Time irreversible copula-based Markov models

Brendan K. Beare and Juwon Seo*

Department of Economics, University of California, San Diego

April 19, 2012

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.

*We thank William McCausland for providing us with the Canadian gasoline price mark-up data used in our empirical section, and Peter Phillips, Dimitris Politis, Hal White, and seminar participants at UC San Diego for helpful comments. The first author thanks Roman Tymkiv for introducing him to circulatory analysis in a poster session at the 2010 CIREQ Time Series Conference in Montreal.

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 2 we discuss this equivalence, and explain how a technique proposed by Genest et al. (1998) may be used to construct parametric families of nonexchangeable copula functions. Our main contributions are provided in Sections 3 and 4. In Section 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 a similar test proposed by Paparoditis and Politis (2002). In Section 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 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 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.

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

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

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

- (a) *For any 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 equations. 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 2.1 or a copula C satisfying property (c) in Proposition 2.1 is said to be *exchangeable*. Nelsen (2007) studied some aspects of nonexchangeable copulas. He proposed to measure the nonexchangeability of a copula C using the following quantity:

$$\delta(C) = 3 \sup_{u,v} \left|C(u,v) - C(v,u)\right|. \tag{2.1}$$

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. 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 3 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. Genest et al. (1998) proposed a particularly convenient method by which this may be achieved. Proposition 2 in their paper states that, if C is an exchangeable copula and $\alpha, \beta \in [0, 1]$, then the following transformation of C is a copula:

$$\tilde{C}(u, v) = u^{1-\alpha} v^{1-\beta} C(u^\alpha, v^\beta). \tag{2.2}$$

We may use (2.2) 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 $\gamma \in [1, \infty)$. If we apply transform (2.2) to the Gumbel copula, then we obtain the family of so-called asymmetric Gumbel copulas:

$$\tilde{C}^{\mathrm{Gmbl}}(u, v) = u^{1-\alpha} v^{1-\beta} \exp\left(-\left((-\alpha \ln u)^\gamma + (-\beta \ln v)^\gamma\right)^{1/\gamma}\right). \tag{2.3}$$

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 $\gamma > 1$. When $\gamma \to \infty$, the asymmetric Gumbel copula reduces to the well-known Marshall-Olkin copula (Nelsen, 2006, p. 53) with parameters α and β.

In Figure 2.1 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 (2.1) 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 2.1 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 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).

