Title
Copula-based Econometric Models of Intertemporal and Cross -sectional Dependence

Author
Seo, Juwon

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Copula-based Econometric Models of Intertemporal and Cross-sectional Dependence

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Economics

by

Juwon Seo

Committee in charge:

Professor Brendan K. Beare, Chair
Professor Graham Elliott
Professor Dmitris N. Politis
Professor Yixiao Sun
Professor Allan Timmermann
Professor Ruth J. Williams

2015

The dissertation of Juwon Seo is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2015

DEDICATION

To Gollum

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

In writing my dissertation, I have incurred many debts which I am glad to have this opportunity to acknowledge. First of all, I wish to record my thanks to Brendan K. Beare, for his invaluable guidance and helpful advices throughout my graduate studies at the University of California, San Diego. He has been willing to be my first reader and caught many errors and awkward expressions. Without his support, none of my achievement would have been possible.

I would like to express my gratitude to Graham Elliott and Yixiao Sun, who from the outset, encouraged in my progress and provided me with many details and suggestions for research and carefully read the manuscript. In addition, I am also indebted to my dissertation committee members, Allan Timmermann, Dimitris N. Politis and Ruth J. Williams for their insightful comments and constructive criticism of my work as it progressed. Endless encouragement was furnished by Jinook Jeong, my former supervisor at Yonsei University. I sincerely thank for his trust in me.

It is a particular pleasure to acknowledge my debt to remarkable cohort of students of the University of California, San Diego. I am grateful to Hyeong Ho Mun, Jungbin Hwang, Myungkyu Shim, Sung Je Byun, Hiejoo Ahn, Heea Jung, Chim Lau, Youjin Hahn, Hee Seung Yang, Suyong Song, Min Seong Kim, Jong-Myun Moon, Hanwool Yoon, Bong Hwan Kim, Dora Xia, Zheng Fang, Zhenting Sun, Matthew Goldman, Yaein Baek, Won-ki Seo, Wonhyong Choi and Eul Noh. And a very special thanks to my officemates in Econ 127, Linchun (Lily) Chen, Anh Pham and Zheng Huang who went through all sorts of happiness, anger, sorrow and pleasure together.

Finally, my highest appreciation goes to my family, Hyeongjoo Seo, Jae Hyun Seo and my parents who have assisted me in innumerable ways, providing all sorts of tangible and intangible support and care, physical and emotional.

Chapter 1 is published in Econometric Theory (2014) and Chapter 2 is published in Journal of Time Series Analysis (2015). I thank Brendan K. Beare for his

permission to use the co-authored papers in my dissertation Chapter 1 and 2.

VITA

2007	B. A. in Economics, Yonsei University, Seoul, Korea
2009	M. A in Economics, Yonsei University, Seoul, Korea
2015	Ph. D. in Economics, University of California, San Diego

PUBLICATIONS

"Time Irreversible Copula-based Markov Models (with Brendan K. Beare)", *Econometric Theory*, **30 (5)**, 923-960.

"Vine Copula Specifications for Stationary Multivariate Markov Chains (with Brendan K. Beare)", *Journal of Time Series Analysis*, **36 (2)**, 228-246.

ABSTRACT OF THE DISSERTATION

Copula-based Econometric Models of Intertemporal and Cross-sectional Dependence

by

Juwon Seo
Doctor of Philosophy in Economics

University of California, San Diego, 2015

Professor Brendan K. Beare, Chair

The modeling of nonlinear and non-Gaussian dependence structures is of great interest to many researchers. Particularly, copula-based models have recently attracted a fair amount of attention due to their applicability and flexibility. This dissertation studies copula-based econometric models of intertemporal and cross-sectional dependence: the first and the third chapters analyze some general dependence types characterized by copulas, time irreversibility and stochastic monotonicity respectively. The second chapter focuses on the development of new copula-based models for stationary multivariate time series.

The first chapter concerns a dependence property called time irreversibility. When we say a model is time irreversible, it means we may expect a plot of the series to exhibit different patterns when time runs forward and backward. We frequently observe time irreversibility in the asymmetric fluctuation of stock market data, unemployment rates, price series or business cycles. In the chapter we show that time reversibility is equivalent to the exchangeability of a copula function, and

suggest a nonparametric test for time irreversibility. The distinguishing feature of our test is that it can detect any arbitrary form of irreversibility. We also show how time irreversible behavior may be described using a function called the circulation density, and propose a nonparametric estimator of this function.

While my first project mainly concerned the first order stationary Markov chains of univariate time series, we turn our attention to higher dimensional cases in the second chapter. We show how to construct flexible models for multivariate time series using a graphical representation of joint distributions called vine copulas. Building on existing studies of copula-based univariate Markov models, our extension is made in two directions: (1) we consider multivariate time series, and (2) we allow Markov chains of any finite order. We propose a vine structure called the M-vine that is particularly well suited to model stationary Markov chains, and convenient to capture some interesting intertemporal and contemporary dependencies. An empirical application to the exchange rates of Korean won (KRW) and the Taiwanese dollar (TWD) is provided.

In the last chapter, we study stochastic monotonicity, a dependence property that can be reframed in terms of the concavity of cross-sections of a copula function. Stochastic monotonicity is a distributional property which says that two variables tend to be positively associated, and it has been of great interest in many areas of economics such as experimental design, information economics, and labor economics. In this chapter, we discuss how to improve the power of the tests by using a modified bootstrap technique to choose a critical value that delivers a limiting rejection rate equal to nominal size over a wide region of the null hypothesis. To show the validity of this approach we draw on recent results on the directional differentiability of the least concave majorant operator, and on bootstrap inference when smoothness conditions sufficient to apply the functional delta method for the bootstrap are not satisfied.

Chapter 1

Time Irreversible Copula-based Markov Models

Abstract. Economic and financial time series frequently exhibit time irreversible dynamics. For instance, there is considerable evidence of asymmetric fluctuations in many macroeconomic and financial variables, and certain game theoretic models of price determination predict asymmetric cycles in price series. In this paper we make two primary contributions to the econometric literature on time reversibility. First, we propose a new test of time reversibility, applicable to stationary Markov chains. Compared to existing tests, our test has the advantage of being consistent against arbitrary violations of reversibility. Second, we explain how a circulation density function may be used to characterize the nature of time irreversibility when it is present. We propose a copula-based estimator of the circulation density, and verify that it is well behaved asymptotically under suitable regularity conditions. We illustrate the use of our time reversibility test and circulation density estimator by applying them to five years of Canadian gasoline price markup data.

1.1 Introduction

A central concern of time series econometrics is modeling the dynamic behavior of random processes over time. Dynamic behavior may be classified as either *time reversible* or *time irreversible*. Loosely speaking, we say that a process is time reversible if its probabilistic structure is unaffected by reversing the direction of time. For instance, if a process is characterized by frequent small decreases and occasional large increases, then if we were to reverse the direction of time we would instead obtain a process characterized by frequent small increases and occasional large decreases. Such a process may therefore be described as time irreversible.

Questions about time reversibility arise naturally in the study of the business cycle. Rothman (1991) refers to the so-called Mitchell-Keynes business cycle hypothesis, which posits that economic expansions are more gradual than economic contractions. In the *General Theory*, Keynes (1936, p. 314) wrote that "the substitution of a downward for an upward tendency often takes place suddenly and violently, whereas there is, as a rule, no such sharp turning point when an upward is substituted for a downward tendency" This quotation appears also in Neftçi (1984) and DeLong and Summers (1986). In these two papers an attempt was made to test empirically for the presence of asymmetry in the business cycle. Neftçi (1984) argued for the importance of asymmetric fluctuations, finding evidence of time irreversibility in the US unemployment rate. DeLong and Summers (1986) concurred with Neftçi's assessment of irreversible dynamics in US unemployment, but found no evidence of time irreversibility in US gross national product or industrial production, or in any of these three variables in five other OECD nations. However, in the 1990's and beyond, more sophisticated econometric techniques were used to identify time irreversible behavior in a wide range of macroeconomic and financial variables; see e.g. Rothman (1991), Ramsey and Rothman (1996), Hinich and Rothman (1998), Chen et al. (2000), Chen and Kuan (2002), Darolles et al. (2004), Racine and Maasoumi (2007), and Psaradakis (2008).

Time irreversible behavior may also arise naturally in models of oligopolistic price setting. Edgeworth price cycles are said to occur when competing firms engage in extended periods of sequential price undercutting, interspersed with occasional short periods of "relenting", during which one firm raises its price significantly and the others follow. This behavior leads to time irreversible price series exhibiting gradual declines and sudden sharp increases, a pattern sometimes referred to as "rockets and feathers" (see e.g. Tappata, 2009). Maskin and Tirole (1988) provided dynamic game-theoretic foundations for the existence of Edgeworth price cycles in Bertrand duopolies. Subsequent extensions were provided by Eckert (2003), who examined the case of asymmetrically sized firms, and Noel (2008), who considered markets with more than two firms, among other scenarios. Empirical researchers (see e.g. Eckert, 2002; Noel, 2007; Wang, 2009; Lewis and Noel, 2011) have found that many retail gasoline markets exhibit prominent Edgeworth price cycles over time. This behavior is not confined to gasoline markets: Peltzman (2000) examined price data for 242 different goods, finding evidence of asymmetric price movements in more than two thirds of them. Edgeworth price cycles have also been reproduced in an experimental setting (Cason et al., 2005).

In this paper we consider the property of time reversibility in the context of copula-based Markov models. This class of models was introduced to the econometric literature by Chen and Fan (2006); subsequent contributions to the subject include Fentaw and Naik-Nimbalkar (2008), Gagliardini and Gouriéroux (2008), Bouyé and Salmon (2009), Chen, Koenker and Xiao (2009), Chen, Wu and Yi (2009), Ibragimov (2009), Beare (2010, 2012), and the recent book by Cherubini et al. (2011). The time series of interest is assumed to be a stationary real valued Markov chain. Model specification involves the selection of a distribution function F to characterize the invariant, or stationary, distribution of the chain, and a copula function C to characterize dynamic dependence. There are two key advantages to this approach. First, complex forms of nonlinear dynamic dependence may easily be introduced with an appropriate choice of C, without any possibility of violating the stationarity condition. Second, there is the possibility of combining a parametric copula C with a nonparametric choice of F, limiting the effect of the curse of dimensionality while maintaining a degree of flexibility not achievable with fully parametric models.

For the class of copula-based Markov models, time reversibility is equivalent to a property of C called exchangeability. In Section 1.2 we discuss this equivalence, and explain how a technique proposed by Khoudraji (1995) may be used to construct parametric families of nonexchangeable copula functions. Our main contributions are provided in Sections 1.3 and 1.4. In Section 1.3 we propose a new test of time reversibility for stationary real valued Markov chains. The key advantage of our test is that it is consistent against *any* violation of time reversibility; existing procedures are typically only able to detect specific forms of time irreversibility. We derive the asymptotic behavior of our test statistic, and explain how asymptotically valid critical values may be obtained using the local bootstrap of Paparoditis and Politis (2002). Finite sample numerical evidence illustrates the primary strength and weakness of our test relative to existing tests. In Section 1.4, building on novel work by McCausland (2007) in the context of finite state Markov chains, we propose to characterize the structure of time irreversibility in a stationary Markov chain using a *circulation density function*. The circulation density function decomposes the total circulation of the chain – the difference between the unconditional probabilities of an increase or decrease – into contributions associated with each quantile of the invariant distribution. This provides us with information about whether the process tends to be more likely to increase or decrease at different quantiles. It turns out that, under mild regularity conditions, the circulation density function is determined by the partial derivatives of C along the main diagonal of the unit square. We propose a nonparametric estimator of the circulation density function and establish consistency and asymptotic normality. Some encouraging finite sample results are provided.

We illustrate the use of our time reversibility test and circulation density estimator in Section 1.5, with an application to five years of weekly Canadian gasoline price markup data. Our results appear to confirm the presence of Edgeworth price cycles in these data. Moreover, our estimated circulation density is suggestive of price undercutting sequences being more prevalent when we are in the lower half of the invariant distribution. This finding is consistent with earlier work by McCausland (2007) using these data.

We offer some concluding thoughts in Section 1.6. The Appendix contains some technical conditions used to demonstrate the validity of the local bootstrap, and proofs of the results given throughout the main body of the paper, along with some supplementary lemmas.

1.2 Nonexchangeable copulas and time irreversibility

Let $\mathscr{X} = \{X_t : t \in \mathbb{Z}\}$ be a stationary real valued Markov chain with invariant cdf $F : \mathbb{R} \to [0, 1]$. Darsow et al. (1992) suggested that copula functions may provide a convenient and powerful way to model the dynamic properties of $\mathscr{X}$. If F is continuous, then Sklar's theorem ensures the existence of a unique copula function $C : [0, 1]^2 \to [0, 1]$ characterizing the relationship between X_t and X_{t+1}, for any $t \in \mathbb{Z}$. Letting $H : \mathbb{R}^2 \to [0, 1]$ denote the joint cdf of X_t and X_{t+1}, we have

$$P(X_t \leq x, X_{t+1} \leq y) = H(x, y) = C(F(x), F(y)) \quad \text{for all } x, y \in \mathbb{R} \text{ and all } t \in \mathbb{Z}.$$

Taken together, C and F jointly determine all finite dimensional distributions of $\mathscr{X}$, with dynamic dependence at lags greater than one determined by the Markov property. Further details on copula functions, Sklar's theorem and related concepts may be found in the monograph of Nelsen (2006).

The following result provides three equivalent formulations of time reversibility for stationary Markov chains. It is well understood and we do not provide a proof.

Proposition 1.2.1. *Suppose $\mathscr{X}$ is a stationary real valued Markov chain with continuous invariant distribution. The following statements are equivalent.*

(a) *For any consecutive integers $t_1 < \cdots < t_n$, we have $(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_n}, \ldots, X_{t_1})$.*

(b) $H(x, y) = H(y, x)$ *for all* $x, y \in \mathbb{R}$.

(c) $C(u, v) = C(v, u)$ *for all* $u, v \in [0, 1]$.

Property (a) is the standard definition of time reversibility for stationary time series. Under the Markov property, time reversibility is equivalent to property (b), sometimes known as the detailed balance condition (McCausland, 2007, p. 308). When F is continuous, the copula C is uniquely defined, and so (b) and (c) are equivalent. Time reversibility of $\mathscr{X}$ is therefore a property of C, the copula characterizing serial dependence. If $\mathscr{X}$ is not time reversible, we say that it is time irreversible.

A joint cdf H satisfying property (b) in Proposition 1.2.1 or a copula C satisfying property (c) in Proposition 1.2.1 is said to be *exchangeable*.[1] Nelsen (2007) studied some aspects of nonexchangeable copulas. He proposed to measure the nonexchangeability of a copula C using the following quantity:

$$\delta(C) = 3 \sup_{u,v} |C(u, v) - C(v, u)| . \tag{1.1}$$

Theorem 2.2 of Nelsen (2007) establishes that $0 \leq \delta(C) \leq 1$ for all copulas C, with the lower and upper bounds attainable.[2] Evidently we have $\delta(C) = 0$ if and only if C is exchangeable. Larger values of $\delta(C)$ signify more substantial nonexchangeability of C – or, in our context, time irreversibility of $\mathscr{X}$. In Section 1.3

[1] An alternative characterization of exchangeability involving canonical U-statistic representations of multivariate distributions and copulas was provided by de la Peña et al. (2006, Theorem 5.5).

[2] A copula C achieves the upper bound $\delta(C) = 1$ if and only if either $C(1/3, 2/3) = 1/3$ and $C(2/3, 1/3) = 0$, or $C(1/3, 2/3) = 0$ and $C(2/3, 1/3) = 1/3$. See Nelsen (2007, Theorem 3.1).

we will use Nelsen's measure of nonexchangeability as the basis for constructing a statistical test of time reversibility.

There are various ways to construct parametric families of nonexchangeable copulas. Khoudraji (1995) proposed a particularly convenient method by which this may be achieved; see also Genest et al. (1998) and Liebscher (2008). If C is an exchangeable copula and $\alpha, \beta \in [0,1]$, then the following transformation of C is a copula:

$$\tilde{C}(u,v) = u^{1-\alpha}v^{1-\beta}C(u^{\alpha}, v^{\beta}). \tag{1.2}$$

We may use (1.2) to generate a family of nonexchangeable copulas using an exchangeable copula. For instance, one well-known family of exchangeable copulas is the Gumbel family; see e.g. Nelsen (2006, Table 4.1, line 4). This is an Archimedean family having generator $u \mapsto (-\ln u)^{\gamma}$, with parameter $\gamma \in [1,\infty)$. All Archimedean copulas are exchangeable by construction, and therefore generate time reversible Markov chains. If we apply transform (1.2) to the Gumbel copula, we obtain the family of so-called asymmetric[3] Gumbel copulas:

$$\tilde{C}^{\text{Gmbl}}(u,v) = u^{1-\alpha}v^{1-\beta}\exp\left(-\left((-\alpha\ln u)^{\gamma} + (-\beta\ln v)^{\gamma}\right)^{1/\gamma}\right). \tag{1.3}$$

The asymmetric Gumbel copula has parameters $(\alpha,\beta,\gamma) \in [0,1]\times[0,1]\times[1,\infty)$, and is nonexchangeable if $\alpha,\beta > 0$, $\alpha \neq \beta$, and $\gamma > 1$. With some tedious but routine calculations involving l'Hôpital's rule, we may verify that the asymmetric Gumbel copula has lower tail dependence coefficient $\lambda_L = 0$ and upper tail dependence coefficient $\lambda_U = \alpha + \beta - (\alpha^{\gamma} + \beta^{\gamma})^{1/\gamma}$. See Nelsen (2006, pp. 214–217) for further discussion of tail dependence. When $\gamma \to \infty$, the asymmetric Gumbel copula reduces to the well-known Marshall-Olkin copula (Nelsen, 2006, p. 53) with parameters (α,β) and upper tail dependence coefficient $\lambda_U = \min\{\alpha,\beta\}$.

[3]Asymmetry here refers to nonexchangeability. It is important not to confuse nonexchangeability with other forms of asymmetry. The ordinary Gumbel copula is asymmetric in the sense of having different upper and lower tail dependence coefficients, but symmetric in the sense of being exchangeable. Nelsen (2006, pp. 36–38) discusses different forms of bivariate symmetry.

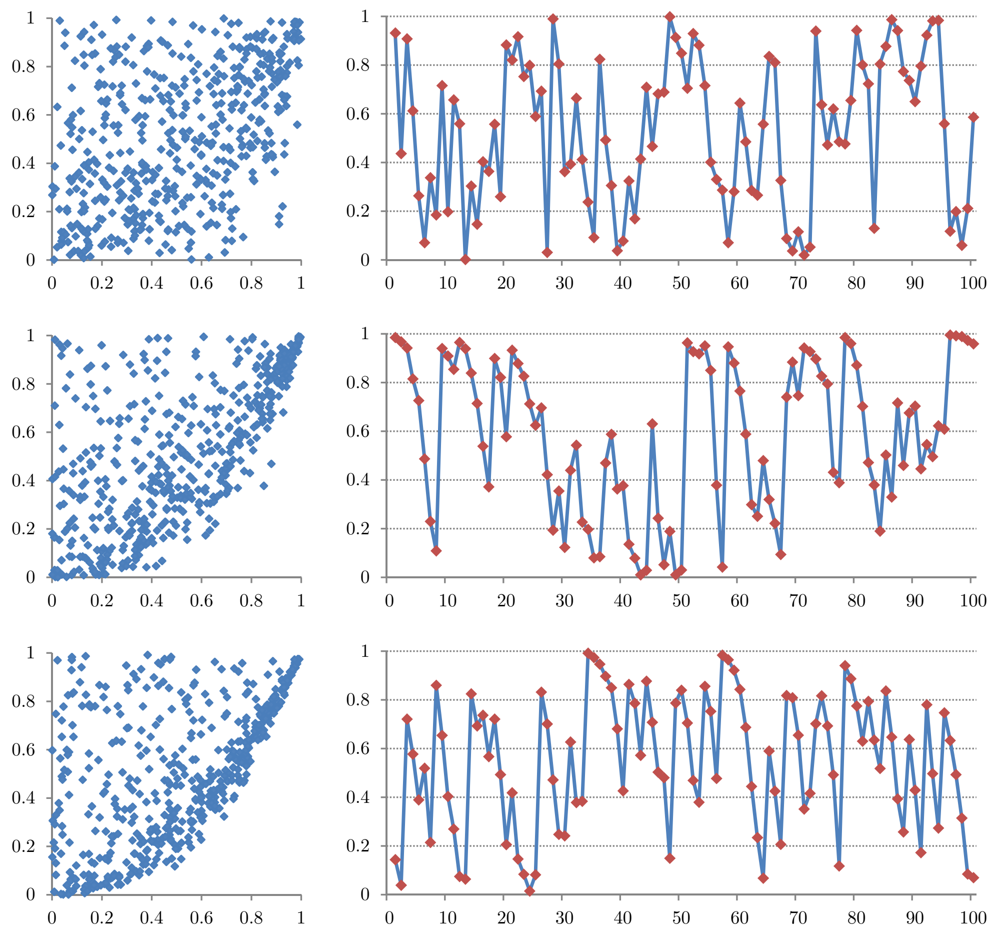


Figure 1.1: Scatterplots and Markov sample paths generated by the asymmetric Gumbel copula
We set $\alpha = 1$, $\beta = 0.5$ and take the invariant distribution to be uniform on $(0, 1)$. γ is equal to 2 in the top row, 5 in the center row, and 10 in the bottom row.

In Figure 1.1 we display several scatterplots and Markov sample paths generated using the asymmetric Gumbel copula. The scatterplots on the left were constructed by drawing from the asymmetric Gumbel copula with $\alpha = 1$, $\beta = 0.5$, and $\gamma = 2, 5, 10$. Nonexchangeability is mildly apparent when $\gamma = 2$, and much more obviously apparent when $\gamma = 5, 10$. The nonexchangeability measure given in (1.1) was numerically calculated to be 0.077 when $\gamma = 2$, 0.1716 when $\gamma = 5$, and 0.2087 when $\gamma = 10$. The Markov sample paths on the right side of Figure 1.1 were generated using the copulas in the corresponding scatterplots to the left.

The invariant distribution of each chain was chosen to be uniform on $(0, 1)$. Casual inspection reveals that decreases in these sample paths tend to be smaller and more frequent than increases. Again, this is much more obvious for larger values of γ. The tendency to exhibit many small decreases and occasional large increases is manifested in, for instance, Edgeworth price cycles. We shall return to the subject of Edgeworth price cycles in our empirical application in Section 1.5. For more details on how to simulate Markov chains using a given copula function and invariant distribution, and on how to empirically estimate models of this kind, we refer the reader to Chen and Fan (2006).

1.3 Testing for time irreversibility

Following the empirical macroeconomic literature on business cycle asymmetry in the 1980s and early 1990s (see e.g. Neftçi, 1984; DeLong and Summers, 1986; Rothman, 1991), a number of authors have proposed statistical tests of time reversibility. Ramsey and Rothman (1996) proposed a test of time reversibility based on symmetric bicovariances, while Chen et al. (2000) proposed a test based on the characteristic function of the differenced process. Chen (2003) proposed a more general class of time reversibility tests subsuming both of the aforementioned tests. Hinich and Rothman (1998) proposed a frequency-domain test involving the bispectrum. Paparoditis and Politis (2002) and Psaradakis (2008) suggested using resampling techniques to test whether the differenced process has median zero. Darolles et al. (2004) proposed a test based on nonlinear canonical correlation analysis. Racine and Maasoumi (2007) proposed an entropy-based test that targets asymmetry in the distribution of the differenced process. Sharifdoost et al. (2009) proposed a test applicable to finite state Markov chains.

In this section we propose a new test of time reversibility. A key advantage of our test is that it is consistent against arbitrary forms of time irreversibility. Most of the tests just mentioned are only consistent against specific forms of time irreversibility. The test of Sharifdoost et al. (2009) does not appear to be subject to this critique, but its applicability is limited by the assumption of a finite state space. In Section 2.3.1 we explain how our test statistic is constructed, and discuss its asymptotic behavior under time reversibility and time irreversibility. In Section 2.3.2 we explain how the local bootstrap of Paparoditis and Politis (2002) can be used to obtain suitable critical values for our test statistic. In Section 2.3.3 we report numerical evidence pertaining to the finite sample performance of our test, using the test of Paparoditis and Politis (2002) as a point of comparison.

1.3.1 Test statistic and limiting distribution

As in the previous section, let $\mathscr{X} = \{X_t : t \in \mathbb{Z}\}$ be a stationary real valued Markov chain with continuous invariant distribution F, joint cdf H for (X_t, X_{t+1}), and corresponding copula function C. Let $\theta \in [0, 1/3]$ be given by

$$\theta = \sup_{x,y} |H(x,y) - H(y,x)| .$$

Since F is continuous, we must have $\theta = \frac{1}{3}\delta(C)$, where $\delta(C)$ is the measure of nonexchangability proposed by Nelsen (2007) and given in (1.1) above. Recalling Proposition 1.2.1(b), we know that $\mathscr{X}$ is time reversible if and only if $\theta = 0$. We therefore propose to test the null hypothesis of time reversibility using a test statistic formed from an empirical analogue to θ. Suppose we observe the T random variables $X_1, \ldots, X_T$. A natural empirical analogue to θ is

$$\theta_T = \sup_{x,y} |H_T(x,y) - H_T(y,x)| ,$$

where H_T is the empirical distribution function

$$H_T(x,y) = \frac{1}{T-1} \sum_{t=1}^{T-1} 1(X_t \leq x, X_{t+1} \leq y).$$

θ_T is the statistic we will use to test the null hypothesis that $\mathscr{X}$ is time reversible.[4] We shall obtain the asymptotic behavior of θ_T under the following conditions on $\mathscr{X}$.

Assumption 1.3.1. The following statements are true.

(a) $\mathscr{X}$ is a stationary real valued Markov chain.

(b) F is continuous.

(c) The α-mixing coefficients of $\mathscr{X}$ satisfy $\alpha_T = O(T^{-\eta})$ for some $\eta > 1$.

Parts (a,b) of Assumption 1.3.1 are basic to our analysis. The mixing condition introduced in part (c) is mild for practical purposes. Chen and Fan (2006), Chen, Wu and Yi (2009) and Beare (2010, 2012) identify conditions on C, satisfied for a wide range of copula functions used in applications, that imply a geometric rate of α-mixing. On the other hand, Example 4.1 of Beare (2012) identifies a family of copula functions that generate α-mixing at a rate no faster than T^{-1}, so part (c) is not automatically satisfied.

Under Assumption 1.3.1 we are able to establish the following result concerning the asymptotic behavior of θ_T under the null and alternative hypotheses. The proof, which may be found in the Appendix, is a straightforward application of results due to Rio (2000) delivering functional central limit theory for weakly dependent processes.

Theorem 1.3.1. *Under Assumption 1.3.1, the following statements are true.*

(a) *If $\mathscr{X}$ is time reversible, then $T^{1/2}\theta_T \to_d \sup_{x,y} |\mathscr{B}(x,y) - \mathscr{B}(y,x)|$ as $T \to \infty$, where $\mathscr{B}$ is a centered Gaussian process on $\mathbb{R}^2$ with covariance kernel*

$$\text{cov}\left(\mathscr{B}(x,y), \mathscr{B}(x',y')\right) = \sum_{t\in\mathbb{Z}} \text{cov}\left(1(X_0 \leq x, X_1 \leq y), 1(X_t \leq x', X_{t+1} \leq y')\right).$$

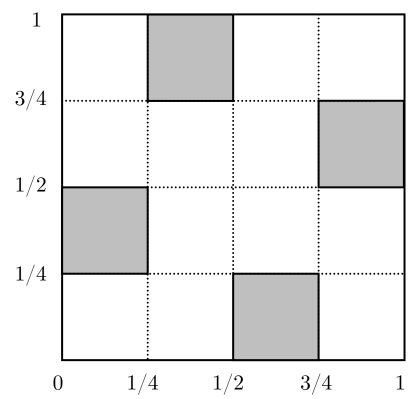
[4] In related work, Genest et al. (2012) proposed using a statistic very similar to θ_T to test for copula exchangeability in an iid bivariate context.

If $\mathscr{X}$ is time irreversible, then for any $c \in \mathbb{R}$ we have $T^{1/2}\theta_T > c$ with probability approaching one as $T \to \infty$.

Theorem 1.3.1(a) gives us the limiting distribution of $T^{1/2}\theta_T$ in terms of the process $\mathscr{B}$ under the null hypothesis that $\mathscr{X}$ is time reversible. A test of time reversibility may be formed by rejecting the null when $T^{1/2}\theta_T$ exceeds the relevant quantile of that limiting distribution. Theorem 1.3.1(b) tells us that, for any fixed critical value c, the probability of $T^{1/2}\theta_T$ exceeding c approaches one when the null hypothesis of time reversibility is false. This means that tests based on $T^{1/2}\theta_T$ will be consistent against any violation of time reversibility.

The covariance structure of the limiting process $\mathscr{B}$ depends on H, which is unknown. Therefore, critical values for our test must be estimated in some fashion. In the following subsection we explain how the local bootstrap procedure of Paparoditis and Politis (2002) may be used to obtain asymptotically valid critical values. We close this subsection with some additional remarks on our test, and on its relation to existing tests of time reversibility.

Remark 1.3.1. Theorem 1.3.1(b) indicates that our test is consistent against any violation of time reversibility. As mentioned at the beginning of this section, most existing tests of time reversibility do not share this property. In particular, the tests of Chen et al. (2000), Paparoditis and Politis (2002), Racine and Maasoumi (2007) and Psaradakis (2008) cannot detect any violation of time reversibility for which the univariate distribution of $X_{t+1} - X_t$ is symmetric about zero. Symmetry of this distribution is a necessary but not sufficient condition for time reversibility. Consider the probability distribution that distributes mass uniformly over the shaded region of the unit square depicted in Figure 1.2. It is easy to see that this distribution has uniform marginals and is asymmetric about the main diagonal of the unit square, implying that it may be represented by a nonexchangeable copula function. Further inspection reveals that, if the joint distribution of (X_t, X_{t+1}) is uniform over the shaded region, then the distribution of $X_{t+1} - X_t$ is symmetric about zero. To see this, note that the sets $\{(x,y) : y \leq x+a\}$ and $\{(x,y) : y \geq x-a\}$ have equal mass



(b)

Figure 1.2: Magic square 1
A pair of random variables distributed uniformly over the shaded region is nonexchangeable, but the distribution of their difference is symmetric about zero.

for all $a \geq 0$. It follows that this form of time irreversibility cannot be detected by the tests just cited, but is consistently identified by the test proposed here.

Remark 1.3.2. Darolles et al. (2004) propose an elegant test for time reversibility based on nonlinear canonical correlation analysis; see e.g. Lancaster (1958). Their procedure works by testing whether a given pair of canonical directions are equal to one another. A drawback of this approach in the context of copula-based Markov models is that the representation of a joint distribution in terms of canonical correlations and canonical directions is valid only when the distribution exhibits finite mean square contingency. As noted by Beare (2010), when C is absolutely continuous, H has finite mean square contingency if and only if C has square integrable density. Theorem 3.3 of Beare (2010) states that this condition rules out the presence of tail dependence in C. Tail dependence is a common property of parametric copula functions used in applications. Thus the test of Darolles et al. (2004) is not always ideally suited to the class of models under consideration. The test proposed here does not suffer from this drawback, as H is not required to have finite mean square contingency.

Remark 1.3.3. It is straightforward to modify our test of time reversibility so that it applies to higher-order Markov processes. If $\mathscr{X}$ is an m^{th}-order Markov chain with $m \geq 2$, then we simply take H and H_T to be the distribution function and empirical distribution function of $(X_t, \ldots, X_{t+m})$, and set $\theta = \sup |H_T(x_0, \ldots, x_m) - H_T(x_m, \ldots, x_0)|$. Theorem 1.3.1 then continues to apply, with the limiting distribution in part (a) replaced by $\sup |\mathscr{B}(x_0, \ldots, x_m) - \mathscr{B}(x_m, \ldots, x_0)|$, where $\mathscr{B}$ is now a centered Gaussian process on $\mathbb{R}^{m+1}$ with $\operatorname{cov}(\mathscr{B}(x_0, \ldots, x_m), \mathscr{B}(x'_0, \ldots, x'_m))$ given by

$$\sum_{t \in \mathbb{Z}} \operatorname{cov}\left(1(X_0 \leq x_0, \ldots, X_m \leq x_m), 1(X_t \leq x'_0, \ldots, X_{t+m} \leq x'_m)\right).$$

1.3.2 Local bootstrap critical values

A difficulty in implementing the test just described is that the law of the process $\mathscr{B}$, and therefore the null limiting distribution of $T^{1/2}\theta_T$ given in Theorem 1.3.1(a), is unknown. We may nevertheless approximate these laws using a bootstrap procedure. Here we propose to apply the local bootstrap of Paparoditis and Politis (2002), a resampling scheme designed specifically for Markovian time series.[5] Further discussion of the local bootstrap may be found in Paparoditis and Politis (1998, 2001).

The local bootstrap may be applied in the following way. We wish to draw a bootstrap sample $X_1^*, \ldots, X_T^*$ based on the observed sample $X_1, \ldots, X_T$. (Strictly speaking we should write $X_{1,T}^*, \ldots, X_{T,T}^*$ for the bootstrap sample, as each bootstrap observation depends on the full sample $X_1, \ldots, X_T$, but we will ignore this notational detail outside of the Appendix.) Suppose for the moment that we have already drawn $X_1^*, \ldots, X_t^*$ for some $t \in \{1, \ldots, T-1\}$. For the $(t+1)^{\text{th}}$ bootstrap observation we set $X_{t+1}^* = X_{J+1}$, where J is a discrete random variable drawn

[5] An alternative possibility would be to use the tapered block multiplier bootstrap of Bucher and Ruppert (2012). This resampling scheme is a time series extension of the multiplier bootstrap used by Scaillet (2005) to test for positive quadrant dependence, and by Rémillard and Scaillet (2009) to test for the equality of copulas. However, the tapered block multiplier bootstrap is not intended to exploit the Markovian structure of $\mathscr{X}$, so we focus here on the local bootstrap.

from the probability mass function

$$P(J = j) = \frac{W_b(X_t^* - X_j)}{\sum_{i=1}^{T-1} W_b(X_t^* - X_i)}, \quad j = 1, \ldots, T-1.$$

Here, $b = b_T$ is a bandwidth parameter, W is a kernel function, and $W_b(\cdot) = b^{-1}W(\cdot/b)$. Our initial bootstrap observation X_1^* is drawn at random from the entire sample $X_1, \ldots, X_T$, with equal probability assigned to each observation. Recursive application of the procedure just described yields the bootstrap sample $X_1^*, \ldots, X_T^*$. Paparoditis and Politis (2002, pp. 314–316) provide some guidelines for the data-based selection of b, which we shall not repeat here.

The idea behind the local bootstrap is that the probability of drawing a particular observation from our sample will be relatively greater if the preceding observation is relatively closer to the most recently drawn bootstrap observation. Given X_t^*, the kernel weights governing the behavior of the random variable J direct us to an observation X_J that is likely to be relatively close to X_t^*, and then we select X_{J+1} as our next bootstrap draw X_{t+1}^*. This has the effect of implicitly estimating the transition probabilities governing $\mathscr{X}$, while restricting the state space of the bootstrap sample to the values taken by the observed sample. For large sample sizes, the transition probabilities governing the bootstrap draws will mimic those governing the underlying process $\mathscr{X}$. Radulović (2002) provides a helpful discussion of bootstrap techniques for Markov chains and other dependent processes, with many additional references.

We wish to use the local bootstrap to approximate the law of the limiting process $\mathscr{B}$. This may be done as follows. Let H_T^* denote the bootstrap analogue to H_T computed from our bootstrap sample:

$$H_T^*(x, y) = \frac{1}{T-1} \sum_{t=1}^{T-1} 1(X_t^* \leq x, X_{t+1}^* \leq y).$$

Let E_T^* denote the expectation operator conditional on the observed sample $X_1, \ldots, X_T$; this is the "bootstrap expectation". Our bootstrap version of the process $\mathscr{B}$ is given by

$$\mathscr{B}_T^*(x,y) = T^{1/2}\left(H_T^*(x,y) - E_T^* H_T^*(x,y)\right).$$

In practice, $E_T^* H_T^*(x,y)$ is computed as the average value of $H_T^*(x,y)$ over a large number of bootstrap samples. This is a little more involved than in the case of the iid bootstrap, where we would simply have $E_T^* H_T^*(x,y) = H_T(x,y)$.

