \le T \end{cases}$$

with $\beta_i \neq \beta_j$ for some $1 \le i, j \le m$. The number and locations of the breaks are unknown. The goal is to construct a testing procedure that enjoys good size and power properties.

The above setting is quite general and includes several important models that are widely studied in econometrics. We list a few leading examples below.

**Example 6.** (Testing for a changing trend in dynamic models with serially correlated errors). In this case, $x_t = 0$ and $z_t$ is a deterministic trend, say a polynomial trend given by $z_t = (1, t/T, \cdots, (t/T)^p)'$. The model then reduces to:

$$y_t = z_t'\beta_t + u_t, \quad t = 1, 2, \cdots, T \tag{14}$$

where the errors $u_t$ are often serially correlated. A special case of (14) that is of particular interest in practice is $z_t = 1$. This model has been widely studied in the literature; see Perron [1991], Vogelsang [1999], Deng and Perron [2008], and Juhl and Xiao [2009]. For the purpose of asymptotic analysis, we assume that there exists a limiting trend function $g(r)$ such that $z_{[Tr]} \to g(r)$, as $T \to \infty$, uniformly in $r \in [0, 1]$. If $z_t = (1, t/T, \cdots, (t/T)^p)'$, then $g(r) = (1, r, \cdots, r^p)'$ with $r \in [0, 1]$.

**Example 7.** (Models with strictly exogenous regressors and serially correlated errors). Hansen [2000] considered the following model ($t = 1, 2, \cdots, T$):

$$\begin{aligned} y_t =& z_t'\beta_t + u_t, \\ z_t =& \sqrt{t/T} v_t \text{ with } v_t \sim i.i.d.\ N(0, 1) \end{aligned}$$

The example, although simple, illustrates an important point. namely, when marginal distributions of the regressors vary over time, the asymptotic distributions of the commonly used tests (say the Sup-W test as in Andrews [1993]) will be in general depend on the second moments of the regressors. And the critical values of the tests need to be tabulated on a case by case basis. For this type of models, bootstrap becomes a necessity.

Our model (13) can include ingredients from both examples 1 and 2.

A large family of tests has been proposed; see Perron et al. [2006] for a comprehensive review. In this paper, we focus on the following tests due to their wide application in practice.

- (i) The Sup-W statistic of Andrews [1993];
- (ii) The Exp-W and Ave-W statistics of Andrews and Ploberger [1994];
- (iii) The multiple-break tests of Bai and Perron [1998];
- (iv) The CUSUM (Kolmogorov-Smirnoff) test with OLS residuals by Ploberger and Krämer [1992];
- (v) The QS (Cramer von-Mises) tests by Perron [1991].

Note that these tests can be broadly divided into two categories, namely the Wald-type tests (i to iii), which require estimation regression coefficients under the alternative hypothesis, and the residual-based test (iv and v), which involve estimation using the full sample. The procedures proposed in this paper apply to all of them. To simplify the notation, we use $S$ to denote an arbitrary statistic among the listed statistics and use $\mathcal{L}$ to denote its limiting distribution under the null hypothesis.

Before proceeding any further, we present some simple simulation results to illustrate the size and power issues associated with these tests. We use the Sup-W and the CUSUM test as examples. The findings carry to other tests as well. For simplicity and without loss of generality, we focus on the case with one break.

## 25.1 Preliminary simulations

We focus on the following Data Generating Process (DGP) with a single structural change in linear trend:

$$y_t = \begin{cases} \alpha + \gamma t + u_t & \text{if } 1 \leq t \leq k_1 \\ \alpha + \gamma t + (t - k_1)\beta + u_t & \text{if } k_1 < t \leq T \end{cases} \tag{15}$$

where $u_t \sim ARMA(1,1)$ (including white noise, $AR(1)$ and $MA(1)$)

$$u_t = \rho u_{t-1} + e_t + \theta e_{t-1},\ e_t \sim i.i.d.N(0,1)$$

We set $\alpha = \gamma = 0$, $k_1 = T/2$ and consider the following specifications $\rho = 0, 0.5, 0.7, 0.9$ and $\theta = 0, 0.2$. The sample size $T = 200$ and the simulation repetition is $5,000$ times. For each case, we vary the values of $\beta$ to examine the size and power of the tests. For the both Sup-W test and CUSUM test, a single break is allowed. We report rejection frequencies at a 5% nominal level.

Suppose our OLS regression is:

$$y_t = \tilde{\alpha} + \tilde{\gamma} t + \tilde{u}_t$$

then the CUSUM statistic is:

$$\max_{[T\varepsilon] \leq t \leq [T - T\varepsilon]} \left| \frac{1}{\hat{\omega}\sqrt{T}} \sum_{s=1}^{t} \tilde{u}_s \right|$$

The trimming proportion is set to 10% hence $\varepsilon = 5\%$. For $T = 200$, it means we take maximization over $[10, 190]$.

For the Sup-W test, suppose the single break is at $t = t_0$, we consider the OLS regression under $H_{t_0}$:

$$y_t = x_t'\gamma + z_t'\beta + z_t'\theta 1(t \leq t_0) + u_t, \quad (t = 1, 2, \cdots, T)$$

Use F test whether $\theta = 0$. The statistics at time $t_0$ is:

$$F_{t_0} = \frac{A_{t_0}^{\top} B_{t_0} A_{t_0}}{\hat{\omega}^2}$$

Here $A_{t_0} = \sum_{t=1}^{t_0} z_t u_t - \sum_{t=1}^{t_0} z_t z_t' \cdot \left( \sum_{t=1}^{T} z_t z_t' \right)^{-1} \cdot \sum_{t=1}^{T} z_t u_t$

$B_{t_0} = \sum_{t=1}^{t_0} z_t z_t' - \sum_{t=1}^{t_0} z_t z_t' \cdot \left( \sum_{t=1}^{T} z_t z_t' \right)^{-1} \cdot \sum_{t=1}^{t_0} z_t z_t'$

Then the Sup-W statistic is:

$$\max_{[T\varepsilon] \leq t \leq [T - T\varepsilon]} \left| \frac{A_{[Tr]}^{\top} B_{[Tr]} A_{[Tr]}}{\hat{\omega}^2} \right|$$

Both Wald type test and residual based test require an estimate for the long run variance of $u_t$, $\hat{\omega}^2$. In practice, we have the options of estimating it under the null or under the alternative hypothesis[20]. It is well known that the choice has important effect on the finite sample size and power of the tests. In our setting, $\hat{\omega}$ is estimated by imposing the null hypothesis. We apply the autoregressive spectral density estimate, with the lag order determined by BIC[21].

*Remark* 23. $\hat{\omega}$ can also be estimated under the alternative hypothesis. In this case, we first find the break date that minimizes the sum of squared residuals from estimation (15). Then, we estimate the residuals conditional on the estimated break date. Finally, we estimate the long run variance from these residuals.

[20] Although for the Sup-W (resp. CUSUM) test, it is more natural to estimate the variance under the alternative (resp. null) hypothesis, the limiting null distribution is invariant to such a choice under the assumption $|\rho| < 1$.

[21] The maximum lag length is set as $K_T = int(12(T/100))^{1/4}$

The limiting distribution of CUSUM statistic is:

$$\sup_{\varepsilon \le r \le 1-\varepsilon} \left| W(r) - \begin{pmatrix} r & r^2 \end{pmatrix}' \times \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix} \times \begin{pmatrix} W(1) & \int_0^1 s \mathrm{d}W(s) \end{pmatrix} \right|$$

The limiting distribution of Sup-W statistic is:

$$\sup_{\varepsilon \le r \le 1-\varepsilon} \left| A^\top(r) B(r) A(r) \right|$$

here $B(r) = \begin{pmatrix} 1 & r \\ r & r^2/2 \end{pmatrix} - \begin{pmatrix} 1 & r \\ r & r^2/2 \end{pmatrix} \times \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix} \times \begin{pmatrix} 1 & r \\ r & r^2/2 \end{pmatrix}$
and $A(r) = \begin{pmatrix} W(r) \\ \int_0^r s\mathrm{d}W(s) \end{pmatrix} - \begin{pmatrix} 1 & r \\ r & r^2/2 \end{pmatrix} \times \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix} \times \begin{pmatrix} W(1) \\ \int_0^1 s\mathrm{d}W(s) \end{pmatrix}$.

We approximate the standard Brownian motion with $20,000$ i.i.d. $N(0,1)$ random variable and repeat the simulation for $20,000$ times to compute critical values. The 95% critical value CUSUM statistic is 0.9019972.