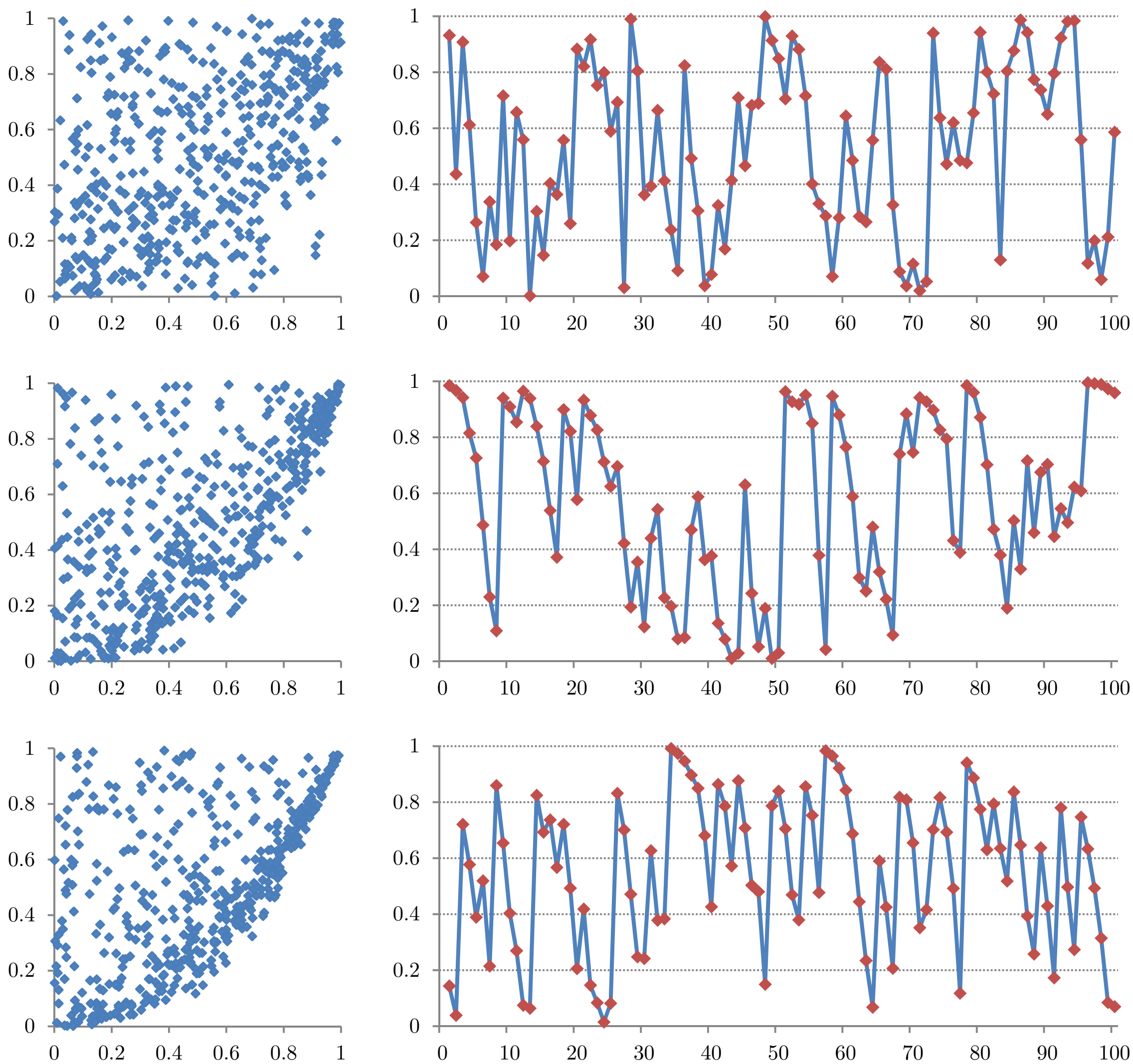

Figure 2.1: Scatterplots and Markov sample paths generated using the asymmetric Gumbel copula. We set $\alpha = 1$, $\beta = 0.5$ and take the invariant distribution to be uniform on $(0, 1)$. γ is equal to 2 in the top row, 5 in the center row, and 10 in the bottom row.

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 3.1 we explain how our test statistic is constructed, and discuss its asymptotic behavior under time reversibility and time irreversibility. In Section 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 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.

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 (2.1) above. Recalling Proposition 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. We shall obtain the asymptotic behavior of θ_T under the following conditions on $\mathscr{X}$.

Assumption 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 3.1 are basic to our analysis. The mixing condition introduced in part (c) is mild for practical purposes. Beare (2010, 2012) identifies 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 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 3.1. *Under Assumption 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*

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

(*b*) *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 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 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 3.1. Theorem 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

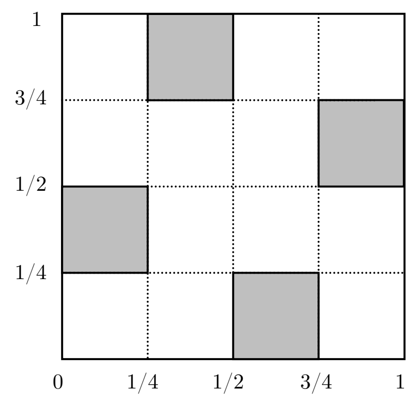

Figure 3.1: A pair of random variables distributed uniformly over the shaded region is nonexchangeable, but the distribution of their difference is symmetric about zero.

(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 3.1. 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 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 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 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 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

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

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 3.1(a), is unknown. We may nevertheless approximate these laws using a bootstrap procedure. Since $\mathscr{X}$ is typically serially dependent, a standard nonparametric bootstrap based on independent resampling from the observed pairs (X_t, X_{t+1}) cannot be expected to yield useful results. On the other hand, a block bootstrap would fail to exploit the Markovian structure of $\mathscr{X}$. Instead, we propose to use the local bootstrap of Paparoditis and Politis (2002), which was designed specifically for Markovian time series. 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 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:

$$H^*_T(x, y) = \frac{1}{T-1} \sum_{t=1}^{T-1} 1(X^*_t \le x, X^*_{t+1} \le y).$$

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 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 3.1. *Under Assumption A.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 A.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 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 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 3.2. *Under Assumption A.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 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 3.2, found in the Appendix, is a straightforward application of Lemma 3.1 and the continuous mapping theorem.