We will demonstrate shortly that the bootstrap distribution (i.e., the distribution conditional on the observed sample) of $\mathscr{B}_T^*$ approximates the distribution of $\mathscr{B}$ when T is large. Theorem 1.3.1(a) states that the limiting distribution of $T^{1/2}\theta_T$ is the distribution of $\sup_{x,y}|\mathscr{B}(x,y) - \mathscr{B}(y,x)|$. Since this distribution is unknown, to obtain a test with approximate size α, we set our critical value c equal to the $(1-\alpha)$-quantile of the bootstrap distribution of $\sup_{x,y}|\mathscr{B}_T^*(x,y) - \mathscr{B}_T^*(y,x)|$. This quantile is calculated in practice by generating a large number of bootstrap processes $\mathscr{B}_T^*$, calculating $\sup_{x,y}|\mathscr{B}_T^*(x,y) - \mathscr{B}_T^*(y,x)|$ for each of them, and then selecting the appropriate order statistic.

Let $\mathscr{L}_T^*(\mathscr{B}_T^*)$ denote the distribution of $\mathscr{B}_T^*$, as an element of $\ell^\infty(\mathbb{R}^2)$, conditional on $X_1, \ldots, X_T$. Here, $\ell^\infty(\mathbb{R}^2)$ denotes the space of bounded real valued functions on $\mathbb{R}^2$, equipped with the uniform metric. $\mathscr{L}_T^*(\mathscr{B}_T^*)$ can be thought of as the "bootstrap distribution" or "bootstrap law" of $\mathscr{B}_T^*$. The following result demonstrates that, under regularity conditions imposed by Paparoditis and Politis (2002), $\mathscr{L}_T^*(\mathscr{B}_T^*)$ approximates the distribution of $\mathscr{B}$ when T is large. Note that this result potentially extends the applicability of the local bootstrap to a much wider range of inferential problems than the time reversibility test considered here. The symbol $\rightsquigarrow$ denotes weak convergence in some metric space; see e.g. van der Vaart and Wellner (1996, Def. 1.3.3).

Lemma 1.3.1. *Under Assumption 1.7.1, as $T \to \infty$ we have $\mathscr{L}_T^*(\mathscr{B}_T^*) \rightsquigarrow \mathscr{B}$ in $\ell^\infty(\mathbb{R}^2)$, with probability one.*

Assumption 1.7.1 may be found in the Appendix, and consists of technical conditions used by Paparoditis and Politis (2002) to establish desirable properties of the local bootstrap procedure. These conditions are not intended to be necessary, and indeed Paparoditis and Politis (2002, Remark 3.2) discuss one direction in which they may be relaxed. Our proof of Lemma 1.3.1, also found in the Appendix, applies Theorem 4.2 of Paparoditis and Politis (2002) to obtain a.s. finite dimensional (fidi) convergence of $\mathscr{B}_T^*$ to $\mathscr{B}$, and Theorem 2.2 of Andrews and Pollard (1994) to establish a.s. stochastic equicontinuity of the sequence of bootstrap processes. Let $\mathscr{L}_T^*(\sup_{x,y} |\mathscr{B}_T^*(x,y) - \mathscr{B}_T^*(y,x)|)$ denote the distribution of $\sup_{x,y} |\mathscr{B}_T^*(x,y) - \mathscr{B}_T^*(y,x)|$ conditional on $X_1, \ldots, X_T$; i.e., its bootstrap distribution. We proposed earlier to approximate the limiting distribution of $T^{1/2}\theta_T$, given in Theorem 1.3.1(a), by $\mathscr{L}_T^*(\sup_{x,y} |\mathscr{B}_T^*(x,y) - \mathscr{B}_T^*(y,x)|)$. The following result justifies this approach.

Theorem 1.3.2. *Under Assumption 1.7.1, for any $c \in \mathbb{R}$ we have*

$$P\left(\sup_{x,y} |\mathscr{B}_T^*(x,y) - \mathscr{B}_T^*(y,x)| > c \,\middle|\, X_1, \ldots, X_T\right) \to P\left(\sup_{x,y} |\mathscr{B}(x,y) - \mathscr{B}(y,x)| > c\right)$$

as $T \to \infty$, with probability one.

Theorem 1.3.2 indicates that, given a critical value c, we may use the local bootstrap to consistently estimate the pointwise asymptotic size of our test. Conversely, we may use the local bootstrap to obtain a critical value c for our test that delivers a given pointwise asymptotic size. The proof of Theorem 1.3.2, found in the Appendix, is a straightforward application of Lemma 1.3.1 and the continuous mapping theorem.

1.3.3 Finite sample performance

Here we report some numerical evidence pertaining to the finite sample performance of our proposed test of time reversibility. We consider two families of bivariate distributions H, each indexed by a single parameter. The first choice of H

is the asymmetric Gumbel copula given in (2.1). We fix $\alpha = 1$, $\beta = 0.5$, and let γ vary over the interval $[1, \infty)$. When $\gamma = 1$, the asymmetric Gumbel copula reduces to the product copula, and so $\mathscr{X}$ is time reversible. $\mathscr{X}$ is time irreversible when $\gamma > 1$, becoming more irreversible as γ increases.

We calculated the rejection rate of our time reversibility test for a range of values of γ, with sample size $T = 150$. For the purpose of comparison, we also calculated rejection rates for three other tests of time reversibility. The first test, proposed by Paparoditis and Politis (2002), is based on the statistic

$$R_T^{\mathrm{PP}} = \frac{1}{T-1} \sum_{t=1}^{T-1} 1\left(X_{t+1} > X_t\right) - \frac{1}{2},$$

the fraction of differenced observations that are positive, minus one half. The second test, proposed by Ramsey and Rothman (1996), is based on the statistic

$$R_T^{\mathrm{RR}} = \frac{1}{T-1} \sum_{t=1}^{T-1} X_{t+1}^2 X_t - \frac{1}{T-1} \sum_{t=1}^{T-1} X_{t+1} X_t^2,$$

the difference of sample bicovariances. The third test, proposed by Chen et al. (2000), is based on the statistic

$$R_T^{\mathrm{CCK}} = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{X_{t+1} - X_t}{1 + \left(X_{t+1} - X_t\right)^2}.$$

As discussed by Chen et al., R_T^{CCK} provides an estimate of $\int_0^\infty E \sin(\omega(X_1 - X_0))g(\omega)\mathrm{d}\omega$, where the weighting function g is taken to be the exponential density with unit decay rate. An earlier study of the finite sample properties of the R_T^{RR} and R_T^{CCK} statistics was reported by Belaire-Franch and Contreras (2004).

For all four statistics θ_T, R_T^{PP}, R_T^{RR} and R_T^{CCK}, we constructed critical values with nominal size 0.05 using the local bootstrap. The local bootstrap was implemented using a Gaussian kernel for W, and smoothing parameter b determined using the data dependent selection rule described by Paparoditis and Politis (2002,

p. 315), with plug-in parameters extracted from an auxiliary first-order autoregression. We employed 400 bootstrap replications. Note that, since our critical values for R_T^{RR} and R_T^{CCK} are determined using the local bootstrap rather than a standard normal approximation, we have dropped the scaling factors used by Ramsey and Rothman (1996) and Chen et al. (2000) to endow their statistics with unit asymptotic variance. This saves us the additional inconvenience of long-run variance estimation.

Our numerical calculations using the asymmetric Gumbel[6] copula are displayed in Figure 1.3(a). The horizontal axis tracks the value of $1 - 1/\gamma$, so we have $\mathscr{X}$ time reversible at the left endpoint of the axis, and increasingly irreversible as we move rightward. All four tests exhibit good size control. As $\gamma \to \infty$, the rejection rates for all tests rise to approximately one. At intermediate values of γ, the test of Paparoditis and Politis generally has the highest power, followed by the test of Ramsey and Rothman, then the test of Chen et al., and finally our own test with the lowest power.

In Remark 1.3.1 we noted that several existing tests of time reversibility are unable to detect forms of irreversibility for which $X_{t+1} - X_t$ is distributed symmetrically about zero. Our second choice of H exploits this fact. We take H to be a convex combination of two copula functions. The first of these is the product copula. The second distributes mass uniformly over the shaded area in Figure 1.2. We assign weight $1 - \lambda$ to the first copula and λ to the second, with $\lambda \in [0, 1]$. Thus $\mathscr{X}$ is time reversible when $\lambda = 0$ and time irreversible when $\lambda > 0$, becoming more irreversible as λ increases. For reasons that will be made clear in Section 1.4.1, we refer to this mixture copula as a zero total circulation copula.

Our numerical calculations using the zero total circulation copula are displayed in Figure 1.3(b). The horizontal axis tracks the value of λ, so we have $\mathscr{X}$ time reversible at the left endpoint of the axis, and increasingly irreversible as we move rightward. All four tests exhibit good size control. As λ increases, the behav-

[6]Results qualitatively similar to those presented in Figure 1.3(a) were obtained using the Clayton and Frank copulas in place of the Gumbel copula.

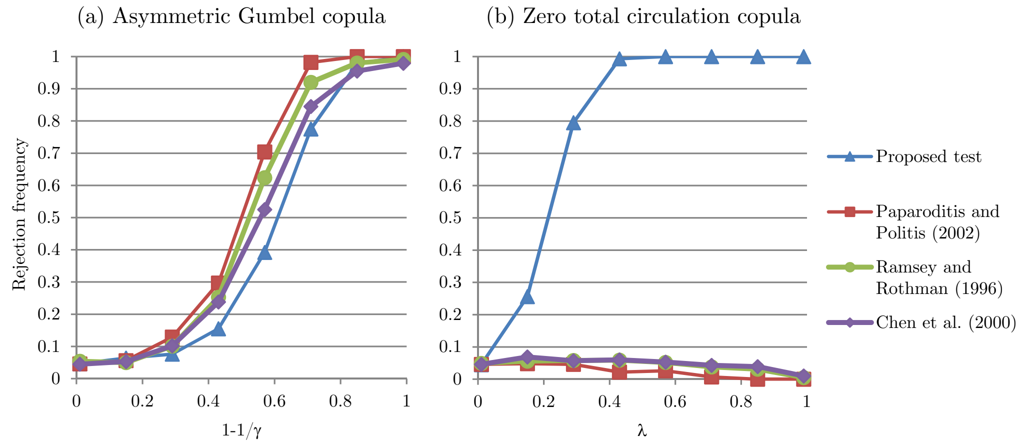


Figure 1.3: Power of the tests
Rejection rates of our time reversibility test, and the tests of Paparoditis and Politis (2002), Ramsey and Rothman (1996) and Chen et al. (2000). Panel (a) displays results for the asymmetric Gumbel copula with $\alpha = 1$, $\beta = 0.5$, and $\gamma \in [1, \infty)$. Panel (b) displays results for a convex combination of the product copula and the copula displayed in Figure 1.2; the weight on the latter is $\lambda \in [0, 1]$. We set $T = 150$ and employed 400 bootstrap replications and 1000 experimental replications. Nominal size is 0.05.

ior of our test is very different to that of the other three tests. The rejection rate of our test rises quickly to one, while the rejection rates of the other tests fall to zero.

Panels (a) and (b) of Figure 1.3 serve to illustrate both the strength and weakness of our approach to testing time reversibility. The key advantage of our test is that, unlike existing tests, it consistently rejects in the presence of any violation of time reversibility. This versatility comes at a price: tests that target specific forms of time irreversibility are likely to be more powerful than our test when irreversibility is indeed of that form. Though we are not aware of any economic models that generate time irreversible dynamics with symmetrically distributed differences, the dynamic behavior of observed time series frequently depart from the predictions of theoretical models. Used as part of a battery of tests, a test that is consistent against arbitrary violations of reversibility provides a safeguard against irreversibility of unexpected form. For instance, in the empirical application reported in Section 1.5, the tests of Ramsey and Rothman, and Chen et al. fail to reject the null of time reversibility, while our own test provides strong evidence of irreversibility. The test of Paparoditis and Politis also identifies irreversibility in this case, but lacks power when applied to irreversible processes with zero median differences.

1.4 Characterizing time irreversibility

In this section we consider a characterization of time irreversibility that may be useful for applications. Building on work by McCausland (2007), we define the *circulation density* for a stationary real valued Markov chain. The circulation density quantifies the net probability upflow at each quantile of the invariant distribution. Visual inspection of the circulation density, a real valued function on the unit interval, provides a convenient way to assess the nature of time irreversibility in a Markov chain.

The circulation density is defined and explained in Section 1.4.1. In Section 1.4.2 we propose a simple copula-based estimator of the circulation density, and investigate its asymptotic and finite sample behavior.

1.4.1 Circulatory analysis of stationary Markov chains

McCausland (2007) introduced the notion of circulation for stationary Markov chains with finite state space. Circulation is intended to measure the direction and intensity of the flow of probability through each state. If a Markov chain is time reversible, then we must necessarily have zero circulation through each state. If it is time irreversible, then the circulation through each state provides information about the nature of that irreversibility. In this section we propose a definition of circulation that is similar in spirit to the definition given by McCausland, but which applies in a natural way when the invariant distribution of $\mathscr{X}$ may not be discrete. We demonstrate a connection between the circulation of $\mathscr{X}$ and the copula function C characterizing its dynamic dependence. At the end of the section we explain how our treatment of circulation builds on McCausland's contribution.

To describe the circulatory behavior of $\mathscr{X}$, we introduce a number of functions from $\mathbb{R}$ to $[0,1]$ which we refer to as flows. The two fundamental flows, denoted $\mathscr{F}_\uparrow$ and $\mathscr{F}_\downarrow$, are defined and referred to as follows.

$$\begin{aligned}
\mathscr{F}_\uparrow(x) &= P(X_{t-1} \le x | X_t = x) && \textit{probability upflow to } x \\
\mathscr{F}_\downarrow(x) &= P(X_{t+1} \le x | X_t = x) && \textit{probability downflow from } x.
\end{aligned}$$

Two additional flows, $\mathscr{F}^\uparrow$ and $\mathscr{F}^\downarrow$, are uniquely determined by the two fundamental flows:

$$\begin{aligned}
\mathscr{F}^\uparrow(x) &= P(X_{t+1} > x | X_t = x) && \textit{probability upflow from } x \\
\mathscr{F}^\downarrow(x) &= P(X_{t-1} > x | X_t = x) && \textit{probability downflow to } x.
\end{aligned}$$

By the law of total probability, our four flows satisfy the identities

$$\mathscr{F}_\uparrow(x) + \mathscr{F}^\downarrow(x) = 1, \quad \mathscr{F}^\uparrow(x) + \mathscr{F}_\downarrow(x) = 1. \tag{1.4}$$

The terms upflow and downflow are evocative of the circulation, or current, of a body of water. Figure 1.4 displays our four flows as arrows pointing toward, or away from, x. Suppose we know that $X_t = x$. The two arrows pointing toward x represent the probabilities that X_{t-1} was less than, or greater than, x. The two arrows pointing away from x represent the probabilities that X_{t+1} will be less than, or greater than, x.

Strictly speaking, conditional probabilities like $P(X_{t+1} \le x | X_t = x)$ are not uniquely defined when F is continuous at x, because we are conditioning on a set of measure zero. Rather, $P(X_{t+1} \le x | X_t = x)$ should be viewed as an equivalence class of functions of x, where any two members of the class must be equal to one another outside a set of F-measure zero. Likewise, the flows $\mathscr{F}_\uparrow(x)$, $\mathscr{F}^\uparrow(x)$, $\mathscr{F}^\downarrow(x)$ and $\mathscr{F}_\downarrow(x)$ should be viewed as being uniquely defined up to a set of F-measure zero. For further discussion of technical issues associated with conditional probabilities of this kind, we refer the reader to Chang and Pollard

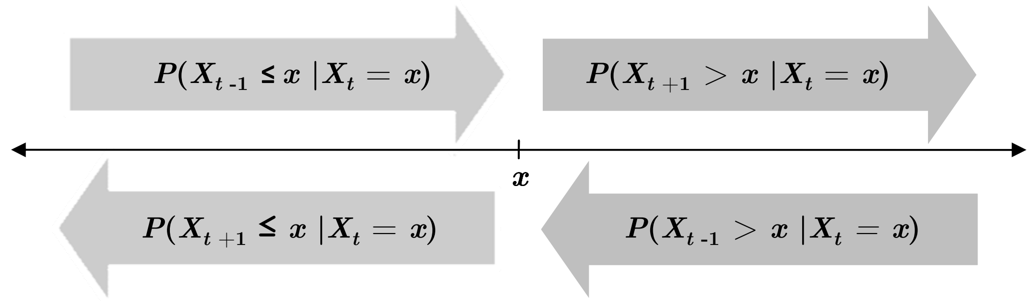


Figure 1.4: Probability upflows and downflows
The circulation density at $u = F(x)$ is equal to the sum of the upward flows minus the sum of the downward flows, divided by two.

(1997).

It may be helpful to introduce some additional terminology to describe certain combinations of our four flows $\mathscr{F}_\uparrow$, $\mathscr{F}^\uparrow$, $\mathscr{F}^\downarrow$ and $\mathscr{F}_\downarrow$:

$\mathscr{F}_\uparrow(x) + \mathscr{F}^\uparrow(x)$	*probability upflow through* x
$\mathscr{F}^\downarrow(x) + \mathscr{F}_\downarrow(x)$	*probability downflow through* x
$\mathscr{F}_\uparrow(x) - \mathscr{F}_\downarrow(x)$	*net probability upflow to* x
$\mathscr{F}^\uparrow(x) - \mathscr{F}^\downarrow(x)$	*net probability upflow from* x
$\mathscr{F}_\uparrow(x) + \mathscr{F}^\uparrow(x) - \mathscr{F}^\downarrow(x) - \mathscr{F}_\downarrow(x)$	*net probability upflow through* x.

A consequence of the identities in (1.4) is that the net probability upflow to x is equal to the net probability upflow from x, which is equal to half the net probability upflow through x. If $\mathscr{X}$ is time reversible, then the flows $\mathscr{F}_\uparrow$, $\mathscr{F}^\uparrow$, $\mathscr{F}^\downarrow$ and $\mathscr{F}_\downarrow$ satisfy two additional identities:

$$\mathscr{F}_\uparrow(x) = \mathscr{F}_\downarrow(x), \quad \mathscr{F}^\uparrow(x) = \mathscr{F}^\downarrow(x).$$

Thus, when $\mathscr{X}$ is time reversible, the net probability upflows to, from, and through x are all equal to zero.

Given $u \in (0,1)$, let $Q(u) = \inf\{y : F(y) \ge u\}$, the u-quantile of the invariant distribution F. We define the *circulation density* of $\mathscr{X}$ to be the function $\psi : (0,1) \to [-1,1]$ given by

$$\psi(u) = \frac{1}{2}\left(\mathscr{F}_{\uparrow}(Q(u)) + \mathscr{F}^{\uparrow}(Q(u)) - \mathscr{F}^{\downarrow}(Q(u)) - \mathscr{F}_{\downarrow}(Q(u))\right), \quad u \in (0,1).$$

That is, $\psi(u)$ is one half of the net probability upflow through $Q(u)$. The circulation density tells us whether, at a given quantile of the invariant distribution, observations tend to be in the middle of an upward or downward string of three observations. If the density is positive, an observation at that quantile is relatively likely to be part of an increasing string, whereas if the density is negative, the observation is more likely to be part of a decreasing string.

As noted earlier, our flows $\mathscr{F}_{\uparrow}$, $\mathscr{F}^{\uparrow}$, $\mathscr{F}^{\downarrow}$ and $\mathscr{F}_{\downarrow}$ are uniquely defined only up to a set of F-measure zero. Consequently, our circulation density $\psi(u)$ may not be uniquely defined for all $u \in (0,1)$. Rather, $\psi(u)$ is uniquely defined up to a set $A \subset (0,1)$, where $A = \{u : Q(u) \in B\}$ for some set $B \subset \mathbb{R}$ of zero F-measure. Since the F-measure of B is precisely the Lebesgue measure of A, we find that $\psi(u)$ is uniquely defined up to a set of u having zero Lebesgue measure. When the invariant distribution of $\mathscr{X}$ is discrete, so that F is a step function, we find that A is empty for any B of zero F-measure, and so $\psi(u)$ is in fact uniquely defined for all $u \in (0,1)$.

Theorem 1.4.1 demonstrates that, under additional smoothness conditions, our circulation density ψ may be expressed in terms of the copula function C describing the dynamic dependence structure of $\mathscr{X}$. More specifically, ψ is the difference between the first partial derivatives of C along the main diagonal of the unit square. The proof of Theorem 1.4.1 may be found in the Appendix.

Theorem 1.4.1. *Let $\mathscr{X}$ be a stationary real valued Markov chain with continuous invariant distribution F, and copula C admitting continuous partial derivatives*

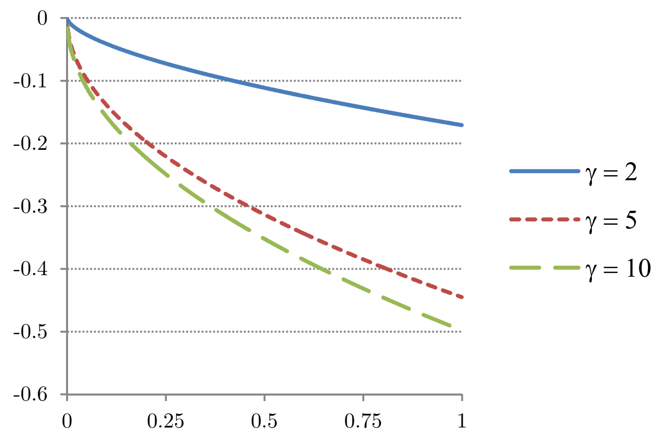


Figure 1.5: Circulation densities for the asymmetric Gumbel copula
Circulation densities for the asymmetric Gumbel copula with $\alpha = 1$, $\beta = 0.5$, and $\gamma = 2, 5, 10$.

$\partial_1 C$ and $\partial_2 C$ everywhere on $(0,1)^2$. Then the circulation density ψ of $\mathscr{X}$ satisfies

$$\psi(u) = \partial_2 C(u,u) - \partial_1 C(u,u)$$

for Lebesgue-a.e. $u \in (0,1)$.

In Figure 1.5 we use the expression for $\psi(u)$ given in Theorem 1.4.1 to graph the circulation density functions corresponding to the asymmetric Gumbel copula given in (2.1), with $\alpha = 1$, $\beta = 0.5$, and $\gamma = 2, 5, 10$. These are the same parameter configurations used to generate the scatterplots and Markov sample paths in Figure 1.1. In each case we see that $\psi(u)$ is negative for all $u \in (0,1)$, indicating a net probability downflow at all quantiles. We also see that $\psi(u)$ is monotone decreasing in each case, rising to zero as $u \downarrow 0$. This is consistent with the pattern of dependence evident in Figure 1.1, where we see many small decreases and occasional large increases – at least when $\gamma = 5, 10$ – with the likelihood of an increase rising as we approach the bottom of the state space. Note that if we were to exchange the values of α and β, the effect would be to multiply each circulation density by -1.

The circulation density tells us whether, at a particular quantile of the invariant distribution, our Markov chain tends to be increasing or decreasing. Integrating the circulation density over the unit interval gives us a single index of circulation, $\Psi = \int_0^1 \psi(u)\mathrm{d}u$. We refer to Ψ as the *total circulation* of $\mathscr{X}$. The following result shows that, defined in this way, the total circulation has a convenient interpretation. The proof may be found in the Appendix.

Theorem 1.4.2. *Let $\mathscr{X}$ be a stationary real valued Markov chain. Then Ψ, the total circulation of $\mathscr{X}$, satisfies $\Psi = P(X_{t-1} \leq X_t) - P(X_{t+1} \leq X_t)$.*

Theorem 1.4.2 reveals that the total circulation measures the overall tendency of $\mathscr{X}$ to increase more frequently than it decreases, or vice-versa. If increases and decreases are equally likely, the total circulation is zero. The circulation density serves to decompose the total circulation into contributions from different quantiles of the invariant distribution. In this sense, it plays a similar role to the spectral density of a covariance stationary process, which decomposes the variance into contributions from cycles of different frequency.

A stationary Markov chain with zero total circulation is not necessarily time reversible. For instance, the copula used to construct the power curves in Figure 1.3(b) generates a time irreversible stationary Markov chain with zero total circulation. In fact, even when a stationary Markov chain has zero circulation density at all quantiles, time reversibility does not necessarily hold. In Figure 1.6 we provide an example of a copula function that generates a time irreversible Markov chain having zero circulation density at all quantiles. This copula function should be understood to distribute mass uniformly over the shaded region. Clearly the shaded region is not symmetric about the $45°$-line, implying that the associated Markov chain $\mathscr{X}$ is time irreversible. The probability upflow to u is equal to the length of the solid part of the line extending between $(0, u)$ and (u, u), while the probability downflow from u is equal to the length of the solid part of the line extending between $(u, 0)$ and (u, u). Careful inspection of Figure 1.6 reveals that these two quantities are

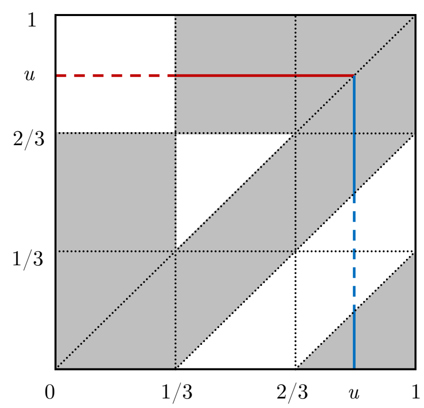


Figure 1.6: Magic square 2
If (X_t, X_{t+1}) is distributed uniformly over the shaded region, then $\mathscr{X}$ is time irreversible, yet has zero circulation density at all quantiles. The probability upflow to u is equal to the length of the solid part of the line extending between $(0, u)$ and (u, u), while the probability downflow from u is equal to the length of the solid part of the line extending between $(u, 0)$ and (u, u).

equal to one another, and continue to be equal for any choice of $u \in (0, 1)$. Thus we find that the circulation density of $\mathscr{X}$ is zero at all quantiles.

Our discussion of circulation in this section has built on prior work by McCausland (2007) for Markov chains with discrete state space. Suppose our stationary real valued Markov chain $\mathscr{X}$ takes only the values $x_1, \ldots, x_n \in \mathbb{R}$. McCausland defined the *circulation through* x_i to be the quantity

$$\frac{1}{2}\left(P(X_t = x_i \text{ and } X_{t+1} > x_i) - P(X_{t-1} > x_i \text{ and } X_t = x_i)\right).$$

With some elementary manipulations, we may rewrite this expression as

$$\frac{1}{4}P(X_t = x_i)\left(\mathscr{F}_{\uparrow}(x_i) + \mathscr{F}^{\uparrow}(x_i) - \mathscr{F}^{\downarrow}(x_i) - \mathscr{F}_{\downarrow}(x_i)\right).$$

Thus, McCausland's circulation through x_i is one quarter of the net probability upflow through x_i, multiplied by the probability assigned by the invariant distribution to x_i. By comparison, as defined here, the circulation density at quantiles corresponding to x_i is half the net probability upflow through x_i, which differs from McCausland's circulation through x_i by a factor of $\frac{1}{2}P(X_t = x_i)$. Dropping the factor $P(X_t = x_i)$ makes sense here because we wish to allow the invariant distribution to be continuous, while dropping the factor of one half appears natural in view of Theorem 1.4.1 and Theorem 1.4.2. The notion of total circulation was also introduced by McCausland, who defined it as half the difference between $P(X_{t-1} \leq X_t)$ and $P(X_{t+1} \leq X_t)$, and showed that this quantity is equal to the sum of state-specific circulations. Theorem 1.4.2 makes it clear that our own definition of total circulation differs from McCausland's definition by a factor of one half.

1.4.2 Estimation of the circulation density

The circulation density function provides a convenient way to quickly assess the nature of time irreversibility in a Markov chain. In this section we consider estimating the circulation density from data. We propose an estimator based on a kernel smoothed version of the empirical copula function, establish its pointwise asymptotic behavior, and assess the finite sample performance of associated inferential procedures using Monte Carlo simulation.

Estimator and asymptotic properties

Theorem 1.4.1 established that, under mild regularity conditions, the circulation density of $\mathscr{X}$ is given by the difference between the partial derivatives of C along the diagonal of the unit square. A natural estimator for the circulation density may therefore be extracted from the partial derivatives of a smooth estimate of C. Let k be a kernel function, let h be a bandwidth parameter, and, for $x \in \mathbb{R}$, let $k_h(x) = h^{-1}k(x/h)$ and $K_h(x) = \int_{-\infty}^{x} k_h(y)\mathrm{d}y$. Given an observed sample

$X_1, \ldots, X_T$, we may construct smooth estimates of H, F, Q and C as follows:

$$\begin{aligned}
\hat{H}_T(x,y) &= \frac{1}{T-1}\sum_{t=1}^{T-1} K_h\left(x - X_t\right) K_h\left(y - X_{t+1}\right) \\
\hat{F}_T(x) &= \frac{1}{T}\sum_{t=1}^{T} K_h\left(x - X_t\right) \\
\hat{Q}_T(u) &= \inf\{y \in \mathbb{R} : \hat{F}_T(y) \geq u\} \\
\hat{C}_T(u,v) &= \hat{H}_T\left(\hat{Q}_T(u), \hat{Q}_T(v)\right).
\end{aligned}$$

A simple nonparametric estimator of ψ is then given by[7]

$$\hat{\psi}_T(u) = \partial_2 \hat{C}_T(u,u) - \partial_1 \hat{C}_T(u,u).$$

Of course, $\hat{\psi}_T$ is not the only possible estimator of ψ. Rémillard and Scaillet (2009) proposed very simple estimators of $\partial_1 C$ and $\partial_2 C$ that could be used to form a uniformly consistent estimator of ψ. But our estimator $\hat{\psi}_T$ appears reasonable when we expect that C and F are smooth. We will establish the pointwise asymptotic properties of $\hat{\psi}_T$ under the following technical conditions.

Assumption 1.4.1. The following statements are true.

(a) $\mathscr{X}$ is a stationary real valued Markov chain.

(b) F is four times continuously differentiable, and C admits continuous mixed partial derivatives to the fourth order.

(c) The α-mixing coefficients of $\mathscr{X}$ satisfy $\alpha_T = O(T^{-\eta})$ for some $\eta > 2$.

(d) The kernel k integrates to one, is even, has compact support, and is four times continuously differentiable.

[7] Our approach to estimating the partial derivatives of C is taken from Fermanian and Scaillet (2003). Rémillard and Scaillet (2009) proposed an alternative estimator formed by differencing the empirical copula function that may also be applicable in the present context.

(e) The bandwidth $h = h_T$ satisfies $Th^3 \to \infty$ and $Th^4 \to c$ for some $c \in [0, \infty)$.

Parts (a,b,c) of Assumption 1.4.1 may be compared to the corresponding parts of Assumption 1.3.1. Note that (b) ensures that H admits continuous mixed partial derivatives to the fourth order. The compact support condition imposed on k in Assumption 1.4.1(d) is mathematically convenient, but may perhaps be replaced by a condition on the rate at which the tails of k decay to zero. Assumption 1.4.1(e) provides the admissible rates of decay for the bandwidth h. It seems likely that the condition $Th^4 \to c \in [0, \infty)$ could be weakened to $Th^5 \to 0$, but we do not pursue this extension here, as it leads to complications in the proof of Lemma 1.7.3 in the Appendix.

Theorem 1.4.3 establishes the asymptotic normality of $\hat{\psi}_T(u)$, giving the asymptotic variance $\sigma^2(u)$ in terms of k, C, Q, and the invariant pdf $f = F'$. A consistent estimator of $\sigma^2(u)$ is provided. In the statement of Theorem 1.4.3, and in its proof, we define $\psi(u) = \partial_2 C(u,u) - \partial_1 C(u,u)$ to avoid ambiguity about the values taken by ψ on sets of Lebesgue measure zero.

Theorem 1.4.3. *Suppose $\mathscr{X}$ satisfies Assumption 1.4.1. Then, for any $u \in (0,1)$ such that $f(Q(u)) > 0$, we have*

$$(Th)^{1/2} \left(\hat{\psi}_T(u) - \psi(u) \right) \to_d N \left(0, \sigma^2(u) \right),$$

where

$$\sigma^2(u) = \frac{\int k(z)^2 dz}{f\left(Q(u)\right)} \cdot \left(\partial_1 C(u,u) \left(1 - \partial_1 C(u,u)\right) + \partial_2 C(u,u) \left(1 - \partial_2 C(u,u)\right) \right).$$

The limiting variance $\sigma^2(u)$ may be consistently estimated by

$$\hat{\sigma}_T^2(u) = \frac{\int k(z)^2 dz}{\hat{f}_T\left(\hat{Q}_T(u)\right)} \cdot \left(\partial_1 \hat{C}_T(u,u) \left(1 - \partial_1 \hat{C}_T(u,u)\right) + \partial_2 \hat{C}_T(u,u) \left(1 - \partial_2 \hat{C}_T(u,u)\right) \right),$$

where $\hat{f}_T = \hat{F}_T'$.

Nonnegativity of the limiting variance $\sigma^2(u)$ appearing in Theorem 1.4.3 follows from the fact that $0 \leq \partial_i C \leq 1$ for $i = 1, 2$; see e.g. Nelsen (2006, Theorem 2.2.7). We may rule out the possibility that $\sigma^2(u) = 0$ if we assume that $0 < \partial_i C(u, u) < 1$ for $i = 1, 2$. If $\sigma^2(u) > 0$, Theorem 1.4.3 can be used to construct pointwise asymptotic confidence bands for $\psi(u)$. Alternatively, the local bootstrap of Paparoditis and Politis (2002) or the robust t-statistic based method of Ibragimov and Müller (2010) can be used to construct confidence bands. We investigate the performance of these approaches in the finite sample simulations reported in Section 1.4.2.

Our proof of Theorem 1.4.3, which may be found in the Appendix, adapts methods employed by Fermanian and Scaillet (2003). Those authors seek to find the joint asymptotic behavior of a single mixed partial derivative of $\hat{C}_T$ evaluated at multiple points in the unit square. Here, our concern is with the joint asymptotic behavior of the two first partial derivatives of $\hat{C}_T$ evaluated at a single point on the main diagonal of the unit square. The application of a result due to Robinson (1983), used also by Fermanian and Scaillet (2003), is central to our argument.

Finite sample performance

Here we report some limited numerical evidence concerning the finite sample performance of confidence bands for the circulation density. Our first set of results pertains to confidence bands formed using the local bootstrap. We consider two sample sizes, $T = 75$ and $T = 150$, and four copula functions C: the product, Gumbel, Clayton and Frank copulas, with parameters for the latter three copulas chosen to make Kendall's rank correlation coefficient equal to 0.5. The associated (lower,upper) tail dependence coefficients are, respectively, $(0, 0)$, $(0, 2 - 2^{1/2})$, $(2^{-1/2}, 0)$ and $(0, 0)$. The invariant distribution F was taken to be standard normal in all cases. For each choice of T and C we computed the circulation density estimator $\hat{\psi}_T(u)$ at the quantiles $u = 0.1, 0.3, 0.5, 0.7, 0.9$. Pointwise nominal 80%, 90% and 95% confidence bands for each circulation density estimate were computed using the local bootstrap with 400 bootstrap replications. For each quantile, we calculated the coverage rate of each confidence band over 1000 randomly generated samples.

Implementation of the circulation density estimator and local bootstrap requires us to choose kernel functions k and W and bandwidth parameters h and b. Both kernels were taken to be Gaussian. For the local bootstrap bandwidth parameter b we used the data dependent selection rule described by Paparoditis and Politis (2002, p. 315), with plug-in parameters extracted from an auxiliary first-order autoregression. For the bandwidth parameter h used to construct the circulation density estimator, we set $h = 0.5T^{-1/4}$, in compliance with Assumption 1.4.1(e). This ad hoc choice of h appears to work well for sample sizes of practical relevance.

The results of our experiment are provided in Table 1.1. For all sample sizes T, copulas C, and quantiles u, the coverage probabilities of our pointwise confidence bands were close to the nominal rates. This suggests that, for the processes under consideration, the local bootstrap does a very good job at approximating the sampling uncertainty associated with our circulation density estimator.

Though confidence bands constructed using the local bootstrap appear to perform well in finite samples, they can be slow to compute. Indeed, we would have liked to include results for larger sample sizes in Table 1.1, but found this impractical from a computational perspective. In situations where a faster and simpler approach to constructing confidence bands is desirable, the robust t-statistic based method of Ibragimov and Müller (2010) may provide an attractive alternative to the local bootstrap. To construct confidence bands using the Ibragimov-Müller method, we first divide our data into q blocks of roughly equal size. We then form q estimates of $\psi(u)$ by applying our circulation density estimator to each block of observations separately. An approximate $(1-\alpha)\%$ confidence band for $\psi(u)$ is given by $\mu_q \pm t_{q-1}(1-\alpha/2)q^{-1/2}s_q$, where μ_q and s_q are the sample average and sample standard deviation of the q estimates of $\psi(u)$, and $t_{q-1}(1-\alpha/2)$ is the

Table 1.1: Pointwise coverage rates of confidence bands for the circulation density Pointwise coverage rates of confidence bands for the circulation density constructed using the local bootstrap. Parameters for the Gumbel, Clayton and Frank copulas were chosen to induce a rank correlation of 0.5. We employed 400 bootstrap replications and 1000 experimental replications.