First we set $\beta$ to zero and consider the size of the tests. When the long run variance is estimated under the null hypothesis, the test will be conservative, especially in the presence of strong positive correlation. For CUSUM test, when $\rho = 0.9$, the rejection rate is 0.04% for $AR(1)$ and 0.08% for $ARMA(1)$.

Table 56: CUSUM size (nominal 5%, $T = 200$)

| rejection rate | $\rho = 0$ | $\rho = 0.5$ | $\rho = 0.7$ | $\rho = 0.9$ |
|---|---|---|---|---|
| $\theta = 0$ | 2.96% | 1.34% | 0.44% | 0.04% |
| $\theta = 0.2$ | 2.04% | 1.04% | 0.66% | 0.08% |

*Remark* 24. For Sup-W test when the long run variance is estimated under the alternative hypothesis, size distortions occur and serious over-rejection is observed and it can be 45% when $\rho = 0.9$ and $T = 100$. The size distortion persists after $T$ is increased to 200.

Next, consider the power of the tests. Figure reports power of the CUSUM test when the long run variance is estimated imposing the null hypothesis. Non-monotonic power presents and clearly the deterioration affects parameter value of practical interest: for $\rho = 0.9$, the power is virtually zero throughout. This is again particularly disturbing since we often expect macro time series to be strongly positively auto-correlated. The above phenomenon has been widely documented and explained: see Perron [1991], Vogelsang [1999], and Crainiceanu and Vogelsang [2007].

Figure 12: CUSUM, no MA term

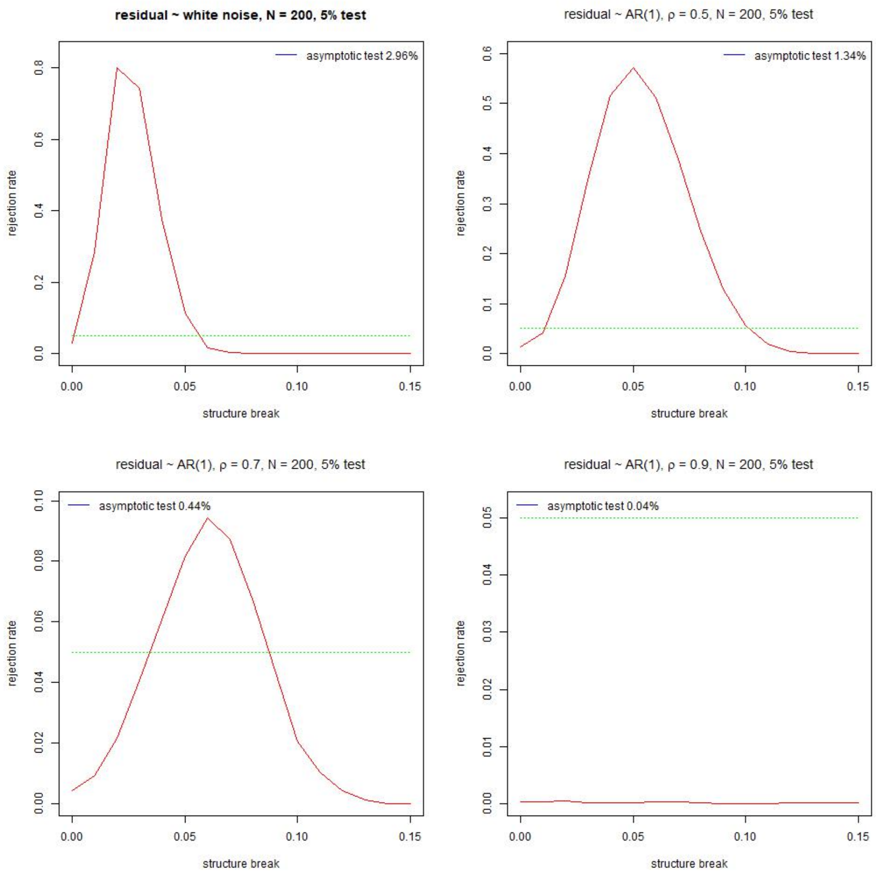
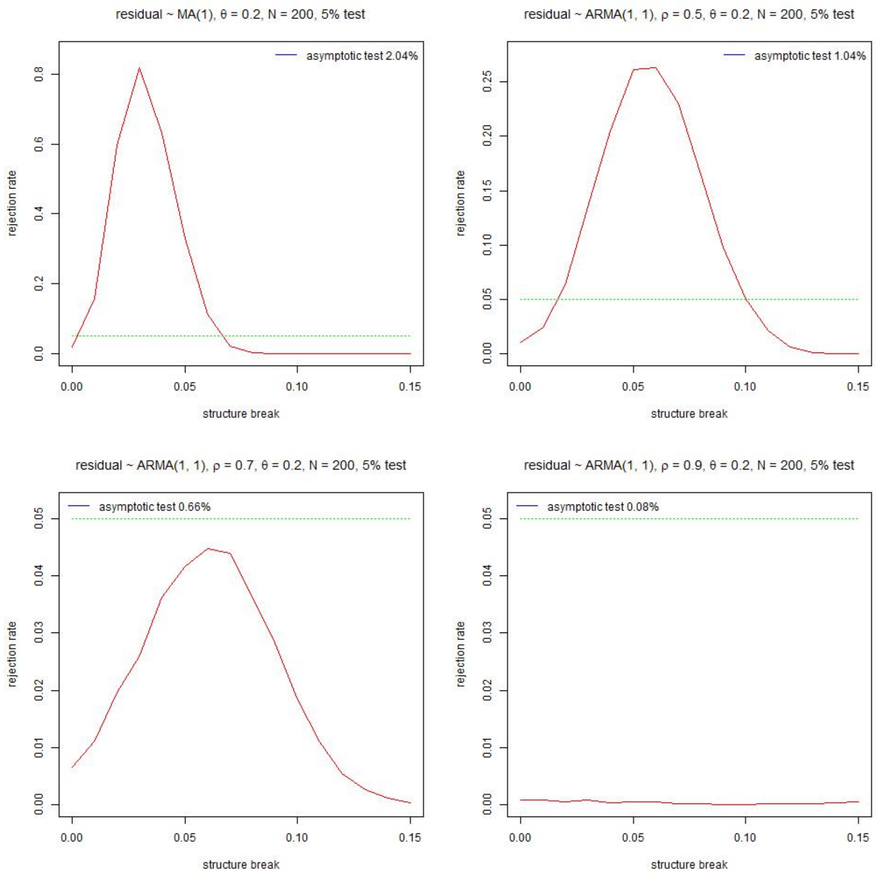


Figure 13: CUSUM, with MA term




To summarize, the result shows that significant size distortions exist when the errors are persistent. This is true irrespective whether the variance is estimated under the null or alternative hypothesis. While the size distortion is smaller if the variance is estimated under the null, the power is also miserable.

The size distortion is because standard asymptotic theory does not provide an adequate approximation in finite samples when the series are persistent. A natural solution is to bootstrap. The non-monotonic power is due to "incorrect"estimation of nuisance parameters (in this case the long run variance) under the alternative hypothesis. The estimate diverges as the size of the break increases. A natural solution is to use an alternative estimate that is bounded even when the break size is large. The estimate also needs to allow for the possibility that an unknown number of breaks may occur under the alternative hypothesis.

In this paper, we will attempt to address two issues simultaneously and in a general framework. We now state the assumptions under which we will be working.

## 25.2 Assumptions

All of the aforementioned tests involve estimating the following regression using a sub-sample or the full sample ($0 < r \leq 1$):

$$y_t = x_t'\gamma + z_t'\beta + u_t,\ t = 1, 2, \cdots, [Tr]$$

Let $w_t = (x_t', z_t')'$ and $\theta = (\gamma', \beta')'$. We may rewrite the above regression as:

$$y_t = w_t'\theta + u_t,\ t = 1, 2, \cdots, [Tr]$$

The OLS estimator of $\theta$ is then given by:

$$\tilde{\theta}(r) = \left( \sum_{t=1}^{[Tr]} w_t w_t' \right)^{-1} \left( \sum_{t=1}^{[Tr]} w_t y_t \right)$$

Under the null hypothesis,

$$\sqrt{T}(\tilde{\theta}(r) - \theta) = \left( \frac{1}{T} \sum_{t=1}^{[Tr]} w_t w_t' \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} w_t u_t \right)$$

We impose the following assumption about the property of the regressors.