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 (2.3). 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, and also the rejection rate of the test of Paparoditis and Politis (2002), for a range of values of γ. In all cases, we set $T=100$ and employed 500 bootstrap replications and 1000 experimental replications. The nominal size of both tests was 0.05. 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.

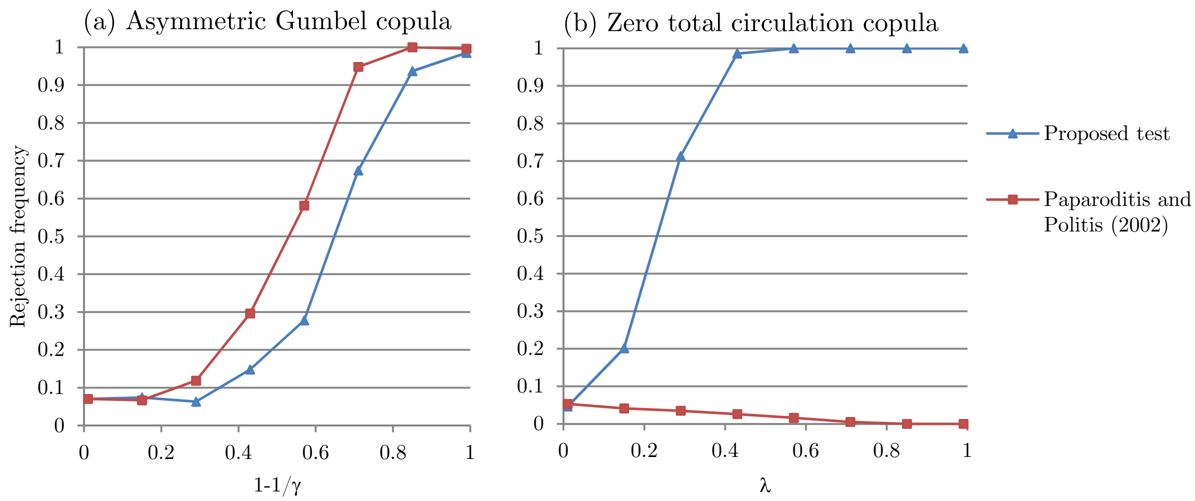

Figure 3.2: Rejection rates of our time reversibility test, and the test of Paparoditis and Politis (2002). Panel (a) displays results for the asymmetric Gumbel copula with $\alpha = 1$, $\beta = 0.5$, and $\gamma \in [1, \infty)$. Panel (b) displays results for a convex linear combination of the product copula and the copula displayed in Figure 3.1; the weight on the latter is $\lambda \in [0, 1]$. We set $T = 100$ and employed 500 bootstrap replications and 1000 experimental replications. The nominal size of the tests is 0.05.

The outcome of our numerical calculations using the asymmetric Gumbel copula is displayed in Figure 3.2(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. Both tests have a rejection rate of 0.070 when $\gamma = 1$, indicating a minor tendency to overreject the null hypothesis of time reversibility. As $\gamma \to \infty$, the rejection rate of both tests rises to approximately one. At intermediate values of γ, the test of Paparoditis and Politis has a uniformly higher rejection rate than the test proposed here. The natural conclusion is that, for this family of distribution functions H, our test is less powerful than the test of Paparoditis and Politis.

In Remark 3.1 we noted that the test of Paparoditis and Politis should be unable to detect deviations from time reversibility that are such that $X_{t+1} - X_t$ is distributed symmetrically about zero. Our second choice of H exploits this fact. We take H to be a convex linear combination of two copula functions. The first of these is the product copula. The second distributes mass uniformly over the shaded area in Figure 3.1. 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 4.1, we refer to this mixture copula as a zero total circulation copula.