Copula	Sample size	Nominal coverage	Coverage at quantile 0.1	0.3	0.5	0.7	0.9
Product	75	0.95	0.971	0.943	0.953	0.952	0.966
		0.90	0.936	0.888	0.906	0.902	0.917
		0.80	0.813	0.782	0.809	0.786	0.799
	150	0.95	0.964	0.951	0.941	0.948	0.962
		0.90	0.908	0.899	0.886	0.893	0.916
		0.80	0.802	0.808	0.773	0.789	0.797
Gumbel	75	0.95	0.961	0.969	0.970	0.965	0.935
		0.90	0.925	0.920	0.922	0.913	0.878
		0.80	0.809	0.805	0.805	0.830	0.751
	150	0.95	0.956	0.952	0.957	0.954	0.968
		0.90	0.904	0.906	0.917	0.901	0.902
		0.80	0.788	0.793	0.808	0.809	0.789
Clayton	75	0.95	0.943	0.971	0.965	0.968	0.961
		0.90	0.873	0.925	0.910	0.920	0.925
		0.80	0.739	0.827	0.807	0.834	0.817
	150	0.95	0.966	0.948	0.957	0.957	0.966
		0.90	0.924	0.898	0.899	0.905	0.911
		0.80	0.792	0.791	0.798	0.797	0.801
Frank	75	0.95	0.964	0.973	0.964	0.971	0.961
		0.90	0.916	0.929	0.911	0.937	0.901
		0.80	0.794	0.842	0.808	0.815	0.782
	150	0.95	0.962	0.952	0.971	0.963	0.962
		0.90	0.911	0.897	0.914	0.923	0.903
		0.80	0.797	0.794	0.819	0.809	0.794

$(1-\alpha/2)$-quantile of the t-distribution with $q-1$ degrees of freedom.

Numerical results on the finite sample performance of confidence bands constructed using the method of Ibragimov and Müller are provided in Table 1.2. The processes considered are the same as in Table 1.1, but we consider larger sample sizes: $T=150$ and $T=500$. We formed confidence bands using $q=5$ equally sized blocks of observations. Compared to the local bootstrap, the pointwise coverage rates of the Ibragimov-Müller confidence bands tended to deviate a little more from the nominal coverage rates, even at larger sample sizes. Nevertheless, the coverage errors are relatively small, and may be acceptable for practical purposes. In view of the ease with which the Ibragimov-Müller bands may be coded and computed, they should be viewed as a convenient alternative to the local bootstrap bands.

We have not reported coverage rates for confidence bands obtained using the first order asymptotic approximation given in Theorem 1.4.3.[8] Confidence bands constructed in this way tended to be excessively conservative for all copulas considered, even with a sample size as large as $T = 5000$. We recommend that the local bootstrap or Ibragimov-Müller method be used to form confidence bands in practice.

1.5 Empirical illustration

In this section we illustrate the use of our time reversibility test and circulation density estimator by applying them to a time series of weekly gasoline price markups in Windsor, Ontario from August 20, 1989 to September 25, 1994. These markups, displayed in Figure 1.7(a), were calculated by dividing the average retail price across a sample of gasoline stations in Windsor by the wholesale price of large scale purchases of unbranded gasoline at the terminal in Toronto, Ontario. The same data were used by Eckert (2002), who studied the asymmetry of price responses to cost increases and decreases, and by McCausland (2007), who divided the markups into six bins and used Bayesian techniques to estimate the circulation through each bin.

Gasoline price dynamics have attracted considerable attention during the last decade due to the presence of Edgeworth cycles in a substantial proportion of markets. Edgeworth cycles involve extended periods of gradual price reduction,

[8] These results are available on request.

Table 1.2: Coverage rates and mean square errors for the circulation density estimator
Coverage rates and mean square errors for our circulation density estimator, with confidence bands constructed using the Ibragimov-Müller method. Parameters for the Gumbel, Clayton and Frank copulas were chosen to induce a rank correlation of 0.5. We employed $q = 5$ blocks of equal length, and 1000 experimental replications.

Copula	Sample size	Nominal coverage	Coverage at quantile 0.1	0.3	0.5	0.7	0.9
Product	150	0.95	0.984	0.936	0.929	0.941	0.987
		0.90	0.942	0.863	0.859	0.880	0.946
		0.80	0.790	0.753	0.757	0.760	0.802
	500	0.95	0.956	0.948	0.927	0.941	0.954
		0.90	0.886	0.883	0.855	0.877	0.872
		0.80	0.760	0.768	0.745	0.776	0.749
Gumbel	150	0.95	0.975	0.957	0.945	0.945	0.989
		0.90	0.918	0.907	0.882	0.886	0.944
		0.80	0.777	0.779	0.782	0.773	0.806
	500	0.95	0.941	0.934	0.919	0.948	0.943
		0.90	0.865	0.870	0.867	0.902	0.879
		0.80	0.724	0.759	0.757	0.797	0.749
Clayton	150	0.95	0.987	0.958	0.954	0.970	0.982
		0.90	0.958	0.902	0.893	0.908	0.920
		0.80	0.838	0.781	0.778	0.780	0.787
	500	0.95	0.970	0.944	0.948	0.949	0.937
		0.90	0.918	0.889	0.890	0.887	0.870
		0.80	0.805	0.776	0.801	0.770	0.752
Frank	150	0.95	0.977	0.962	0.963	0.957	0.980
		0.90	0.921	0.906	0.911	0.899	0.911
		0.80	0.769	0.775	0.800	0.782	0.781
	500	0.95	0.934	0.925	0.940	0.929	0.943
		0.90	0.876	0.873	0.889	0.874	0.895
		0.80	0.747	0.776	0.772	0.760	0.780

followed by shorter periods of rapid price increase. Game theoretic foundations for Edgeworth cycles were provided by Maskin and Tirole (1988), who showed that Edgeworth price cycles emerge naturally as a Markov perfect equilibrium in a dynamic model of Bertrand competition between two firms. Extensions of this result have been provided by Eckert (2003) and Noel (2008). Other key papers on

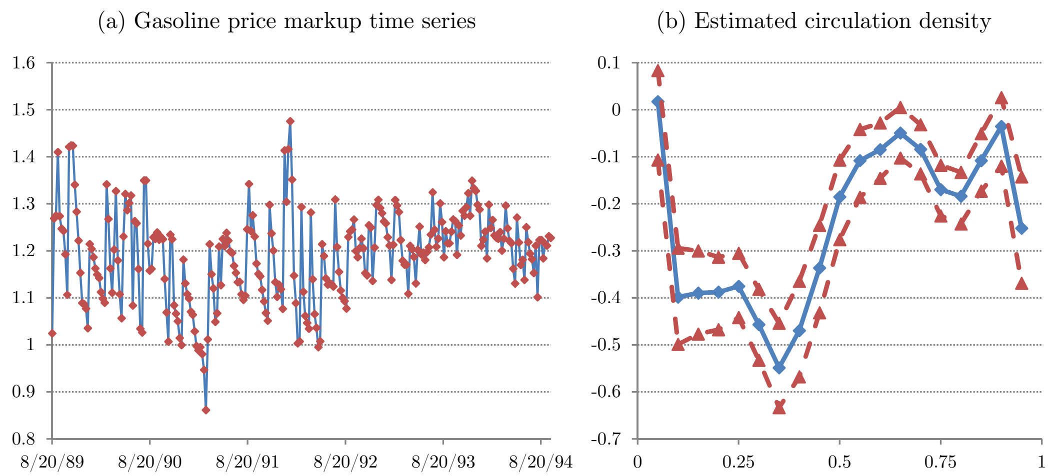


Figure 1.7: Empirical example: gasoline price markups
Panel (a) displays the average weekly gasoline price markups in Windsor, Ontario from 8/20/1989 to 9/25/1994. Panel (b) displays the circulation density estimated using these data, with pointwise 95% confidence bands constructed using the local bootstrap.

Edgeworth cycles in gasoline markets include Noel (2007), Wang (2009) and Lewis and Noel (2011); further references may be found in Noel (2011).

On casual inspection, the time series of price markups in Figure 1.7(a) seems to contain a large number of long decreasing strings of observations, consistent with the presence of Edgeworth cycles. Applying our test of time reversibility to this series yields a p-value of 0.00, indicating overwhelming rejection of reversibility. The test of Paparoditis and Politis (2002) also yields a p-value of 0.00, while the tests of Ramsey and Rothman (1996) and Chen et al. (2000) yield p-values of 0.16 and 0.47 respectively. In Figure 1.7(b) we display our estimated circulation density for the price markup time series, including 95% pointwise confidence bands obtained using the local bootstrap. The circulation density estimate is negative everywhere above the 0.05 quantile, and the 95% confidence bands mostly exclude zero at quantiles 0.1 and higher. This pattern is consistent with the presence of Edgeworth cycles, under which downward price movements are more likely than upward price movements unless the markup is very low. Further, the circulation density appears to dip substantially in the lower half of the state space, achieving its minimum value near the 0.35 quantile of the invariant distribution. In the language of Section 1.4.1, we say that there is a significant net probability downflow through this region. This suggests that sequences of price undercutting may be most likely to occur when the markup is near the 0.35 quantile.

Our estimated circulation density is broadly consistent with the pattern of circulation estimated by McCausland (2007) using the same data. After dividing the markups into six bins, McCausland estimated the circulation through each interior bin. (The circulation through the first and last bins is necessarily zero.) Table 4 of McCausland (2007) reveals that, while the estimated circulation through each bin is negative, the estimated circulation through the third bin is at least six times as large as the estimated circulation through any of the other bins. This third bin corresponds to markups between 1.1 and 1.2; the corresponding empirical quantiles are 0.22 and 0.56. Our circulation density estimate exhibits a similar pattern, but provides us with a more precise idea of where the tendency for downward price movement is strongest, and avoids the loss of information inherent to methods that classify observations into discrete bins.

It is apparent from Figure 1.7(a) that our price markup series is somewhat more volatile in the first half of the sample than it is in the second half. This may be due in part to the Iraqi invasion of Kuwait in August 1990, which created a spike in the price of crude oil lasting for the better part of a year (Hamilton, 2009, pp. 220–223). Our price markup series dips below unity during this period. Dropping observations prior to August 1991 changes the shape of our estimated circulation density at quantiles below 0.2: instead of rising to 0.02 as we move left toward the 0.05 quantile, the estimated circulation density falls to -0.44. At quantiles above 0.2, the shape of the estimated circulation density is mostly unaffected by excluding the earlier part of the sample.

1.6 Conclusion

In this paper we have made two primary contributions to the literature on time reversibility. First, we proposed a new test of time reversibility, applicable to stationary Markov chains. Compared to existing tests, our test has the advantage of being consistent against arbitrary violations of reversibility. Second, building on work by McCausland (2007), we proposed a new way to characterize the nature of time irreversibility when it is present. Our circulation density estimator was shown to be well behaved asymptotically under suitable regularity conditions, and numerical evidence suggests that associated inferential methods perform well in finite samples.

Our work here may be extended in several directions. On the technical side, it may be interesting to consider the problem of bandwidth selection for our circulation density estimator in more detail. The bandwidth decay rates permitted under Assumption 1.4.1(e) imply zero asymptotic bias. Therefore, Theorem 1.4.3 provides no guidance about how one might choose h to optimize the asymptotic mean square error of our circulation density estimator. Presumably, such optimization would entail a bandwidth decay rate of $T^{-1/5}$. Relaxation of the condition $Th^4 \to c \in [0, \infty)$ in Assumption 1.4.1(e), so that the asymptotic bias given in Theorem 1.4.3 is potentially nonzero, may be required in order to deal rigorously with the problem of bandwidth selection. Extending Theorem 1.4.3 in this way involves a number of technical difficulties and goes beyond the scope of the present paper.

On the more practical end, a priority for future work is to systematically apply our time reversibility test and circulation density estimator to a range of macroeconomic time series. Business cycle asymmetry is by now fairly well established for many variables of interest, but the study of circulation densities may perhaps yield new insights into the nature of this asymmetry. It may also be of interest to investigate whether the asymmetric Gumbel copula, or other nonexchangeable copula families, may be used to improve the empirical modeling and forecasting of macroeconomic and financial variables exhibiting asymmetric cyclical behavior. We leave these matters to future research.

1.7 Mathematical appendix

1.7.1 Technical conditions for local bootstrap validity

To formally establish the applicability of the local bootstrap to our testing procedure, we build on some of the results in Paparoditis and Politis (2002). Those authors obtain their results under a number of technical conditions. We shall employ the same conditions here.

Assumption 1.7.1. The following statements are true.

(a) $\mathscr{X}$ is an aperiodic, stationary, geometrically ergodic, real valued Markov chain.

(b) The invariant distribution $F(\cdot)$ and one-step transition distributions $F(\cdot|x)$, $x \in \mathbb{R}$, satisfy the following conditions.

  (i) $F(\cdot)$ and $F(\cdot|x)$, $x \in \mathbb{R}$, are absolutely continuous, with bounded densities $f(\cdot)$ and $f(\cdot|x)$, $x \in \mathbb{R}$.

  (ii) There exists $L \in (0, \infty)$ such that, for all $x_1, x_2 \in \mathbb{R}$ and $y \in \bar{\mathbb{R}}$,

$$|F(y|x_2)f(x_2) - F(y|x_1)f(x_1)| \leq L|x_2 - x_1|.$$

  (iii) There exists $L' \in (0, \infty)$ such that, for all $x, y_1, y_2 \in \mathbb{R}$,

$$|f(y_2|x) - f(y_1|x)| \leq L'|y_2 - y_1|.$$

(c) There exists a compact set $S \subset \mathbb{R}$ such that $P(X_0 \in S) = 1$ and $f(\cdot|x) > 0$ for all $x \in S$.

(d) The kernel W is a bounded, Lipschitz continuous, even pdf on $\mathbb{R}$ satisfying $W(x) > 0$ for all $x \in \mathbb{R}$, and $\int |x|W(x)\mathrm{d}x < \infty$.

(e) The bandwidth $b = b_T$ satisfies $b \asymp T^{-\delta}$ for some $\delta \in (0, 1/2)$. That is, there exist $a_1, a_2 \in (0, \infty)$ such that $a_1 \leq bT^{\delta} \leq a_2$ for all sufficiently large T.

1.7.2 Proofs

The following preliminary result is used in our proofs of Theorem 1.3.1 and Lemma 1.3.1.

Lemma 1.7.1. *Suppose Assumption 1.3.1 holds. Then as* $T \to \infty$ *we have* $T^{1/2}(H_T - H) \rightsquigarrow \mathscr{B}$ *in* $\ell^{\infty}(\mathbb{R}^2)$*. This continues to be true if* $\mathscr{X}$ *is not a Markov chain.*

Proof of Lemma 1.7.1. H_T is the empirical distribution function of a sample of size $T-1$ drawn from the bivariate process $\{(X_t, X_{t+1}) : t \in \mathbb{Z}\}$. This bivariate process inherits the stationarity and α-mixing rate of the univariate process $\mathscr{X}$. Therefore, since H is continuous when F is continuous, results due to Rio (2000, ch. 7) imply that $T^{1/2}(H_T - H) \rightsquigarrow \mathscr{B}$. □

Proof of Theorem 1.3.1. If $\mathscr{X}$ is time reversible, then $H(x, y) = H(y, x)$ for all $x, y \in \mathbb{R}$, and so

$$T^{1/2}\theta_T = \sup_{x,y} \left| T^{1/2}(H_T(x, y) - H(x, y)) - T^{1/2}(H_T(y, x) - H(y, x)) \right|.$$

Since $T^{1/2}(H_T - H) \rightsquigarrow \mathscr{B}$ by Lemma 1.7.1, part (a) now follows from an application of the continuous mapping theorem. If $\mathscr{X}$ is time irreversible, then we may choose $x, y \in \mathbb{R}$ such that $H(x, y) \neq H(y, x)$. Since $H_T(x, y) - H_T(y, x) = H(x, y) - H(y, x) + O_p(T^{-1/2})$ by Lemma 1.7.1, we find that

$$T^{1/2}\theta_T \geq T^{1/2}|H_T(x, y) - H_T(y, x)| = T^{1/2}|H(x, y) - H(y, x)| + O_p(1).$$

Divergence of $T^{1/2}|H(x,y) - H(y,x)|$ to infinity establishes part (b). $\square$

Proof of Lemma 1.3.1. Let $\mathscr{B}_T = \sqrt{T}(H_T - H)$ and recall that $\mathscr{B}_T^* = \sqrt{T}(H_T^* - E^* H_T^*)$. Let $\mathscr{L}^*(\mathscr{B}_T^*)$ denote the law of $\mathscr{B}_T^*$ conditional on $\mathscr{X}$. Noting that $\mathscr{L}_T^*(\mathscr{B}_T^*) = \mathscr{L}^*(\mathscr{B}_T^*)$ a.s., we see that it suffices for us to show that $\mathscr{L}^*(\mathscr{B}_T^*) \rightsquigarrow \mathscr{B}$ a.s. We will do this by verifying a.s. fidi convergence and a stochastic equicontinuity condition; see e.g. Theorem 10.2 of Pollard (1990).

First, Theorem 4.2 of Paparoditis and Politis (2002) will be used to show a.s. fidi convergence. Fix s pairs $(x_1, y_1), \ldots, (x_s, y_s) \in \mathbb{R}^2$. Let $g : \mathbb{R}^2 \to \{0,1\}^s$ be given by

$$g(v,w) = (1(v \leq x_1, w \leq y_1), \ldots, 1(v \leq x_s, w \leq y_s)) \,.$$

We may now write

$$(\mathscr{B}_T(x_1,y_1), \ldots, \mathscr{B}_T(x_s,y_s)) = \frac{\sqrt{T}}{T-1} \sum_{t=1}^{T-1} (g(X_t, X_{t+1}) - Eg(X_t, X_{t+1})) \quad (1.5)$$

and

$$(\mathscr{B}_T^*(x_1,y_1), \ldots, \mathscr{B}_T^*(x_s,y_s)) = \frac{\sqrt{T}}{T-1} \sum_{t=1}^{T-1} \left(g(X_t^*, X_{t+1}^*) - E^* g(X_t^*, X_{t+1}^*)\right) . \quad (1.6)$$

The assumptions of Theorem 4.2 of Paparoditis and Politis (2002) are satisfied[9] under Assumption 1.7.1. Applying this result in combination with (1.5) and (1.6) we obtain

$$d_{KS}\left(\mathscr{L}^*\left(\mathscr{B}_T^*(x_1,y_1), \ldots, \mathscr{B}_T^*(x_s,y_s)\right), \mathscr{L}\left(\mathscr{B}_T(x_1,y_1), \ldots, \mathscr{B}_T(x_s,y_s)\right)\right) \to 0$$

a.s., where d_{KS} is the Kolmogorov-Smirnov metric on the space of probability dis-

[9] In fact, Paparoditis and Politis (2002) require g to be continuous, which is not the case here. However, inspection of their proofs of Theorems 4.1 and 4.2 reveals that it suffices for g to be continuous on a subset of $\mathbb{R}^2$ of full H-measure. Continuity of H ensures that this condition is satisfied here.

tributions on $\mathbb{R}^s$. In view of Lemma 1.7.1, it follows that

$$\mathscr{L}^* \left(\mathscr{B}_T^*(x_1, y_1), \ldots, \mathscr{B}_T^*(x_s, y_s)\right) \rightsquigarrow \left(\mathscr{B}(x_1, y_1), \ldots, \mathscr{B}(x_s, y_s)\right)$$

a.s. This proves a.s. fidi convergence of $\mathscr{L}^*(\mathscr{B}_T^*)$ to $\mathscr{B}$.

It remains to verify stochastic equicontinuity. To this end we shall apply Theorem 2.2 of Andrews and Pollard (1994). In this paragraph it will be helpful to explicitly recognize that the bootstrap draws are properly viewed as a triangular array, so we shall write $X_{1,T}^*, \ldots, X_{T,T}^*$ for the bootstrap sample constructed from $X_1, \ldots, X_T$. Also, we condition on $\mathscr{X}$ throughout, and omit a.s. qualifiers. Now, for any $x, y \in \mathbb{R}$ we may write

$$\mathscr{B}_T^*(x, y) = \frac{\sqrt{T}}{T-1} \sum_{t=1}^{T-1} \left(f(Y_{t,T-1}^*) - E^* f(Y_{t,T-1}^*)\right), \tag{1.7}$$

where f (not to be confused with the pdf of X_0) is the indicator of $(-\infty, x] \times (-\infty, y]$, and $Y_{t,T-1}^* = (X_{t,T}^*, X_{t+1,T}^*)$. Let $\mathcal{F}$ be the collection of all such f as (x, y) varies over $\mathbb{R}^2$. Comparing Theorem 2.2 of Andrews and Pollard (1994) with (1.7), we see that $\mathscr{B}_T^*$ satisfies stochastic equicontinuity if, for some even integer $Q \geq 2$ and some $\gamma > 0$, we have (i) $\sum_{j=1}^{\infty} j^{Q-2} \alpha_j^{\gamma/(Q+\gamma)} < \infty$, and (ii) $\int_0^1 x^{-\gamma/(2+\gamma)} N(x, \mathcal{F})^{1/Q} \mathrm{d}x < \infty$. Here, the α_j's are α-mixing coefficients corresponding to the array $\{Y_{t,T}^* : t \leq T, T = 1, 2, \ldots\}$, while $N(x, \mathcal{F})$ is a bracketing number for $\mathcal{F}$; see Andrews and Pollard (1994, p. 120) for details. Theorem 3.4 of Paparoditis and Politis (2002) implies that the ρ-mixing coefficients for the array $\{Y_{t,T}^* : t \leq T, T = 1, 2, \ldots\}$ decay at a geometric rate. It follows from the well-known inequality between ρ- and α-mixing coefficients (see e.g. Proposition 3.11 in Bradley, 2007) that the α-mixing coefficients must also decay at a geometric rate, and so condition (i) holds for any permissible Q and γ. Further, it is known (see e.g. Examples 2.5.4 and 2.5.7 in van der Vaart and Wellner, 1996) that $N(x, \mathcal{F})$ increases at a polynomial rate as $x \downarrow 0$, so we may choose Q and γ such that condition (ii) is satisfied. Theorem 2.2 of Andrews and Pollard (1994) therefore yields stochastic equicontinuity of $\mathscr{B}_T^*$.

We have established that, conditional on $\mathscr{X}$, $\mathscr{B}_T^*$ satisfies fidi convergence and stochastic equicontinuity with probability one. The weak convergence to be proved now follows from Theorem 10.2 of Pollard (1990) or Corollary 2.3 of Andrews and Pollard (1994). □

Proof of Theorem 1.3.2. We know from Lemma 1.3.1 that $\mathscr{L}_T^*(\mathscr{B}_T^*) \rightsquigarrow \mathscr{B}$ a.s. An application of the continuous mapping theorem yields

$$\mathscr{L}_T^*\left(\sup_{x,y}|\mathscr{B}_T^*(x,y)-\mathscr{B}_T^*(y,x)|\right) \rightarrow_d \sup_{x,y}|\mathscr{B}(x,y)-\mathscr{B}(y,x)|$$

a.s. The statement to be proved follows from the continuity of this limiting distribution. □

Proof of Theorem 1.4.1. Since F and $\partial_2 C$ are continuous, we may define a regular family of conditional cdfs for X_t given X_{t+1} by writing $P(X_t \leq x|X_{t+1}=y) = \partial_2 C(F(x),F(y))$ for all $x \in \mathbb{R}$ and F-a.e. $y \in \mathbb{R}$. Continuity of F ensures that $F(Q(u)) = u$ for all $u \in (0,1)$, so we have $\mathscr{F}_\uparrow(Q(u)) = \partial_2 C(u,u)$ for a.e. $u \in (0,1)$. Similarly, $\mathscr{F}_\downarrow(Q(u)) = \partial_1 C(u,u)$ for a.e. $u \in (0,1)$. Our desired result follows by noting that the identities in (1.4) allow us to write $\psi(u) = \mathscr{F}_\uparrow(Q(u)) - \mathscr{F}_\downarrow(Q(u))$ for a.e. $u \in (0,1)$.

□

Proof of Theorem 1.4.2. In view of (1.4) and the definitions of $\mathscr{F}_\uparrow$ and $\mathscr{F}_\downarrow$, we have

$$\begin{aligned}\int\psi(u)\mathrm{d}u &= \int P(X_{t-1}\leq Q(u)|X_t=Q(u))\mathrm{d}u - \int P(X_{t+1}\leq Q(u)|X_t=Q(u))\mathrm{d}u \\ &= \int P(X_{t-1}\leq x|X_t=x)\mathrm{d}F(x) - \int P(X_{t+1}\leq x|X_t=x)\mathrm{d}F(x).\end{aligned}$$

The law of iterated expectations allows us to write

$$\int P(X_{t-1}\leq x|X_t=x)\mathrm{d}F(x) = \int P(X_{t-1}\leq X_t|X_t=x)\mathrm{d}F(x) = P(X_{t-1}\leq X_t).$$

Similarly, we have $\int P(X_{t+1} \leq x | X_t = x) \mathrm{d}F(x) = P(X_{t+1} \leq X_t)$. $\square$

To prove Theorem 1.4.3, the following two preliminary results will be useful.

Lemma 1.7.2. *Suppose Assumption 1.4.1 holds. Then for any $x \in \mathbb{R}$, as $T \to \infty$ the random vector*

$$(Th)^{1/2} \cdot \begin{bmatrix} \partial_1 \hat{H}_T(x,x) - \partial_1 H(x,x) \\ \partial_2 \hat{H}_T(x,x) - \partial_2 H(x,x) \\ \hat{f}_T(x) - f(x) \end{bmatrix}$$

converges in distribution to the trivariate normal distribution with zero mean and covariance matrix

$$\Sigma = \int k(z)^2 dz \cdot f(x) \cdot \begin{bmatrix} \partial_1 C(u,u) & \partial_1 C(u,u) \partial_2 C(u,u) & \partial_1 C(u,u) \\ \partial_1 C(u,u) \partial_2 C(u,u) & \partial_2 C(u,u) & \partial_2 C(u,u) \\ \partial_1 C(u,u) & \partial_2 C(u,u) & 1 \end{bmatrix},$$

where $u = F(x)$.

Proof of Lemma 1.7.2. Our proof of this result bears some resemblance to the proof of Theorem 7 of Fermanian and Scaillet (2003). Like those authors, we establish our result by applying Lemma 7.1 of Robinson (1983). In view of the Cramér-Wold theorem it suffices for us to show that, for any $\lambda = (\lambda_1, \lambda_2, \lambda_3)^\top \in \mathbb{R}^3$,

$$(Th)^{1/2} \left(\sum_{i=1}^{2} \lambda_i (\partial_i \hat{H}_T(x,x) - \partial_i H(x,x)) + \lambda_3 (\hat{f}_T(x) - f(x)) \right) \to_d N(0, \lambda^\top \Sigma \lambda).$$

Using integration by parts and a change of variables, we may show that

$$E\hat{f}_T(x) = \int k_h(x-y) f(y) \mathrm{d}y = \int f(x - hr) k(r) \mathrm{d}r.$$

Applying a Taylor expansion to f and exploiting the fact that k is even, we obtain $E\hat{f}_T(x) = f(x) + O(h^2)$. Similar arguments yield $E\partial_i \hat{H}_T(x,x) = \partial_i H(x,x) +$

$O(h^2)$ for $i = 1, 2$. Since $Th^5 \to 0$, the bias in our estimators is asymptotically negligible, and now we need only show that

$$(Th)^{1/2}\left(\sum_{i=1}^{2}\lambda_i(\partial_i\hat{H}_T(x,x) - E\partial_i\hat{H}_T(x,x)) + \lambda_3(\hat{f}_T(x) - E\hat{f}_T(x))\right) \to_d N(0, \lambda^\top\Sigma\lambda). \tag{1.8}$$

We now apply Lemma 7.1 of Robinson (1983). For $t = 0, \ldots, T$ let

$$\begin{aligned} V_{1tT} &= \lambda_1 h\left(k_h(x - X_{t+1})K_h(x - X_{t+2}) - Ek_h(x - X_{t+1})K_h(x - X_{t+2})\right), \\ V_{2tT} &= \lambda_2 h\left(K_h(x - X_t)k_h(x - X_{t+1}) - EK_h(x - X_t)k_h(x - X_{t+1})\right), \\ V_{3tT} &= \lambda_3 h\left(k_h(x - X_{t+1}) - Ek_h(x - X_{t+1})\right). \end{aligned}$$

The term on the left-hand side of (1.8) is equal to

$$(Th)^{1/2}\left(\frac{1}{T-1}\sum_{t=0}^{T-2}h^{-1}V_{1tT} + \frac{1}{T-1}\sum_{t=1}^{T-1}h^{-1}V_{2tT} + \frac{1}{T}\sum_{t=0}^{T-1}h^{-1}V_{3tT}\right).$$

Boundedness of k ensures that the random variables V_{itT} are bounded uniformly in i, t and T, so we may rewrite this quantity as $S_T + O(T^{-1/2}h^{-1/2}) = S_T + o(1)$, where $S_T = T^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{3}h^{-1/2}V_{itT}$. If applicable, Lemma 7.1 of Robinson (1983) establishes the asymptotic normality of S_T; we now verify its assumptions, which are labeled A3.1 and A7.1–A7.4. A3.1 is implied by our condition[10] on the α-mixing rate of $\mathscr{X}$. A7.1 holds with $q = 2$ due to the stationarity of $\mathscr{X}$. A7.2 holds since $Th \to \infty$.

A7.3 is satisfied if we can identify constants σ_{ij}, $i, j = 1, 2, 3$, such that $h^{-1}EV_{itT}V_{jtT} \to \lambda_i\lambda_j\sigma_{ij}$. Let $\kappa_2 = \int k(x)^2\mathrm{d}x$. Arguments given in the proof of Theorem 7 in Fermanian and Scaillet (2003, pp. 49–51) establish that for $i = 1, 2$ we may take $\sigma_{ii} = \kappa_2\partial_i H(x, x)$, $\sigma_{33} = \kappa_2 f(x)$ and $\sigma_{i3} = \sigma_{3i} = \kappa_2\partial_i H(x, x)$. It remains for us to identify $\sigma_{12} = \sigma_{21}$. Fermanian and Scaillet (2003, pp. 48–49)

[10] In fact, Lemmas 1.7.2 and 1.7.3 and Theorem 1.4.3 remain true if our Assumption 1.4.1(c) is replaced with A3.1 of Robinson (1983), which requires that $\sum_{j=T}^{\infty}\alpha_j = o(T^{-1})$.

establish that $Ek_h(x - X_{t+1})K_h(x - X_{t+2}) = O(1)$ and $EK_h(x - X_t)k_h(x - X_{t+1}) = O(1)$, so we have

$$h^{-1}EV_{1tT}V_{2tT} = \lambda_1\lambda_2 hE\left(K_h(x - X_t)k_h(x - X_{t+1})^2 K_h(x - X_{t+2})\right) + O(h). \tag{1.9}$$

Since $\mathscr{X}$ is a Markov chain, the joint cdf of X_t and X_{t+2} conditional on X_{t+1} is of the form

$$P(X_t \leq w, X_{t+2} \leq z | X_{t+1} = y) = \partial_2 C(F(w), F(y))\partial_1 C(F(y), F(z));$$

see e.g. Darsow et al. (1992). We may therefore write

$$\begin{aligned} & E\left(K_h(x - X_t)K_h(x - X_{t+2}) | X_{t+1} = y\right) \\ = & \left(\int K_h(x - w)\partial_2 C(F(\mathrm{d}w), F(y))\right)\left(\int K_h(x - z)\partial_1 C(F(y), F(\mathrm{d}z))\right). \end{aligned} \tag{1.10}$$

Integration by parts and a change of variables yield

$$\int K_h(x - w)\partial_2 C(F(\mathrm{d}w), F(y)) = \int \partial_2 C(F(x - hr), F(y))k(r)\mathrm{d}r.$$

Applying a Taylor expansion to $\partial_2 C(F(\cdot), F(y))$ and exploiting the symmetry of k, we find that this last term is equal to $\partial_2 C(F(x), F(y)) + O(h^2)$, with the order of the remainder term holding uniformly in y over any set on which $f(y)$ is bounded away from zero. We may show in similar fashion that

$$\int K_h(x - z)\partial_1 C(F(y), F(\mathrm{d}z)) = \partial_1 C(F(y), F(x)) + O(h^2),$$

with the order of the remainder term again holding uniformly in y over any set on which $f(y)$ is bounded away from zero. Returning to (1.10), we now have

$$E\left(K_h(x - X_t)K_h(x - X_{t+2}) | X_{t+1} = y\right) = \partial_2 C(F(x), F(y))\partial_1 C(F(y), F(x)) + R_T(y),$$

where the remainder term $R_T(y)$ satisfies $\sup_{f(y)>\varepsilon} |R_T(y)| = O(h^2)$ for any $\varepsilon > 0$.

Applying the law of iterated expectations and making another change of variables, we obtain

$$
\begin{aligned}
& E\left(K_h(x-X_t)k_h(x-X_{t+1})^2K_h(x-X_{t+2})\right) \\
= & \int \partial_2 C(F(x),F(y))\partial_1 C(F(y),F(x))k_h(x-y)^2 f(y)\mathrm{d}y + \int R_T(y)k_h(x-y)^2 f(y)\mathrm{d}y \\
= & \; h^{-1}\int \partial_2 C(F(x),F(x-hr))\partial_1 C(F(x-hr),F(x))k(r)^2 f(x-hr)\mathrm{d}r + O(h).
\end{aligned}
$$

Here, to obtain the order of the approximation error, we note that

$$
\int R_T(y)k_h(x-y)^2 f(y)\mathrm{d}y = h^{-1}\int R_T(x-hr)f(x-hr)k(r)^2\mathrm{d}r,
$$

which is $O(h)$ since k has compact support, f and k are bounded, and R_T is uniformly $O(h^2)$ in a neighborhood of x. Next, taking a Taylor expansion and once again exploiting the symmetry of k, we find that

$$
\begin{aligned}
& E\left(K_h(x-X_t)k_h(x-X_{t+1})^2K_h(x-X_{t+2})\right) \\
= & \; h^{-1}\partial_1 C(F(x),F(x))\partial_2 C(F(x),F(x))f(x)\int k(r)^2\mathrm{d}r + O(h),
\end{aligned}
$$

and so (1.9) allows us to set $\sigma_{12}=\sigma_{21}=\kappa_2\partial_1 C(F(x),F(x))\partial_2 C(F(x),F(x))f(x)$. Thus A7.3 of Robinson (1983) is satisfied.

To verify A7.4 we will demonstrate that $EV_{itT}V_{j,t+s,T}=O(h^2)$ for $i,j=1,2,3$ and $s\geq 1$. Boundedness of K_h allows us to write

$$
|EV_{itT}V_{j,t+s,T}| \leq ah^2Ek_h(x-X_{t+1})k_h(x-X_{t+s+1}) + O(h^2)
$$

for some $a<\infty$. Let H^s denote the joint cdf of X_{t+1} and X_{t+s+1}. Integration by parts and a change of variables yield

$$
\begin{aligned}
h^2Ek_h(x-X_{t+1})k_h(x-X_{t+s+1}) &= h^2\iint k_h(x-y)k_h(x-z)H^s(\mathrm{d}y,\mathrm{d}z) \\
&= \iint k'(v)k'(w)H^s(x-hv,x-hw)\mathrm{d}v\mathrm{d}w \quad (1.11)
\end{aligned}
$$

Using the Markov property of $\mathscr{X}$ and smoothness of H, one may show without difficulty that H^s is twice continuously differentiable in a neighborhood of (x, x). Therefore, since $\int k' = 0$, we may use a Taylor expansion to show that the right-hand side of (1.11) is $O(h^2)$. We conclude that $EV_{itT}V_{j,t+s,T} = O(h^2)$, and so A7.4 is satisfied. Lemma 7.1 of Robinson (1983) thus implies that (1.8) holds, with Σ having $(i, j)^{\text{th}}$ element σ_{ij}. This completes the proof. □

Lemma 1.7.3. *Suppose Assumption 1.4.1 holds. Then for any $x \in \mathbb{R}$, as $T \to \infty$ we have*

(*i*) $(Th)^{1/2}\left(\partial_1\hat{H}_T(\hat{x}_T, \hat{x}_T) - \partial_1\hat{H}_T(x, x)\right) \to_p 0$,

(*ii*) $(Th)^{1/2}\left(\partial_2\hat{H}_T(\hat{x}_T, \hat{x}_T) - \partial_2\hat{H}_T(x, x)\right) \to_p 0$, *and*

(*iii*) $(Th)^{1/2}\left(\hat{f}_T(\hat{x}_T) - \hat{f}_T(x)\right) \to_p 0$,

where $\hat{x}_T = \hat{Q}_T(u)$ and $u = F(x)$.