**Assumption 25.** *The partial sum processes of the regressors and their second moment satisfy the following conditions:*

$$\frac{1}{T} \sum_{t=1}^{[Tr]} w_t \Rightarrow N(r)$$

$$\frac{1}{T} \sum_{t=1}^{[Tr]} w_t w_t' \Rightarrow M(r)$$

*where $N(r)$ is a vector of limiting function of the regressors, and $M(r)$ is a positive definite matrix function.*

The regressors can include stationary as well as trending regressors. Let $w_{t,j}$ denote the $j^{th}$ component of $w_t$. If $w_{t,j}$ is a stationary stochastic process with mean zero, then the corresponding component in $N(r)$ is zero. If $w_{t,j}$ is a deterministic trend, then the corresponding component in $N(r)$ is the limiting trend function. Assumption 25 is more general than what is typically adopted in the structural change literature, under which models with trending and stationary regressors are usually treated separately because they lead to different limiting distributions (c.f. Bai and Perron [1998]). A notable exception is Hansen [2000].

The sequence of errors $\{u_t\}$ satisfy following conditions.

**Assumption 26.** *$u_t = C(L)\varepsilon_t$, where $C(L) = \sum_{j=0}^{+\infty} c_j L^j$, $c_0 = 1$, and $L$ is the lag operator, with $C(z) \neq 0$ for all $z$ inside the unit circle ($|z| \leq 1$) and $\sum_{j=0}^{+\infty} j^s |c_j| < \infty$ for some $s \geq 1$, and $\varepsilon_t$ is i.i.d. with $\mathbb{E}\varepsilon_t^2 = \sigma_\varepsilon^2$, $\mathbb{E}\varepsilon_t^4 < +\infty$. $\mathbb{E}u_t w_s = 0$ for all $t$ and $s$.*

We denote the long run variance of $u_t$ by $\omega^2$ and its short run variance by $\sigma^2$, respectively.

In **Assumption** 26, the errors are generated by a linear process, which covers a wide range of time series and includes stationary $ARMA(p,q)$ process as a special case. The summability and moment conditions ensure an invariance principle for the sieve bootstrap that we use in our proposed procedure. Under these assumptions,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t \Rightarrow \sigma_\varepsilon W(r)$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t \Rightarrow \omega W(r)$$

where $W$ is a standard Wiener process.

The following assumption is concerned with the relation between the regressors and the errors.

**Assumption 27.** *$\mathbb{E}w_t u_t = 0$ for all $t$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} w_t u_t \Rightarrow G(r)$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} w_t w_t' u_t^2 \Rightarrow \sigma^2 M(r)$. Here $G(r)$ is a mean-zero Gaussian process.*

*Remark* 25. Assumptions 2 and 3 are conventional assumptions assumed in stationary time series analysis. See Hansen [2000] and Park [2002] for similar assumptions.

Under the stated assumptions, we have

$$\sqrt{T}(\tilde{\theta}(r)-\theta) = \left(\frac{1}{T}\sum_{t=1}^{[Tr]} w_t w_t'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} w_t u_t\right) \Rightarrow M(r)^{-1}G(r)$$

We now consider implications of Assumptions 1-3 for the two examples considered. The analysis helps to pinpoint aspects of the model that determine the null limiting distributions of the tests.

**Example.** 1(continued). We have $z_{[Tr]} \to g(r)$ uniformly in $r \in [0,1]$. Thus,

$$\frac{1}{T}\sum_{t=1}^{[Tr]} w_t \Rightarrow N(r) = \int_0^r g(s)\mathrm{d}s$$

$$\frac{1}{T}\sum_{t=1}^{[Tr]} w_t w_t' \Rightarrow M(r) = \int_0^r g(s)g(s)'\mathrm{d}s$$

Specifically,

$$N(r) = \begin{bmatrix} r \\ \int_0^r s\mathrm{d}s \\ \cdots \\ \int_0^r s^p\mathrm{d}s \end{bmatrix}$$

$$M(r) = \begin{bmatrix} r & \int_0^r s\mathrm{d}s & \cdots & \int_0^r s^p\mathrm{d}s \\ \int_0^r s\mathrm{d}s & \int_0^r s^2\mathrm{d}s & \cdots & \int_0^r s^{p+1}\mathrm{d}s \\ \cdots & & \ddots & \\ \int_0^r s^p\mathrm{d}s & \int_0^r s^{p+1}\mathrm{d}s & \cdots & \int_0^r s^{2p}\mathrm{d}s \end{bmatrix}$$

If $z_t = 1$, then $g(r) = 1$, $N(r) = r$, and $M(r) = r$. Under Assumption 2 and 3,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} w_t u_t \Rightarrow G(r) = \omega\int_0^r g(s)\mathrm{d}W(s)$$

and

$$\frac{1}{T}\sum_{t=1}^{[Tr]} w_t w_t' u_t^2 \Rightarrow \sigma^2 M(r) = \sigma^2\int_0^r g(s)g'(s)\mathrm{d}s$$

Thus

$$\sqrt{T}(\tilde{\theta}(r)-\theta) \Rightarrow M(r)^{-1}G(r) = \left[\int_0^r g(s)g'(s)\mathrm{d}s\right]^{-1}\cdot\omega\int_0^r g(s)\mathrm{d}W(s)$$

Let $\tilde{u}_t$ denote the regression residuals obtained imposing the null hypothesis, we have

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\tilde{u}_t &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}(y_t - x_t'\tilde{\gamma} - z_t'\tilde{\beta}) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}u_t - \frac{1}{T}\sum_{t=1}^{[Tr]}w_t' \cdot \sqrt{T}(\tilde{\theta}-\theta) \\
&\Rightarrow \omega W(r) - \int_0^r g(s)'\mathrm{d}s \cdot \left[\int_0^r g(s)g'(s)\mathrm{d}s\right]^{-1} \cdot \omega \int_0^r g(s)\mathrm{d}W(s) \\
&= \omega \underline{\mathrm{W}}(r) \qquad (16)
\end{aligned}
$$

The first equation is the main ingredient of Wald-based tests and the second equation plays a similar role in residuals based tests. The above result demonstrates what are expected to enter the limiting distributions of the tests.

**Example.** 2(continued) $w_t = z_t = \sqrt{t/T}v_t$, $v_t \sim N(0,1)$. Thus, as shown as Hansen [2000],

$$
\frac{1}{T}\sum_{t=1}^{[Tr]} w_t \Rightarrow N(r) = 0
$$

$$
\frac{1}{T}\sum_{t=1}^{[Tr]} w_t w_t' \Rightarrow M(r) = r^2/2
$$

Also,

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} w_t u_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} \sqrt{\frac{t}{T}} u_t v_t \Rightarrow G(r) = \sigma \int_0^r \sqrt{s}\mathrm{d}W_1(s)
$$

because $v_t$ is i.i.d. $N(0,1)$ and independent of $u_t$, long run variance of $u_t v_t$ is just variance of $u_t v_t$ which is $\sigma^2 = \mathbb{E}u_t^2$. Specifically notice that $W_1(r)$ is the Wiener process independent of $W(r)$ as $v_t u_t$ is uncorrelated with $u_t$.

Hence,

$$
\sqrt{T}(\tilde{\theta}(r) - \theta) \Rightarrow M(r)^{-1}G(r) = \frac{2\sigma}{r^2}\int_0^r \sqrt{s}\mathrm{d}W_1(s)
$$

Further,

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\tilde{u}_t &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}u_t - \frac{1}{T}\sum_{t=1}^{[Tr]}w_t' \cdot \sqrt{T}(\tilde{\theta}-\theta) \\
&\Rightarrow \omega W(r) - G(1)M(1)^{-1}N(r) \\
&= \omega W(r)
\end{aligned}
$$

We can also include a time trend $x_t = t/T$ here. Then $w_t = (x_t, z_t) = (t/T, \sqrt{t/T}v_t)$.

$$
\frac{1}{T}\sum_{t=1}^{[Tr]} w_t \Rightarrow N(r) = \begin{pmatrix} r/2 \\ 0 \end{pmatrix}
$$

$$
\frac{1}{T}\sum_{t=1}^{[Tr]} w_t w_t' \Rightarrow M(r) = \begin{pmatrix} r^3/3 & 0 \\ 0 & r^2/2 \end{pmatrix}
$$

Also,

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} w_t u_t \Rightarrow \begin{pmatrix} \omega W(r) \\ \sigma \int_0^r \sqrt{s}\mathrm{d}W_1(s) \end{pmatrix}
$$

notice that $W_1(r)$ is the Wiener process independent of $W(r)$ as $v_t u_t$ is uncorrelated with $u_t$.