The outcome of our numerical calculations using the zero total circulation copula is displayed in Figure 3.2(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. Both tests exhibit good size control: when $\lambda = 0$, the rejection rate of our test is 0.046, and the rejection rate of the test of Paparoditis and Politis is 0.053. As λ increases, the behavior of the two tests is very different. The rejection rate of our test rises quickly to one, while the rejection rate of the test of Paparoditis and Politis decreases to zero.

The test statistic used by Paparoditis and Politis is $\theta_T^{\mathrm{PP}} = \frac{1}{T-1}\sum_{t=1}^{T-1} 1(X_{t+1} > X_t)$, the proportion of differenced observations that are positive. Time reversibility is rejected when $|\theta_T^{\mathrm{PP}} - \frac{1}{2}|$ exceeds a critical value generated using the local bootstrap. Since the zero total circulation copula was specifically constructed so that $P(X_{t+1} > X_t) = \frac{1}{2}$, it is not surprising that the test of Paparoditis and Politis does not achieve a rejection rate in excess of 5%. The fact that the rejection rate declines to zero as $\lambda \to 1$ is less obvious and merits further explanation. Consider the extreme case where $\lambda = 1$. In view of the periodic nature of the distribution in Figure 3.1, the statistic θ_T^{PP} will be exactly equal to one half whenever $T-1$ is a multiple of four. Therefore θ_T^{PP} converges to one half at the rate T^{-1}, and our asymptotic rejection rate will be zero unless our critical value decays to zero at the rate T^{-1} or faster. Apparently critical values obtained from the local bootstrap do not decay to zero at this rate; this may be due to the kernel smoothing used in the construction of bootstrap samples. Consequently, the asymptotic rejection rate falls to zero as $\lambda \to 1$.

Panels (a) and (b) of Figure 3.2 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 any violation of time reversibility. This versatility comes at a price: tests that are constructed to target specific forms of time irreversibility are likely to be more powerful than our test when irreversibility is indeed of that form. Therefore, our test serves to complement existing procedures.

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 4.1. In Section 4.2 we propose a simple copula-based estimator of the circulation density, and investigate its asymptotic and finite sample behavior.

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.

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

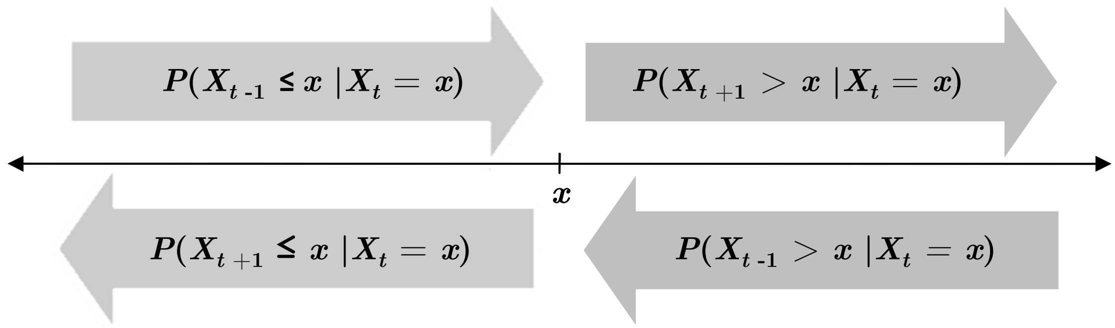

Figure 4.1: Probability upflows and downflows to and from x. The circulation density at $u = F(x)$ is equal to the sum of the upward flows minus the sum of the downward flows, divided by two.

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

$$\mathscr{F}_{\uparrow}(x) + \mathscr{F}^{\downarrow}(x) = 1, \quad \mathscr{F}^{\uparrow}(x) + \mathscr{F}_{\downarrow}(x) = 1. \tag{4.1}$$