Proof of Lemma 1.7.3. We begin by establishing that $\hat{x}_T = x + O_p(T^{-1/2})$. Following the argument of Fermanian and Scaillet (2003, pp. 44–45), we find that

$$\sup_{y\in\mathbb{R}}\left|T^{1/2}(\hat{F}_T(y) - F(y)) - \mathscr{G}(y) - T^{1/2}\int\left(F(y - hr) - F(y)\right)k(r)\mathrm{d}r\right| = o_p(1), \tag{1.12}$$

for some centered Gaussian process $\mathscr{G}$ on $\mathbb{R}$ with a.s. continuous sample paths. Applying a second-order Taylor expansion to F and exploiting the fact that k is even, we obtain

$$\sup_{y\in\mathbb{R}}\left|T^{1/2}\int\left(F(y - hr) - F(y)\right)k(r)\mathrm{d}r - \frac{1}{2}T^{1/2}h^2 f'(y)\int r^2k(r)\mathrm{d}r\right| = O_p(T^{1/2}h^3). \tag{1.13}$$

Under Assumption 1.4.1(e), $Th^4 \to c \in [0, \infty)$. Consequently, from (1.12) and (1.13) we have

$$T^{1/2}(\hat{F}_T - F) \rightsquigarrow \mathscr{G} + \frac{1}{2}c^{1/2}f'\int r^2k(r)\mathrm{d}r,$$

and an application of the functional delta method using the inversion operator yields $\hat{x}_T = x + O_p(T^{-1/2})$, as claimed.

We now prove part (i) of Lemma 1.7.3. Using a third-order Taylor expansion, we find that

$$\partial_1 \hat{H}_T(\hat{x}_T, \hat{x}_T) - \partial_1 \hat{H}_T(x, x) = \sum_{j=1}^{3} \frac{1}{j!}(\hat{x}_T - x)^j \frac{\mathrm{d}^j}{\mathrm{d}z^j} \partial_1 \hat{H}_T(z, z)|_{z=x} + R_T, \quad (1.14)$$

where the remainder term R_T is equal to

$$R_T = \frac{1}{24}(\hat{x}_T - x)^4 \frac{\mathrm{d}^4}{\mathrm{d}z^4} \partial_1 \hat{H}_T(z, z)|_{z=\tilde{x}_T}$$

for some $\tilde{x}_T$ between $\hat{x}_T$ and x. Boundedness of k and its first four derivatives ensures that

$$\sup_{\tilde{x} \in \mathbb{R}} \left| \frac{\mathrm{d}^4}{\mathrm{d}z^4} \partial_1 \hat{H}_T(z, z)|_{z=\tilde{x}} \right| = O(h^{-5}).$$

Therefore, since $Th^3 \to \infty$, we have $R_T = O_p(T^{-2}h^{-5}) = o_p(T^{-1/2}h^{-1/2})$. To demonstrate that the right-hand side of (1.14) is $o_p(T^{-1/2}h^{-1/2})$, it now suffices for us to show that

$$\frac{\mathrm{d}^j}{\mathrm{d}z^j} \partial_1 \hat{H}_T(z, z)|_{z=x} = o_p(T^{(j-1)/2}h^{-1/2})$$

for $j = 1, \ldots, 3$. This will be true if

$$\frac{1}{T} \sum_{t=1}^{T} k_h^{(i)}(x - X_t) K_h^{(j-i)}(x - X_{t+1}) = o_p(T^{(j-1)/2}h^{-1/2}) \quad (1.15)$$

for $j = 1, \ldots, 3$ and $i = 0, \ldots, j$, where parenthesized superscripts signify higher-order differentiation. Using integration by parts and a change of variables, we find that

$$E k_h^{(i)}(x - X_t) K_h^{(j-i)}(x - X_{t+1}) = \iint k(v)k(w) H^{(i+1, j-i)}(x - hv, x - hw) \mathrm{d}v \mathrm{d}w = O(1).$$

It follows that

$$\frac{1}{T}\sum_{t=1}^{T} k_h^{(i)}(x-X_t)K_h^{(j-i)}(x-X_{t+1}) = T^{(j-1)/2}h^{-1/2}S_T + O(1), \tag{1.16}$$

where, suppressing the dependence of S_T and V_{tT} on i and j in our notation[11], we define $S_T = T^{-1/2}\sum_{t=1}^{T}(T^{j/2}h^{2j})^{-1/2}V_{tT}$ and

$$V_{tT} = T^{-j/4}h^{j+1/2}\left(k_h^{(i)}(x-X_t)K_h^{(j-i)}(x-X_{t+1}) - Ek_h^{(i)}(x-X_t)K_h^{(j-i)}(x-X_{t+1})\right).$$

In view of (1.16) and the fact that $T^{(j-1)/2}h^{-1/2} \to \infty$, we may verify (1.15) by showing that $S_T = o_p(1)$. We shall do this by verifying that S_T and V_{tT} satisfy assumptions A3.1, A7.1-A7.4 of Lemma 7.1 of Robinson (1983), with $\sigma^2 = 0$. A3.1 holds under our assumption on the mixing rate of $\mathscr{X}$. A7.1 holds with $q = 1$ due to the stationarity of $\mathscr{X}$. A7.2 holds since $Th^3 \to \infty$. A7.3 holds with $\sigma^2 = 0$ if $EV_{tT}^2 = o(T^{j/2}h^{2j})$. Using a change of variables, we may show that

$$\begin{aligned} & Ek_h^{(i)}(x-X_t)^2K_h^{(j-i)}(x-X_{t+1})^2 \\ = \; & h^{-2j}\iint k^{(i)}(v)^2K^{(j-i)}(w)^2H^{(1,1)}(x-hv,x-hw)\mathrm{d}v\mathrm{d}w = O(h^{-2j}). \end{aligned}$$

Therefore, we have

$$EV_{tT}^2 \leq 2T^{-j/2}h^{2j+1}Ek_h^{(i)}(x-X_t)^2K_h^{(j-i)}(x-X_{t+1})^2 = O(T^{-j/2}h) = o(T^{j/2}h^{2j}), \tag{1.17}$$

where the first inequality follows from the fact that the variance of any random variable is no greater than twice its expected square. Thus A7.3 holds. A7.4 holds if (a) $\sup_{1\leq t\leq T}|V_{tT}| = O(1)$, (b) $E|V_{tT}V_{t+1,T}| = o(T^{j/2}h^{2j})$, and (c) $E|V_{tT}V_{t+s,T}| = O(T^jh^{4j})$ for $s \geq 2$. Boundedness of $k^{(i)}$ and $K^{(j-i)}$ may be used to show that $\sup_{1\leq t\leq T}|V_{tT}| = O(T^{-j/4}h^{-1/2}) = O(1)$, yielding (a). Parts (b) and (c) follow from (1.17) using the Cauchy-Schwarz inequality. We have now verified all as-

[11] As defined here, S_T and V_{tT} differ from S_T and V_{itT} as defined in the proof of Lemma 1.7.2, but play the same role in the application of Lemma 7.1 of Robinson (1983).

sumptions of Lemma 7.1 of Robinson (1983), which allows us to conclude that $S_T = o_p(1)$. Thus, (1.15) holds for any $j = 1, \dots, 3$ and $i = 0, \dots, j$, and so the right-hand side of (1.14) is $o_p(T^{-1/2}h^{-1/2})$. This proves part (i) of Lemma 1.7.3. Parts (ii) and (iii) may be proved using the same approach. □

Proof of Theorem 1.4.3. Lemma 1.7.2 and Lemma 1.7.3 jointly imply that

$$(Th)^{1/2} \cdot \begin{bmatrix} \partial_1 \hat{H}_T(\hat{x}_T, \hat{x}_T) - \partial_1 H(x,x) \\ \partial_2 \hat{H}_T(\hat{x}_T, \hat{x}_T) - \partial_2 H(x,x) \\ \hat{f}_T(\hat{x}_T) - f(x) \end{bmatrix} \to_d N(0, \Sigma), \tag{1.18}$$

where $x = Q(u)$ and $\hat{x}_T = \hat{Q}_T(u)$. Noting that

$$\hat{\psi}_T(u) = \frac{\partial_2 \hat{H}_T(\hat{x}_T, \hat{x}_T) - \partial_1 \hat{H}_T(\hat{x}_T, \hat{x}_T)}{\hat{f}_T(\hat{x}_T)} \quad \text{and} \quad \psi(u) = \frac{\partial_2 H(x,x) - \partial_1 H(x,x)}{f(x)},$$

we can use the delta method to obtain $(Th)^{1/2}(\hat{\psi}_T(u) - \psi(u)) \to_d N(0, \sigma^2(u))$. Let

$$a_1 = \frac{-1}{f(x)}, \quad a_2 = \frac{1}{f(x)}, \quad a_3 = \frac{\partial_1 H(x,x) - \partial_2 H(x,x)}{f(x)^2}.$$

Then, applying the delta method, $\sigma^2(u)$ is given by

$$\begin{aligned} \sigma^2(u) &= \sum_{i=1}^{3} \sum_{j=1}^{3} a_i a_j \Sigma_{ij} \\ &= \frac{\int k(z)^2 \mathrm{d}z}{f(Q(u))} \cdot \left(\partial_1 C(u,u) \left(1 - \partial_1 C(u,u)\right) + \partial_2 C(u,u) \left(1 - \partial_2 C(u,u)\right) \right). \end{aligned}$$

That $\hat{\sigma}_T^2(u) \to_p \sigma^2(u)$ follows easily from (1.18). □

Acknowledgements

Chapter 1 is published in *Econometric Theory (2014).* I thank Brendan K. Beare for his permission to use the co-authored paper in my dissertation Chapter.

Chapter 2

Vine Copula Specifications for Stationary Multivariate Markov Chains

Abstract. Vine copulae provide a graphical framework in which multiple bivariate copulae may be combined in a consistent fashion to yield a more complex multivariate copula. In this paper we discuss the use of vine copulae to build flexible semiparametric models for stationary multivariate higher-order Markov chains. We propose a new vine structure, the M-vine, that is particularly well suited to this purpose. Stationarity may be imposed by requiring the equality of certain copulae in the M-vine, while the Markov property may be imposed by requiring certain copulae to be independence copulae.

2.1 Introduction

Copula functions or copulae[1] have become a popular tool for modeling potentially nonlinear stochastic relationships between two or more variables, particularly in economics and finance. See, for instance, either of the textbooks by Cherubini et al. (2004, 2012), or any of the three review articles by Patton (2009, 2012, 2013). In a time series context we may distinguish between the use of copulae to model contemporaneous relationships between multiple time series, and the use of copulae to model the dynamics of an individual time series. While the former approach is more common, the latter has become increasingly popular in recent years, with applications including studies of air quality (Joe, 1997), blood pressure and antidepressants (Lambert and Vandenhende, 2002), order durations (Savu and Ng, 2005), automobile claims (Frees and Wang, 2006), coffee prices (Abegaz and Naik-Nimbalkar, 2008), higher education funding reforms (Dearden et al., 2008), nursing home utilization (Sun et al., 2008), earnings mobility (Bonhomme and Robin, 2009), foreign exchange (Bouyé and Salmon, 2009), stock prices (Domma et al., 2009), electricity load (Smith et al., 2010), and gasoline prices (Beare and Seo, 2014).

Recent contributions by Yi and Liao (2010) and Rémillard et al. (2012) bridge the gap between methods that use copulae to model contemporaneous dependence between variables, and methods that use copulae to model univariate serial dependence. In the more general approach of the latter set of authors, we specify a parametric $2m$-variate copula function to model the dependence between the $2m$ random variables associated with consecutive realizations of a stationary m-variate Markov chain. The m univariate marginal distributions of the chain are left unspecified, and estimated nonparametrically. This generalizes the approach of Chen and Fan (2006), who considered the univariate case $m = 1$. In the multivariate case $m \geq 2$, additional care is needed when selecting an appropriate $2m$-variate copula,

[1]Refer to Nelsen (2006) for an introduction to copula functions and their properties.

as some of the most common parametric families can be very restrictive. With a multivariate Student copula, we require that all pairs of variables have the bivariate Student copula with the same degrees of freedom, while with a multivariate Archimedean copula, all pairs of variables must have the same bivariate copula. In both cases we impose symmetric dependence between any pair of variables.[2] These may not be desirable restrictions to impose in applications; in particular, we typically have no reason to expect the bivariate copula characterizing serial dependence in one time series to resemble the bivariate copula characterizing contemporaneous dependence between two time series. It may therefore be useful to investigate the use of more flexible multivariate copulae to simultaneously capture serial and contemporaneous dependence.

In this paper we explore the use of vine copulae in the multivariate framework of Rémillard et al. (2012). Vine copulae provide a graphical framework in which multiple bivariate copulae may be combined in a consistent fashion to yield a more complex multivariate copula. They are also referred to as pair copula constructions. Vine copulae were first proposed by Bedford and Cooke (2001, 2002), though some of the core ideas can be found in earlier work by Joe (1996, 1997). Their use in stationary multivariate Markov models was suggested by Rémillard et al. (2012, p. 34) as an area for future research. Smith et al. (2010) used vine copulae to model the serial dependence in time inhomogeneous univariate Markov chains, but we are the first to consider the stationary multivariate case. We consider not only first-order but also higher-order Markov chains, studied previously by Ibragimov (2009) through the lens of copula theory.

Our primary contribution is the invention of a new vine structure which we call the M-vine. We argue that the M-vine is particularly well suited for modeling stationary multivariate higher-order Markov chains. It is possible to specify an M-vine copula for the entire mn-variate distribution of a sample of n consecu-

[2] Generalized multivariate Student and Archimedean copulae (Demarta and McNeil, 2005; Fischer, 2011) may provide more flexibility. We do not consider these here, instead focusing on vine copulae.

tive realizations of an m-variate chain. Stationarity is easily imposed by requiring the equality of certain bivariate copulae linked by the M-vine, while the Markov property may be imposed by requiring certain bivariate copulae in the M-vine to be independence copulae. We discuss the application of the semiparametric estimator of Rémillard et al. (2012) to our M-vine model, and also a computationally faster stepwise estimator studied recently by Hobæk Haff (2012, 2013).

The remainder of our paper is organized as follows. In Section 2.2 we provide a brief review of the essential features of vine copulae. Our main contributions are in Section 2.3, where we study the use of different vine structures for stationary multivariate Markov models, and propose our new M-vine structure. In Section 2.4 we discuss the estimation of M-vine models, and present an illustrative application to exchange rate data. Section 2.5 closes with some remarks on possible directions for future research.

2.2 Essentials of vine copulae

We commence with a very brief introduction to the essential features of vine copulae, drawing mostly on the pioneering articles of Bedford and Cooke (2001, 2002) as well as discussions by Kurowicka and Cooke (2006) and Czado (2010). The graphical structure underlying a vine copula is called a regular vine or R-vine, and takes the form of a nested collection of trees. A tree is a connected acyclic graph; that is, a collection N of $q \geq 2$ elements called nodes, and a collection $E \subset \binom{N}{2}$ of $q-1$ unordered pairs of nodes called edges,[3] with the edges chosen such that there is a unique sequence of edges connecting any two nodes.

Definition 2.2.1. Given a totally ordered set N_1 with $q \geq 2$ elements, a *regular vine on* N_1 is a collection of $q-1$ trees $\mathcal{V} = (T_1, \ldots, T_{q-1})$ satisfying the following conditions.

(i) T_1 has nodes N_1 and a set of edges denoted E_1.

[3] We use $\binom{N}{2}$ to denote the collection of all subsets of N that contain exactly two elements.

(ii) For $k = 2, \ldots, q-1$, T_k has nodes $N_k = E_{k-1}$ and a set of edges denoted E_k.

(iii) *Proximity condition*: For $k = 2, \ldots, q-1$, if $u = \{u_1, u_2\}$ and $v = \{v_1, v_2\}$ are nodes of T_k connected by an edge, then $u \cap v$ is a singleton.

To each edge e in any of the trees of a regular vine, we assign three subsets of the nodes of the first tree: U_e, the complete union of e; D_e, the conditioning set of e; and $\{a_e, b_e\}$, the conditioned set of e, which necessarily has two elements.

Definition 2.2.2. Let $\mathcal{V} = (T_1, \ldots, T_{q-1})$ be a regular vine on N_1. Given an edge e_k of T_k for some k, the *complete union of* e_k is the set

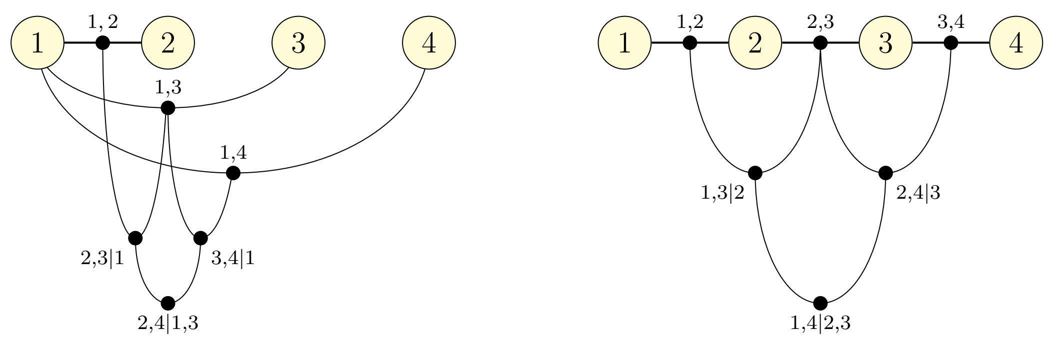
$$U_{e_k} = \{i \in N_1 : i \in e_1 \in e_2 \in \cdots \in e_{k-1} \in e_k \text{ for some } (e_1, \ldots, e_{k-1}) \in E_1 \times \cdots E_{k-1}\}$$

if $k \geq 2$, or $U_{e_k} = e_k$ if $k = 1$. The *conditioning set of* e_k, denoted D_{e_k}, is the intersection of the complete unions of the edges of T_{k-1} connected by e_k if $k \geq 2$, or the empty set if $k = 1$. The *conditioned set of* e_k, denoted $\{a_{e_k}, b_{e_k}\}$, is the symmetric difference of the complete unions of the edges of T_{k-1} connected by e_k if $k \geq 2$, or simply e_k if $k = 1$. The conditioned set of e_k contains exactly two elements of N_1, while the conditioning set of e_k contains exactly $k-1$ elements of N_1 (Bedford and Cooke, 2002, Lemma 4.2). We always label the elements of the conditioned set of e_k such that $a_{e_k} \preceq b_{e_k}$, where $\preceq$ is the total ordering on N_1.[4]

It is often convenient to label the edges of a regular vine according to their conditioned and conditioning sets. For instance, if $N_1 = \{1, 2, 3, 4\}$ and $\{\{1, 2\}, \{2, 3\}\}$ is an edge of T_2, then this edge has conditioned set $\{1, 3\}$ and conditioning set $\{2\}$, and we may label it as 1,3|2. Similarly, if $N_1 = \{1, 2, 3, 4\}$ and $\{2, 3\}$ is an edge of T_1, then this edge has conditioned set $\{2, 3\}$ and empty conditioning set, and we may label it as 2,3. Such a labelling is always unambiguous: no two edges will have the same conditioned and conditioning sets.

---

[4] If $N_1 \subset \mathbb{R}$ we take $\preceq$ to be the usual inequality relation $\leq$ on the real line.



(a) C-vine

(b) D-vine

Figure 2.1: Regular vines on four elements
On the left we display a C-vine on four elements. Node 1 is the root of the first tree: it has degree 3, while other nodes have degree 1. On the right we display a D-vine on four elements. Nodes 1 and 4 have degree 1, while nodes 2 and 3 have degree 2.

In Figure 2.1, we display two examples of regular vines on four elements. First consider the vine on the left. The first tree of this vine is given by the nodes $N_1 = \{1, 2, 3, 4\}$ and the edges 1,2, 1,3 and 1,4. In the second tree, these three edges are connected by the edges 2,3|1 and 2,4|1. In the third tree, these two edges are connected by the edge 2,4|1,3. Next consider the vine on the right. The first tree of this vine is given by the nodes $N_1 = \{1, 2, 3, 4\}$ and the edges 1,2, 2,3 and 3,4. In the second tree, these three edges are connected by the edges 1,3|2 and 2,4|3. In the third tree, these two edges are connected by the edge 1,4|2,3. Notice that in both vines the proximity condition is satisfied at all times: only edges that share a node may be connected by an edge of the following tree.

The two vines depicted in Figure 2.1 are representative members of two important families of regular vines: the C-vines (C for canonical) and D-vines (D for drawable). These families are defined as follows.

Definition 2.2.3. Let $\mathcal{V} = (T_1, \ldots, T_{q-1})$ be a regular vine on N_1. The *degree* of a node of some tree T_k is the number of edges of T_k that connect to it. $\mathcal{V}$ is said to be a *C-vine* if each tree T_k has a node of degree $q - k$. That node is called the *root* of T_k. $\mathcal{V}$ is said to be a *D-vine* if each node of the first tree has degree no greater than two.

When $q \leq 3$ all regular vines are C-vines and D-vines, and when $q \leq 4$ all regular vines are C-vines or D-vines, but when $q \geq 5$ regular vines arise that belong to neither family (Joe, 2011). An important property of D-vines is that all trees beyond the first are uniquely determined by the first tree, due to the need to satisfy the proximity condition. This property will prove useful to us later.

The utility of regular vines derives from the fact that, when we assign a univariate cumulative distribution function (cdf) to each node of the first tree, and a bivariate copula to each edge of each tree, we obtain a unique q-dimensional multivariate distribution. The pairing of the regular vine with the assignment of univariate cdfs and bivariate copulae is called a vine copula specification.

Definition 2.2.4. $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ is a *vine copula specification* on N_1 if

(i) $\mathcal{F} = \{F_i : i \in N_1\}$ is a set[5] of absolutely continuous invertible univariate cdfs.

(ii) $\mathcal{V}$ is a regular vine on N_1.

(iii) $\mathcal{C} = \{C_{a_e,b_e|D_e} : e \in \cup_{k=1}^{q-1} E_k\}$ is a set of absolutely continuous bivariate copulae.

We will write f_i for the probability density function (pdf) associated with $F_i \in \mathcal{F}$, and $c_{a_e,b_e|D_e}$ for the copula density associated with $C_{a_e,b_e|D_e} \in \mathcal{C}$. The existence of these densities is assumed in our definition of a vine copula specification. This is not necessary, but is convenient for our purposes.

The assignment of the cdfs in $\mathcal{F}$ to the nodes in N_1 pins down the n univariate margins of the multivariate distribution determined by our vine copula specification, while the assignment of copulae to edges determines certain conditional copulae associated with the corresponding conditioned and conditioning sets. A

[5]Formally, $\mathcal{F}$ and $\mathcal{C}$ map the nodes and edges of $\mathcal{V}$ to cdfs and copulae. We follow convention and refer to $\mathcal{F}$ and $\mathcal{C}$ as sets of cdfs and copulae that are assigned to the nodes and edges of $\mathcal{V}$.

multivariate distribution that is consistent with the specified marginal distributions and conditional copulae is said to realize the copula vine specification.

Definition 2.2.5. The joint cdf F of q random variables X_i, $i \in N_1$, is said to *realize* the vine copula specification $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ on N_1 if

(i) For each $i \in N_1$, F_i is the marginal cdf of X_i.

(ii) For each $e \in \cup_{k=1}^{q-1} E_k$, $C_{a_e,b_e|D_e}$ is the conditional copula of X_{a_e} and X_{b_e} given the random variables X_i, $i \in D_e$.

For a definition and discussion of conditional copulae, see Patton (2006). Note that in condition (ii) of Definition 2.2.5, if D_e is empty, $C_{a_e,b_e|D_e}$ is simply the copula of X_{a_e} and X_{b_e}. If D_e is nonempty, the conditional copula of X_{a_e} and X_{b_e} given X_i, $i \in D_e$, is required to be a member of $\mathcal{C}$ - an ordinary bivariate copula - meaning that it cannot depend on the particular values taken by the conditioning variables X_i, $i \in D_e$. This is an important restriction, and we will return to it shortly.

The following result gives the precise form of the multivariate distribution associated with a vine copula specification. It appears in slightly modified form as Theorem 3 in Bedford and Cooke (2001), or Theorem 2.5 in Dißmann et al. (2013).

Theorem 2.2.1. *Let $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ be a vine copula specification on N_1. There is a unique joint cdf F that realizes $(\mathcal{F}, \mathcal{V}, \mathcal{C})$. F is absolutely continuous, with pdf f given by*

$$f = \left(\prod_{i \in N_1} f_i\right) \left(\prod_{k=1}^{q-1} \prod_{e \in E_k} c_{a_e,b_e|D_e}\left(F_{a_e|D_e}, F_{b_e|D_e}\right)\right).$$

The conditional cdfs $F_{a_e|D_e}$ and $F_{b_e|D_e}$ are determined by $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ in the following recursive fashion. If $e \in E_1$, so that $D_e = \varnothing$, then $F_{a_e|D_e} = F_{a_e}$ and $F_{b_e|D_e} = F_{b_e}$. If $e \in E_k$ for some $k \geq 2$, then

$$F_{a_e|D_e} = \frac{\partial C_{a_{e'},b_{e'}|D_{e'}}(F_{a_{e'}|D_{e'}}, F_{b_{e'}|D_{e'}})}{\partial F_{b_{e'}|D_{e'}}} \quad \text{and} \quad F_{b_e|D_e} = \frac{\partial C_{a_{e''},b_{e''}|D_{e''}}(F_{a_{e''}|D_{e''}}, F_{b_{e''}|D_{e''}})}{\partial F_{a_{e''}|D_{e''}}},$$

where e' and e'' are the edges of T_{k-1} connected by e, with $a_e = a_{e'}$ and $b_e = b_{e''}$.

The recursive method of constructing conditional cdfs given in Theorem 2.2.1 was proposed in an early contribution by Joe (1996). An important part of the statement of Theorem 2.2.1 is that, given any vine copula specification $(\mathcal{F}, \mathcal{V}, \mathcal{C})$, we can always find a joint cdf F that realizes $(\mathcal{F}, \mathcal{V}, \mathcal{C})$. This means that the cdfs and copulae assigned to the nodes and edges of our regular vine can never be mutually incompatible. The properties defining the structure of a regular vine rule out such incompatibilities. Joe (1997, ch. 3) discusses the problem of determining whether a given collection of marginal distributions is compatible in his treatment of Fréchet classes.

While part (ii) of Definition 2.2.5 indicates that we may view the copulae assigned to the edges of trees $T_2, \ldots, T_{q-1}$ as conditional copulae, part (iii) of Definition 2.2.4 requires them to be ordinary bivariate copulae. In general we would expect that, for $e \in E_k$ with $k \geq 2$, the conditional copula $C_{a_e,b_e|D_e}(u, v|w)$ of X_{a_e} and X_{b_e} given X_i, $i \in D_e$, should depend not only on the quantiles u and v associated with X_{a_e} and X_{b_e}, but also on the vector of values $w \in \mathbb{R}^{k-1}$ taken by the conditioning variates. But because we require the members of $\mathcal{C}$ to be bivariate copulae, $C_{a_e,b_e|D_e}$ can be a function only of the variables u and v, so that $C_{a_e,b_e|D_e}(u, v|w) = C_{a_e,b_e|D_e}(u, v)$. Torgovitsky (2012) refers to this condition as *copula invariance*. Although in principle we could relax the copula invariance condition by allowing the members of $\mathcal{C}$ to be arbitrary conditional copulae, in practice this eliminates much of the appeal of the vine copula specification as a dimension reduction device. Some authors use the expression *simplified pair copula construction* to emphasize the role of the copula invariance condition in a vine copula specification. Hobæk Haff et al. (2010) argue that a vine copula specification can deliver a good approximation to multivariate distributions even when the copula invariance condition is not satisfied, while Acar et al. (2012) provide a more skeptical perspective. Stöber et al. (2013) show that the members of $\mathcal{C}$ will automatically satisfy copula invariance whenever $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ is realized by a distribution with a multivariate Gaussian, Student, or Clayton copula.

2.3 Vine copulae and stationary Markov chains

In this section we consider the use of vine copulae to model an $\mathbb{R}^m$-valued stationary Markov chain of order p. Since the finite dimensional distributions of such a chain are uniquely determined by the $m \times (p+1)$ dimensional joint distribution of $p+1$ consecutive realizations, it is natural to consider vines on the $m \times n$ array of nodes $N_1 = \{1, \ldots, m\} \times \{1, \ldots, n\}$, with $n = p+1$. Care must be taken when building a vine copula specification on this array in order to ensure that the stationarity condition is satisfied. An alternative approach is to let n be the number of time periods in a sample to which we wish to fit our vine copula specification, with $n > p+1$. In this case we need not only take care to ensure that stationarity holds, but also that the Markov property is satisfied.

In Definition 2.2.1 we required that the nodes of a regular vine constitute a totally ordered set. This was done to ensure that the labeling of nodes in the conditioned set $\{a_e, b_e\}$ of some edge e can be made unambiguous by requiring that $a_e \preceq b_e$. For the remainder of the paper, we adopt the convention that the nodes of the $m \times n$ array N_1 are endowed with the following total order: $(i, s) \preceq (j, t)$ if and only if $(s-1)m + i \leq (t-1)m + j$.

To provide motivation and build intuition, we begin in Section 2.3.1 by considering stationary bivariate first-order Markov chains, and proceed to stationary bivariate second-order Markov chains in Section 2.3.2. In Section 2.3.3 we consider the general case where $m \geq 1$ and $p \geq 1$. We propose a new regular vine, the M-vine, and argue that it is particularly well suited for our purpose.

2.3.1 Bivariate first-order Markov chains

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a stationary first-order Markov chain taking values in $\mathbb{R}^2$, and for $i \in \{1, 2\}$ and $t \in \mathbb{N}$ let X_{it} denote the i^{th} element of $\mathbf{X}_t$. The finite dimensional distributions of our Markov chain are determined by the joint distribution of $(X_{11}, X_{21}, X_{12}, X_{22})$, which we wish to model using a vine copula specification. To this end we define the node set $N_1 = \{1,2\}^2$, which can be visualized as a 2×2 array of indices it, with i indicating the row and t indicating the column. In Figure 2.2, we display four trees on N_1. We shall consider each of these trees in turn as candidates for the first tree of a regular vine on N_1.

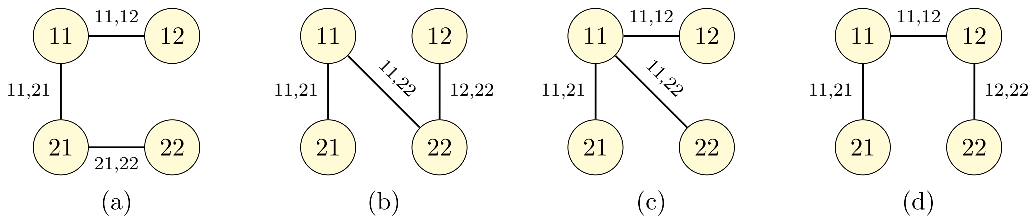


Figure 2.2: Four trees on a 2×2 array of nodes
The trees in panels (a) and (c) are less suitable for modeling a stationary bivariate first-order Markov chain because they do not include both of the edges 11,21 and 12,22. We would like to assign the same copula to these two edges in order to ensure stationarity.

First consider the tree displayed in Figure 2.2. Each node in this tree has degree of at most two, so there exists a unique regular vine on N_1 commencing with this tree, and it is a D-vine. The edges in the second tree are 12,21|11 and 11,22|21, and the edge in the third tree is 12,22|11,21. This vine is poorly suited to model the behavior of our Markov chain. The reason is that we have assumed $\mathbf{X}_1, \mathbf{X}_2, \ldots$ to be stationary, which requires the bivariate distributions of (X_{11}, X_{21}) and (X_{12}, X_{22}) to be identical. This will be the case if and only if (i) $F_{11} = F_{12}$ and $F_{21} = F_{22}$, and (ii) $C_{11,21} = C_{12,22}$. Condition (i) is easily satisfied with a suitable choice of $\mathcal{F}$ in our vine copula specification, but condition (ii) is problematic, because our vine does not include the edge 12,22. Satisfaction of (ii) when we do not directly specify $C_{11,21}$ and $C_{12,22}$ typically entails a complicated restriction on those copulae we do specify. This problem can easily be avoided with a more appropriate choice of vine.

Next consider the tree in Figure 2.2. Like the tree in panel (a), this tree defines a D-vine on N_1, with edges 21,22|11 and 11,12|22 in the second tree, and edge 21,12|11,22 in the third tree. Since the first tree includes both of the edges

11,21 and 12,22, we can easily impose stationarity by assigning the same copula to both edges, the same cdf to nodes 11 and 12, and the same cdf to nodes 21 and 22. This makes the vine a good candidate for a stationary Markov model.

The tree in Figure 2.2 (c) is the first tree of a C-vine. It does not include both of the edges 11,21 and 12,22 and is therefore poorly suited to our purpose. The same is true for any C-vine on N_1, regardless of which node is chosen to be the root of the first tree.

The tree in Figure 2.2 (d) defines a D-vine on N_1, with edges 21,12|11 and 11,22|12 in the second tree, and edge 21,22|11,12 in the third tree. Like the tree in panel (b), it includes both of the edges 11,21 and 12,22, so stationarity is easily imposed. It also includes the edge 11,12, whereas the tree in panel (b) instead includes 11,22. It seems to us that the choice of edges in panel (d) is likely to be more useful in applications. The copula assigned to 11,12 controls the serial dependence in X_{1t}. Interesting patterns of serial dependence may be generated using a suitable copula; for instance, a nonexchangeable copula may be used to model Edgeworth price cycles or other time irreversible dynamics, as in Beare and Seo (2014). We cannot explicitly specify the univariate dynamics of either component of $\mathbf{X}_t$ if we commence our vine with the tree in panel (b). Instead, they are implicitly determined by the copulae assigned to other edges.

Another potential advantage of commencing our vine with the tree in panel (d) is that the following tree will include the edge 21,12|11. The copula associated with this edge controls the Granger-causal effect (Granger, 1969) of X_{2t} upon X_{1t}. In particular, X_{2t} will not Granger-cause X_{1t} if and only if we assign the independence copula to the edge 21,12|11. There does not seem to be a simple way to directly impose or exclude Granger causality if our vine commences with the tree in panel (b).

To illustrate the use of vine copula specifications based on the initial tree in Figure 2.1 (d), we randomly generated two stationary bivariate Markov chains of length 50 with particular dependence structures. These chains are displayed in

Figure 2.3. The first chain, shown in panel (a), is a stationary vector autoregressive process of order one, or VAR(1) process, with Gaussian innovations. It follows the recursion $\mathbf{X}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$, where $\boldsymbol{\varepsilon}_t \sim \text{iid}\mathcal{N}(0, \boldsymbol{\Omega})$, and parameters are given by

$$\boldsymbol{\mu} = \begin{bmatrix} 3.5 \\ -0.5 \end{bmatrix}, \quad \boldsymbol{\Phi} = \begin{bmatrix} 0.3 & 0.2 \\ 0.1 & 0.1 \end{bmatrix}, \quad \boldsymbol{\Omega} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

For this process, the joint distribution of $(X_{11}, X_{21}, X_{12}, X_{22})$ is Gaussian. Therefore, in the corresponding vine copula specification[6] based on the initial tree in Figure 2.2 (d), all cdfs in $\mathcal{F}$ and copulae in $\mathcal{C}$ are Gaussian. With some routine calculations we determine their precise form: F_{11} and F_{12} are $\mathcal{N}(5, 1.6638^2)$ cdfs, F_{21} and F_{22} are $\mathcal{N}(0, 1.2377^2)$ cdfs, and the correlation parameters determining the copulae in $\mathcal{C}$ are given by

$$\rho_{11,21} = 0.7243, \quad \rho_{11,12} = 0.4078, \quad \rho_{12,22} = 0.7243,$$
$$\rho_{21,12|11} = 0.1123, \quad \rho_{11,22|12} = -0.1406, \quad \rho_{21,22|11,12} = -0.0144.$$

To obtain the Markov chain shown in panel (b) of Figure 2.3, we made a single modification to the vine copula specification of $(X_{11}, X_{21}, X_{12}, X_{22})$ in panel (a): we replaced the Gaussian copula assigned to edge 11,12, which controls the serial dependence in X_{1t}, with an asymmetric Gumbel copula of the form

$$C(u, v) = u^{1-\alpha}v^{1-\beta}\exp\left(-\left((-\alpha\ln u)^\gamma + (-\beta\ln v)^\gamma\right)^{1/\gamma}\right), \tag{2.1}$$

with parameters $\alpha = 0.99$, $\beta = 0.5$ and $\gamma = 50$. This family of copulae was used by Beare and Seo (2014) to generate stationary univariate time irreversible Markov chains characterized by frequent small decreases and less frequent large increases, a feature typical of many price series. We see in panel (b) that the time path of X_{1t} does indeed exhibit this feature, as does the time path of X_{2t} to a somewhat lesser

[6] Since the distribution of $(X_{11}, X_{21}, X_{12}, X_{22})$ is multivariate Gaussian, the result of Stöber et al. (2013) mentioned at the end of Section 2.2 ensures that it can be reproduced using a vine copula.

extent. X_{2t} inherits the irreversible nature of X_{1t} due to the dependence between the two variables.