Hence,

$$\sqrt{T}(\tilde{\theta}(r) - \theta) \Rightarrow M(r)^{-1} G(r) = \begin{pmatrix} \frac{3\omega}{r^3} W(r) \\ \frac{2\sigma}{r^2} \int_0^r \sqrt{s} \mathrm{d}W_1(s) \end{pmatrix}$$

Further,

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \tilde{u}_t &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t - \frac{1}{T} \sum_{t=1}^{[Tr]} w_t' \cdot \sqrt{T}(\tilde{\theta} - \theta) \\ &\Rightarrow \omega W(r) - G(1) M(1)^{-1} N(r) \\ &= \omega \left[ W(r) - 3rW(1)/2 \right] \end{aligned}$$

The proposed procedure involves nonparametric estimation of nuisance parameters. We impose the following assumptions on the kernel $K(\cdot)$ and the bandwidth $h$.

**Assumption 28.** *$K(\cdot)$ is a bounded, non-negative, symmetric Lipschitz continuous density such that $\int_{-\infty}^{+\infty} |uK(u)| \mathrm{d}u < +\infty$. The bandwidth $h$ satisfies $h \to 0$ and $Th^4 \to \infty$ and $Th^6 \to 0$ as $T \to \infty$.*

This assumption is the similar as Assumptions 5 and 6 in Juhl and Xiao [2009]. Our bandwidth requirement is stronger than Juhl and Xiao [2009] as we include the case where regressors are stationary while Juhl and Xiao [2009] only consider fixed time trend regressors. The conditions, along with Assumptions 1 to 3, ensure the nuisance parameter estimates have good properties under both the null and the alternative hypothesis.

When $x_t$ or $z_t$ are stationary regressors, we need some technical constraints on its dependence:

**Assumption 29.** *The stationary regressors part of $w_t$ need to satisfy the following conditions:*

*(i). $w_t = (x_t', z_t')$ has finite eight order moment: $\sup_t \mathbb{E} \|w_t\|^8 < +\infty$.*

*(ii). $u_t$ is independent of data up to fourth order:*

$$\mathbb{E}[\|U\|^i | X, Z] = \mathbb{E} \|U\|^i \quad i = 1, 2, 3, 4$$

*where $U := (u_1, u_2, \cdots, u_T)'$, $X := (x_1', x_2', \cdots, x_T')$ and $Z := (z_1', z_2', \cdots, z_T')$.*

*(iii).$w_t = (x_t', z_t')$ is strong mixing with coefficient decaying in polynomial rate:*

$$\alpha_m \leq cm^{-3}$$

*for some constants $c > 0$.*

*Remark* 26. $\alpha_m$ is defined as:

$$\alpha_m = \sup\{ | \mathbb{P}(A \cap B) - \mathbb{P}(A) \mathbb{P}(B) | : A \in \mathscr{F}_1^i, B \in \mathscr{F}_{i+m}^{\infty}, i \in \mathbb{Z}_+ ) \}$$

$\mathscr{F}_1^i$ denotes the $\sigma-$algebra generated by $w_1, w_2, \cdots, w_i$.

$\mathscr{F}_{i+m}^{\infty}$ denotes the $\sigma-$algebra generated by $w_{i+m}, w_{i+m+1}, \cdots$.

*Remark* 27. For stationary regressors, we need the strong mixing rate is at least with cubic polynomial order. This is required to ensure uniform convergence. Cubic order is very mild, see Hansen [2008]. It will be satisfied if the strong mixing rate is exponentially decayed. Specifically, i.i.d. sequence automatically fulfills the condition.

*Remark* 28. For Hansen [2000] special case (see Example 2), if the stochastic regressor is as $f(t/T) v_t$ where $f(\cdot)$ denotes a deterministic time trend ($f(x) = \sqrt{x}$ in Example 2), we need $v_t$ to meet all above requirements listed in **Assumption** 29.

Having laid out the model and the assumptions needed, we now turn to bootstrap procedures for improved inference. We first investigate the appropriateness of procedures along the line of conventional bootstrap method.

## 26 Conventional bootstrap tests

We consider the following procedure, which we label as the conventional sieve bootstrap test. We assume the errors of the model are serially correlated. The case with lagged dependent variables and martingale difference errors can be handled along the same lines, the findings are similar.

1. Estimate (13) and construct the test, say the Sup-W test or the CUSUM test, where the long run variance is estimated under the null hypothesis using an autoregressive approximation. The lag order is determined using BIC and denoted by $k$.

2. (Generate the bootstrap sample). First, obtain residuals imposing the null hypothesis:

$$\tilde{u}_t = y_t - x_t'\tilde{\gamma} - z_t'\tilde{\beta}$$

   Then, estimate an $AR(k)$ model for $\tilde{u}_t$:

$$\tilde{u}_t = \sum_{j=1}^{k} d_j \tilde{u}_{t-j} + e_t$$

   Denote the estimated parameters and residuals as $\tilde{d}_j$ and $\tilde{e}_t$. Next, sample with replacement from the re-centered empirical distribution of $\{\tilde{e}_t\}_{t=1}^T$ to obtain $\{\tilde{e}_t^*\}_{t=1}^T$.
   Finally, generate $\tilde{u}_t^*$ recursively as

$$\tilde{u}_t^* = \sum_{j=1}^{k} \tilde{d}_j \tilde{u}_{t-j}^* + \tilde{e}_t^*$$

   and generate $y_t^*$ as $y_t^* = x_t'\tilde{\gamma} + x_t'\tilde{\beta} + \tilde{u}_t^*$. Note that the regressors $x_t$ and $z_t$ are fixed across bootstrap samples, as in Hansen [2000].

3. Estimate (13) using $\{y_t^*\}_{t=1}^T$ and construct the test as in Step 1.

4. Repeat step 2 and 3 for $B$ times and obtain the bootstrap critical values.

5. Report a rejection if the value of the statistic in step 1 exceeds the critical value.

The above procedure can be shown to have correct size asymptotically. Indeed simulations show that it improves upon asymptotic tests. To this end, we again consider the model in Section (25.1). When error correlation is mild, size of conventional bootstrap test is very close to the nominal size of 5%. However, when error correlation is very strong ($\rho = 0.9$), it is still conservative, 2.8% for $AR(1)$ and 2.6% for $ARMA(1,1)$.