The terms upflow and downflow are evocative of the circulation, or current, of a body of water. Figure 4.1 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 $P(X_{t+1} \leq x | X_t = x)$ are not uniquely defined when F is continuous at x, because we are conditioning on a set of measure zero. Rather, $P(X_{t+1} \leq x | X_t = x)$ 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 (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 $\mathscr{F}_\downarrow$:

$$\begin{aligned}
\mathscr{F}_\uparrow(x) + \mathscr{F}^\uparrow(x) &\qquad \textit{probability upflow through } x \\
\mathscr{F}^\downarrow(x) + \mathscr{F}_\downarrow(x) &\qquad \textit{probability downflow through } x \\
\mathscr{F}_\uparrow(x) - \mathscr{F}_\downarrow(x) &\qquad \textit{net probability upflow to } x \\
\mathscr{F}^\uparrow(x) - \mathscr{F}^\downarrow(x) &\qquad \textit{net probability upflow from } x \\
\mathscr{F}_\uparrow(x) + \mathscr{F}^\uparrow(x) - \mathscr{F}^\downarrow(x) - \mathscr{F}_\downarrow(x) &\qquad \textit{net probability upflow through } x.
\end{aligned}$$

A consequence of the identities in (4.1) 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 $Q(u) = \inf\{y : F(y) \geq u\}$, 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 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 4.1 may be found in the Appendix.

Theorem 4.1. *Let $\mathscr{X}$ be a stationary real valued Markov chain with continuous invariant distribution F, and copula C admitting continuous partial derivatives $\partial_1 C$ 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 4.2 we use the expression for $\psi(u)$ given in Theorem 4.1 to graph the circulation density functions corresponding to the asymmetric Gumbel copula given in (2.3), 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 2.1. 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 2.1, 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

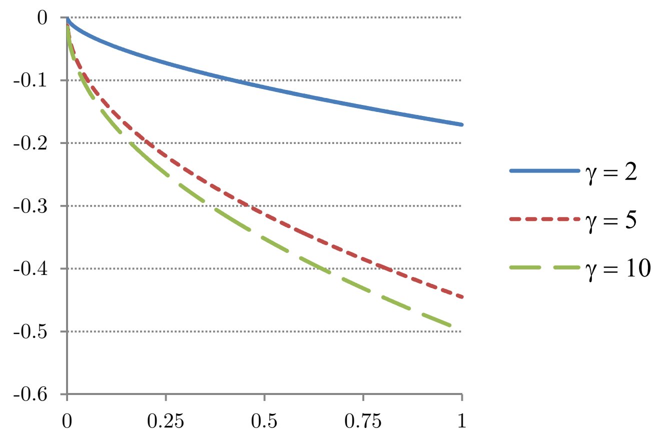

Figure 4.2: Circulation densities for the asymmetric Gumbel copula with $\alpha = 1$, $\beta = 0.5$, and $\gamma = 2, 5, 10$.

way, the total circulation has a convenient interpretation. The proof may be found in the Appendix.

Theorem 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 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 3.2(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 4.3 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

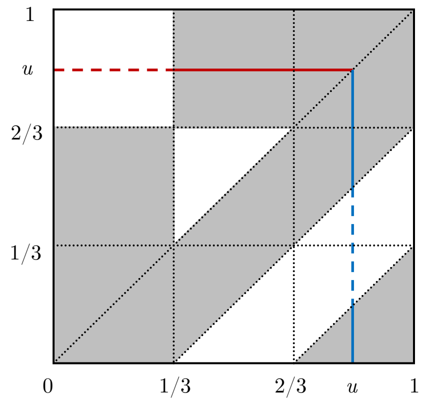

Figure 4.3: If (X_t, X_{t+1}) is distributed uniformly over the shaded region, then $\mathscr{X}$ is time irreversible, yet has zero circulation density at all quantiles. The probability upflow to u is equal to the length of the solid part of the line extending between $(0, u)$ and (u, u), while the probability downflow from u is equal to the length of the solid part of the line extending between $(u, 0)$ and (u, u).

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 4.3 reveals that these two quantities are 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 4.1 and Theorem 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 4.2 makes it clear that our own definition of total circulation differs from McCausland's definition by a factor of one half.

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 its finite sample performance using Monte Carlo simulation.

4.2.1 Estimator and asymptotic properties

Theorem 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

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

We will establish the pointwise asymptotic properties of $\hat{\psi}_T$ under the following technical conditions.

Assumption 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.

(e) The bandwidth $h = h_T$ satisfies $Th^3 \to \infty$ and $Th^4 \to 0$.

Parts (a,b,c) of Assumption 4.1 may be compared to the corresponding parts of Assumption 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 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 4.1(e) provides the admissible rates of decay for the bandwidth h. The requirement that $Th^4 \to 0$ could likely be weakened if we were to allow nonzero bias in the asymptotic distribution of $\hat{\psi}_T(u)$, but we do not pursue this extension here.