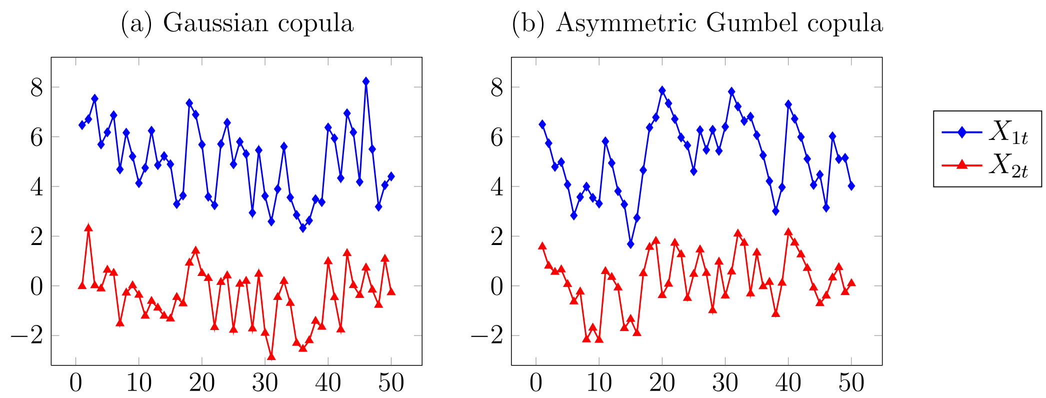


Figure 2.3: Two stationary bivariate first-order Markov chains
The bivariate chain in panel (a) is a stationary VAR(1) process with Gaussian innovations. To obtain the bivariate chain in panel (b), we first determined the vine copula specification for $(X_{11}, X_{21}, X_{12}, X_{22})$ in panel (a), with initial tree chosen as in Figure 2.2 (d). Then we modified the copula assigned to the edge 11,12, changing it from a Gaussian copula to an asymmetric Gumbel copula. Parameter values are given in the main text.

Though the bivariate chains in panels (a) and (b) of Figure 2.3 exhibit distinct patterns of serial dependence, the differences are attributable to a single change in our vine copula specification for the joint distribution of $(X_{11}, X_{21}, X_{12}, X_{22})$. Aspects of this distribution determined by parts of the vine copula specification other than the copula assigned to edge 11,12 are identical in the two panels. In particular, the bivariate stationary distributions of the two Markov chains are equal to one another. In effect, the vine copula specification has allowed us to inject a strongly nonlinear form of serial dependence into a stationary VAR(1) model without affecting its stationary distribution. Other forms of nonlinear dependence could be introduced using different copula functions, or by modifying the copulae assigned to other edges.

2.3.2 Bivariate second-order Markov chains

Suppose now that our stationary bivariate process $\mathbf{X}_1, \mathbf{X}_2, \ldots$ is second-order Markov rather than first-order Markov. In this case the finite dimensional distributions of the chain are determined by the joint distribution of $(X_{11}, X_{21}, X_{12}, X_{22}, X_{13}, X_{23})$, which we can model using a vine copula specification on the set of nodes $N_1 = \{1,2\} \times \{1,2,3\}$. We will focus attention on the regular vine $\mathcal{V}$ displayed in Figure 2.4. The first two trees of $\mathcal{V}$ resemble an extension of those obtained in the 2×2 case when we commence with the initial tree shown in Figure 2.2 (d). The nodes in the second tree of $\mathcal{V}$ have degree no greater than two, so trees three through five are uniquely determined by the proximity condition for a regular vine. Clearly $\mathcal{V}$ is neither a C-vine nor D-vine. We shall refer to it as an M-vine, because - with a bit of imagination - the first and second trees displayed in Figure 2.4 are roughly M-shaped. Another useful mnemonic may be *M for Markov*, though not all vine structures suitable for Markov models are M-vines. A general definition and discussion of M-vines on $m \times n$ arrays of nodes will be given in Section 2.3.3.

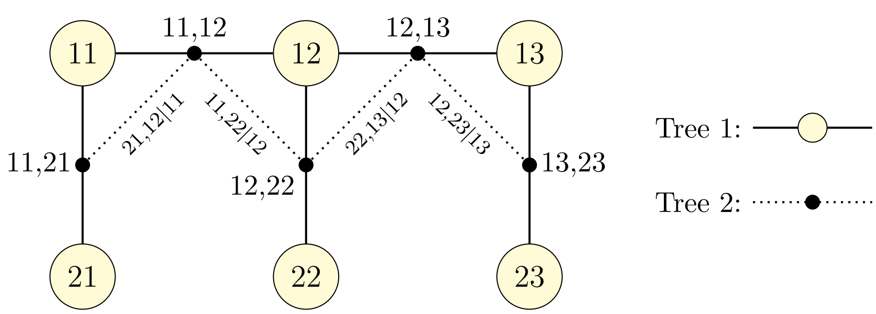


Figure 2.4: M-vine on a 2×3 array of nodes
Each node in the second tree of the M-vine has degree of at most two, so the remaining three trees are determined by the proximity condition. The third tree has edges 21,22|11,12, 11,13|12,22 and 22,23|12,13. The fourth tree has edges 21,13|11,12,22 and 11,23|12,22,13. The fifth tree has the single edge 21,23|11,12,22,13.

Though the structure of $\mathcal{V}$ may appear arbitrary, it is in fact the unique regular vine on N_1 satisfying three useful properties: (i) it includes the edge 11,21, so that we may directly specify the contemporaneous relationship between X_{1t} and

X_{2t}; (ii) it includes the edge 11,12, so that we may directly specify the serial dependence between consecutive realizations of X_{1t}; and (iii) we can impose stationarity in a vine copula specification built on $\mathcal{V}$ by requiring various edges of $\mathcal{V}$ to be assigned the same copula.[7] To see why this is true, we reason as follows. Let $\mathcal{V}'$ be a regular vine on N_1 satisfying properties (i) through (iii). Since the first tree of $\mathcal{V}'$ includes the edge 11,21, it must also include the edges 12,22 and 13,23, because stationarity requires that the copulae associated with these three edges are equal. Similarly, since the first tree of $\mathcal{V}'$ includes the edge 11,12, it must also include the edge 12,13, because stationarity requires that the copulae associated with these two edges are equal. This pins down all five edges in the first tree of $\mathcal{V}'$, and they are precisely those in the first tree of $\mathcal{V}$. In the second tree of $\mathcal{V}'$, we are forced to include the edges 21,12|11 and 12,23|13 due to the proximity condition. But if we include 21,12|11 then we must also include 22,13|12, because stationarity requires that the copulae associated with these two edges are equal, and similarly if we include 12,23|13 then we must also include 11,22|12 for the same reason. This pins down all four edges in the second tree of $\mathcal{V}'$, and they are precisely those in the second tree of $\mathcal{V}$. Each node in the second tree has degree of at most two, so the proximity condition pins down the edges in all remaining trees of $\mathcal{V}'$ and $\mathcal{V}$. Thus $\mathcal{V}$ is the unique regular vine on N_1 satisfying properties (i) through (iii).

2.3.3 Multivariate higher-order Markov chains

In the previous two subsections we explored the use of vine copulae in models of stationary bivariate first- and second-order Markov chains. We proceed now to the general case where our stationary sequence $\mathbf{X}_1, \mathbf{X}_2, \ldots$ takes values in $\mathbb{R}^m$ and is Markov of order p. Our main results are given here. We will be somewhat more formal than in the previous two subsections, which were primarily motiva-

[7] More precisely, condition (iii) means that if $\mathcal{V}$ includes an edge e with conditioned and conditioning sets contained in the first two columns of N_1, then it should also include another edge e' whose conditioned and conditioning sets are the same as those of e, but translated one column to the right.

tional in intent.

Throughout this subsection we consider vines on the $m \times n$ array of nodes $N_1 = \{1, \ldots, m\} \times \{1, \ldots, n\}$ for some $m, n \in \mathbb{N}$ with $n \geq 2$. Our first task is to extend the M-vine structure introduced earlier in the 2×3 context to the more general $m \times n$ context. To do this, we first need to introduce the notion of the restriction of a regular vine to some subset of nodes $\tilde{N}_1 \subset N_1$.

Definition 2.3.1. Let $\mathcal{V}$ be a regular vine on N_1 with trees $T_k = (N_k, E_k)$, $k = 1, \ldots, mn-1$, and let $\tilde{N}_1$ be a subset of N_1 with $\tilde{q} \geq 2$ elements. Let $\tilde{E}_1 = E_1 \cap \binom{\tilde{N}_1}{2}$ and $\tilde{T}_1 = (\tilde{N}_1, \tilde{E}_1)$, and if $\tilde{q} \geq 3$ then for $k = 2, \ldots, \tilde{q}-1$ let $\tilde{N}_k = \tilde{E}_{k-1}$, $\tilde{E}_k = E_k \cap \binom{\tilde{N}_k}{2}$ and $\tilde{T}_k = (\tilde{N}_k, \tilde{E}_k)$. We refer to $\tilde{\mathcal{V}} = (\tilde{T}_1, \ldots, \tilde{T}_{\tilde{q}-1})$ as the *restriction of* $\mathcal{V}$ *to* $\tilde{N}_1$.

Intuitively, the restriction of $\mathcal{V}$ to a subset $\tilde{N}_1$ of N_1 is what we are left with after deleting all nodes in the first tree that do not belong to $\tilde{N}_1$, all edges in the first tree that connect to one of the deleted nodes, and all edges in subsequent trees whose complete union includes one of the deleted nodes. This may or may not be a regular vine on $\tilde{N}_1$; indeed, each $\tilde{T}_k$ need not be a tree, and may have empty edge and (for $k \geq 2$) node sets.

With Definition 2.3.1 in hand, we proceed to our definition of an M-vine on N_1.

Definition 2.3.2. A regular vine $\mathcal{V}$ on N_1 with trees $T_k = (N_k, E_k)$, $k = 1, \ldots, mn-1$, is said to be an *M-vine* if it satisfies the following two conditions.

(i) $E_1 = \Big\{\{(i,s),(j,t)\} \in \binom{N_1}{2} : (i = j-1 \text{ and } s = t) \text{ or } (i = j = 1 \text{ and } s = t-1)\Big\}$.

(ii) For each pair of adjacent columns $A_t = \cup_{i=1}^m \{(i,t),(i,t+1)\}$, $t = 1, \ldots, n-1$, the restriction of $\mathcal{V}$ to A_t is a D-vine.

A vine copula specification $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ on N_1 is said to be an *M-vine copula specification on* N_1 if $\mathcal{V}$ is an M-vine.

Remark 2.3.1. It is clear from Definition 2.3.2 that if $\mathcal{V}$ is an M-vine on N_1, then for any collection of adjacent columns $A_{s,u} = \{(i,t) \in N_1 : s \leq t \leq u\}$, $1 \leq s < u \leq n$, the restriction of $\mathcal{V}$ to $A_{s,u}$ is also an M-vine.

Theorem 2.3.1. *There exists a unique M-vine on* N_1.

In Figure 2.5 we display the first three trees of the M-vine on a 3×3 array of nodes. We recommend that in a first attempt at reading the following proof of Theorem 2.3.1, the reader mentally substitutes $m = n = 3$, and refers closely to Figure 2.5 at all stages. It may also be useful to consult Figure 2.4 for the case where $m = 2$ and $n = 3$.

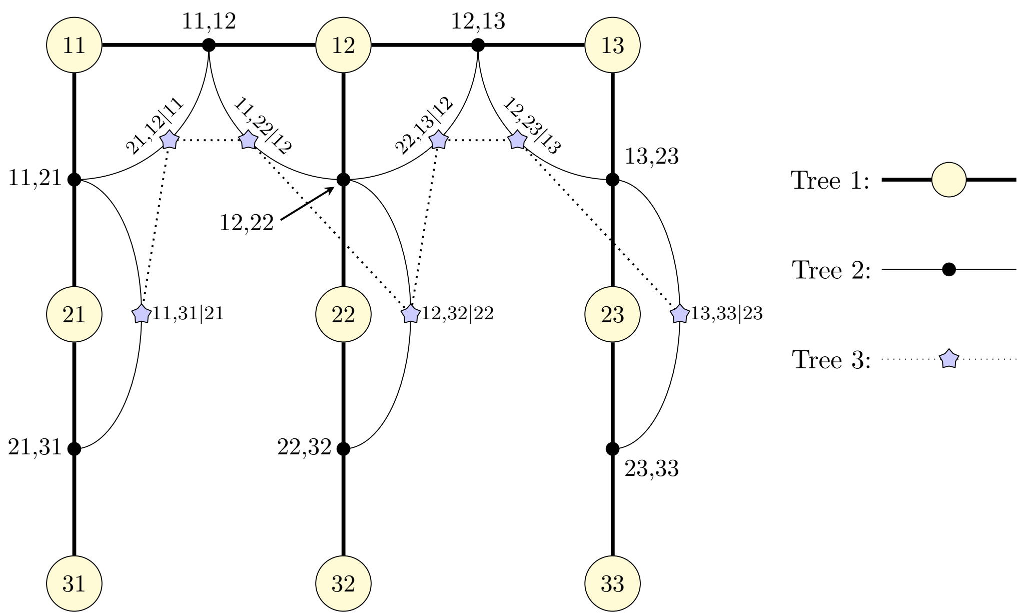


Figure 2.5: M-vine on a 3×3 array of nodes
We display the first three trees of the unique M-vine on a 3×3 array of nodes. Since each node in the third tree has degree of at most two, the remaining five trees in the M-vine are uniquely determined by the proximity condition for a regular vine. Trees three through eight collectively form a D-vine.

Proof of Theorem 2.3.1. We assume that $m \geq 2$ and $n \geq 3$, since otherwise the unique M-vine on N_1 is simply the unique D-vine on N_1 with edges E_1 in the first tree. For $t = 1, \ldots, n-1$, let $\mathcal{V}_t$ denote the restriction of $\mathcal{V}$ to A_t. We need to show that the trees $T_2, \dots, T_{mn-1}$ are uniquely determined by the connectedness and proximity conditions for a regular vine and the requirement that each $\mathcal{V}_t$ is a D-vine.

For $k = 2, \dots, m$, the requirement that $\mathcal{V}_1$ is a D-vine determines $2m - k$ of the edges in T_k. The requirement that $\mathcal{V}_2$ is a D-vine also determines $2m - k$ of the edges in T_k, but $m - k$ of these edges are shared with $\mathcal{V}_1$, so $\mathcal{V}_1$ and $\mathcal{V}_2$ jointly determine $3m - k$ of the edges in T_k. If $n \geq 4$ then each of the remaining D-vines $\mathcal{V}_3$ through $\mathcal{V}_{n-1}$ determine an additional m edges in T_k, so in total we have determined $mn - k$ of the edges in T_k. For $\mathcal{V}$ to be a regular vine, T_k must contain exactly $mn - k$ edges, so we have uniquely determined T_k. Moreover, T_k is connected as it is composed of $n - 1$ overlapping D-vine trees (the k^{th} trees of $\mathcal{V}_t$ and $\mathcal{V}_{t+1}$ have $m - k + 1$ common nodes) and the edges in T_k must satisfy the proximity condition since each belongs to at least one of the D-vines $\mathcal{V}_1$ through $\mathcal{V}_{n-1}$.

The nodes in T_m have degree of at most two. To see this, note that all but $n - 2$ of the nodes in T_m belong to exactly one of the D-vines $\mathcal{V}_t$, while the remaining $n - 2$ nodes belong to exactly two of the D-vines $\mathcal{V}_t$. Since all nodes in a D-vine have degree of at most two, the nodes in T_k belonging to a single $\mathcal{V}_t$ have degree of at most two. Denote the $n - 2$ nodes shared by two of the D-vines by a_t, $t = 1, \dots, n - 2$. The node a_t belongs to the m^{th} trees of $\mathcal{V}_t$ and $\mathcal{V}_{t+1}$. Viewed as a node in the m^{th} tree of $\mathcal{V}_t$ or of $\mathcal{V}_{t+1}$, a_t includes the node $(m, t + 1)$ in its conditioned set, which is a node of degree one, and so a_t itself has degree one.[8] Therefore, viewed as a node in T_m, a_t has degree two. We conclude that all nodes in T_m have degree of at most two. Consequently, the proximity condition uniquely determines all of the remaining trees $T_{m+1}, \dots, T_{mn-1}$, and in fact $(T_m, \dots, T_{mn-1})$ is the unique D-vine with initial tree T_m. $\square$

We see from the proof of Theorem 2.3.1 that the practical construction of

[8] Here we apply the following easily demonstrated property of D-vines: if a node in tree $k \geq 2$ of a D-vine has a node of degree one in its conditioned set, then the node itself has degree one.

M-vines is straightforward. The first tree is given explicitly, the requirement that the restriction of the M-vine to each pair of adjacent columns be a D-vine uniquely determines trees 2 through m, and the remaining trees are uniquely determined by the proximity condition. The structure of the first tree allows us to directly specify the serial dependence between consecutive realizations of X_{1t}, but not between consecutive realizations of other elements of $\mathbf{X}_t$, so in applications it may be important to consider the ordering of the elements of $\mathbf{X}_t$.

It is a simple matter to impose stationarity on the distribution realized by an M-vine copula specification. We have seen this already in Section 2.3.2 in the 2×3 case, and in Section 2.3.1 in the 2×2 case; note that the vine with first tree displayed in panel (d) of Figure 2.2 is in fact an M-vine. For the more general $m \times n$ case, it will be convenient to introduce two new concepts. The first is an obvious extension of Definition 2.3.1.

Definition 2.3.3. Let $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ be a vine copula specification on N_1, and let $\tilde{N}_1$ be a subset of N_1 with $\tilde{q} \geq 2$ elements. The *restriction of* $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ *to* $\tilde{N}_1$ is the triple $(\tilde{\mathcal{F}}, \tilde{\mathcal{V}}, \tilde{\mathcal{C}})$, where $\tilde{\mathcal{F}}$ consists of those members of $\mathcal{F}$ associated with members of $\tilde{N}_1$, $\tilde{\mathcal{V}}$ is the restriction of $\mathcal{V}$ to $\tilde{N}_1$, and $\tilde{\mathcal{C}}$ consists of those members of $\mathcal{C}$ associated with edges in $\tilde{\mathcal{V}}$.

The second new concept is translation invariance. We will say that a vine copula specification is translation invariant if it assigns the same cdf to any two nodes in the same row of N_1, and if it assigns the same copula to any two edges whose conditioned and conditioning sets differ only by translation across columns.

Definition 2.3.4. A vine copula specification $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ on N_1 is said to be *translation invariant* if it satisfies the following two conditions.

(i) $F_{i,t} = F_{i,t+1}$ for all $i = 1, \ldots, m$ and all $t = 1, \ldots, n-1$.

(ii) For any $e, e' \in \cup_{k=1}^{mn-1} E_k$ such that $D_e = D_{e'} + (0, s)$ and $\{a_e, b_e\} = \{a_{e'}, b_{e'}\} + (0, s)$ for some $s \geq 1$, we have $C_{a_e, b_e | D_e} = C_{a_{e'}, b_{e'} | D_{e'}}$.

Remark 2.3.2. For an M-vine copula specification, translation invariance limits the free assignment of copulae to the edges of $\mathcal{V}$. To the $mn-k$ edges in T_k, $k=1,\dots,mn-1$, we may freely assign $\min\{m, mn-k\}$ copulae, with no two assignments going to edges whose conditioned and conditioning sets differ only by translation across columns. The remaining copulae are determined by translation invariance. For instance, in Figure 2.5 we see that when $m=n=3$ we may freely assign copulae to the edges 11,21, 21,31 and 11,12 in the first tree. Translation invariance then requires that edges 12,22 and 13,23 are assigned the same copula as edge 11,21, that edges 22,32 and 23,33 are assigned the same copula as edge 21,31, and that edge 12,13 is assigned the same copula as edge 11,12. Similarly, in the second tree we may freely assign copulae to the edges 11,31|21, 12,21|11 and 11,22|12. Translation invariance then requires that edges 12,32|22 and 13,33|23 are assigned the same copula as edge 11,31|21, that edge 22,13|12 is assigned the same copula as edge 21,12|11, and that edge 12,23|13 is assigned the same copula as edge 11,22|12.

A translation invariant M-vine copula specification on N_1 is realized by the distribution of a stationary array of random variables. To be clear, we define stationarity for an $m \times n$ array of random variables.

Definition 2.3.5. The $m \times n$ array of random variables $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ is said to be *stationary* if, for any $t=1,\dots,n-1$, any $u=t,\dots,n-1$, and any $s=1,\dots,n-u$, the joint distributions of the $m \times (u-t+1)$ arrays $(\mathbf{X}_t, \dots, \mathbf{X}_u)$ and $(\mathbf{X}_{t+s}, \dots, \mathbf{X}_{u+s})$ are equal.

The following result demonstrates the equivalence of translation invariance and stationarity for M-vine copula specifications.

Theorem 2.3.2. *Suppose the joint distribution of the $m \times n$ array of random variables $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ realizes the M-vine copula specification $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ on N_1. Then $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ is stationary if and only if $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ is translation invariant.*

Proof of Theorem 2.3.2. It is clear that stationarity of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ implies translation invariance of $(\mathcal{F}, \mathcal{V}, \mathcal{C})$. To show the reverse implication, consider the restrictions of $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ to $A_{1,n-1}$ and to $A_{2,n}$, the first and last $n-1$ columns of N_1. Recalling Remark 2.3.1, each of these restrictions forms an M-vine copula specification on an $m \times (n-1)$ array of nodes, with identical vine structure in view of the uniqueness demonstrated in Theorem 2.3.1. Therefore, due to the translation invariance of $(\mathcal{F}, \mathcal{V}, \mathcal{C})$, the two restrictions are identical other than a trivial difference in the labeling of columns, and so are realized by the same joint distribution. Consequently, $(\mathbf{X}_1, \ldots, \mathbf{X}_{n-1})$ and $(\mathbf{X}_2, \ldots, \mathbf{X}_n)$ have the same joint distribution, and it follows easily that $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is stationary. $\square$

Remark 2.3.3. Stationarity and translation invariance may not be equivalent for vine structures other than the M-vine. Consider the D-vine on a 2×2 array of nodes whose first tree is displayed in Figure 2.2 (a). This vine does not include any two edges e, e' satisfying the condition in part (ii) of Definition 2.3.4. Therefore, any vine copula specification built on this vine will satisfy translation invariance provided that $F_{11} = F_{12}$ and $F_{21} = F_{22}$. Clearly this is not enough to ensure stationarity of the associated quadrivariate distribution.

We noted at the beginning of this section that, for a stationary $\mathbb{R}^m$-valued Markov chain $\mathbf{X}_1, \mathbf{X}_2, \ldots$ of order p, all finite dimensional distributions are uniquely determined by the joint distribution of $(\mathbf{X}_1, \ldots, \mathbf{X}_{p+1})$. Any translation invariant M-vine copula specification on an $m \times (p+1)$ array of nodes can be used to model this distribution, which uniquely determines a transition law we can use to simulate a stationary Markov chain of arbitrary length. Alternatively, if we seek to simulate n consecutive realizations $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, where $n \geq p+1$, we can instead use an M-vine copula specification to model the entire joint distribution of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$. This facilitates simulation of the Markov chain using existing algorithms for sampling from regular vine copula specifications; see e.g. Dißmann et al. (2013, Algorithm 2.2). For such an approach to be valid, we need to impose conditions beyond translation invariance on our M-vine so that $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is not only stationary, but also Markov of order p.

Definition 2.3.6. The $m \times n$ array of random variables $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is said to be *Markov of order* p, or p*-Markov* if, for any $t = p+1, \ldots, n$, the joint distribution of $\mathbf{X}_t$ conditional on $\mathbf{X}_1, \ldots, \mathbf{X}_{t-1}$ is equal to the joint distribution of $\mathbf{X}_t$ conditional on $\mathbf{X}_{t-p}, \ldots, \mathbf{X}_{t-1}$.

We shall refer to the condition we impose on an M-vine copula specification to ensure that our model has the desired Markov property as p-independence.

Definition 2.3.7. An M-vine copula specification $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ on N_1 is said to be p-*independent* if it assigns independence copulae to all edges in $\mathcal{V}$ that do not belong to the restriction of $\mathcal{V}$ to some collection of $p+1$ adjacent columns in N_1.

Remark 2.3.4. Every M-vine copula specification on N_1 is p-independent for $p = n-1$. For $p < n-1$, the p-independence property forces us to assign independence copulae to $k(n-p-1)$ edges in tree T_{mp+k} for $k = 1, \ldots, m-1$, and to all edges in trees $T_{mp+m}, \ldots, T_{mn-1}$. These are precisely the edges whose conditioned elements are separated by p or more adjacent columns of N_1, which are necessarily contained in the corresponding conditioning set.

Remark 2.3.5. Let $\tilde{N}_1 = \{1, \ldots, m\} \times \{1, \ldots, p+1\}$. Any translation invariant M-vine copula specification on $\tilde{N}_1$ is the restriction to $\tilde{N}_1$ of a unique translation invariant p-independent M-vine copula specification on N_1. Conversely, any translation invariant p-independent M-vine copula specification on N_1 is uniquely determined by its restriction to $\tilde{N}_1$. Since that restriction is translation invariant, it is (recall Remark 2.3.2) itself uniquely determined by the m cdfs $F_{1,1}, \ldots, F_{m,1}$, and the assignment of copulae to $\min\{m, mp+m-k\}$ edges in tree T_k, $k = 1, \ldots, mp+m-1$, with no two assignments going to edges whose conditioned and conditioning sets differ only by translation across columns.

The following result identifies the link between p-independence and the Markov property.

Theorem 2.3.3. *Suppose the joint distribution of the $m \times n$ array of random variables $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ realizes the M-vine copula specification $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ on N_1. Then $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is p-Markov if and only if $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ is p-independent.*

Proof of Theorem 2.3.3. If $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ is not p-independent, then there is an edge in $\mathcal{V}$ to which we do not assign the independence copula, and that has (recall Remark 2.3.4) its conditioned elements (i, s) and (j, t) separated by p or more adjacent columns of N_1 in its conditioning set (so that $t > s + p$). But then we do not have conditional independence of $\mathbf{X}_s$ and $\mathbf{X}_t$ given $\mathbf{X}_{s+1}, \ldots, \mathbf{X}_{t-1}$, and so $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ cannot be p-Markov. Suppose instead that $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ is p-independent. For $s = 1, \ldots, n-1$ and $t = s+1, \ldots, n$, let $B_{s,t}$ denote the collection of edges in $\mathcal{V}$ that have one conditioned element in column t of N_1, and the other conditioned element in one of columns s through $t-1$. In view of Theorem 2.2.1, for any $t = p+1, \ldots, n$, the conditional density of $\mathbf{X}_t$ given $\mathbf{X}_1, \ldots, \mathbf{X}_{t-1}$ can be written as

$$f_{\mathbf{X}_t|\mathbf{X}_1,\ldots,\mathbf{X}_{t-1}} = \frac{f_{\mathbf{X}_1,\ldots,\mathbf{X}_t}}{f_{\mathbf{X}_1,\ldots,\mathbf{X}_{t-1}}} = f_{\mathbf{X}_t} \prod_{e \in B_{1,t}} c_{a_e,b_e|D_e}\left(F_{a_e|D_e}, F_{b_e|D_e}\right),$$

while the conditional density of $\mathbf{X}_t$ given $\mathbf{X}_{t-p}, \ldots, \mathbf{X}_{t-1}$ can be written as

$$f_{\mathbf{X}_t|\mathbf{X}_{t-p},\ldots,\mathbf{X}_{t-1}} = \frac{f_{\mathbf{X}_{t-p},\ldots,\mathbf{X}_t}}{f_{\mathbf{X}_{t-p},\ldots,\mathbf{X}_{t-1}}} = f_{\mathbf{X}_t} \prod_{e \in B_{t-p,t}} c_{a_e,b_e|D_e}\left(F_{a_e|D_e}, F_{b_e|D_e}\right).$$

The ratio of these two conditional densities is therefore equal to the product of all of the $c_{a_e,b_e|D_e}(F_{a_e|D_e}, F_{b_e|D_e})$ associated with edges that have one conditioned element in column t of N_1, and the other conditioned element in one of columns 1 through $t-p-1$. Any such edges cannot belong to the restriction of $\mathcal{V}$ to p adjacent columns of N_1, and so the associated copulae must be independence copulae owing to the p-independence property of $(\mathcal{F}, \mathcal{V}, \mathcal{C})$. Since the independence copula has density identically equal to one, we find that $f_{\mathbf{X}_t|\mathbf{X}_1,\ldots,\mathbf{X}_{t-1}} = f_{\mathbf{X}_t|\mathbf{X}_{t-p},\ldots,\mathbf{X}_{t-1}}$, which demonstrates that $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is p-Markov. We conclude that $(\mathcal{F}, \mathcal{V}, \mathcal{C})$ is p-

independent if and only if $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is p-Markov. □

2.4 Estimation of an M-vine copula specification

In this section we briefly discuss and illustrate the application of existing semiparametric estimation procedures to Markov models specified using M-vines. Semiparametric estimation for stationary copula-based Markov models was introduced by Chen and Fan (2006) for univariate time series, and extended by Rémillard et al. (2012) to the multivariate case. Both papers draw on earlier work by Genest et al. (1995) in the iid context. The semiparametric procedure for stationary multivariate Markov chains leaves the specification of the m univariate marginal distributions $F_{1,1}, \ldots, F_{m,1}$ nonparametric, and requires a parametric model for the copula of the $m \times (p+1)$ array[9] of random variables $(\mathbf{X}_1, \ldots, \mathbf{X}_{p+1})$. M-vines provide a convenient and flexible way to obtain such a parametric model.

Let $(\tilde{\mathcal{F}}, \tilde{\mathcal{V}}, \tilde{\mathcal{C}}_{\boldsymbol{\theta}})$ be a translation invariant M-vine copula specification for the joint distribution of $(\mathbf{X}_1, \ldots, \mathbf{X}_{p+1})$, with $\boldsymbol{\theta} \in \mathbb{R}^d$ a vector of parameters to be estimated. The number of parameters d depends on the functional form of the bivariate copulae in $\tilde{\mathcal{C}}_{\boldsymbol{\theta}}$. If each copula is determined by a single parameter, then in view of the translation invariance condition (recall Remark 2.3.2) we will have $d = m\left(mp + \frac{1}{2}(m-1)\right)$ parameters to estimate in total. Let $\boldsymbol{\theta}_0 \in \mathbb{R}^d$ denote the unknown true parameter vector. We wish to estimate $\boldsymbol{\theta}_0$ using the observed $m \times n$ array $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, where $n > p+1$. Let $(\mathcal{F}, \mathcal{V}, \mathcal{C}_{\boldsymbol{\theta}})$ be the unique (recall Remark 2.3.5) translation invariant p-independent M-vine copula specification on $N_1 = \{1, \ldots, m\} \times \{1, \ldots, n\}$ whose restriction to $\tilde{N}_1 = \{1, \ldots, m\} \times \{1, \ldots, p+1\}$ is $(\tilde{\mathcal{F}}, \tilde{\mathcal{V}}, \tilde{\mathcal{C}}_{\boldsymbol{\theta}})$. In view of Theorem 2.2.1, we see that the log-likelihood of our sample

[9] Rémillard et al. (2012) in fact take $p = 1$, but it is clear that an obvious modification of their procedure may be applied to higher-order Markov processes.

is

$$\begin{aligned}\ell_n(\boldsymbol{\theta}) &= \sum_{i=1}^{m}\sum_{t=1}^{n} \ln f_{i,t}(X_{i,t}) \\ &+ \sum_{k=1}^{mp+m-1}\sum_{e\in E_k} \ln c_{a_e,b_e|D_e}(F_{a_e|D_e}(X_{a_e}|\mathbf{X}_{D_e};\boldsymbol{\theta}), F_{b_e|D_e}(X_{b_e}|\mathbf{X}_{D_e};\boldsymbol{\theta});\boldsymbol{\theta}),\end{aligned} \tag{2.2}$$

where E_k is the collection of edges of T_k, the k^{th} tree of $\mathcal{V}$, and $\mathbf{X}_{D_e}$ denotes the random variables $(X_{i,t} : (i,t) \in D_e)$. The parameter vector $\boldsymbol{\theta}$ enters our log-likelihood function through the copula densities $c_{a_e,b_e|D_e}$, and also through the conditional cdfs $F_{a_e|D_e}$ and $F_{b_e|D_e}$ determined by the members of $\mathcal{F}$ and $\mathcal{C}_{\boldsymbol{\theta}}$ as described in Theorem 2.2.1. Note that p-independence of $(\mathcal{F}, \mathcal{V}, \mathcal{C}_{\boldsymbol{\theta}})$ implies (recall Remark 2.3.4) that all copulae assigned to the edges of trees $T_{mp+m}, \ldots, T_{mn-1}$ are independence copulae, so in (2.2) we sum only over $k = 1, \ldots, mp + m - 1$.

The log-likelihood given in (2.2) does not provide a feasible means to estimate $\boldsymbol{\theta}_0$ because we wish to be nonparametric in our treatment of the marginal cdfs in $\mathcal{F}$. Following Rémillard et al. (2012) we may commence by estimating these cdfs using the empirical distributions of the m observed univariate time series:

$$\hat{F}_{i,t}(\cdot) = \frac{1}{n+1}\sum_{s=1}^{n} 1(X_{i,s} \leq \cdot), \quad i = 1, \ldots, m, \quad t = 1, \ldots, n. \tag{2.3}$$

We then estimate $\boldsymbol{\theta}_0$ by maximizing the second stage log-likelihood

$$\ell_n^*(\boldsymbol{\theta}) = \sum_{k=1}^{mp+m-1}\sum_{e\in E_k} \ln c_{a_e,b_e|D_e}(\hat{U}_{a_e|D_e}(\boldsymbol{\theta}), \hat{V}_{b_e|D_e}(\boldsymbol{\theta});\boldsymbol{\theta}). \tag{2.4}$$

Here, $\hat{U}_{a_e|D_e}(\boldsymbol{\theta})$ and $\hat{V}_{b_e|D_e}(\boldsymbol{\theta})$ are determined in the following recursive fashion. If $e \in E_1$, so that $D_e = \varnothing$, we set $\hat{U}_{a_e|D_e}(\boldsymbol{\theta}) = \hat{F}_{a_e}(X_{a_e})$ and $\hat{V}_{b_e|D_e}(\boldsymbol{\theta}) = \hat{F}_{b_e}(X_{b_e})$, which do not depend on $\boldsymbol{\theta}$. If $e \in E_k$ for some $k \geq 2$ then we set

$$\hat{U}_{a_e|D_e}(\boldsymbol{\theta}) = \frac{\partial}{\partial v} C_{a_{e'},b_{e'}|D_{e'}}\left(\hat{U}_{a_{e'}|D_{e'}}(\boldsymbol{\theta}), v; \boldsymbol{\theta}\right)\Big|_{v=\hat{V}_{b_{e'}|D_{e'}}(\boldsymbol{\theta})}$$

and

$$\hat{V}_{b_e|D_e}(\boldsymbol{\theta}) = \frac{\partial}{\partial u} C_{a_{e''},b_{e''}|D_{e''}}\left(u, \hat{V}_{b_{e''}|D_{e''}}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)\bigg|_{u=\hat{U}_{a_{e''}|D_{e''}}(\boldsymbol{\theta})}$$

where e' and e'' are the edges of T_{k-1} connected by e, with $a_e = a_{e'}$ and $b_e = b_{e''}$. The value of $\boldsymbol{\theta}$ at which $\ell_n^*(\boldsymbol{\theta})$ is maximized provides our estimate $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_0$. Rémillard et al. (2012) provide technical conditions under which $n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ has a limiting multivariate normal distribution centered at zero. As discussed by those authors, the limiting covariance matrix takes a complicated form due to the non-parametric estimation of marginal cdfs, but may be consistently estimated using a parametric bootstrap technique. Implementation of the bootstrap using our M-vine model may be achieved using the algorithm developed by Dißmann et al. (2013) for sampling from general regular vines.