Figure 14: CUSUM, no MA term

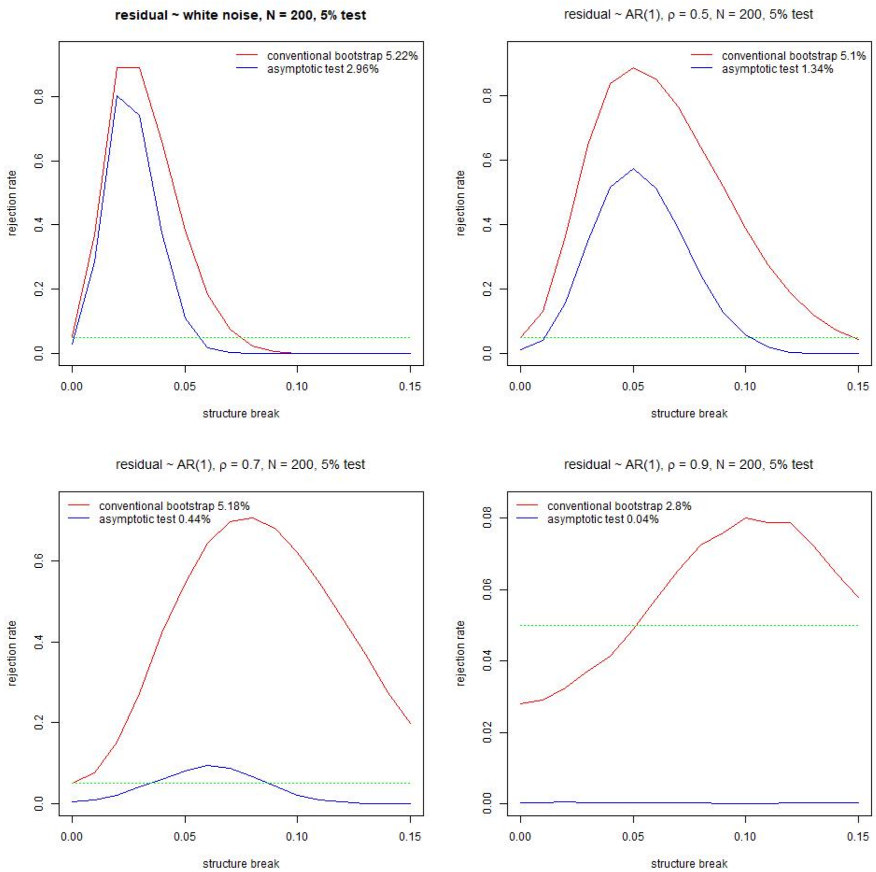
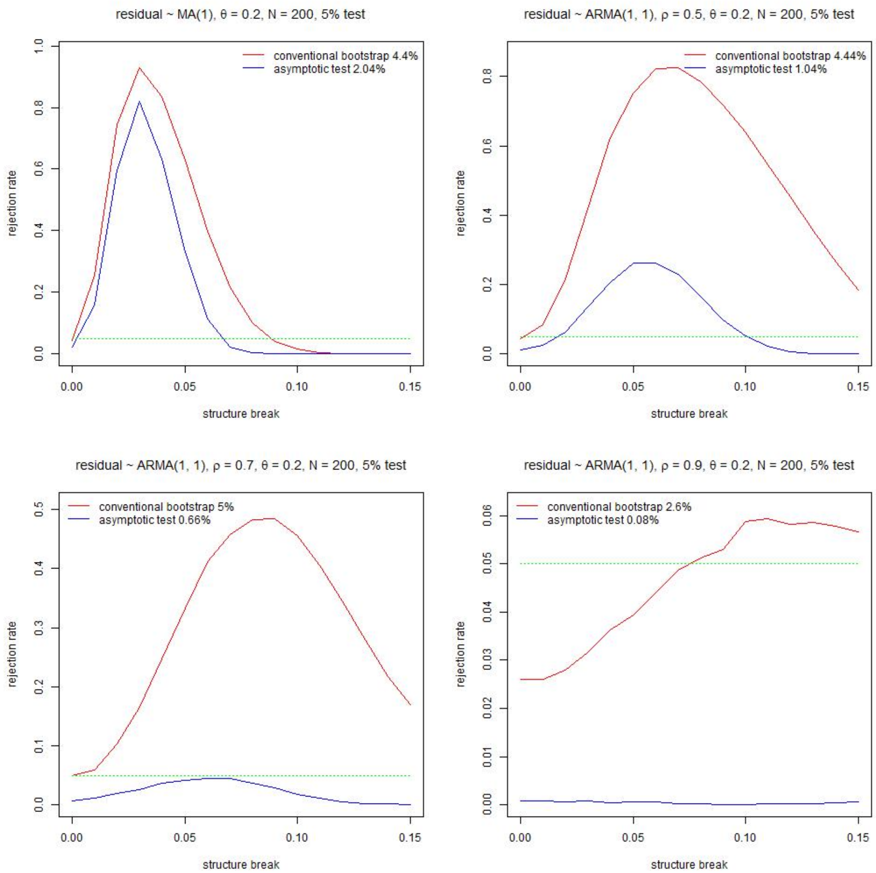


Figure 15: CUSUM, with MA term




The power property awaits to be explored. Some interesting pattern emerges from above figures. First, the test shows non-monotonic power. The second and slightly surprising result is that the power of the bootstrap test improves upon the asymptotic test, with power decreasing at a latter stage. And the improvement is relatively more significant when the autocorrelation is mild. This is because when a break is present, the sum of the estimated autoregressive coefficients is upward biased. Hence the estimated long run variance is exploding. As a result, the bootstrap critical values are smaller than the asymptotic ones. Finally, and more importantly, the power improvement is limited. When $\rho = 0.9$, the power is below 10% globally, for both AR and ARMA residuals. This is disturbing since this corresponds to a region of parameter values of particular importance to economics.

If it is known *a prior* that at most one break occurs under the alternative hypothesis. Then, we can modify the first and third step by estimating the long run variance under the alternative hypothesis. Specifically, we first find the break date that minimizes the sum of squared residuals. Then, we estimate the residuals conditional on the estimated break date and use them to construct an estimate for the long run variance. We conducted some simulations and the results shows that this indeed delivers significant power improvements over the procedure discussed above. However, the improvement vanishes if multiple breaks occur of if the change is smooth, and similar non-monotonic pattern as in above figures emerges. Because in practice we rarely know the number of breaks before looking at the data, it is desirable to have a procedure that can adapt to multiple changes. We now consider such procedures.

Specifically we propose bootstrap based testing procedures using a nuisance parameter estimator with nonparametrically detrended data. The proposed procedures have subtle differences depending on whether lagged dependent variables are present or absent in the regression. We treat these two cases sequentially.

## 27 Bootstrap procedures for models with dependent errors

Lagged dependent variables are not present and Assumption 2(i) applies. We first propose a procedure without bootstrap bias correction. We name this procedure as bootstrap with nonparametrically estimated residuals. We label this procedure as NB(bootstrap with nonparametrically estimated residuals).

1. (Construct the test statistic). *We estimate different regressions.* The first regression is a regression of (13) under the null, denote the corresponding parameter estimates as $\tilde{\gamma}$ and $\tilde{\beta}$, and residuals as $\tilde{u}_t$. Next, we construct a nuisance parameter estimator that have good properties even under the alternative with structural breaks (non-constant $\beta_t$). In particular, we consider the following semiparametric regression with partially varying coefficients:

$$y_t = x_t'\hat{\gamma} + z_t'\hat{\beta}\left(\frac{t}{T}\right) + \hat{u}_t,\ t = 1, 2, \cdots, T \tag{17}$$

   let $\hat{u}_t$ be the corresponding residuals. Based on $\hat{u}_t$, estimate the following autoregressive model for $\hat{u}_t$ with lag order determined by AIC or BIC:

$$\hat{u}_t = \rho_1 \hat{u}_{t-1} + \rho_2 \Delta \hat{u}_{t-1} + \cdots + \rho_k \Delta \hat{u}_{t-k+1} + e_t$$

   Denote the parameter estimates as $\hat{\rho} = (\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_k)$ and residuals as $\hat{e}_t$. We can then use $\hat{e}_t$ and $\hat{\rho}$ to construct an estimate for the long-run variance of $u_t$. Denote the estimate by $\hat{\sigma}^2$. We construct the test statistic in the conventional way as the standard procedure, as in Bai and Perron [1998] *except that we use $\hat{\omega}^2$ as the variance estimate.* For example, for the CUSUM or QS tests, the partial sum process will be constructed based on the residuals from the (null) restricted regression in the standard way but $\hat{\omega}^2$ will be used for standardization. Denote the testing statistic as $S$.

2. (Generate the bootstrap sample). Next, sample with replacement from the re-centered empirical distribution of $\{\hat{e}_t\}_{t=1}^T$ to obtain $\{\hat{e}_t^*\}_{t=1}^T$. Simulate samples under the null (using $\tilde{\gamma}$ and $\tilde{\beta}$). Specifically, we generate

$$\begin{aligned} u_t^* &= \hat{\rho}_1 u_{t-1}^* + \hat{\rho}_2 u_{t-2}^* + \cdots + \hat{\rho}_k u_{t-k}^* + \hat{e}_t^* \\ y_t^* &= x_t'\tilde{\gamma} + z_t'\tilde{\beta} + u_t^* \end{aligned}$$

   Based on the bootstrapped data $\{y_t^*\}$, we construct the bootstrapped test statistic as Step 1, i.e. we again estimate two sets of regressions based on $\{u_t^*\}$, and construct an estimate $\hat{\sigma}^*$ for the long-run variance of $\{u_t^*\}$ based on the semiparametric regression, and construct the test statistics as the standard procedure, except that we use $\hat{\sigma}^{*2}$ as the variance estimate. Notice that we use, in the bootstrap stage, the same bandwidth and lag order in Step 1. Denote the bootstrapped test statistic as $S^*$.

3. Repeat step 2 for $N$ times, denote the test statistic for each simulated sample as $S_b^*$, $b = 1, 2, \cdots, N$. The limiting null distribution of the test statistic can then be approximated by the empirical distribution of $S_b^*$. Let $C_\alpha^*$ be the $(1-\alpha)$-th quantile of $\{S_b^*\}_{b=1}^N$, i.e.

$$\mathbf{P}^*\left(S_b^* \leq C_\alpha^*\right) = 1 - \alpha$$

4. Compare the testing statistic $S$ in step 1 with the bootstrapped critical value $C_\alpha^*$. The null hypothesis will be rejected at the level $\alpha$ if $S \geq C_\alpha^*$.