Theorem 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 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 4.3. *Suppose $\mathscr{X}$ satisfies Assumption 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 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 4.3 can be used to construct pointwise asymptotic confidence intervals for $\psi(u)$. Alternatively, the local bootstrap of Paparoditis and Politis (2002) could be used to construct confidence intervals. We investigate this possibility in the finite sample simulations reported in the following subsection.

Our proof of Theorem 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.

4.2.2 Finite sample performance

Here we report some limited numerical evidence pertaining to the finite sample performance of our circulation density estimator. For $T = 75$ and $T = 150$, we generated 1500 samples of T iid standard normal random variables. For each sample 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 of Paparoditis and Politis (2002), with 600 bootstrap replications. For each quantile, we calculated the coverage rate of each confidence band over the 1500 randomly generated samples, and also the mean squared error for the circulation density estimator.

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 followed the Silverman rule of thumb and set $h = 1.06\hat{s}_T T^{-1/5}$, where $\hat{s}_T$ is the sample standard deviation.

The results of our experiment are provided in Table 4.1. For both sample sizes T and all quantiles u, the coverage probabilities of our pointwise confidence bands were extremely close to the nominal rate. This suggests that, in this context, the local bootstrap procedure does a very good job at approximating the sampling uncertainty associated with our estimators. The mean square errors for our estimators were also very small, peaking at only 0.0035 when $T = 75$ and 0.0023 when $T = 150$.

We have not reported coverage rates for confidence intervals obtained using the first order asymptotic approximation given in Theorem 4.3, and variance estimator $\hat{\sigma}_T^2(u)$. Con-

Sample size: $T = 75$					
	Quantile				
	0.1	0.3	0.5	0.7	0.9
95% coverage	0.955	0.949	0.939	0.955	0.944
90% coverage	0.897	0.899	0.895	0.902	0.892
80% coverage	0.789	0.789	0.803	0.793	0.776
Mean square error	0.0023	0.0033	0.0035	0.0033	0.0023
Sample size: $T = 150$					
	Quantile				
	0.1	0.3	0.5	0.7	0.9
95% coverage	0.933	0.943	0.951	0.948	0.953
90% coverage	0.884	0.895	0.905	0.901	0.884
80% coverage	0.770	0.817	0.792	0.810	0.782
Mean square error	0.0016	0.0022	0.0023	0.0022	0.0016

Table 4.1: Coverage rates and mean square errors for our circulation density estimator, with confidence bands constructed using the local bootstrap. We employed 600 bootstrap replications and 1500 experimental replications. Samples were iid standard normal.

fidence intervals constructed in this way tended to be excessively conservative. With $T = 150$, the coverage rate for nominal 80% confidence intervals was above 95% at all quantiles. Even with $T = 1500$, the coverage rate remained above 91%. The discrepancy between asymptotic and finite sample results may be explained by the fact that the automatic bandwidth selection rules used in our simulations generate bandwidths that decay to zero at the rate $T^{-1/5}$, whereas Assumption 4.1(e) requires our bandwidths to decay faster than $T^{-1/4}$. Assumption 4.1(e) eliminates bias in the first-order asymptotic distribution of $(Th)^{1/2}(\hat{\psi}_T(u) - \psi(u))$; however, a general principle for optimal bandwidth selection is that one seeks to achieve an ideal balance between asymptotic bias and variance, and such a balance would typically entail nonzero bias. We recommend that the local bootstrap be used to form confidence bands in situations where a bandwidth selection procedure not satisfying Assumption 4.1(e) is used.