In practice the numerical maximization of $\ell_n^*(\boldsymbol{\theta})$ can be computationally challenging when d is moderately large. A less efficient but computationally much simpler estimate of $\boldsymbol{\theta}_0$ can be obtained using the method of stepwise semiparametric (SSP) estimation, which was proposed by Aas et al. (2009) and studied in detail in a recent pair of papers by Hobæk Haff (2012, 2013). To apply SSP estimation we partition our parameter vector as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m+mp-1})$ in such a way that the copulae associated with edges of T_k depend only on the subvector of parameters $\boldsymbol{\theta}_k$. We then proceed as follows. First, for each $e \in E_1$, we set $\check{U}_{a_e|D_e} = \hat{F}_{a_e}(X_{a_e})$ and $\check{V}_{b_e|D_e} = \hat{F}_{b_e}(X_{b_e})$, with $\hat{F}_{a_e}$ and $\hat{F}_{b_e}$ given by (2.3). Next, iterating over $k = 1, \ldots, m + mp - 1$, we obtain $\check{\boldsymbol{\theta}}_{n,k}$ as the value of $\boldsymbol{\theta}_k$ that maximizes

$$\ell_{n,k}^*(\boldsymbol{\theta}_k) = \sum_{e \in E_k} \ln c_{a_e,b_e|D_e}(\check{U}_{a_e|D_e}, \check{V}_{b_e|D_e}; \boldsymbol{\theta}_k),$$

and then (if $k < m + mp - 1$) for each $e \in T_{k+1}$ compute

$$\check{U}_{a_e|D_e} = \frac{\partial}{\partial v} C_{a_{e'},b_{e'}|D_{e'}}\left(\check{U}_{a_{e'}|D_{e'}}, v; \check{\boldsymbol{\theta}}_{n,k}\right)\bigg|_{v=\check{V}_{b_{e'}|D_{e'}}}$$

and

$$\check{V}_{b_e|D_e} = \frac{\partial}{\partial u} C_{a_{e''},b_{e''}|D_{e''}} \left(u, \check{V}_{b_{e''}|D_{e''}}; \check{\boldsymbol{\theta}}_{n,k}\right)\Big|_{u=\check{U}_{a_{e''}|D_{e''}}}$$

where e' and e'' are the edges of T_k connected by e, with $a_e = a_{e'}$ and $b_e = b_{e''}$. After completing the iteration over k we are left with our SSP estimate $\check{\boldsymbol{\theta}}_n = (\check{\boldsymbol{\theta}}_{n,1}, \ldots, \check{\boldsymbol{\theta}}_{n,m+mp-1})$ of $\boldsymbol{\theta}_0$. This can be used as our final estimate of $\boldsymbol{\theta}_0$, or else used as a starting value in the numerical maximization of $\ell_n^*(\boldsymbol{\theta})$ in (2.4).

As an empirical illustration of the methods just discussed we fit a bivariate first-order Markov model to two exchange rate series: the Korean won (KRW) and the Taiwanese dollar (TWD), both denominated in US dollars (USD). We confine our analysis to the period spanning February 24, 2013, to October 20, 2013, containing $n = 239$ daily observations in total. The two series are displayed in Figure 2.6. We used a one-sided Hodrick-Prescott filter (see e.g. Mehra, 2004) to purge predictable trending behavior from the two series, and fit our vine copula model to the filtered series, which may be interpreted as unanticipated shocks to the exchange rates.

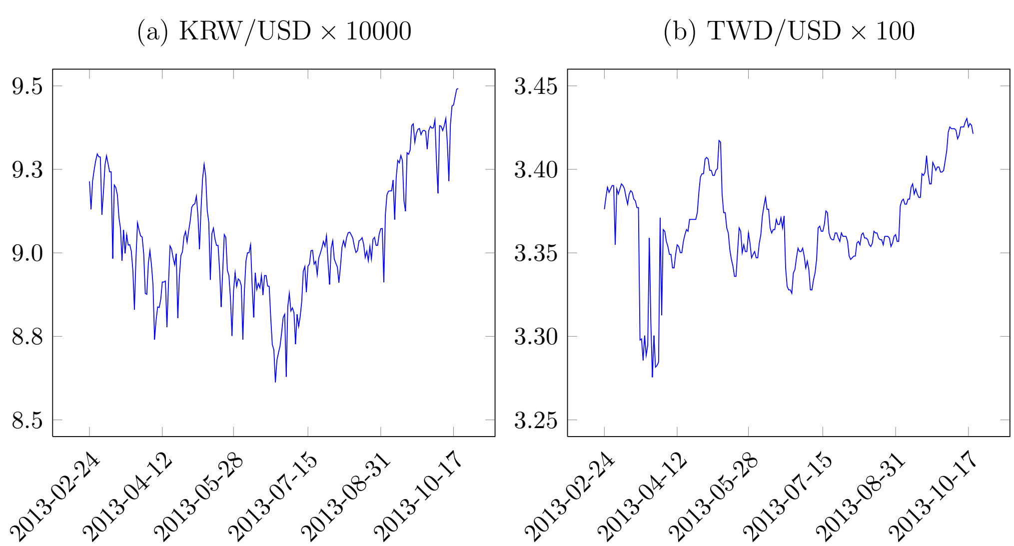


Figure 2.6: Daily US dollar values of the Korean won (a) and Taiwanese dollar (b)

As a guide to specifying the copulae in the first tree of our M-vine, in panels (a), (b) and (c) of Figure 2.7 we display scatter plots of pairs of empirical probability integral transforms of the filtered exchange rate series and their lagged values. In panel (a) we see that the contemporaneous relationship between our two series is strikingly nonexchangeable: large positive shocks to the KRW/USD rate are rarely accompanied by large negative shocks to the TWD/USD rate, while large negative shocks to the KRW/USD rate more often coincide with large positive shocks to the TWD/USD rate. We capture this asymmetry by assigning an asymmetric Clayton copula to this pair of variables.[10] This copula is of the form

$$C(u, v; \alpha, \beta, \gamma) = u^{1-\alpha} v^{1-\beta} \left(u^{-\alpha\gamma} + v^{-\beta\gamma} - 1\right)^{-1/\gamma},$$

with parameters $\alpha, \beta \in [0, 1]$ and $\gamma \in (0, \infty)$. Panel (b) of Figure 2.7 reveals that the pattern of serial dependence in the KRD/USD rate is approximately exchangeable, but exhibits another form of asymmetry: the data cluster relatively tightly toward the top right of the panel, but are more disperse toward the lower left. We could model this feature of the data using a Gumbel copula, which can feature positive upper tail dependence but has zero lower tail dependence. The Gumbel copula is given by equation (2.1) with $\alpha = \beta = 1$, and has parameter $\gamma \in [0, \infty)$; the upper tail dependence coefficient is $2 - 2^{1/\gamma}$ (see e.g. Nelsen, 2006, p. 215). In panel (c) we see that there is strong serial dependence in the TWD/USD rate. The data in this panel are best fit using a Student copula. We find that the likelihood-based fit of the Student copula to the data in panel (c) is substantially better than the likelihood-based fit of the Gumbel copula (and also of the Gaussian, Student, Frank and Clayton copulae) to the data in panel (b), and for this reason elect to label the filtered TWD/USD series as $X_{1,t}$, the first variable in our M-vine. The filtered KRW/USD series is labeled $X_{2,t}$. Since our M-vine does not include the edge $21{,}22$, the serial dependence in the KRW/USD rate is modeled implicitly by the assignment of copulae to other edges of the vine, and not using the Gumbel copula. For

[10] The likelihood-based fit of the asymmetric Clayton copula to these data is slightly better than the fit of the asymmetric Gumbel copula given in (2.1).

the sake of simplicity we assign Gaussian copulae to the edges of trees 2 and 3.

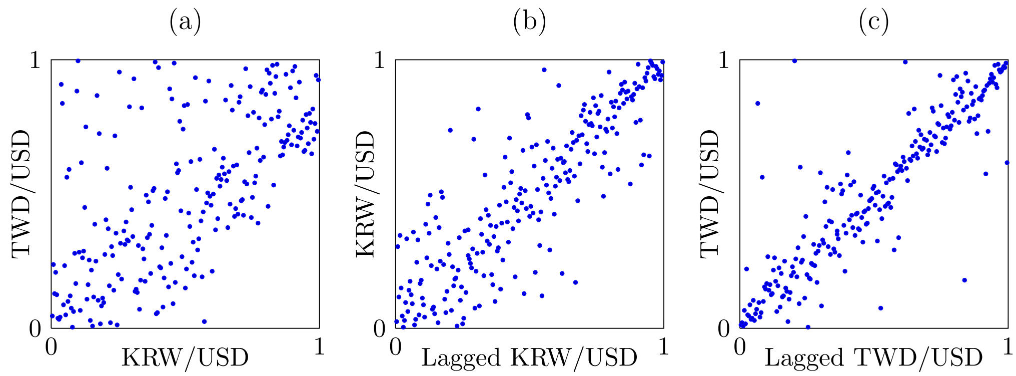


Figure 2.7: Bivariate dependencies in the filtered exchange rate series
The scatter plots in panels (a), (b) and (c) display pairs of empirical probability integral transforms of the filtered exchange rate series and their lagged values.

SSP estimates of the parameter values for our selected model, with bootstrap confidence intervals, are given in Table 2.1. We fixed the parameter β in the asymmetric Clayton copula assigned to edge $11{,}21$ equal to the upper bound of one, because without this restriction our estimate was extremely close to one, as were the estimates obtained using all bootstrapped samples. The estimates of α and γ indicate significant asymmetric dependence between contemporaneous realizations of the filtered exchange rate series, as expected. The Student copula assigned to edge $11{,}12$ has a large estimated correlation parameter of 0.936 with a 95% confidence interval of $(0.920, 0.955)$, and a small estimated degrees of freedom parameter of 2.025 with a 95% confidence interval of $(0.123, 3.067)$. The estimated correlation parameters for the Gaussian copulae assigned to the edges of tree 2 are small and insignificantly different from zero, while the estimated correlation parameter for the Gaussian copula assigned to the sole edge of tree 3 is moderately small but significantly negative.

We can characterize the fit of our estimated M-vine specification using the BIC, computed as $-2\ell_n^*(\check{\boldsymbol{\theta}}_n) + d\log n$. The BIC value for our estimated model is -580.01. If we had reversed the order of our variables, so that $X_{1,t}$ was the

Table 2.1: Estimated M-vine copula specification for the filtered exchange rate series.
Our translation invariant M-vine copula specification for $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is 1-independent with copulae assigned to edges as shown in the table. Parameters were estimated by SSP estimation, with the exception of β which was set equal to one. Confidence intervals were constructed using the bootstrap.

Tree	Edge	Copula	Parameter	Estimate	95% Conf. Int.
1	11,21	Asym. Clayton	α	0.441	$(0.264, 0.528)$
			β	1.000	–
			γ	6.616	$(2.545, 10.077)$
	11,12	Student	corr.	0.936	$(0.920, 0.955)$
			d.f.	2.025	$(0.123, 3.067)$
2	21,12$\vert$11	Gaussian	corr.	-0.113	$(-0.230, 0.000)$
	11,22$\vert$12	Gaussian	corr.	-0.034	$(-0.166, 0.098)$
3	21,22$\vert$11,12	Gaussian	corr.	-0.325	$(-0.478, -0.115)$

filtered KRW/USD series and $X_{2,t}$ was the filtered TWD/USD series, and assigned a Gumbel copula to the edge 11,12, we would have obtained the inferior BIC value -411.39. The improved fit of our selected model can be largely attributed to the superior fit of the Student copula in panel (c) relative to the fit of the Gumbel copula in panel (b); this accounts for more than 90% of the difference in overall BIC values.

2.5 Final remarks

In this paper we have explored the use of vine copulae to model stationary multivariate Markov chains. We proposed a new vine structure, the M-vine, with which it is simple to impose stationarity and the Markov property upon the induced probabilistic model for the observed data. Estimation of an M-vine copula specification is straightforward using existing semiparametric techniques that are nonparametric with respect to univariate marginal behaviour and parametric with respect to dependence between variables.

There are many potential avenues for further research on vine copula models for stationary Markov chains. One subject we have not addressed is the link between the bivariate copulae assigned to the edges of an M-vine, and the mixing and ergodic properties of the associated Markov chain. Results of this kind are available for univariate first-order Markov chains (Chen and Fan, 2006; Gagliardini and Gouriéroux, 2008; Chen, Wu and Yi, 2009; Beare, 2010, 2012; Longla and Peligrad, 2012), as are some limited results for multivariate first-order Markov chains (Rémillard et al., 2012), but their extension to vine models for multivariate higher-order Markov chains remains an open problem. Another area for future research is the development of automated selection algorithms for vine copula specifications. Recent advances in this direction have been made by Dißmann et al. (2013), who considered the case where we observe repeated samples from a multivariate distribution for which we desire a vine copula specification. It may be useful to adapt their methods to the stationary Markov case considered here.

Acknowledgements

Chapter 2 is published in *Journal of Time Series Analysis (2015)*. I thank Brendan K. Beare for his permission to use the co-authored paper in my dissertation Chapter.

Chapter 3

Tests of Stochastic Monotonicity with Improved Size and Power Properties

Abstract. We develop improved statistical procedures for testing the null hypothesis of stochastic monotonicity. Stochastic monotonicity can be reformulated in terms of the concavity of cross-sections of a copula function; our test statistic is based on a empirical measure of departures from concavity. While existing tests of stochastic monotonicity deliver a limiting rejection rate equal to the nominal significance level at one point and below the nominal significance level elsewhere in the null, our test raises the limiting rejection rate to the nominal significance level over a wide region of the null. This improves power against relevant local alternatives. Implementation of our procedure is based on preliminary estimation of a contact set, similar to procedures developed recently in other contexts. To show the validity of our approach we draw on recent results on the directional differentiability of the least concave majorant operator, and on bootstrap inference when smoothness conditions sufficient to apply the functional delta method for the bootstrap are not satisfied. An application to intergenerational income mobility is provided.

3.1 Introduction

In stochastic modeling, a variety of orderings can be used to compare the 'magnitude' of random variables, such as stochastic dominance, mean residual life ordering, likelihood ratio ordering, positive dependence ordering, and so on. In this paper, we focus on *stochastic monotonicity*, an ordering of random variables based on conditional distributions. For two random variables X and Y, with $F_{Y|X}(\cdot|x)$ denoting the cumulative distribution function of Y conditional on $X = x$, we say Y *is stochastically increasing in* X (or equivalently, Y *is positive regression dependent on* X) if and only if $F_{Y|X}(y|x)$ is a nonincreasing function of x for all y. In what follows, we denote by $\mathcal{X}$ and $\mathcal{Y}$ the supports of X and Y respectively, and define the conditional distribution $F_{Y|X}$ on $\mathcal{Y} \times \mathcal{X}$. This paper studies statistical methods to test stochastic monotonicity with the null hypothesis of Y being stochastically increasing in X.

Stochastic monotonicity can be of interest in many applications. Suppose, for example, we want to determine empirically whether a son's social status is positively related to that of his parents. This is a question about intergenerational mobility, one of the classic subjects in sociology and labor economics (Becker and Tomes, 1979; Becker and Tomes, 1986; Mulligan, 1999; Han and Mulligan, 2001; Restuccia and Urrutia, 2004). The conventional approach to the problem has been to investigate the dependence between the son's and parents' status measured by wage or social class and verify, for instance, that they have positive correlation or positive quadrant dependence. On the other hand, stochastic monotonicity can provide more information on the aspect of intergenerational mobility because it also identifies nonmonotone aspects of the relationship between the son's and parent's income that would be undetectable with a test of positive correlation or positive quadrant dependence. For instance, if a son of very wealthy parents has a higher probability of earning a very low income than a son of moderately wealthy parents, perhaps due to perverse incentives arising from the anticipation of inheritance, then this would violate stochastic monotonicity but may be consistent with positive correlation or positive quadrant dependence.

Such questions about the monotonicity or nonmonotonicity of relationships between variables also arise naturally in many other fields. In entry deterrence games, evidence that a firm's investment level is not increasing in market size can be taken as evidence of strategic behavior with the intent to deter market entrants (Ellison and Ellison, 2007). In signaling games, monotonicity also plays an important role when we want to examine if firms invest more in advertising when their product has better quality (Milgrom and Roberts, 1984), or if more talented workers tend to have more education or work faster (Spence, 1973; Akerlof, 1976). In adverse selection models, we may be interested to see if an individual with high probability of loss buys more comprehensive insurance coverage (Wilson, 1977), or in auction theory, we may wonder if a buyer's bid increases when his reservation price increases (Vickrey, 1961).

Stochastic monotonicity also appears as an assumption in various economic models. In the argument of Lucas and Stokey (1989) or Hopenhayn and Prescott (1992), stochastic monotonicity is one of the key conditions for certain Markovian models to have a stationary distribution. This property is crucial in dynamic programming to ensure a unique equilibrium solution and is thus assumed in many economic models concerning Markov perfect equilibrium. For instance, Ericson and Pakes (1995) postulate the transition function that describes a firm's industry dynamic is formed from a distribution which exhibits stochastic monotonicity. See also Pakes (1986) and Olley and Pakes (1996) for a similar argument with applications in the patent market and telecommunications equipment industry. Balbus, Reffett and Woźnyc (2012) also have assumed stochastic monotonicity to analyze the equilibrium of infinite horizon stochastic overlapping generations models. In the discussion of Small et al. (2014), stochastic monotonicity is introduced to relax monotonicity restrictions for IV estimation, while Blundell et al. (2007) assume stochastic monotonicity to weaken the exclusion restriction.

Although stochastic monotonicity is an important concept across a range of fields, not much attention has been paid to research on statistical tests of stochastic monotonicity until recently. Lee, Linton and Whang (2009) were the first to propose a test for stochastic monotonicity, using a kernel smoothed U-statistic to assess the monotonicity of the conditional distribution in the conditioning variable. In a more recent contribution Delgado and Escanciano (2012) have suggested a test that does not require smoothed estimation and is based on the distance between the empirical copula and its least concave majorant in the explanatory variable coordinate. The method in Delgado and Escanciano (2012) has advantages over Lee, Linton and Whang (2009) in that the test is implementable under the minimal assumption of continuity of the marginal distributions without requiring the selection of a kernel function or bandwidth, and the statistic is invariant under strictly increasing transformations of the data.

A limitation of the test of Delgado and Escanciano (2012) is that it is conservative over much of the null, suggesting that power against relevant alternatives may be limited. The null hypothesis of stochastic monotonicity is composite, meaning that there are many possible conditional distribution functions that are consistent with stochastic monotonicity. Delgado and Escanciano (2012) derive the asymptotic null distribution of their test statistic assuming the independence of X and Y. This turns out to be the least favorable case (lfc) which means that the limit distribution of their test statistic is largest (in the sense of stochastic dominance) at this point in the null. Since Delgado and Escanciano (2012) compare their test statistic to a critical value drawn from the limit distribution at the lfc, their test delivers a limiting rejection rate equal to the nominal significance level at the lfc, but below the nominal significance level at other points in the null.

In this paper we show how to improve the power of the test of Delgado and Escanciano (2012) by using a modified bootstrap technique to raise the limiting rejection rate of the test to the nominal significance level over a wide region of the null hypothesis, and not merely at the lfc. To show the validity of our approach we draw on recent results on the directional differentiability of the least concave majorant operator (Beare and Moon, 2015; Beare and Shi, 2015) and on the application of the functional delta method with directionally differentiable operators (Fang and Santos, 2015).

The remainder of this paper is organized as follows. In Section 2, we propose our test statistics and explain how they are constructed. The partial least concave majorant operator is introduced for the construction of the statistics. In Section 3, we provide a brief review of differentiability concepts and establish the asymptotic distributions of our statistics at each null point. In Section 4, a valid bootstrap procedure based on the estimator of the directional derivative of partial least concave majorant operator is proposed. Lastly, the finite sample performance of our tests will be examined with an application to intergenerational income mobility in Section 5 and Section 6.

3.2 Null hypothesis and test statistic

Let X and Y be continuous random variables, and let $C(u, v)$ denote the copula of Y and X. The null hypothesis of stochastic monotonicity can be reformulated in terms of the shape of this copula function. Theorem 5.2.10 and Corollary 5.2.11 in Nelsen (2006) state that Y is stochastically increasing in X if and only if $C(u, v)$ is concave in v for any $u \in [0, 1]$. We shall therefore write our null hypothesis as

$$\Theta_0 = \{C \in \Theta : C(u, \cdot) \text{ is concave for each fixed } u \in [0, 1]\},$$

where Θ denotes the collection of bivariate copula functions on $[0, 1]^2$ with continuous partial derivatives. The alternative hypothesis is $\Theta_1 = \Theta \setminus \Theta_0$.

It is clear that the partial concavity introduced in the preceding paragraph is not as strong as the general notion of concavity of a bivariate function. While it

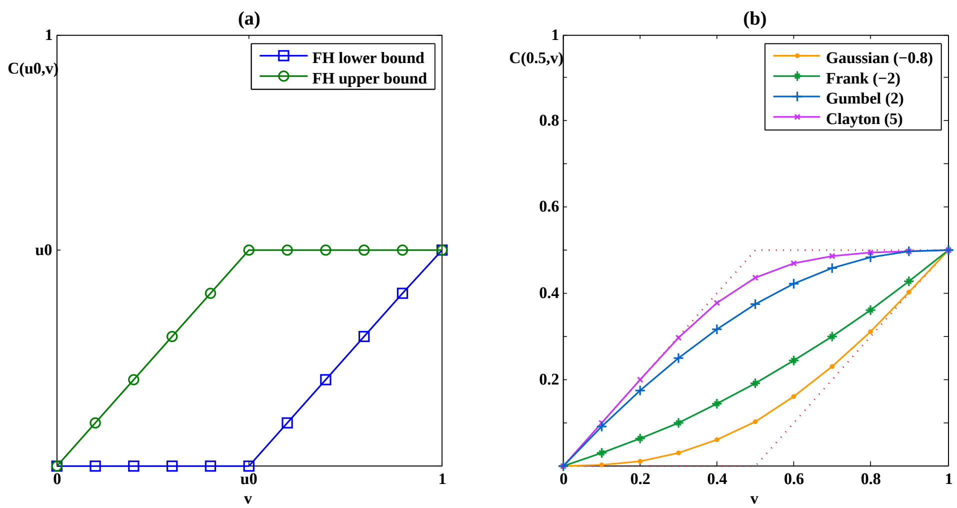


Figure 3.1: Examples of vertical sections of copulas
The vertical section of any copula at $u = u_0$ is a nondecreasing and 1-Lipschitz function between the Fréchet-Hoeffding lower bound, $\max(u_0 + v - 1, 0)$, and the Fréchet-Hoeffding upper bound, $\min(u_0, v)$, shown in Panel (a). Panel (b) displays the vertical sections of the Gaussian copula (with the parameter, -0.8), Frank copula (-2), Gumbel copula (2), and Clayton copula (5) at $u = 0.5$ within the two Fréchet-Hoeffding boundaries.

is well known that the only copula that is concave is the Fréchet-Hoeffding upper bound $C(u, v) = \min(u, v)$, there are many copulas which have concave vertical section (see Def 2.2.5 and Cor 2.2.6 in Nelsen, 2006) for a fixed $u \in [0, 1]$. Vertical sections of copulas, in fact, can be any functions that are nondecreasing and 1-Lipschitz, provided that they stay between the Fréchet-Hoeffding lower and upper bounds. Thus, they can be concave, convex, or otherwise. In Figure 3.1, we display (a) the Fréchet-Hoeffding lower and upper bounds and (b) some of the vertical sections of parametric copulas broadly used. The conditions for those parametric copulas to be in Θ_0 are summarized in Table 3.1.[1]

Having clarified the null hypothesis, we shall now proceed to the construc-

[1]Following the conventional notation, Φ^{-1} is the quantile function of the standard normal distribution and N_ρ is the joint cumulative normal distribution function with mean zero and correlation ρ. t_ν^{-1} denotes the inverse cumulative distribution of the univariate t-distribution with the degree of freedom ν and $t_{\rho,\nu}$ is the multivariate t-distribution with the degree of freedom ν, the scale parameter ρ and the location parameter zero.

tion of our test statistics. Suppose we observe n independent and identically distributed copies of (X, Y), denoted by (X_i, Y_i), $i = 1, \ldots, n$. Define the empirical cdfs of X and Y, and the empirical copula of Y and X as

$$F_{X,n}(x) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq x), \quad F_{Y,n}(y) = \frac{1}{n}\sum_{i=1}^{n} 1(Y_i \leq y) \quad \text{for } (x, y) \in \mathbb{R}^2,$$

$$C_n(u, v) = \frac{1}{n}\sum_{i=1}^{n} 1\{F_{Y,n}(Y_i) \leq u, F_{X,n}(X_i) \leq v\} \quad \text{for } (u, v) \in [0, 1]^2.$$

Our test statistic is of the form[2]

$$M_n^p = n^{1/2} \left\| \tilde{\mathcal{M}} C_n - C_n \right\|_p \tag{3.1}$$

where $\| \cdot \|_p$ is the L^p-norm with respect to the Lebesgue measure on $[0, 1]^2$ given $p \in [1, \infty]$, and $\tilde{\mathcal{M}}$ is the partial least concave majorant (hereby, partial lcm) operator applied to the second argument. In order to provide a formal definition of $\tilde{\mathcal{M}}$, we shall begin by reviewing the definition of the least convave majorant (lcm) operator $\mathcal{M}$, and also of the restricted lcm operator $\mathcal{M}_{[a,b]}$ in Beare and Moon (2015). We denote by $\ell^\infty([a, b])$ the collection of uniformly bounded real valued functions on $[a, b]$ equipped with the uniform metric, and let $\ell^\infty_{\text{co}}([a, b]) = \{f \in \ell^\infty([a, b]) : f \text{ is concave}\}$. The following Definition 3.2.1 is taken from Beare and Moon (2015).

Definition 3.2.1. Given a closed interval $[a, b] \subseteq [0, 1]$, the lcm over $[a, b]$ is the operator $\mathcal{M}_{[a,b]} : \ell^\infty([0, 1]) \to \ell^\infty([a, b])$ that maps each $f \in \ell^\infty([0, 1])$ to the function

$$\mathcal{M}_{[a,b]} f(u) = \inf\{g(u) : g \in \ell^\infty_{\text{co}}([a, b]) \text{ and } f \leq g \text{ on } [a, b]\}, \quad u \in [a, b].$$

We write $\mathcal{M}$ as shorthand for $\mathcal{M}_{[0,1]}$, and refer to $\mathcal{M}$ as the lcm operator.

[2]This statistic has been studied in Delgado and Escanciano (2012) for $p = \infty$.

Table 3.1: Concavity and convexity of vertical sections of some copula families
This table summarizes the shape of the vertical sections of popular parametric copulas. The vertical sections of the Gaussian and Student t-copulas are concave when the correlation parameter is nonnegative, and otherwise convex. Frank copulas also have concave vertical sections with positive copula parameters, and convex with negative parameters while the vertical sections of Clayton copulas are concave when the copula parameter is positive, and convex when the parameter belongs to the interval $[-1, 0)$. The vertical sections of the Gumbel copulas are always concave.

Copula family	$C(u, v)$	Vertical sections
Gaussian (ρ)	$N_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$	$\rho \in [0, 1]$: concave $\rho \in [-1, 0]$: convex
Student (ρ, ν)	$t_{\rho,\nu}(t_\nu^{-1}(u), t_\nu^{-1}(v))$	$\rho \in [0, 1]$: concave $\rho \in [-1, 0]$: convex
Frank (θ)	$-\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$\theta > 0$: concave $\theta < 0$: convex
Clayton(θ)	$\left[\max\left(u^{-\theta} + v^{-\theta} - 1, 0\right)\right]^{-1/\theta}$	$\theta > 0$: concave $\theta \in [-1, 0)$: convex
Gumbel (θ)	$\exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right)$	$\theta \geq 1$: concave
F-H Lower bound	$\max(u + v - 1, 0)$	- : convex
F-H Upper bound	$\min(u, v)$	- : concave
Independence	uv	- : linear

The lcm operator has been studied in Carolan (2002) and employed in many other econometric and statistic applications. For instance, Durot (2003) applies the lcm operator in the context of testing the monotonicity of regression curves. Carolan and Tebbs (2005), Beare and Moon (2015), Beare and Schmidt (2015) and Beare and Shi (2015) use test statistics formed by the lcm operator for testing density ratio ordering. Here we introduce a new operator, the partial lcm operator $\tilde{\mathcal{M}}$, which is an extension of the lcm operator for the functions of higher dimension. Intuitively, the partial lcm operator is the lcm operator for each $u \in [0,1]$ applied to the function $f(u,\cdot)$.

Definition 3.2.2. Given a closed interval $[a,b] \subseteq [0,1]$, the partial lcm over $[0,1] \times [a,b]$ is the operator $\tilde{\mathcal{M}}_{[a,b]} : \ell^\infty([0,1]^2) \to \ell^\infty([0,1] \times [a,b])$ that maps each $f \in \ell^\infty([0,1]^2)$ to the function

$$\tilde{\mathcal{M}}_{[a,b]} f(u,v) = \mathcal{M}_{[a,b]}(f(u,\cdot))(v) \qquad (u,v) \in [0,1] \times [a,b]$$

We write $\tilde{\mathcal{M}}$ as shorthand for $\tilde{\mathcal{M}}_{[0,1]}$, and refer to $\tilde{\mathcal{M}}$ as the partial lcm operator.

To establish the asymptotic results in the next section, we require the empirical copula process $n^{1/2}(C_n - C)$ to admit a weak limit. Weak convergence of the empirical copula process was first demonstrated by Deheuvels (1981a, 1981b) under independence, and generalized to nonindependence by Gaenssler and Stute (1987) and Fermanian, Radulović and Wegkamp (2004) in the Skorokhod space $D([0,1]^2)$ and $\ell^\infty([0,1]^2)$ respectively. Later Segers (2012) established weak convergence under a milder assumption that does not require smoothness of the copula at the boundary of the unit square. For more detailed discussion see also Stute (1984), van der Vaart and Wellner (1996, 2007) and Tsukahara (2005). Here we adopt the following assumption taken from Segers (2012).

Assumption 3.2.1. (i) The random variables X and Y have continuous cumulative distribution functions, and (ii) the copula of Y and X, $C(u,v)$ admits continuous partial derivatives on $(0,1)^2$.

Assumption 3.2.1 (i) ensures the uniqueness of the copula of Y and X on $[0,1]^2$. We additionally require Assumption 3.2.1 (ii), a smoothness condition, to ensure weak convergence of the empirical copula process. In what follows, let $\rightsquigarrow$ denote weak convergence as in Definition 1.3.3 in van der Vaart and Wellner (1996), and $\partial_i C(u,v)$ the partial derivative of $C(u,v)$ with respect to the i-th argument for $i = 1,2$. Under Assumption 2.1, Theorem 3.2.1 follows by Segers (2012).

Theorem 3.2.1. *Suppose Assumption 3.2.1 holds. Then as $n \to \infty$, we have*

$$n^{1/2}(C_n - C) \rightsquigarrow \mathcal{G}_C \quad \text{in } \ell^\infty\left([0,1]^2\right),$$

where $\mathcal{G}_C$ is a tight Gaussian process on $[0,1]^2$ which can be written as

$$\mathcal{G}_C(u,v) = \mathcal{B}_C(u,v) - \partial_1 C(u,v)\mathcal{B}_C(u,1) - \partial_2 C(u,v)\mathcal{B}_C(1,v),$$

with $\mathcal{B}_C$ being a Brownian bridge on $[0,1]^2$ with covariance function

$$E(\mathcal{B}_C(u_1,v_1)\mathcal{B}_C(u_2,v_2)) = C(u_1 \wedge u_2, v_1 \wedge v_2) - C(u_1,v_1)C(u_2,v_2).$$

Theorem 2.1 may continue to hold in modified form if the independence condition on $(X_i, Y_i)_{i\in Z}$ is replaced with a general weak dependence condition. As an extension of the independent case, Doukhan, Fermanian and Lang (2009) demonstrate the weak convergence of the empirical copula process for stationary η-dependent sequences (Dedecker et al., 2007). The assumption of η-dependence can also be replaced with strong mixing, with mixing coefficients satisfying $\alpha_n = O(n^{-k})$ for some $k > 1$, using the functional central limit theorem in Rio (2000). In this case Theorem 2.1 continues to hold with $\mathcal{B}_C$ replaced by the Gaussian process $\mathcal{B}^*_C$ with covariance kernel $E(\mathcal{B}^*_C(u_1,v_1)\mathcal{B}^*_C(u_2,v_2))$ equal to

$$\sum_{i\in Z} \text{cov}(1\{F_X(X_1) \le u_1, F_Y(Y_1) \le v_1\}, 1\{F_X(X_i) \le u_2, F_Y(Y_i) \le v_2\}).$$

We close this section with some remarks on our statistics.

Remark 3.2.1. $\|\tilde{\mathcal{M}}C_n - C_n\|_p$ can be interpreted as an estimate of $\mu_C^p := \|\tilde{\mathcal{M}}C - C\|_p$, a measure of the extent to which C violates partial concavity. Since copulas are invariant under strictly increasing transformations, (i.e., for any increasing functions λ_1 and λ_2, the transformed variables $(\lambda_1(X), \lambda_2(Y))$ have the same copula as (X, Y)), our measure μ_C^p also possess a scale invariance property as it is purely determined by the copula. Empirical copulas have the same type of scale invariance property, and so inference about stochastic monotonicity based on the statistic M_n^p is not affected by strictly increasing transformations of the variables.

Remark 3.2.2. μ_C^p will differ depending on whether C is the copula of X and Y, or the copula of Y and X. This is as it should be, because stochastic monotonicity is not a symmetric property in general, meaning that "Y is stochastically increasing in X" does not necessarily imply "X is stochastically increasing in Y". If the copula of X and Y is exchangeable, however, partial concavity of the copula of X and Y implies partial concavity of the copula of Y and X, and in consequence, stochastic monotonicity in one direction implies stochastic monotonicity in the other direction.

Remark 3.2.3. Given the data, the test statistic M_n^p can be easily computed in practice by evaluating C_n at grid points on $\{0, 1/n, 2/n, ..., 1\} \times \{0, 1/n, 2/n, ..., 1\}$. For instance, our L^∞ statistic can be calculated as

$$M_n^\infty = n^{1/2} \max_{1 \le i \le n} \max_{1 \le j \le n} \left\{ \tilde{\mathcal{M}}C_n \left(\frac{i}{n}, \frac{j}{n} \right) - C_n \left(\frac{i}{n}, \frac{j-1}{n} \right) \right\}.$$

Computation of C_n is straightforward from (3.1), and the MATLAB command `convhull` provides $\tilde{\mathcal{M}}C_n$ evaluated at each grid point. Likewise, when $p = 1$, our L^1 statistic can be calculated as

$$M_n^1 = n^{-2/3} \sum_{1 \le i \le n-1} \sum_{1 \le j \le n} \left[\frac{1}{2} \left\{ \tilde{\mathcal{M}}C_n \left(\frac{i}{n}, \frac{j}{n} \right) + \tilde{\mathcal{M}}C_n \left(\frac{i}{n}, \frac{j-1}{n} \right) \right\} - C_n \left(\frac{i}{n}, \frac{j-1}{n} \right) \right].$$

For $p \in (1, \infty)$, the exact computation of M_n^p is feasible but complicated, and we recommend using Theorem 4.1 below to obtain a convenient and numerically accurate approximation to M_n^p.

Remark 3.2.4. An extension of our test procedure beyond the bivariate case is straightforward to achieve by reformulating the null hypothesis as

$$H_0 : F_{Y,Z|X}(y, z|x) \text{ is nonincreasing in } x \text{ for each } y \in \mathcal{Y} \text{ and } z \in \mathcal{Z}.$$

H_0 is in this case equivalent to the concavity of cross sections of the copula of (X, Y, Z) fixing the last two arguments[3], and thus we need instead to estimate the empirical copula of (X, Y, Z) with a modified partial lcm operator in which the lcm operator is applied to the first argument with the last two arguments of the function being fixed.

3.3 Asymptotic null distribution

In this section we establish the limit distribution of our test statistic M_n^p at all points in the null hypothesis Θ_0. We achieve this by establishing a differentiability property of the operator $\tilde{\mathcal{M}}$, and applying the functional delta method and the continuous mapping theorem.

Let $\mathcal{I}$ be the identity operator on $\ell^\infty([0,1]^2)$ and let $\tilde{\mathcal{D}} \equiv \tilde{\mathcal{M}} - \mathcal{I}$. By construction, $\tilde{\mathcal{M}}C = C$ (or $\tilde{\mathcal{D}}C = 0$) whenever C is in Θ_0 and accordingly, we can rewrite our test statistic M_n^p in (3.1) as

$$M_n^p = n^{1/2} \left\| \tilde{\mathcal{D}}C_n - \tilde{\mathcal{D}}C \right\|_p \quad \text{when } C \in \Theta_0.$$

Recall that Theorem 3.2.1 states the quantity $n^{1/2}(C_n - C)$ weakly converges to the continuous random element $\mathcal{G}_C$. Then, as is widely studied in the literature, the law

[3]This may be seen by rewriting H_0 as

$$H_0 : \int_0^t F_{Y,Z|X}(F_Y^{-1}(u), F_Z^{-1}(v)|F_X^{-1}(\bar{t}))d\bar{t} \text{ is concave in } t \text{, for each pair of } (u, v) \in [0,1]^2,$$

and noting that $C(t, u, v) = \int_0^t F_{Y,Z|X}(F_Y^{-1}(u), F_Z^{-1}(v)|F_X^{-1}(\bar{t}))d\bar{t}$ is the copula of (X, Y, Z).

of $n^{1/2}(\tilde{\mathcal{D}}C_n - \tilde{\mathcal{D}}C)$ can be obtained by invoking the functional delta method. The key requirement for the functional delta method is generally known as Hadamard differentiability, which has been weakened to Hadamard directional differentiability (Shapiro, 1991; Fang and Santos, 2015). The following definitions of Hadamard differentiability and Hadamard directional differentiability are taken from Fang and Santos (2015). See also Shapiro (1990, 1991) and Bonnans and Shapiro (2000) for more general definitions on topological vector spaces.