In the leading case of testing for a changing trend in dynamic models, $x_t = 0$ and $z_t$ is a deterministic trend, the semiparametric regression (17) reduces to a varying coefficient nonparametric estimation

$$y_t = z_t'\hat{\beta}\left(t/T\right) + \hat{u}_t,\ t = 1, 2, \cdots, T$$

which can be estimated by (using a local constant estimator)

$$\hat{\beta}_t = \arg\min_{\beta} \sum_{s=1}^{T} \left(y_s - z_s'\beta\right)^2 \cdot K\left(\frac{t-s}{Th}\right)$$

Consider the special case with $z_t = (1, t/T)$. Then, the preceding equation becomes

$$\left(\hat{\beta}_{t,0}, \hat{\beta}_{t,1}\right) = \arg\min_{\beta_0,\beta_1} \sum_{s=1}^{T} \left(y_s - \beta_0 - \beta_1 \cdot \frac{s}{T}\right)^2 \cdot K\left(\frac{t-s}{Th}\right)$$

This can also be written as

$$\left(\hat{\beta}_{t,0}, \hat{\beta}_{t,1}\right) = \arg\min_{\beta_0,\beta_1} \sum_{s=1}^{T} \left[y_s - \left(\beta_0 + \beta_1 \cdot \frac{t}{T}\right) - \beta_1 \cdot \frac{s-t}{T}\right]^2 \cdot K\left(\frac{t-s}{Th}\right)$$

which is equivalent to

$$\left(\hat{\beta}^*_{t,0}, \hat{\beta}^*_{t,1}\right) = \arg\min_{\beta^*_0,\beta^*_1} \sum_{s=1}^{T} \left[y_s - \beta^*_0 - \beta^*_1 \cdot \frac{s-t}{T}\right]^2 \cdot K\left(\frac{t-s}{Th}\right)$$

and the local estimate for $z_t'\beta\left(t/T\right)$ is then given by $\hat{\beta}^*_{t,0}$. If the regression (13) has only a constant term, i.e. $z_t = 1$, the the problem becomes

$$\hat{\beta}_{t,0} = \arg\min_{\beta_0} \sum_{s=1}^{T} \left(y_s - \beta_0\right)^2 \cdot K\left(\frac{t-s}{Th}\right)$$

and this is simply the Nadaraya-Watson estimator.

The following result shows that the bootstrap procedure consistently estimates the null limiting distribution. Note that the convergence is in the sense of Giné and Zinn [1990].

**Theorem 3.** *Under the null hypothesis and Assumptions 1–5,* $S^* \xrightarrow{d} \mathcal{L}$ *in* $\mathbf{P}$.

**Definition 1.** We say $\xi^*_T = o_{p^*}(1)$ in $\mathbf{P}$ if $P^*(|\xi^*_T| > \varepsilon) = o_p(1)$ for all $\varepsilon > 0$.

Here $P^*(\cdot) := P(\cdot|X, Y, Z)$ is bootstrap probability conditional on data and $E^*(\cdot) := E(\cdot|X, Y, Z)$ is bootstrap expectation conditional on data.

**Definition 2.** We say $\xi^*_T = O_{p^*}(1)$ in $\mathbf{P}$ if $\delta_T \xi^*_T = o_{p^*}(1)$ for any positive sequence $\delta_T = o(1)$.

## 27.1 A modified bootstrap procedure

In many macroeconomic applications, the serial correlation can be quite strong and the largest autoregressive root may be close to 1. In this case, estimate of $\rho_1$ may be downward biased under the null hypothesis. As a result, the bootstrap procedure discussed above may have size distortions. Based on such a consideration, we consider a modification that using an additional layer of bootstrap to correct the bias. We label this procedure as modified bootstrap procedure.

1. (Construct the test statistic). Same as before, denote the testing statistic as $S$.

2. (Bias correction). Re-sample with replacement from the re-centered empirical distribution of $\hat{e}_t$ and simulate samples under the null (using $\tilde{\gamma}$ and $\tilde{\beta}$) in the same way, i.e. we generate

$$\begin{aligned} u_t^* &= \hat{\rho}_1 u_{t-1}^* + \hat{\rho}_2 \Delta u_{t-1}^* + \cdots + \hat{\rho}_k \Delta u_{t-k+1}^* + \hat{e}_t^* \\ y_t^* &= x_t' \tilde{\gamma} + z_t' \tilde{\beta} + u_t^* \end{aligned}$$

Based on the re-sampled data $\{y_t^*\}$, following step 1, we estimate $(\rho_1, \rho_2, \cdots, \rho_k)$. Repeat this for $B$ times (again, with the same bandwidth and lag order), and denote the estimated autoregressive coefficients for $j$-th sample as $(j = 1, 2, \cdots, B)$

$$\hat{\rho}^{*(j)} = (\hat{\rho}_1^{*(j)}, \cdots, \hat{\rho}_k^{*(j)})$$

Estimate the bias of the largest autoregressive root:

$$\hat{b} = \hat{\rho}_1 - \frac{1}{B} \sum \hat{\rho}_1^{*(j)}$$

Finally, construct the bias corrected estimate for $\rho_1$ as

$$\hat{\rho}_1^c = \begin{cases} \hat{\rho}_1 + \hat{b} & \text{if } |\hat{\rho}_1 + \hat{b}| < 1 \\ \hat{\rho}_1 & \text{otherwise} \end{cases}$$

3. (Generate the bootstrap sample). Re-sample with replacement from the re-centered empirical distribution of $\hat{e}_t$ and generate the bootstrap sample using bias-corrected estimates, i.e. we generate

$$\begin{aligned} u_t^{c*} &= \hat{\rho}_1^c u_{t-1}^* + \hat{\rho}_2 \Delta u_{t-1}^* + \cdots + \hat{\rho}_k \Delta u_{t-k+1}^* + e_t^* \\ y_t^{c*} &= x_t' \tilde{\gamma} + z_t' \tilde{\beta} + u_t^{c*} \end{aligned}$$

and construct the test statistic in the same way as step 1 (again the same bandwidth and lag order are used).

4. Repeat step 3 for $N$ times, denote the testing statistic for each simulated sample as $S_b^{c*}$, $b = 1, 2, \cdots, N$. The limiting null distribution of the test statistic can then be approximated by the empirical distribution of $S_b^{c*}$. Let $C_\alpha^{c*}$ be the $(1-\alpha)$-th quantile of $\{S_b^{c*}\}_{b=1}^N$, i.e.,

$$\mathbf{P}^* \left( S_b^{c*} \leq C_\alpha^{c*} \right) = 1 - \alpha$$

5. Compare the testing statistic $S$ in step 1 with the bootstrapped critical value $C_\alpha^{c*}$. The null hypothesis will be rejected at the level $\alpha$ if $S \geq C_\alpha^{c*}$.

*Remark* 29. The application of nonparametric estimates ensure the tests have monotonic power. The bootstrap fixes the size. For persistent data, it is crucial to apply the bias correction before bootstrapping. It is important to note that the correction acknowledges that the bias depends on both the data generating process and the bandwidth, and this is why the bandwidth are fixed throughout. Also, we correct the bias only if this does not violate the stationarity condition, because we do not really want to move the estimate if it is already above the true value.

**Theorem 4.** *Under the null and Assumptions 1–5, conditional on the data and for almost all sample paths,* $S^{c*} \xrightarrow{D} \mathcal{L}$ *in* $\mathbf{P}$.

## 27.2 Simulation

We still use the same data generating process and consider the same specifications as in section 25.1. We report results for the CUSUM test. (Results for Sup-W test are very similar.) The reported values are based on 5,000 replications with 200 bootstrap samples for each replication. Lag orders are determined by BIC. For nonparametric estimation, we consider Epanechnikov kernel and apply a local-constant estimator. The bandwidth is set to

$$h = cT^{-1/5}$$

where $T$ is the sample size and $c$ is a constant. We consider $c = 1.0$ and $2.0$. Results are quite similar and our reports are based on $c = 1.0$. Note that for $T = 200$, $h$ then takes values 0.347 and 0.693, for $c = 1.0$ and $2.0$, respectively. Thus, $c = 2.0$ could be viewed as large bandwidth choice for sample sizes ($T = 200$) typically encountered in macroeconomics.

We first consider NB without size correction. The power property is attractive as all of them are monotonic to structural break. The results show that the size becomes more stable and close to the nominal level except when the series is very persistent. When error correlation is $\rho = 0.9$, the test is quite aggressive, 8.34% for $AR(1)$ and 9.08% for $ARMA(1,1)$.