Definition 3.3.1. Let $\mathbb{D}$ and $\mathbb{E}$ be Banach spaces, and $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$.

(i) The map ϕ is said to be *Hadamard differentiable* at $\theta \in \mathbb{D}_\phi$ *tangentially* to a set $\mathbb{D}_0 \subseteq \mathbb{D}$ if there is a continuous linear map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ such that

$$\lim_{n\to\infty} \left\| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h) \right\|_{\mathbb{E}} = 0$$

for all sequences $\{h_n\} \subset \mathbb{D}$ and $\{t_n\} \subset \mathbb{R}$ satisfying $t_n \to 0$ and $h_n \to h \in \mathbb{D}_0$ as $n \to \infty$, and $\theta + t_n h_n \in \mathbb{D}_\phi$ for all n.

(ii) The map ϕ is said to be *Hadamard directionally differentiable* at $\theta \in \mathbb{D}_\phi$ *tangentially* to a set $\mathbb{D}_0 \subseteq \mathbb{D}$ if there is a continuous map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ such that

$$\lim_{n\to\infty} \left\| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h) \right\|_{\mathbb{E}} = 0$$

for all sequences $\{h_n\} \subset \mathbb{D}$ and $\{t_n\} \subset \mathbb{R}_+$ satisfying $t_n \downarrow 0$ and $h_n \to h \in \mathbb{D}_0$ as $n \to \infty$, and $\theta + t_n h_n \in \mathbb{D}_\phi$ for all n.

The main difference between Hadamard differentiability and Hadamard directional differentiability is that Hadamard directional differentiability does not require linearity of ϕ'_θ while Hadamard differentiability does. Consequently, Hadamard directional differentiability is weaker than Hadamard differentiability. Proposition 2.1 in Fang and Santos (2015) makes clear the connection between the two differentiability concepts: Hadamard differentiability always implies Hadamard directional differentiability, while Hadamard directional differentiability implies Hadamard differentiability only when the Hadamard directional derivative is linear. This distinction between the two differentiability concepts is critical for the issue being discussed because our operator $\tilde{\mathcal{D}}$ turns out to be Hadamard directionally differentiable but not Hadamard differentiable. Our expression for the weak limit of M_n^p will involve the Hadamard directional derivative of $\tilde{\mathcal{D}}$, and consequently it becomes critical for us to determine its explicit form.

In the following, let $\mathcal{C}([0,1]^2)$ be the space of continuous real valued functions on $[0,1]^2$ equipped with the uniform norm. Our Lemma 3.3.1 establishes the Hadamard directional differentiability of $\tilde{\mathcal{M}}$ at $C \in \Theta_0$ tangentially to $\mathcal{C}\left([0,1]^2\right)$. We shall denote the directional derivative $\tilde{\mathcal{M}}'_C$.

Lemma 3.3.1. *If $C \in \Theta_0$ then $\tilde{\mathcal{M}}$ is Hadamard directionally differentiable at C tangentially to $\mathcal{C}\left([0,1]^2\right)$. Given $u \in [0,1]$, $v \in (0,1)$ and $h \in \mathcal{C}([0,1]^2)$, if $C(u,\cdot)$ is affine in a neighborhood of v, then we have $\tilde{\mathcal{M}}'_C h(u,v) = \mathcal{M}_{[a^u_{C,v}, b^u_{C,v}]}(h(u,\cdot))(v)$, where*

$$
\begin{aligned}
a^u_{C,v} &= \sup\{v' \in (0,v] : C(u,\cdot) \textit{ is not affine in a neighborhood of } v'\},\\
b^u_{C,v} &= \inf\{v' \in [v,1) : C(u,\cdot) \textit{ is not affine in a neighborhood of } v'\},
\end{aligned}
$$

and we define $\inf \varnothing = 1$ and $\sup \varnothing = 0$. If $C(u,\cdot)$ is not affine in a neighborhood of $v \in (0,1)$, or if $v \in \{0,1\}$, then $\tilde{\mathcal{M}}'_C h(u,v) = h(u,v)$.

Since $\ell^\infty([0,1]^2)$ is a metric space, the continuity of $\tilde{\mathcal{M}}'_C$ directly follows from Proposition 3.1 in Shapiro (1990). On the other hand, we see that our directional derivative is not linear in general. For instance, let $C(u,v) = uv$, the independence copula, and let $h_1(u,v) = uv^2$ and $h_2(u,v) = -uv^2$ for $(u,v) \in [0,1]^2$. Then Lemma 3.1 implies that the directional derivatives are $\tilde{\mathcal{M}}'_C h_1(u,v) = uv$, $\tilde{\mathcal{M}}'_C h_2(u,v) = -uv^2$ while $\tilde{\mathcal{M}}'_C (h_1 + h_2)(u,v) = 0$.[4] Therefore we observe

[4]For any given $u \in [0,1]$, $C(u,\cdot)$ is affine with $a^u_{C,v} = 0$ and $b^u_{C,v} = 1$. Therefore from Lemma 3.3.1 we obtain $\tilde{\mathcal{M}}'_C h_1(u,v) = \mathcal{M}_{[0,1]}(h_1(u,\cdot))(v)$=$uv$ by applying the lcm operator on the convex function $h_1(u,\cdot)$ for each fixed $u \in [0,1]$. On the other hand, $h_2(u,v) = -uv^2$ is concave in v for any fixed $u \in [0,1]$ and thus, $\tilde{\mathcal{M}}'_C h_2$ is the function h_2 itself.

$\tilde{\mathcal{M}}'_C(h_1 + h_2) \neq \tilde{\mathcal{M}}'_C h_1 + \tilde{\mathcal{M}}'_C h_2$, and $\tilde{\mathcal{M}}'_C h$ is not linear in h. In fact $\tilde{\mathcal{M}}'_C h$ is linear in h for $C \in \Theta_0$ (equivalently $\tilde{\mathcal{M}}$ is Hadamard differentiable at $C \in \Theta_0$ in direction h), if and only if [5] $C(u, \cdot)$ is strictly concave for all $u \in [0, 1]$, which is the case that the derivative is the identity map. We see, however, from the property of copulas that $C(u, \cdot)$ is always linear when $u \in \{0, 1\}$. Consequently $\tilde{\mathcal{M}}$ cannot be Hadamard differentiable at any properly defined copula in Θ_0.

Building on Lemma 3.3.1, we now obtain the Hadamard directional differentiability of $\tilde{\mathcal{D}}$ at $C \in \Theta_0$ tangentially to $\mathcal{C}([0, 1]^2)$. We will denote the derivative by $\tilde{\mathcal{D}}'_C$.

Lemma 3.3.2. *If $C \in \Theta_0$ then $\tilde{\mathcal{D}}$ is Hadamard directionally differentiable at C tangentially to $\mathcal{C}([0, 1]^2)$ with derivative $\tilde{\mathcal{D}}'_C = \tilde{\mathcal{M}}'_C - \mathcal{I}$.*

From the corresponding property of $\tilde{\mathcal{M}}'_C$ we deduce immediately that $\tilde{\mathcal{D}}'_C$ is continuous but not linear, and thus $\tilde{\mathcal{D}}$ is Hadamard directionally differentiable but not Hadamard differentiable at any $C \in \Theta_0$.

With an application of the functional delta method for directionally differentiable operators (Shapiro, 1991, Theorem 2.1) and the continuous mapping theorem, we finally establish the weak convergence of our test statistic M_n^p.

Theorem 3.3.1. *Under Assumption 3.2.1, for $p \in [1, \infty]$ and $C \in \Theta_0$,*

$$M_n^p = \|n^{1/2}(\tilde{\mathcal{D}}C_n - \tilde{\mathcal{D}}C)\|_p \rightsquigarrow \|\tilde{\mathcal{D}}'_C \mathcal{G}_C\|_p$$

as $n \to \infty$.

Since the limiting process $\tilde{\mathcal{D}}'_C \mathcal{G}_C$ in Theorem 3.3.1 depends on the copula C, the limiting distribution of M_n^p is different at different points (copulas) in the

[5] If $C(u, \cdot)$ is strictly concave for all $u \in [0, 1]$, it is clear from Lemma 3.1 that $\tilde{\mathcal{M}}'_C = \mathcal{I}$, and so $\tilde{\mathcal{M}}'_C$ is linear. On the other hand, if there is some u_0 for which $C(u_0, \cdot)$ is affine over an interval $(a, b) \subseteq [0, 1]$, then $\tilde{\mathcal{M}}'_C(h_1 + h_2) \neq \tilde{\mathcal{M}}'_C h_1 + \tilde{\mathcal{M}}'_C h_2$ on $\{u_0\} \times (a, b)$ with, for instance, $h_1(u, v) = uv^2$ and $h_2(u, v) = -uv^2$. Thus $\tilde{\mathcal{M}}'_C$ cannot be linear.

null. Delgado and Escanciano (2012) particularly consider the independence copula, $C(u,v) = uv$ among all the null points. It turns out that this is the lfc when $p = \infty$ and the test in Delgado and Escanciano (2012) at the α-level of significance is ensured to have a limiting rejection rate of no greater than α at all points in the null. However, as shown in their simulation, the rejection rate of their test falls below α in the limit at all points in the null other than the lfc, and in fact our Remark 3.2 implies that the rejection rate actually reaches to zero whenever all interior cross-sections of the copula are strictly concave. This is not desirable because power will be poor against alternatives that are close to points in the null other than the lfc. Since the null hypothesis for our problem is its own boundary, every point in the null can be approximated by a sequence of local alternatives, and the problem is particularly serious. Lee, Linton and Whang (2009) also consider a test in which the critical value is chosen to control size at the lfc, and we expect this test suffers from the same problem.

The bootstrap procedure we develop in the next section alleviates this problem by approximating the actual limiting distribution at each point of the null in Theorem 3.3.1. By using data dependent critical values, we expect the rejection rate at each point in the null to approach α as the sample size gets larger whenever the limit distribution is nondegenerate. On the other hand, when the null is false, we expect to have higher rejection rates than the test in Delgado and Escanciano (2012) which uses critical values from the lfc.

Remark 3.3.1. When C is the independence copula, i.e. $C(u,v) = uv$, the Hadamard directional derivative of $\tilde{\mathcal{M}}$ at C is identical to $\tilde{\mathcal{M}}$. Therefore, for a given $h \in \mathcal{C}([0,1]^2)$, $\tilde{\mathcal{D}}'_C h$ represents the difference between the function h and its partial lcm. In addition, our test statistic $M_n^p = n^{1/2}\|\tilde{\mathcal{M}}C_n - C_n\|_p$ can be expressed in terms of the directional derivative as $n^{1/2}\|\tilde{\mathcal{D}}'_C C_n\|_p$.

Remark 3.3.2. Theorem 3.3.1 implies that we have a degenerate limit at those copulas $C \in \Theta_0$ for which $C(u,\cdot)$ is strictly concave for all $u \in (0,1)$. In this case we have $\tilde{\mathcal{M}}'_C \mathcal{G}_C = \mathcal{G}_C$ on $(u,v) \in (0,1) \times [0,1]$ by Lemma 3.3.1. Although the vertical sections of the copula are linear when $u \in \{0,1\}$, this does not change the degeneracy of the limit distribution because $\mathcal{G}_C$ in Theorem 3.2.1 is a tucked Gaussian process on $[0,1]^2$.

3.4 Bootstrap procedure

The limiting distribution of our test statistic in Theorem 3.3.1 depends on the random process $\mathcal{G}_C$ and $\tilde{\mathcal{M}}'_C$, which in turn is determined by the underlying copula C. There are a variety of ways to construct a bootstrap approximation to the law of $\mathcal{G}_C$, as we will discuss shortly. However, due to the lack of differentiability of the partial least concave majorant operator $\tilde{\mathcal{M}}$, a good bootstrap approximation to the law of $\mathcal{G}_C$ will not lead directly to a good bootstrap approximation to the law of our test statistic. This is because the functional delta method for the bootstrap fails to apply with nondifferentiable operators. The first part of this section will address this problem, and the rest of the section will show how a valid bootstrap approximation to the limiting distribution of our test statistic can be constructed.

We start out by assuming that there is a good bootstrap version of $\mathcal{G}_C$, $\mathcal{G}^*_{C,n}$, in the sense that the conditional law of $\mathcal{G}^*_{C,n}$ provides a valid approximation of the weak limit of the empirical copula process. For instance, we may use pseudo samples to construct the marginal empirical distributions

$$F^*_{X,n}(x) = \frac{1}{n}\sum_{i=1}^{n} W_i 1(X_i \le x), \quad F^*_{Y,n}(y) = \frac{1}{n}\sum_{i=1}^{n} W_i 1(Y_i \le y),$$

where the weights $(W_1, \ldots, W_n)$ are drawn from the multinomial distribution with success probabilities $1/n$, independent of the data. Then, $\mathcal{G}^*_{C,n} = n^{1/2}(C^*_n - C_n)$ suitably approximates $\mathcal{G}_C$ with $C^*_n = \frac{1}{n}\sum_{i=1}^{n} W_i 1(F^*_{Y,n}(Y_i) \le u, F^*_{X,n}(X_i) \le v)$ in large samples. This resampling method has been adopted in Fermanian et al. (2004) to approximate the weak limit of the empirical copula process.

Alternative choices of resampling scheme are also available. The multiplier bootstrap of Scaillet (2005) and Remillard and Scaillet (2008) has been used to test the equivalence of two copulas, and later adopted for goodness of fit tests of copula based models (Kojadinovic and Yan, 2011; Kojadinovic et al., 2011). Bucher and Dette (2010) suggest the direct multiplier bootstrap, which is more computationally convenient than the ordinary multiplier bootstrap because it does not require estimates of the partial derivatives of the copula function. The direct multiplier bootstrap has been applied in many studies (e.g., in Genest (2012) for the test of exchangeability of copulas and in Bouzebda and Cherfi (2012) for the tests of radial symmetry), but tends to be less accurate than the multiplier bootstrap as demonstrated in the simulation results in Bucher and Dette (2010).

We clarify our assumption on $\mathcal{G}^*_{C,n}$ as follows. Let $\mathrm{BL}_1(\ell^\infty([0,1]^2))$ denote the class of all real valued functions on $\ell^\infty([0,1]^2)$ that are bounded by one in absolute value and are Lipschitz continuous with Lipschitz constant no greater than one. Our Assumption 3.4.1 requires that the bounded Lipschitz distance between the law of $\mathcal{G}^*_{C,n}$ conditional on the data and the law of $\mathcal{G}_C$ converges in probability to zero on $\ell^\infty([0,1]^2)$.

Assumption 3.4.1. As $n \to \infty$, $\mathcal{G}^*_{C,n}$ satisfies

$$\sup_{f \in \mathrm{BL}_1(\ell^\infty([0,1]^2))} \left| E\left(f(\mathcal{G}^*_{C,n}) \right| (X_1, Y_1), \ldots, (X_n, Y_n) \right) - Ef(\mathcal{G}_C) \Big| \rightsquigarrow 0,$$

where $\mathcal{G}_C$ is the weak limit of $n^{1/2}(C_n - C)$ in Theorem 3.2.1.

Even when the bootstrap process $\mathcal{G}^*_{C,n}$ provides a valid approximation to $\mathcal{G}_C$, however, it is not straightforward in our case to obtain a valid approximation to $\|\tilde{\mathcal{D}}'_C \mathcal{G}_C\|_p$, the limiting distribution of our test statistic at $C \in \Theta_0$. Fang and Santos (2015) warn that although the assumption of Hadamard differentiability can be replaced with Hadamard directional differentiability for the application of the functional delta method with some minor conditions, the linearity of Hadamard derivatives plays a crucial role in establishing the consistency of bootstrap approximations. This makes the application of standard bootstrap techniques in our framework to be problematic and in fact, Theorem 3.1 and Remark 3.3 in Fang and Santos (2015) suggest that we cannot simulate $n^{1/2}(\tilde{\mathcal{D}}C_n^* - \tilde{\mathcal{D}}C_n)$ using the bootstrap version of the empirical copula to approximate the process $n^{1/2}(\tilde{\mathcal{D}}C_n - \tilde{\mathcal{D}}C)$ in the usual way.

Our approximation of $\tilde{\mathcal{D}}'_C\mathcal{G}_C$ is instead achieved by estimating the operator $\tilde{\mathcal{D}}'_C$ applied to $\mathcal{G}^*_{C,n}$. A similar approach has been taken in Beare and Shi (2015) for the test of likelihood ratio ordering, and our proposal in this section extends their technical suggestion to the case when the partial lcm is involved in the test statistic. We begin by introducing an operator $\mathcal{S}$, a mapping from a bivariate function to a quadravariate function which measures the difference between the function and the convex combinations of two points on the function. Using this operator, we next define the contact set B_f that identifies linear segments of the function f. Implementation of our alternative bootstrap procedure is based on preliminary estimation of this contact set B_f, similar to procedures developed recently in other contexts such as in Linton, Song and Whang (2010), Anderson, Linton and Whang (2012) and Lee, Song and Whang (2014).

Definition 3.4.1 (Operator $\mathcal{S}$)**.** Let $A = \{(u, v, w_1, w_2) \in [0,1]^4 : w_1 \leq v \leq w_2$ and $u \in [0,1]\}$. For $f \in \ell^\infty([0,1]^2)$, we define the map $\mathcal{S} : \ell^\infty([0,1]^2) \to \ell^\infty(A)$ by

$$\mathcal{S}f(u, v, w_1, w_2) = \frac{(w_2 - v)f(u, w_1) + (v - w_1)f(u, w_2)}{w_2 - w_1} - f(u, v) \qquad (3.2)$$

when $w_1 < w_2$. When $w_1 = w_2$, $\mathcal{S}f(u, v, w_1, w_2) = 0$.

Definition 3.4.2 (Contact set)**.** Given $f \in \ell^\infty([0,1]^2)$, we call the set

$$B_f = \{(u, v, w_1, w_2) \in A : \mathcal{S}f(u, v, w_1, w_2) = 0\}$$

the contact set, and for each fixed $(u, v) \in [0,1]^2$ we call the set

$$B_f(u, v) = \{(w_1, w_2) \in [0,1]^2 : (u, v, w_1, w_2) \in B_f\}$$

a cross section of B_f.

The first term in (3.2) expressed as a ratio corresponds to the function value at (u, v) of the affine function that linearly interpolates two points $f(u, w_1)$ and $f(u, w_2)$. We therefore find that for a fixed $u \in [0,1]$, $f(u,\cdot)$ is concave (convex) on an interval $[a, b] \subseteq [0,1]$ if and only if $\mathcal{S}f(u, v, w_1, w_2)$ is nonpositive (nonnegative) for all $w_1 \le v \le w_2$ in $[a, b]$. On the other hand, if $\mathcal{S}f = 0$ for all $w_1 \le v \le w_2$ in $[a, b]$, it must be the case that $f(u,\cdot)$ is affine on $[a, b]$. Accordingly, $B_C(u, v)$ identifies the largest closed interval containing v over which $C(u,\cdot)$ is affine in the form of $[a^u_{C,v}, v] \times [v, b^u_{C,v}]$, using the notation in Lemma 3.3.1. When C is the independence copula, for instance, $B_C(u, v)$ includes all the points in $[0, v] \times [v, 1]$ and therefore we have $B_C = A$. We also notice that $B_C(u, v)$ is nonempty as it always contains the elements on the diagonal of unit square $\{(w_1, w_2) \in [0,1]^2 : w_1 = w_2\}$.

Using Definitions 3.4.1 and 3.4.2, we can derive alternative expressions for our test statistic M^p_n and the directional derivative of $\tilde{\mathcal{D}}$ that are more practical and useful for computational purposes. In Theorem 3.4.1, we provide a new expression for M^p_n using the operator $\mathcal{S}$ and in Lemma 4.1, we rewrite $\tilde{\mathcal{D}}'_C$ in terms of the operator $\mathcal{S}$ and the contact set B_C.

Theorem 3.4.1. *The test statistic* $M^p_n = n^{1/2}\|\tilde{\mathcal{M}}C_n - C_n\|_p$ *can be written as*

$$M^p_n = n^{1/2}\left\| \sup_{(w_1,w_2)\in[0,v]\times[v,1]} \mathcal{S}C_n(\cdot, w_1, w_2)\right\|_p .$$

In particular, when $p = \infty$ *the test statistic is reduced to*

$$M^\infty_n = n^{1/2} \sup_{(u,v,w_1,w_2)\in A} \mathcal{S}C_n(u, v, w_1, w_2)$$

where A is the set defined in Definition 3.4.1.

Lemma 3.4.1. *The Hadamard directional derivative of $\tilde{\mathcal{D}}$ at $C \in \Theta_0$ in direction $h \in C([0,1]^2)$ can be written as*

$$\tilde{\mathcal{D}}'_C h(u,v) = (\tilde{\mathcal{M}}'_C - \mathcal{I})h(u,v) = \sup_{(w_1,w_2)\in B_C(u,v)} \mathcal{S}h(u,v,w_1,w_2), \quad (u,v) \in [0,1]^2.$$

When C is the independence copula, $C(u,v) = uv$, $B_C(u,v)$ is the set $[0,v] \times [v,1]$ as pointed out earlier. Therefore, our findings in Theorem 3.4.1 and Lemma 3.4.1 are consistent with Remark 3.3.1 in the previous section.

Lemma 3.4.1 implies that we may estimate the asymptotic distribution in Theorem 3.3.1 by a consistent estimate of $B_C(u,v)$ along with $\mathcal{G}^*_{C,n}$ that satisfies Assumption 3.4.1. With suitably chosen κ_n, natural estimates of B_C and $B_C(u,v)$ might be,

$$B_{C,n} = \{(u,v,w_1,w_2) \in A : |\mathcal{S}C_n(u,v,w_1,w_2)| \leq \kappa_n\}$$

and

$$B_{C,n}(u,v) = \{(w_1,w_2) \in [0,1]^2 : (u,v,w_1,w_2) \in B_{C,n}\}, (u,v) \in [0,1]^2$$

using the empirical copula C_n as a nonparametric estimate of C. Employing these estimates, we suggest to approximate the limiting distribution of M_n^p by the law of

$$M_n^* = \left\| \sup_{(w_1,w_2)\in B_{C,n}(\cdot)} \mathcal{S}\mathcal{G}^*_{C,n}(\cdot,w_1,w_2) \right\|_p .$$

The proposal is justified by Theorem 3.4.2 with some conditions on κ_n stated therein. Since consistency in the bounded Lipschitz metric on $\mathbb{R}$ implies convergence in distribution, our test with approximate rejection rate α is implemented by setting the critical value equal to the $(1-\alpha)$ quantile of the bootstrap distribution of M_n^*.[6]

[6]See Remark 3.1 in Fang and Santos (2015).

Theorem 3.4.2. *Let $\mathcal{G}^*_{C,n}$ be a sequence that satisfies Assumption 3.4.1 and M^*_n be the bootstrap statistics obtained by κ_n chosen to satisfy (i) $\kappa_n \to 0$ and (ii) $n^{1/2}\kappa_n \to \infty$, as $n \to \infty$. Then we have*

$$\sup_{f \in BL_1(\mathbb{R})} \left| E\left(f\left(M^*_n\right) | \left(X_1, Y_1\right), \ldots, \left(X_n, Y_n\right)\right) - Ef\left(\|\tilde{\mathcal{M}}'_C \mathcal{G}_C - \mathcal{G}_C\|_p\right)\right| \rightsquigarrow 0$$

for $C \in \Theta_0$.

3.5 Simulation

In this section, we report some numerical evidence concerning the finite sample performance of our tests. We have considered the following four data generating processes,

(N1) $Y_i = \varepsilon_i$

(A1) $Y_i = -0.1X_i + \varepsilon_i$

(A2) $Y_i = -0.1X_i^2 + \varepsilon_i$

(A3) $Y_i = -0.1X_i^5 + \varepsilon_i$

(A4) $Y_i = 0.2X_i - 0.2\exp(-250(X_i - 0.5)^2) + \varepsilon_i$

where X_i is drawn from the uniform distribution $U[0, 1]$, and ε_i is generated from $N(0, 0.1^2)$ and independent of X_i in each Monte Carlo simulation. Model (N1) corresponds to the independence copula, i.e., the lfc, and models (A1)-(A4) are used to evaluate the performance of the tests under the alternative. The vertical sections of copulas in models (A1)-(A3) are smooth convex functions on the unit interval while the vertical sections of the copula in model (A4) have both concave and convex segments.

Table 3.2 reports the rejection rates in 1000 Monte Carlo replications of each model. The nominal level is set to 5%, and the tuning parameter κ_n for our tests are chosen to deliver a 5% rejection rate at the lfc (N1). Along with the results of our test, we also report the performance of the tests in Delgado and Escanciano (2012)

Table 3.2: Rejection frequencies of tests of stochastic monotonicity

The table shows the rejection frequencies of our tests using the L^1, L^2, and L^∞ statistics, the test of Delgado and Escanciano (2012), and the test of Lee, Linton and Whang (2009). The nominal level is set to 5%.

Model	n	L^1	L^2	L^∞	DE stat	$LLW_{0.4}$	$LLW_{0.5}$	$LLW_{0.6}$
N1	100	0.055	0.048	0.049	0.046	0.030	0.034	0.035
	200	0.054	0.053	0.053	0.052	0.031	0.031	0.034
	300	0.062	0.053	0.044	0.042	0.032	0.036	0.039
A1	100	0.877	0.828	0.653	0.634	0.258	0.408	0.542
	200	0.988	0.980	0.911	0.880	0.541	0.749	0.853
	300	0.999	1.000	0.995	0.980	0.752	0.911	0.964
A2	100	0.874	0.806	0.620	0.599	0.314	0.469	0.587
	200	0.990	0.981	0.938	0.906	0.617	0.805	0.892
	300	1.000	1.000	0.995	0.981	0.827	0.938	0.972
A3	100	0.670	0.568	0.399	0.314	0.283	0.372	0.434
	200	0.898	0.882	0.739	0.686	0.571	0.685	0.742
	300	0.971	0.960	0.851	0.846	0.782	0.859	0.895
A4	100	0.003	0.030	0.154	0.032	0.013	0.012	0.013
	300	0.004	0.178	0.539	0.382	0.025	0.012	0.009
	1000	0.210	0.997	0.999	0.997	0.141	0.039	0.016

and Lee, Linton and Whang (2009) for comparison. For Lee, Linton and Whang (2009)'s test, we use the Epanechnikov kernel and bandwidth parameters $h = 0.4$, 0.5,and 0.6 following their choice.

We summarize the simulation results as follows. Firstly in models (A1)-(A3), we clearly observe the power improvement with all of our tests with the L^1, L^2, and L^∞ statistics over the tests in Delgado and Escanciano (2012) and Lee, Linton and Whang (2009). The data generating process in model (A4) seems to require large sample size to satisfactorily detect the negation of the null hypothesis using our tests with the L^1, L^2, and L^∞ statistics and Delgado and Escanciano (2012)'s test. Under this alternative, our test with the L^∞ statistic outperforms the tests with the L^1 and L^2 statistics, while with the same L^∞ statistic our test shows better performance than Delgado and Escanciano (2012)'s test. On the other hand, the test by Lee, Linton and Whang (2009) obtains no higher rejection rates as sample size gets larger. This can be also seen in the simulations of Delgado and Escanciano (2012) even with the sample size as large as 500.

3.6 Application

In this section, we revisit the empirical example of intergenerational mobility discussed in the beginning of the paper. Our null hypothesis particularly concerns the income mobility from parents to the next generation. In the previous studies, Solon (1992, 1999, 2002) and Minicozzi (2003) have found positive correlation between the son's income and parental income. Lee, Linton and Whang (2009) and Delgado and Escanciano (2012) made similar conclusion but in the stochastic monotonicity sense.

We applied our test to the same data set used in Minicozzi (2003), Lee, Linton and Whang (2009) and Delgado and Escanciano (2012). The data is from the Panel Study of Income Dynamics (PSID) where the variable X is the logarithm of parental predicted income while Y is the logarithm of the son's averaged full time

Table 3.3: The results of our tests for intergenerational income mobility
The results of our tests with the L^1, L^2, and L^∞-statistics for the application to intergenerational income mobility.

	test statistics	5% critical value	10% critical value	p-values
L^1-statistic	0.099	0.216	0.147	0.997
L^2-statistic	0.103	0.282	0.189	1.000
L^∞-statistic	0.403	0.922	0.645	0.993

real income at the age of 28 and 29. We have $n = 616$ observations after we drop some of the samples with a censoring issue as is done in Lee, Linton and Whang (2009). The data set is downloadable on the *Journal of Applied Econometrics* website.

Table 3.3 summarizes the results of our test with the L^1, L^2, and L^∞ statistics. The bootstrap critical values and bootstrap p-values are obtained with 1000 bootstrap replications. All of the three tests fail to reject the null hypothesis at both the 5% and 10% significance level, and hence we conclude that stochastic monotonicity is observed. In fact, we observe that the bootstrap p-values of the statistics are close to one, which suggests very strong evidence of stochastic monotonicity between the son's income and parental income. The conclusion is consistent with those of Lee, Linton and Whang (2009) and Delgado and Escanciano (2012). However, we observe that the critical values for our test are smaller than those of Delgado and Escanciano (2012), for example, the critical value of their test is between 0.811 and 0.813 at the 10% significance level while ours is 0.645.

3.7 Conclusion

We have developed improved statistical procedures for testing the null hypothesis of stochastic monotonicity. While existing tests of stochastic monotonicity deliver a limiting rejection rate equal to the nominal significance level at one point and below the nominal significance level elsewhere in the null, our test raises the limiting rejection rate to the nominal significance level over a wide region of the null. This improves power against relevant local alternatives.

Although we have been concerned with testing the null of stochastic monotonicity in this paper, the differential properties of the partial lcm operator we have established may be useful for constructing new tests of other null hypotheses of interest. Hypotheses expressed in terms of functional inequalities or monotonicity constraints when the function has more than two dimensions may often be reframed as tests of partial concavity.[7] Our approach to estimating the directional derivative of the partial lcm operator may be applicable for developing bootstrap tests of such hypotheses with good size and power properties. We leave the investigation of this subject to future research.

3.8 Appendix

Proof of Lemma 3.3.1. Our proof extends the arguments used by Beare and Moon (2015) to establish Hadamard directional differentiability of the lcm on univariate functions. Before developing the argument, we enumerate some useful properties of the lcm operator $\mathcal{M}$, and the partial lcm operator $\tilde{\mathcal{M}}$.

(i) $\mathcal{M}(f+g)= \mathcal{M}f+g$ for any $f, g \in \ell^\infty$ with g affine.

(ii) $\mathcal{M}$ is positive homogeneous of degree one, i.e., $c\mathcal{M}f= \mathcal{M}(cf)$ for any $f \in \ell^\infty$ and $c \in R^+$.

(iii) $\tilde{\mathcal{M}}$ is positive homogenious of degree one.

(iv) $\|\tilde{\mathcal{M}}f - \tilde{\mathcal{M}}g\|_\infty \leq \|f-g\|_\infty$ for any $f, g \in \ell^\infty([0,1]^2)$.

The properties of the lcm operator $\mathcal{M}$ in **(i)** and **(ii)** are well known in the literature (see e.g. Durot and Tocquet (2003, Lemma 2.1) and Beare and Moon (2015))

[7]For instance, the null hypothesis of conditional stochastic dominance can be formulated in terms of an inequality between two conditional distribution functions, or in terms of the partial concavity of a related bivariate function. See Delgado and Escanciano (2013) for details.

whereas the positive homogeneity of $\tilde{\mathcal{M}}$ in **(iii)** can be simply inferred from the positive homogeneity of $\mathcal{M}$, and the last property **(iv)** follows from Marshall's lemma (Durot and Tocquet, 2003, Lemma 2.2).

As a starting point, we notice that for each n,

$$\begin{aligned} & \left\| \frac{\tilde{\mathcal{M}}(C+t_n h_n) - \tilde{\mathcal{M}}C}{t_n} - \tilde{\mathcal{M}}'_C h \right\|_\infty \\ \leq \; & \left\| \frac{\tilde{\mathcal{M}}(C+t_n h) - \tilde{\mathcal{M}}C}{t_n} - \tilde{\mathcal{M}}'_C h \right\|_\infty + \left\| \frac{\tilde{\mathcal{M}}(C+t_n h_n) - \tilde{\mathcal{M}}(C+t_n h)}{t_n} \right\|_\infty \end{aligned}$$

where the second term is bounded by the $\|h_n - h\|_\infty$ that vanishes as $n \to \infty$. We can apply **(iii)** and **(iv)** to obtain,

$$\begin{aligned} \left\| \frac{\tilde{\mathcal{M}}(C+t_n h_n) - \tilde{\mathcal{M}}(C+t_n h)}{t_n} \right\|_\infty &= \left\| \tilde{\mathcal{M}}(t_n^{-1}C + h_n) - \tilde{\mathcal{M}}(t_n^{-1}C + h) \right\|_\infty \\ &\leq \|h_n - h\|_\infty . \end{aligned}$$

As a consequence, we need only to prove that for any sequences $t_n \downarrow 0$ and $h \in C[0,1]^2$, $t_n^{-1}(\tilde{\mathcal{M}}(C+t_n h) - \tilde{\mathcal{M}}C) \to \tilde{\mathcal{M}}'_C h$ in which $\tilde{\mathcal{M}}C$ can be replaced by C for $C \in \Theta_0$. Since $C(u,\cdot)$ is concave for each $u \in [0,1]$, Minkowski's hyperplane theorem ensures that for any fixed $(u,v) \in [0,1]^2$ there exists an affine function $\xi_{u,v} \in \mathcal{C}([0,1])$ such that $\xi_{u,v}(\cdot) \geq C(u,\cdot)$ and $\xi_{u,v}(v) = C(u,v)$. By introducing $\xi_{u,v}$, we can rewrite

$$\begin{aligned} t_n^{-1}(\tilde{\mathcal{M}}(C+t_n h)(u,v) - C(u,v)) &= t_n^{-1}(\tilde{\mathcal{M}}(C+t_n h)(u,v) - \xi_{u,v}(v)) \\ &= t_n^{-1}(\mathcal{M}(C(u,\cdot) + t_n h(u,\cdot))(v) - \xi_{u,v}(v)) \\ &= t_n^{-1}(\mathcal{M}(C(u,\cdot) + t_n h(u,\cdot) - \xi_{u,v}(\cdot))(v)) \\ &= \mathcal{M}\left(h(u,\cdot) + t_n^{-1}(C(u,\cdot) - \xi_{u,v}(\cdot)\right)(v) \qquad (3.3) \end{aligned}$$

where the second equation follows from Definition 3.2.2, and the rest of the equations are by **(i)** and **(ii)**. Observing that the sequence (3.3) is monotone, the uniform convergence is implied by the pointwise convergence by virtue of Dini's theorem. Thus we can finish the proof by showing the pointwise convergence of (3.3) toward $\tilde{\mathcal{M}}_{[a^u_{C,v}, b^u_{C,v}]} h(u,v)$.