Figure 16: CUSUM, no MA term

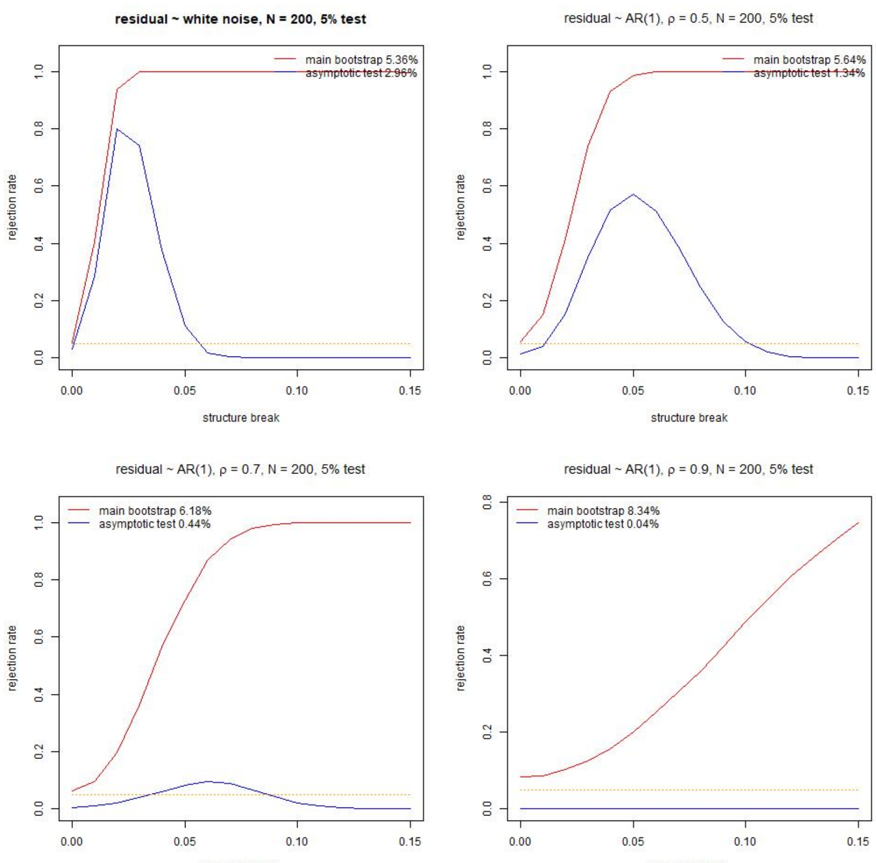


Figure 17: CUSUM, with MA term

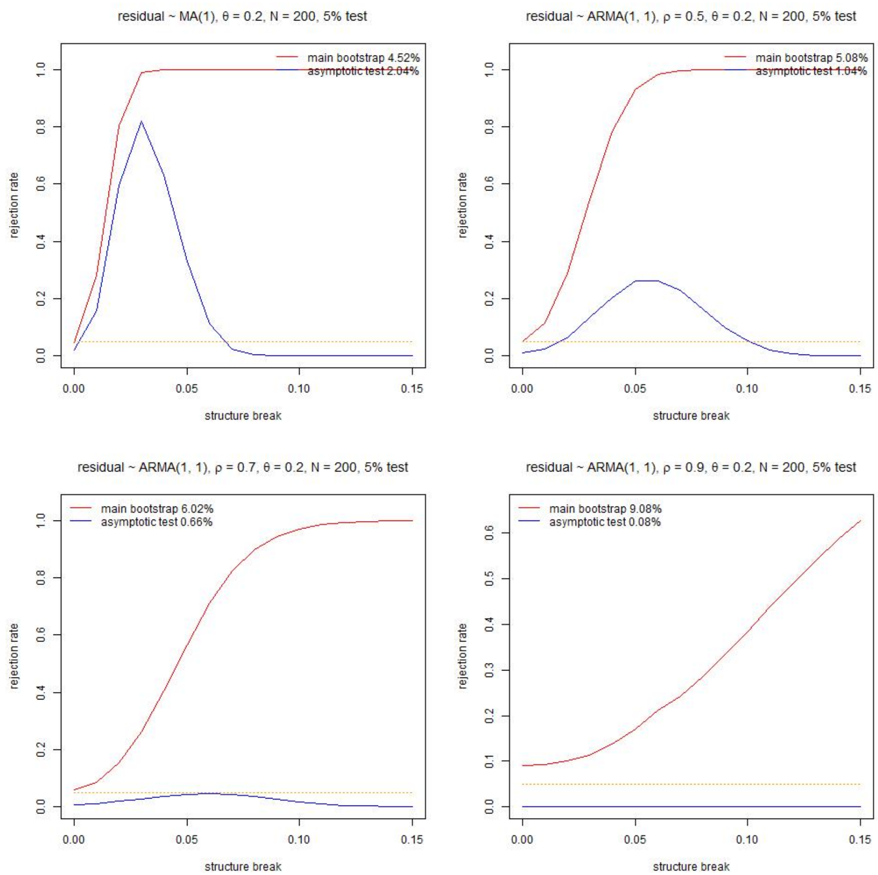


Figure 18: CUSUM, no MA term

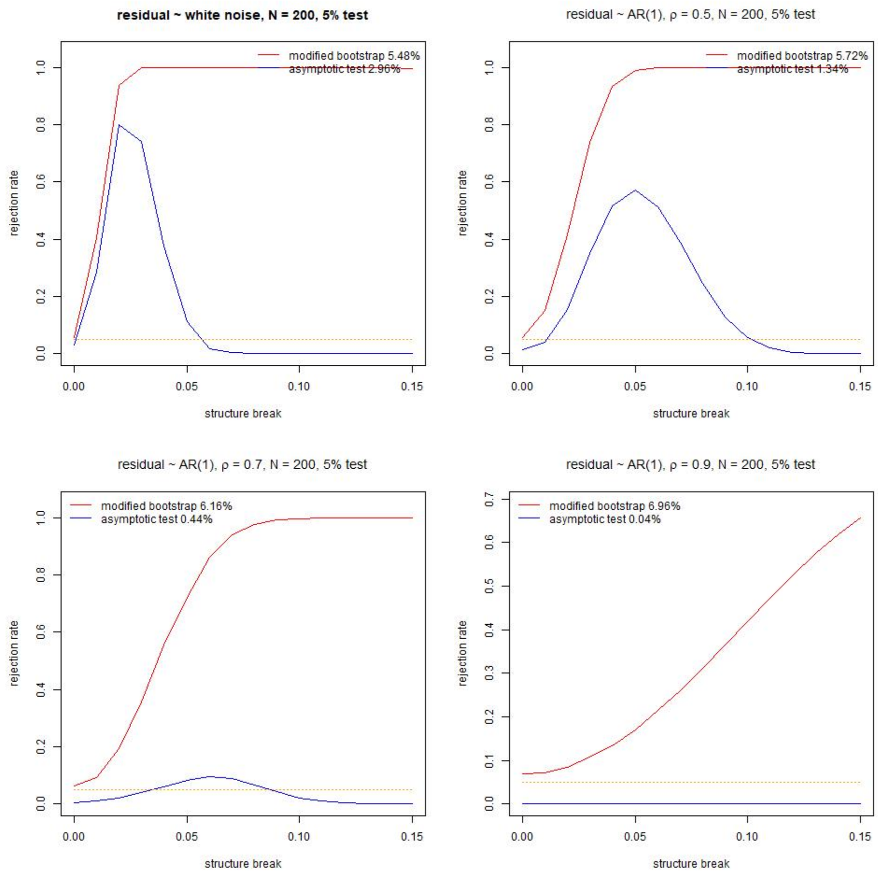
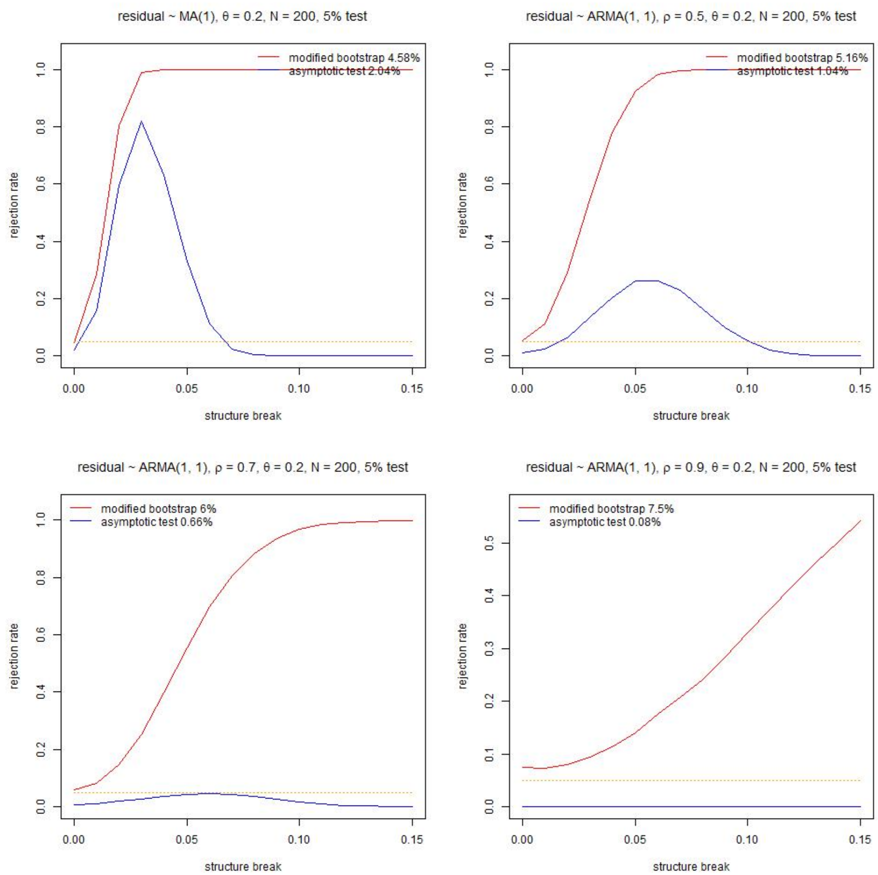