Let $h_{n,u,v} \equiv h(u,\cdot) + t_n^{-1}(C(u,\cdot) - \xi_{u,v}(\cdot))$. Then Lemma 1 in Carolan (2002) implies,

$$\mathcal{M}h_{n,u,v}(v) = \sup_{v' \in [0,v]} \sup_{v'' \in [v,1]} \frac{(v''-v)h_{n,u,v}(v') + (v-v')h_{n,u,v}(v'')}{v''-v'} \text{ when } v'' \neq v' \tag{3.4}$$

with $\mathcal{M}h_{n,u,v}(v) = h_{n,u,v}(v) = h(u,v)$ when $v'' = v'$. By substituting the expression for $h_{n,u,v}$ and applying Lemma 1 in Carolan (2002), we can show that

$$\begin{aligned} & \frac{(v''-v)h_{n,u,v}(v') + (v-v')h_{n,u,v}(v'')}{v''-v'} \\ \leq \; & \tilde{\mathcal{M}}h(u,v) + t_n^{-1}\left[\frac{(v''-v)C(u,v') + (v-v')C(u,v'')}{v''-v'} - C(u,v)\right] \end{aligned} \tag{3.5}$$

Now recall that $C(u,\cdot)$ is concave and not affine in the left neighborhood of $a^u_{C,v}$ nor the right-neighborhood of $b^u_{C,v}$. Therefore for any fixed $\delta > 0$, (3.5) diverges to negative infinity whenever (v', v'') is not in $[(a^u_{C,v} - \delta) \vee 0, v] \times [v, (b^u_{C,v} + \delta) \wedge 1]$. In result, the supremums in (3.5) are not taken over the complement of $[(a^u_{C,v} - \delta) \vee 0, v] \times [v, (b^u_{C,v} + \delta) \wedge 1]$, and thus for n sufficiently large

$$\mathcal{M}h_{n,u,v}(v) = \sup_{v' \in [(a^u_{C,v} - \delta) \vee 0, v]} \sup_{v'' \in [v, (b^u_{C,v} + \delta) \wedge 1]} \frac{(v''-v)h_{n,u,v}(v') + (v-v')h_{n,u,v}(v'')}{v''-v'}$$

which is equivalent to $\mathcal{M}_{[a^u_{C,v} - \delta) \vee 0, (b^u_{C,v} + \delta) \wedge 1]} h_{n,u,v}(v)$. The rest of the proof is the same as Beare and Moon (2015). By letting $\delta \downarrow 0$ and using continuity of h, we conclude the convergence,

$$\mathcal{M}h_{n,u,v}(v) \to \tilde{\mathcal{M}}_{[a^u_{C,v}, b^u_{C,v}]} h(u,v).$$

We refer to the proof in Beare and Moon (2015) for details. $\square$

Proof of Theorem 3.4.1. For $g \in \ell^\infty([0,1])$, Lemma 1 in Carolan (2002) implies that the restricted lcm on $[a,b]$ can be written as

$$\mathcal{M}_{[a,b]}g(u) = \sup_{(w_1,w_2)\in[a,u]\times[u,b]} \frac{(w_2-u)g(w_1)+(u-w_1)g(w_2)}{w_2-w_1}. \tag{3.6}$$

By Definition 3.2.2 and (3.6), we may write

$$\tilde{\mathcal{M}}h(u,v) = \mathcal{M}(h(u,\cdot))(v) = \sup_{(w_1,w_2)\in[0,v]\times[v,1]} \frac{(w_2-v)h(u,w_1)+(v-w_1)h(u,w_2)}{w_2-w_1}$$

from which we obtain

$$n^{1/2}(\tilde{\mathcal{M}}C_n - C_n)(u,v) \;=\; n^{1/2} \sup_{(w_1,w_2)\in[0,v]\times[v,1]} \mathcal{S}C_n(u,v,w_1,w_2).$$

We therefore have

$$M_n^p = n^{1/2}\left\| \sup_{(w_1,w_2)\in[0,v]\times[v,1]} \mathcal{S}C_n(\cdot,w_1,w_2)\right\|_p.$$

□

Proof of Lemma 3.3.2. The result follows from Definition 3.3.1, Lemma 3.3.1 and the triangular inequality. We want to show,

$$\lim_{n\to\infty}\left\|\frac{\tilde{\mathcal{D}}(C+t_nh_n)-\tilde{\mathcal{D}}C}{t_n} - (\tilde{\mathcal{M}}'_C - \mathcal{I})(h)\right\|_\infty = 0$$

for $C \in \Theta_0$ and $h \in \mathcal{C}([0,1]^2)$. Since $\tilde{\mathcal{D}} = \tilde{\mathcal{M}} - \mathcal{I}$,

$$\begin{aligned}
&\left\|\frac{\tilde{\mathcal{D}}(C+t_nh_n)-\tilde{\mathcal{D}}C}{t_n} - (\tilde{\mathcal{M}}'_C - \mathcal{I})(h)\right\|_\infty \\
=\;& \left\|\frac{(\tilde{\mathcal{M}}-\mathcal{I})(C+t_nh_n)-(\tilde{\mathcal{M}}-\mathcal{I})(C)}{t_n} - (\tilde{\mathcal{M}}'_C - \mathcal{I})(h)\right\|_\infty \\
\le\;& \left\|\frac{\tilde{\mathcal{M}}(C+t_nh_n)-\tilde{\mathcal{M}}C}{t_n} - \tilde{\mathcal{M}}'_C h\right\|_\infty + \left\|\frac{(C+t_nh_n)-C}{t_n} - h\right\|_\infty
\end{aligned}$$

The proof is done because limit of the first term and second term both converge to zero by the Hadamard directional differentiability of $\tilde{\mathcal{M}}$ and the convergence of $h_n \to h$. $\square$

Proof of Lemma 3.4.1. First, suppose that $C(u,\cdot)$ is affine in a neighborhood of v. In this case Lemma 3.3.1 implies that $\tilde{\mathcal{M}}'_C h(u,v) = \mathcal{M}_{[a^u_{C,v},b^u_{C,v}]}(h(u,\cdot))(v)$, or equivalently

$$\tilde{\mathcal{M}}'_C h(u,v) = \sup_{(w_1,w_2)\in[a^u_{C,v},v]\times[v,b^u_{C,v}]} \frac{(w_2-v)h(u,w_1)+(v-w_1)h(u,w_2)}{w_2-w_1}$$

by (3.6). Since $C(u,\cdot)$ is not affine outside the interval $(a^u_{C,v},b^u_{C,v})$, we notice that $B(u,v)=\{(w_1,w_2)\in[a^u_{C,v},v]\times[v,b^u_{C,v}]\}$. Therefore

$$\tilde{\mathcal{M}}'_C h(u,v) - h(u,v) = \sup_{(w_1,w_2)\in B(u,v)} \frac{(w_2-v)h(u,w_1)+(v-w_1)h(u,w_2)}{w_2-w_1} - h(u,v)$$

as we claimed. Next suppose that $C(u,\cdot)$ is not affine in a neighborhood of v. We find that this is the case when for all $(w_1,w_2)\in B_C(u,v)$ we have either $w_1=v$ or $w_2=v$, or both. In all these cases $\sup_{(w_1,w_2)\in B_C(u,v)} \mathcal{S}h(t,u,v,w)=0$, or equivalently, $\tilde{\mathcal{M}}'_C h(u,v)=h(u,v)$. The proof is done. $\square$

Lemma 3.8.1. *Under Assumptions 3.2.1 and 3.4.2, we have* $P\{B_C \subseteq B_{C,n} \subseteq B^\delta_C\} \to 1$ *for any fixed* $\delta>0$ *as* $n\to\infty$ *where*

$$B^\delta_C = \left\{(u,v,w_1,w_2)\in A : \inf_{b\in B_C} d_E(b,(u,v,w_1,w_2)) \le \delta\right\}$$

is the δ*-neighborhood of the contact set* B_C*. Here,* d_E *is the usual Euclidean distance on* $[0,1]^4$.

Proof of Lemma 3.8.1. We prove the lemma by showing (1) $P\{B_C\subseteq B_{C,n}\}\to 1$ and (2) $P\{B_{C,n}\subseteq B^\delta_C\}\to 1$.

(1) First we observe that $\mathcal{S}C(u,v,w_1,w_2)=0$ on the contact set B_C. Since

$\mathcal{S}$ is linear, we may write

$$\sup_{(u,v,w_1,w_2)\in B_C} |\mathcal{S}C_n(u,v,w_1,w_2)| = n^{-1/2} \sup_{(u,v,w_1,w_2)\in B_C} \left|\mathcal{S}\left(n^{1/2}(C_n - C)\right)((u,v,w_1,w_2))\right|.$$

Under the Assumption 3.4.2, $\kappa_n^{-1}\sup_{(u,v,w_1,w_2)\in B_C}|\mathcal{S}C_n(u,v,w_1,w_2)| \to 0$ in probability. Hence we notice that $P\{\sup_{(u,v,w_1,w_2)\in B}|\mathcal{S}C_n(u,v,w_1,w_2)| > \kappa_n\} \to 0$ and therefore $P\{B_C \subseteq B_{C,n}\} \to 1$.

(2) Next we show $P\{B_{C,n} \subseteq B_C^\delta\} \to 1$ by showing that for any fixed $\delta > 0$, $B_{C,n}$ and $A \setminus B_C^\delta$ are eventually disjoint events. Since the contact set B_C is defined with the condition $\mathcal{S}C$ is equal to zero, $|\mathcal{S}C| > 0$ outside the contact set B_C. Exploiting the continuity of $\mathcal{S}$, we have $\inf_{(u,v,w_1,w_2)\in A\setminus B_C^\delta}|\mathcal{S}C(u,v,w_1,w_2)| > 0$. We also note that $\sup_{(u,v,w_1,w_2)\in B_{C,n}}|\mathcal{S}C(u,v,w_1,w_2)|$ approaches to zero as $n \to \infty$, and therefore we have
$P(\sup_{(u,v,w_1,w_2)\in B_{C,n}}|\mathcal{S}C(u,v,w_1,w_2)| < \inf_{(u,v,w_1,w_2)\in A\setminus B_C^\delta}|\mathcal{S}C(u,v,w_1,w_2)|) \to 1$. Thus it must be the case that $B_{C,n}$ and $A \setminus B_C^\delta$ are disjoint with probability approaching one. □

Lemma 3.8.2. *For any $g \in \ell^\infty([0,1]^2)$, we have the following inequality*

$$\sup_{(u,v,w_1,w_2)\in A} |\mathcal{S}g(u,v,w_1,w_2)| \le 2 \sup_{(u,v)\in[0,1]^2} |g(u,v)|.$$

Proof of Lemma 3.8.2. From the definition of $\mathcal{S}$ in Definition 3.4.1, and the property of supremum operator,

$$\begin{aligned}
&\sup_{(u,v,w_1,w_2)\in A} |\mathcal{S}g(u,v,w_1,w_2)| \\
=\ &\sup_{(u,v,w_1,w_2)\in A} \left|\frac{(w_2-v)g(u,w_1)+(v-w_1)g(u,w_2)}{w_2-w_1} - g(u,v)\right| \\
\le\ &\sup_{(u,v,w_1,w_2)\in A} \left|\frac{(w_2-v)g(u,w_1)+(v-w_1)g(u,w_2)}{w_2-w_1}\right| + \sup_{(u,v)\in[0,1]^2}|g(u,v)| \\
\le\ &2\sup_{(u,v)\in[0,1]^2}|g(u,v)|.
\end{aligned}$$

The last inequality follows from that the ratio $\{(w_2-v)g(u,w_1)+(v-w_1)g(u,w_2)\}$

$/(w_2 - w_1)$ is a convex combination of the two points $g(u, w_1)$ and $g(u, w_2)$ so the supremum is bounded by $\sup_{(u,v)\in[0,1]^2} |g(u, v)|$. ☐

Proof of Theorem 3.4.2. For notational convenience let us define

$$\mathcal{J}'_C \mathcal{G}_C = \|\tilde{\mathcal{M}}'_C \mathcal{G}_C - \mathcal{G}_C\|_p.$$

Also for each n, let $\hat{\mathcal{J}}'_n : \ell^\infty([0, 1]^2) \to \mathbb{R}$ be the operator

$$\hat{\mathcal{J}}'_n f = \left\| \sup_{(w_1,w_2)\in B_{C,n}(\cdot)} \mathcal{S}f(\cdot, w_1, w_2) \right\|_p, \quad f \in \ell^\infty([0, 1]^2).$$

The proof is done by applying results from Fang and Santos (2015) in Theorem 3.3, Lemma A.6 and Remark 3.6. Here we verify their key conditions 1) uniform Lipschitz continuity of $\hat{\mathcal{J}}'_n$ in n and 2) the weak convergence $\hat{\mathcal{J}}'_n h \rightsquigarrow \mathcal{J}'_C h$ as $n \to \infty$ for every h in $\mathcal{C}\left([0, 1]^2\right)$.

(1) From the inequality, $|\,\|f_1\|_p - \|f_2\|_p\,| \le \|f_1 - f_2\|_p$, we have

$$\left| \hat{\mathcal{J}}'_n f_1 - \hat{\mathcal{J}}'_n f_2 \right| \le \left\| \sup_{(w_1,w_2)\in B_{C,n}(\cdot)} \mathcal{S}f_1(\cdot, w_1, w_2) - \sup_{(w_1,w_2)\in B_{C,n}(\cdot)} \mathcal{S}f_2(\cdot, w_1, w_2) \right\|_p$$

for $f_1, f_2 \in \ell^\infty([0, 1]^2)$. Now exploiting Hölder's inequality and the linearity of $\mathcal{S}$,

$$\begin{aligned}
&\left\| \sup_{(w_1,w_2)\in B_{C,n}(\cdot)} \mathcal{S}f_1(\cdot, w_1, w_2) - \sup_{(w_1,w_2)\in B_{C,n}(\cdot)} \mathcal{S}f_2(\cdot, w_1, w_2) \right\|_p \\
\le\ & \sup_{(u,v,w_1,w_2)\in B_{C,n}} |\mathcal{S}f_1((u, v, w_1, w_2)) - \mathcal{S}f_2((u, v, w_1, w_2))| \\
\le\ & \sup_{(u,v,w_1,w_2)\in A} |\mathcal{S}f_1((u, v, w_1, w_2)) - \mathcal{S}f_2((u, v, w_1, w_2))| \\
\le\ & \sup_{(u,v,w_1,w_2)\in A} |\mathcal{S}(f_1 - f_2)(u, v, w_1, w_2)| \\
\le\ & 2 \sup_{(u,v)} |f_1(u, v) - f_2(u, v)|\,.
\end{aligned}$$

The last inequality follows from Lemma 3.8.2. So the uniform Lipschitz condition holds.

(2) By the uniform continuity of $\mathcal{S}h$, for any given $\epsilon > 0$ there exists $\delta > 0$ such that

$$\left| \sup_{(u,v,w_1,w_2)\in B_C^\delta} \mathcal{S}h(u,v,w_1,w_2) - \sup_{(u,v,w_1,w_2)\in B_C} \mathcal{S}h(u,v,w_1,w_2) \right| \leq \epsilon.$$

Applying Lemma 3.8.2 with this property, we have the inequalities

$$\begin{aligned} \left| \hat{\mathcal{J}}_n' h - \mathcal{J}_C' h \right| &\leq \left\| \sup_{(v,w)\in B_{C,n}(\cdot)} \mathcal{S}h(\cdot,v,w) - \sup_{(v,w)\in B_C(\cdot)} \mathcal{S}h(\cdot,v,w) \right\|_p \\ &\leq \left\| \sup_{(v,w)\in B_C^\delta(\cdot)} \mathcal{S}h(\cdot,v,w) - \sup_{(v,w)\in B_C(\cdot)} \mathcal{S}h(\cdot,v,w) \right\|_p \\ &\leq \|\epsilon\|_p \leq \epsilon \end{aligned}$$

with probability tending to one as $n \to \infty$. This establishes the weak convergence result $\hat{\mathcal{J}}_n' h \rightsquigarrow \mathcal{J}_C' h$ for h in $\mathcal{C}\left([0,1]^2\right)$ and we finish the proof. □

References

AAS K., CZADO C., FRIGESSI, A. AND BAKKEN, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* **44** 182–198.

ABEGAZ, F. AND NAIK-NIMBALKAR, U. V. (2008). Dynamic copula-based Markov time series. *Communications in Statistics – Theory and Methods* **37** 2447–2460.

ACAR, E. F., GENEST, C. AND NEŠLEHOVÁ, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis* **110** 74–90.

AKERLOF, G. (1976). The economics of caste, the rat race, and other woeful tales. *Quarterly Journal of Economics* **90** 599–617.

ANDERSON, G., LINTON, O. AND WHANG, Y. -J. (2012). Nonparametric estimation and inference about the overlap of two distributions. *Journal of Econometrics* **171** 1–23.

ANDREWS, D. W. K. AND POLLARD, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review* **62** 119–132.

BALBUS, Ł., REFFETT, K., AND WOŹNYC, Ł. (2012). Stationary Markovian equilibrium in altruistic stochastic OLG models with limited commitment. *Journal of Mathematical Economics* **48** 115–132.

BEDFORD, T. AND COOKE, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence* **32** 245–268.

BEDFORD T. AND COOKE R.M. (2002). Vines - a new graphical model for dependent random variables. *Annals of Statistics* **30** 1031–1068.

BEARE, B. K. (2010). Copulas and temporal dependence. *Econometrica* **78** 395–410.

BEARE, B. K. (2012). Archimedean copulas and temporal dependence. *Econometric Theory* **28** 1165–1185.

BEARE, B. K. AND MOON, J. -M. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory* 31 (3), 471-492.p

BEARE, B. K. AND SCHMIDT, L. D. W. (2015). An empirical test of pricing kernel monotonicity. To appear in the *Journal of Applied Econometrics*.

BEARE, B. K. AND SEO, J. (2014). Time irreversible copula-based Markov models. *Econometric Theory* **30** 923–960.

BEARE, B. K. AND SHI, X. (2015). An improved bootstrap test of density ratio ordering. Unpublished manuscript.

BECKER, G. S. AND TOMES, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy* **87** 1153–1189.

BECKER, G. S. AND TOMES, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics* **4** S1–S39.

BELAIRE-FRANCH, J. AND CONTRERAS, D. (2004). A power comparison among tests for time reversibility. *Economics Bulletin* **3** 1–17.

BLUNDELL, R., GOSLING, A., ICHIMURA, H. AND MEGHIR, C. (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica* **75** 323–363.

BONHOMME, S. AND ROBIN, J. -S. (2009). Assessing the equalizing force of mobility using short panels: France 1990–2000. *Review of Economic Studies* **76** 63–92.

BONNANS, J. F. AND SHAPIRO, A. (2000). *Perturbation Analysis of Optimization Problems.* Springer-Verlag, New York.

BOUYÉ, E. AND SALMON, M. (2009). Dynamic copula quantile regressions and tail area dynamic dependence in Forex markets. *European Journal of Finance* **15** 721–750.

BOUZEBDA, S., AND CHERFI, M. (2012). Test of symmetry based on copula function. *Journal of Statistical Planning and Inference* **142** 1262–1271.

BRADLEY, R. C. (2007). *Introduction to Strong Mixing Conditions,* v. 1. Kendrick Press, Heber City.

BÜCHER, A. AND DETTE, D. (2010). A note on bootstrap approximations for the empirical copula process. *Statistics and Probability Letters* **80** 1925–1932.

BÜCHER, A. AND RUPPERT, M. (2012). Consistent testing for a constant copula under strong mixing based on the tapered block multiplier technique. arXiv:1206.1675 [math.ST].

CAROLAN, C. A. (2002). The least concave majorant of the empirical distribution function. *Canadian Journal of Statistics* **30** 317–328.

CAROLAN, C. A. AND TEBBS, J. M. (2005). Nonparametric tests for and against likelihood ratio ordering in the two sample problem. *Biometrika* **92** 159–171.

CASON, T. N., FRIEDMAN, D. AND WAGENER, F. (2005). The dynamics of price dispersion, or Edgeworth variations. *Journal of Economic Dynamics and Control* **29** 801-??22.

CHANG, J. T. AND POLLARD, D. (1997). Conditioning as disintegration. *Statistica Neerlandica* **51** 287–317.

CHEN, T. -K. (2003). Testing serial independence against time irreversibility. *Studies in Nonlinear Dynamics and Econometrics* **7** 1–28.

CHEN, T. -K., CHOU, R. Y. AND KUAN, C. -M. (2000). Testing time reversibility without moment restrictions. *Journal of Econometrics* **95** 199–218.

CHEN, T. -K. AND KUAN, C. M. (2002). Time irreversibility and EGARCH effects in US stock index returns. *Journal of Applied Econometrics* **17** 565–578.

CHEN, X. AND FAN, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics* **130** 307–335.

CHEN, X., KOENKER, R. AND XIAO, Z. (2009). Copula-based nonlinear quantile autoregression. *Econometrics Journal* **12** S50–S67.

CHEN, X., WU, W. B. AND YI, Y. (2009). Efficient estimation of copula-based semiparametric Markov models. *Annals of Statistics* **37** 4214–4253.

CHERUBINI, U., LUCIANO, E. AND VECCHIATO, W. (2004). *Copula Methods in Finance*. Wiley, Chichester.

CHERUBINI, U., MULINACCI, S., GOBBI, F. AND ROMAGNOLI, S. (2011). *Dynamic Copula Methods in Finance*. Wiley Finance Series.

CHERUBINI, U., GOBBI, F., MULINACCI, S. AND ROMAGNOLI, S. (2012). *Dynamic Copula Methods in Finance*. Wiley, Chichester.

CZADO, C. (2010). Pair-copula constructions of multivariate copulas. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T. (Eds.), *Copula Theory and its Applications,* pp. 93–109. Springer-Verlag, Berlin Heidelberg.

DAROLLES, S., FLORENS, J. -P. AND GOURIÉROUX, C. (2004). Kernel-based nonlinear canonical analysis and time reversibility. *Journal of Econometrics* **119** 323–353.

DARSOW, W. F., NGUYEN, B. AND OLSEN, E. T. (1992). Copulas and Markov processes. *Illinois Journal of Mathematics* **36** 600–642.

DARDANONI, V., FIORINI, M. AND FORCINA, A. (2012). Stochastic monotonicity in intergenerational mobility tables. *Journal of Applied Econometrics* **27** 85–107.

DEARDEN, L., FITZSIMONS, E., GOODMAN, A. AND KAPLAN, G. (2008). Higher education funding reforms in England: the distributional effects and the shifting balance of costs. *Economic Journal* **118** F100–F125.

DEHEUVELS, P. (1981a). A Kolmogorov–Smirnov type test for independence and multivariate samples. *Revue Roumaine de Mathématiques Pures et Appliquées* **26** 213–226.

DEHEUVELS, P. (1981b). A nonparametric test for independence. *Publications de l'Institut de Statistique de l'Université Paris* **26** 29–50.

DELGADO, M. A. AND ESCANCIANO, J. C. (2012). Distribution-free tests of stochastic monotonicity. *Journal of Econometrics* **170** 68–75.

DE LA PEÑA, V., IBRAGIMOV, R. AND SHARAKHMETOV, S. (2006). Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series. In Rojo, J. (Ed.), *2nd Erich L. Lehmann Symposium – Optimality. IMS Lecture Notes -??Monograph Series* **49** 183–209.

DELONG, J. B. AND SUMMERS, L. H. (1986). Are business cycles symmetrical? In Gordon, R. J. (Ed.), *The American Business Cycle: Continuity and Change.* University of Chicago Press, Chicago.

DEMARTA, S. AND MCNEIL, A. J. (2005). The t copula and related copulas. *International Statistical Review* **73** 111–129.

DIßMANN, J., BRECHMANN, E. C., CZADO, C. AND KUROWICKA, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis* **59** 52–69.

DOMMA, F., GIORDANO, S. AND PERRI, P. F. (2009). Statistical modeling of temporal dependence in financial data via a copula function. *Communications in Statistics – Simulation and Computation* **38** 703–728.

DOUKHAN, P., FERMANIAN J. -D. AND LANG, G. (2009). An empirical central limit theorem with applications to copulas under weak dependence. *Statistical Inference for Stochastic Processes* **12** 65–87.

ECKERT, A. (2002). Retail price cycles and response asymmetry. *Canadian Journal of Economics* **35** 52–77.

ECKERT, A. (2003). Retail price cycles and presence of small firms. *International Journal of Industrial Organization* **21** 151–170.

ELLISON, G. AND ELLISON, S. F. (2011). Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *American Economic Journal: Microeconomics* **3** 1–36.

ERICSON, R., PAKES, A. (1995). Markov-perfect industry dynamics: a framework for empirical work. *Review of Economic Studies* **62** 53–82.

FANG, Z. AND SANTOS, A. (2015). Inference on directionally differentiable functions. Unpublished manuscript.

FERMANIAN, J. D., RADULOVIĆ, D. AND WEGKAMP, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10** 847–860.

FENTAW, A. AND NAIK-NIMBALKAR, U. V. (2008). Dynamic copula-based Markov time series. *Comm. Statist. Theory Methods* **37** 2447–2460.

FERMANIAN, J. -D. AND SCAILLET, O. (2003). Nonparametric estimation of copulas for time series. *Journal of Risk* **5** 25–54.

FISCHER, M. (2011). Multivariate copulae. In Kurowicka, D. and Joe, H. (Eds.), *Dependence Modelling: Vine Copula Handbook,* pp. 19–36. World Scientific Publishing Co., Singapore.

FREES, E. W. AND WANG, P. (2006). Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics* **38** 360–373.

GAENSSLER, P. AND STUTE, W. (1987). *Seminar on Empirical Processes, DMV Sem*. **9**. Basel: Birkhäuser.

GAGLIARDINI, P. AND GOURIÉROUX, C. (2008). Duration time-series models with proportional hazard. *Journal of Time Series Analysis* **29** 74–124.

GENEST, C., GHOUDI, K. AND RIVEST, L. -P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82** 543–552.

GENEST, C., NEŠLEHOVÁ, J. AND QUESSY, J. -F. (2012). Tests of symmetry for bivariate copulas. *Annals of the Institute of Statistical Mathematics* **64** 811–834.

GENEST, C., GHOUDI, K. AND RIVEST, L. -P. (1998). Comment on the article of E. W. Frees and E. A. Valdez entitled “Understanding relationships using copulas”. *North American Actuarial Journal* **2** 143–149.

GENEST, C., NEŠLEHOVÁ, J. AND QUESSY, J. -F. (2012). Tests for symmetry for bivariate copulas. *Annals of the Institute of Statistical Mathematics* **64** 811–834.

GRANGER, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37** 424–438.

HAMILTON, J. D. (2009). Causes and consequences of the oil shock of 2007–08. *Brookings Papers on Economic Activity* Spring 2009, 215–259.

HAN, S. AND MULLIGAN, C. B. (2001). Human capital, heterogeneity and estimated degrees of intergenerational mobility. *Economic Journal* **111** 207–243.

HINICH, M. J. AND ROTHMAN, P. (1998). Frequency-domain test of time reversibility. *Macroeconomic Dynamics* **2** 72–88.

HOBÆK HAFF, I., AAS, K., AND FRIGESSI, A. (2010). On the simplified pair copula construction – simply useful or too simplistic? *Journal of Multivariate Analysis* **101** 1296–1310.

HOBÆK HAFF, I. (2012). Comparison of estimators for pair-copula constructions. *Journal of Multivariate Analysis* **110** 91–105.

HOBÆK HAFF, I. (2013). Parameter estimation for pair-copula constructions. *Bernoulli* **19** 462–491.

HOPENHAYN, H. A. AND PRESCOTT, E. C. (1992). Stochastic monotonicity and stationary distributions for dynamic economics. *Econometrica* **60** 1387–1406.

IBRAGIMOV, R. (2009). Copula-based characterizations for higher-order Markov processes. *Econometric Theory* **25** 819–846.

IBRAGIMOV, I. AND MÜLLER, U. K. (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business and Economic Statistics* **28** 453–468.

JOE, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In Rüschendorf, L., Schweizer, B. and Taylor, M. D. (Eds.), *Distributions with fixed marginals and related topics,* pp. 120-??41. Institute of Mathematical Statistics, Hayward.

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

JOE, H. (2011). Dependence comparisons of vine copulae with four or more variables. In Kurowicka, D. and Joe, H. (Eds.), *Dependence Modelling: Vine Copula Handbook,* pp. 139–164. World Scientific Publishing Co., Singapore.

KEYNES, J. M. (1936). *The General Theory of Employment, Interest and Money.* Macmillan, London.

KHOUDRAJI, A. (1995). *Contributions à l'étude des copules et à la modélisation de valeurs extrêmes bivariées*. Ph.D. thesis, Université Laval, Québec City.

KLEMENT, E. P., KOLESÁROVÁ, A., MESIAR, R. AND SEMPI, C. (2007). Copulas constructed from horizontal sections.*Communications in Statistics - Theory and Methods* **36** 2901–2911.

KOJADINOVIC, I. AND YAN, J.(2011). A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing* **21** 17–30.

KOJADINOVIC, I., YAN, J. AND HOLMES, M. (2011). Fast large-sample goodness-of-fit tests for copulas. *Statistica Sinica* **21** 841–871.

KUROWICKA, D. AND COOKE, R. M. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley, Chichester.

LAMBERT, P. AND VANDENHENDE, F. (2002). A copula-based model for multivariate non-normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. *Statistics in Medicine* **21** 3197–3217.

LANCASTER, H. O. (1958). The structure of bivariate distributions. *Annals of Mathematical Statistics* **29** 719–736.

LEE, S., SONG, K. AND WHANG, Y. -J. (2014). Testing for a general class of functional inequalities. CeMMAP Working Paper No. 09/14, Institute for Fiscal Studies.

LEE, S., LINTON, O. B. AND WHANG, Y. -J. (2009). Testing for stochastic monotonicity. *Econometrica* **77** 585–602.

LEWIS, M. S. AND NOEL, M. D. (2011). The speed of gasoline price response in markets with and without Edgeworth cycles. *Review of Economics and Statistics* **93** 672–682.

LIEBSCHER, E. (2008). Construction of asymmetric multivariate copulas. *Journal of Multivariate Analysis* **99** 2234–2250.

LINTON, O. B., SONG, K. AND WHANG, Y. -J. (2010). An improved bootstrap test of stochastic dominance. *Journal of Econometrics* **154** 186–202.

LONGLA, M. AND PELIGRAD, M. (2012). Some aspects of modeling dependence in copula-based Markov chains. *Journal of Multivariate Analysis* **111** 234–240.

LUCAS, R. E. AND STOKEY, N. L. (1989). *Recursive Methods in Dynamics Economics.* Harvard University Press, Cambridge, MA.

MASKIN, E. AND TIROLE, J. (1988). A theory of dynamic oligopoly II: price competition, kinked demand curves and Edgeworth cycles. *Econometrica* **56** 571–599.

MCCAUSLAND, W. J. (2007). Time reversibility of stationary regular finite-state Markov chains. *Journal of Econometrics* **136** 303–318.

MEHRA, Y. P. (2004). The output gap, expected future inflation and inflation dynamics: another look. *B.E. Journal of Macroeconomics* **4**, Article 17.

MILGROM, P. R. AND ROBERTS, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy* **94** 796–821.

MINICOZZI, A. L. (2003). Estimation of sons' intergenerational earnings mobility in the presence of censoring. *Journal of Applied Econometrics* **18** 291–314.

MULLIGAN, C. B. (1999). Galton versus the human capital approach to inheritance. *Journal of Political Economy* **107** S184–S224.

NEFTÇI, S. N. (1984). Are economic time series asymmetric over the business cycle? *Journal of Political Economy* **92** 307–328.

NELSEN, R. B. (2006). *An Introduction to Copulas.* 2nd edition. Springer-Verlag, New York.

NELSEN, R. B. (2007). Extremes of nonexchangeability. *Statistical Papers* **48** 329–336.

NOEL, M. D. (2007). Edgeworth price cycles, cost-based pricing and sticky pricing in retail gasoline markets. *Review of Economics and Statistics* **89** 324–334.

NOEL, M. D. (2008). Edgeworth cycles and focal prices: computational dynamic Markov equilibria. *Journal of Economics and Management Strategy* **17** 345–377.

NOEL, M. D. (2011). Edgeworth price cycles. In Durlauf, S. N. and Blume, L. E. (Eds.), *New Palgrave Dictionary of Economics.* Palgrave Macmillan.

OLLEY, G. S. AND PAKES, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica* **64** 1263–1297.

PAKES, A. (1986). Patents as options: some estimates of the value of holding European patent stocks. *Econometrica* **54** 755–784.

PAPARODITIS, E. AND POLITIS, D. N. (1998). The backward local bootstrap for conditional predictive inference in nonlinear time series. In Lipitakis, E. A. (Ed.), *Proceedings of the 4th Hellenic European Conference on Computer Mathematics and its Applications.* Lea Publishing, Athens.

PAPARODITIS, E. AND POLITIS, D. N. (2001). A Markovian local resampling scheme for nonparametric estimators in time series analysis. *Econometric Theory* **17** 540–566.

PAPARODITIS, E. AND POLITIS, D. N. (2002). The local bootstrap for Markov processes. *Journal of Statistical Planning and Inference* **108** 301–328.

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

PATTON, A. J. (2009). Copula-based models for financial time series. In Anderson, T. G., Davis, R. A., Kreiß, J. -P. and Mikosch, T. (Eds.), *Handbook of Financial Time Series*, pp. 767–785. Springer-Verlag, Berlin Heidelberg.

PATTON, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis* **110** 4–18.

PATTON, A. J. (2013). Copula methods for forecasting multivariate time series. In Elliott, G. and Timmermann, A. (Eds.), *Handbook of Economic Forecasting* vol. 2, pp. 899–960. Access Online via Elsevier.

PELTZMAN, S. (2000). Prices rise faster than they fall. *Journal of Political Economy* **108** 466–502.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* Institute of Mathematical Statistics, Hayward.

PSARADAKIS, Z. (2008). Assessing time reversibility under minimal assumptions. *Journal of Time Series Analysis* **29** 881–905.

RACINE, J. S. AND MAASOUMI, E. (2007). A versatile and robust metric entropy test of time-reversibility, and other hypotheses. *Journal of Econometrics* **138** 547–567.

RADULOVIĆ, D. (2002). On the bootstrap and empirical processes for dependent sequences. In Dehling, H., Mikosch, T. and Sørensen, M. (Eds.), *Empirical Process Techniques for Dependent Data.* Birkhäuser, Boston.

RAMSEY, J. B. AND ROTHMAN, P. (1996). Time irreversibility and business cycle asymmetry. *Journal of Money, Credit and Banking* **28** 1–21.

RÉMILLARD, B. AND SCAILLET, O. (2009). Testing for equality between two copulas. *Journal of Multivariate Analysis* **100** 377–386.

RÉMILLARD, B., PAPAGEORGIOU, N. AND SOUSTRA, F. (2012). Copula-based semiparametric models for multivariate time series. *Journal of Multivariate Analysis* **110** 30–42.

RESTUCCIA, D. AND URRUTIA, C. (2004). Intergenerational persistence of earnings: the role of early and college education. *American Economic Review* **94** 1354–1378.

RIO, E. (2000). *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants.* Springer-Verlag, Berlin.

ROBINSON, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis* **4** 185–207.

ROTHMAN, P. (1991). Further evidence on the asymmetric behavior of unemployment rates over the business cycle. *Journal of Macroeconomics* **13** 291–298.

SAVU, C. AND NG, W. L. (2005). The SCoD model: analyzing durations with a semiparametric copula approach. *International Review of Finance* **5** 55–74.

SCAILLET, O. (2005). A Kolmogorov-Smirnov type test for positive quadrant dependence. *Canadian Journal of Statistics* **33** 415–427.

SEGERS, J. (2012). Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli* **18** 764–782.

SHAPIRO, A. (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications* **66** 477–487.

SHAPIRO, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research* **30** 169–186.

SHARIFDOOST, M., MAHMOODI, S. AND PASHA, E. (2009). A statistical test for time reversibility of stationary finite state Markov chains. *Applied Mathematical Sciences* **52** 2563–2574.

SMALL, D., TAN, Z., LORCH, S., AND BROOKHART, A. (2014). Instrumental variable estimation when compliance is not deterministic: the stochastic monotonicity assumption. Unpublished manuscript.

SMITH, M., MIN, A., ALMEIDA, C. AND CZADO, C. (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association* **105** 1467–1479.

SOLON, G. (1992). Intergenerational income mobility in the United States. *American Economic Review* **82** 393–408.

SOLON, G. (2002). Cross-country differences in intergenerational earnings mobility. *Journal of Economic Perspectives* **16** 59–66.

SPENCE, A. M. (1973). Job market signaling. *Quarterly Journal of Economics* **87** 355–374.

STÖBER, J., JOE, H. AND CZADO, C. (2013). Simplified pair copula constructions - limitations and extensions. *Journal of Multivariate Analysis* **119** 101–118.

STUTE, W. (1984). The oscillation behavior of empirical processes: the multivariate case. *Annals of Probabability* **12** 361–379.

SUN, J., FREES, E. W. AND ROSENBERG, M. A. (2008). Heavy-tailed longitudinal data modeling using copulas. *Insurance: Mathematics and Economics* **42** 817–830.

TAPPATA, M. (2009). Rockets and feathers: understanding asymmetric pricing. *RAND Journal of Economics* **40** 673–687.

TORGOVITSKY, A. (2012). *Essays in Econometric Theory*. Doctoral thesis, Yale University.

TSUKAHARA, H. (2005). Semiparametric estimation in copula models. *Canadian Journal of Statistics* **33** 357–375.

VAN DER VAART, A. W. AND WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

VICKREY, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance* **16** 8–37.

WANG, Z. (2009). (Mixed) strategy in oligopoly pricing: evidence from gasoline price cycles before and under a timing regulation. *Journal of Political Economy* **117** 987–1030.

WILSON, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory* **16** 167–207.

YI, W. AND LIAO, S. S. (2010). Statistical properties of parametric estimators for Markov chain vectors based on copula models. *Journal of Statistical Planning and Inference* **140** 1465–1480.