Figure 19: CUSUM, with MA term




The bias correction, employed in modified bootstrap procedure, offers a further refinement. The size becomes uniformly closer to the nominal level with the improvement being more significant when the series is persistent. Notice the power of modified test is a little worse than bootstrap without size correction, when residuals is persistent ($\rho = 0.9$). We view this as an acceptable price paid to have sizes uniformly closer to nominal level (and to be able to handle a wide range of models with trending moments).

# 28 Conclusion

This paper studies procedures for testing structural changes with good size and power properties. We focus on dynamic models and the analysis covers a wide range of important inference problems. Existing tests either suffer from substantial size distortions or exhibit non-monotonic power. We propose to address these two issues simultaneously by constructing estimates for nuisance parameters using nonparametrically detrended residuals to achieve good power and an appropriate bootstrap procedure to improve the size. The core of the construction is a modified bootstrap procedure. It is of sieve type and it differs from the conventional bootstrap procedure in two aspects: (1) it uses estimates from the nonparametric regression to generate bootstrap samples, and (2) it uses simulations to correct for the bias associated with the estimates for the largest auto-regressive root. We show that the procedure yields tests with adequate size and good power against a broad class of structural changes, including one time discrete change, smooth change and multiple structural changes.

**Extension**: this paper concentrates on the model where dynamics are relegated to the error term. In other contexts, it may be desirable to model the dynamics directly. For example, a lagging term $y_{t-1}$ could be included:

$$y_t = \rho y_{t-1} + x_t'\gamma + z_t'\beta + u_t$$

Here $u_t$ is assumed to be a martingale difference sequence to avoid endogeneity. We want to test whether there is a structure break in $\beta$. Furthermore, the lagging term could be added to the test, that is, whether there is a structural break in $\rho$. Extending our algorithms (with good size and power properties) to this dynamic model is of practical value and will be analyzed in another project.

# References


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

Donald WK Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856, 1993.

Donald WK Andrews and Hong-Yuan Chen. Approximately median-unbiased estimation of autoregressive models. *Journal of Business & Economic Statistics*, 12(2):187–204, 1994.

Donald WK Andrews and Werner Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, pages 1383–1414, 1994.

Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78, 1998.

Brendan K Beare. Copulas and temporal dependence. *Econometrica*, 78(1):395–410, 2010.

Brendan K Beare. Archimedean copulas and temporal dependence. *Econometric Theory*, 28(6): 1165–1185, 2012.

Alon Brav and Reuven Lehavy. An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics. *The Journal of Finance*, 58(5):1933–1968, 2003.

Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach.* Springer Science & Business Media, 2003.

Xiaohong Chen. Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics*, 6:5549–5632, 2007.

Xiaohong Chen and Yanqin Fan. Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335, 2006.

Xiaohong Chen and Xiaotong Shen. Sieve extremum estimates for weakly dependent data. *Econometrica*, pages 289–314, 1998.

Xiaohong Chen and Zhijie Xiao. Copula-based time series with filtered nonstationarity. 2016.

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

Xiaohong Chen, Wei Biao Wu, and Yanping Yi. Efficient estimation of copula-based semiparametric markov models. *The Annals of Statistics*, pages 4214–4253, 2009.

Ciprian M Crainiceanu and Timothy J Vogelsang. Nonmonotonic power for tests of a mean shift in a time series. *Journal of Statistical Computation and Simulation*, 77(6):457–476, 2007.

William F Darsow, Bao Nguyen, Elwood T Olsen, et al. Copulas and markov processes. *Illinois Journal of Mathematics*, 36(4):600–642, 1992.

Robert M De Jong. Nonlinear minimization estimators in the presence of cointegrating relations. *Journal of Econometrics*, 110(2):241–259, 2002.

Ai Deng and Pierre Perron. The limit distribution of the cusum of squares test under general mixing conditions. *Econometric Theory*, 24(3):809–822, 2008.

Francis X Diebold and Celia Chen. Testing structural stability with endogenous breakpoint a size comparison of analytic and bootstrap procedures. *Journal of Econometrics*, 70(1):221–241, 1996.

Paul Doukhan, Jean-David Fermanian, Gabriel Lang, et al. *Copulas of a vector-valued stationary weakly dependent process*. INSEE, 2005.

Ray C Fair. Computing median unbiased estimates in macroeconometric models. *Journal of Applied Econometrics*, 11(4):431–435, 1996.

Evarist Giné and Joel Zinn. Bootstrapping general empirical measures. *The Annals of Probability*, pages 851–869, 1990.

Clive WJ Granger. Time series concepts for conditional distributions. *Oxford Bulletin of Economics and Statistics*, 65:689–701, 2003.

Bruce E Hansen. Testing for structural change in conditional models. *Journal of Econometrics*, 97 (1):93–115, 2000.

Bruce E Hansen. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(03):726–748, 2008.

Rustam Ibragimov. Copula-based characterizations for higher order markov processes. *Econometric Theory*, 25(03):819–846, 2009.

Harry Joe. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

Eric Jondeau and Michael Rockinger. The copula-garch model of conditional dependencies: An international stock market application. *Journal of international money and finance*, 25(5):827–853, 2006.

Ted Juhl and Zhijie Xiao. Tests for changing mean with monotonic power. *Journal of Econometrics*, 148(1):14–24, 2009.

Lutz Kilian. Accounting for lag order uncertainty in autoregressions: the endogenous lag order bootstrap algorithm. *Journal of Time Series Analysis*, 19(5):531–548, 1998.

Martial Longla and Magda Peligrad. Some aspects of modeling dependence in copula-based markov chains. *Journal of Multivariate Analysis*, 111:234–240, 2012.

RB Nelson. An introduction to copulas, lectures notes in statistics vol. 39, 1999.

Whitney K Newey and Kenneth D West. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, pages 777–787, 1987.

Cuong C Nguyen and M Ishaq Bhatti. Copula model dependency between oil prices and stock markets: Evidence from china and vietnam. *Journal of International Financial Markets, Institutions and Money*, 22(4):758–773, 2012.

Joon Y Park. An invariance principle for sieve bootstrap in time series. *Econometric Theory*, 18 (02):469–490, 2002.

Andrew J Patton. A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18, 2012.

Pierre Perron. *A test for changes in a polynomial trend function for a dynamic time series*. Université de Montréal, Centre de recherche et développement en économique, 1991.

Pierre Perron et al. Dealing with structural breaks. *Palgrave handbook of econometrics*, 1(2):278–352, 2006.

Werner Ploberger and Walter Krämer. The cusum test with ols residuals. *Econometrica: Journal of the Econometric Society*, pages 271–285, 1992.

Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954, 1988.

M Sklar. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8: 229–231, 1959.

Charles J Stone. Optimal global rates of convergence for nonparametric regression. *The annals of statistics*, pages 1040–1053, 1982.

H Victor, Rustam Ibragimov, Shaturgun Sharakhmetov, et al. Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series. In *Optimality*, pages 183–209. Institute of Mathematical Statistics, 2006.

Timothy J Vogelsang. Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics*, 88(2):283–299, 1999.

Xiaoqian Wen, Yu Wei, and Dengshi Huang. Measuring contagion between energy market and stock market during financial crisis: A copula approach. *Energy Economics*, 34(5):1435–1446, 2012.