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 5.1(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, followed by shorter periods of

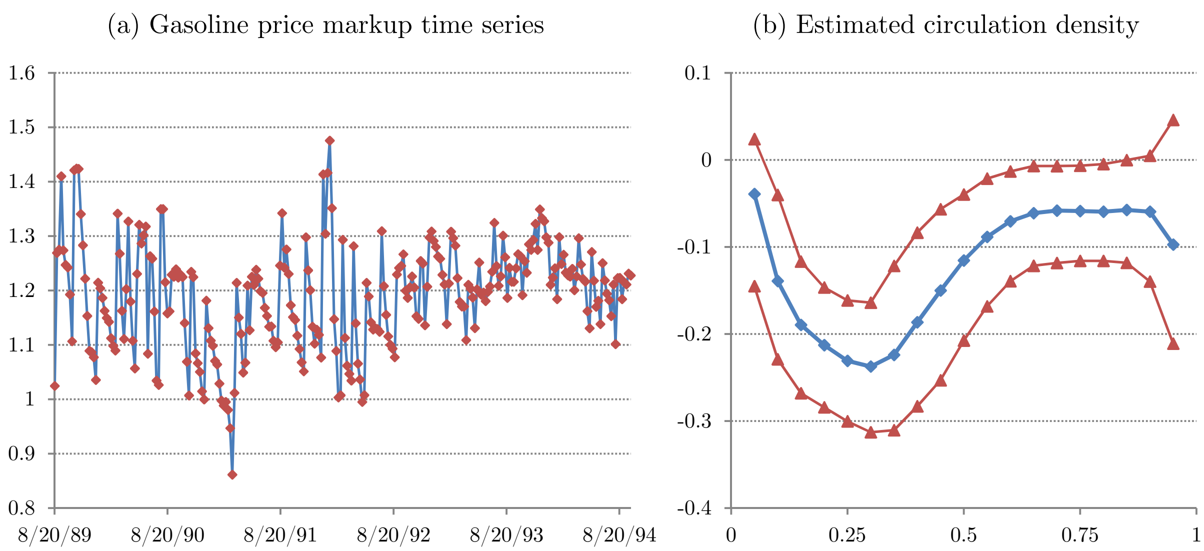

Figure 5.1: Panel (a) displays the average weekly gasoline price markups in Windsor, Ontario from 8/20/1989 to 9/25/1994. Panel (b) displays the circulation density estimated using these data, with pointwise 95% confidence bands constructed using the local bootstrap.

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 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 5.1(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 0.000, indicating overwhelming rejection of reversibility. In Figure 5.1(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, and the 95% confidence bands exclude zero at all quantiles between 0.1 and 0.85. 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 0.3 quantile of the invariant distribution. In the language of Section 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 0.3 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 5.1(a) that our price markup series is somewhat more volatile in the first half of the sample than it is in the second half. The class of models considered in this paper permits conditional heteroskedasticity, but of a form limited by the assumption that our series is Markovian. In particular, ARCH-type conditional heteroskedasticity (Engle, 1982) may be accommodated in our framework, but this is not true in general for GARCH-type conditional heteroskedasticity (Bollerslev, 1986), as the former is Markovian while the latter is not. Of course, unconditional heteroskedasticity violates our stationary condition and therefore falls outside the scope of our analysis. Dividing our sample in two, we continued to obtain negative circulation density estimates at all quantiles using either half of the sample. In the first half of the sample, the circulation density estimate appears more symmetric than it does in the full sample, achieving a minimum of -0.22 at the 0.45 quantile. In the latter half of the sample, the estimated circulation density is significantly negative and below -0.1 when we are at or below the 0.3 quantile, but above -0.1 and insignificantly different from zero at higher quantiles.

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 it also performs 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 rules used in our numerical simulations converge to zero too slowly to satisfy Assumption 4.1(e). Moreover, they are not explicitly designed to minimize the mean square error of our circulation density estimator. Relaxation of the condition $Th^4 \to 0$ in Assumption 4.1(e), so that the asymptotic bias given in Theorem

4.3 is potentially nonzero, may be required in order to deal rigorously with the problem of bandwidth selection. Extending Theorem 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.

A Mathematical appendix

A.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 A.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.

A.2 Proofs

The following preliminary result is used in our proofs of Theorem 3.1 and Lemma 3.1.

Lemma A.1. *Suppose Assumption 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 A.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 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 A.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 A.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 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

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

We may now write

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

and

$$(\mathscr{B}_T^*(x_1,y_1),\ldots,\mathscr{B}_T^*(x_s,y_s)) = \frac{\sqrt{T}}{T-1}\sum_{t=1}^{T-1}\left(g(X_t^*,X_{t+1}^*) - E^*g(X_t^*,X_{t+1}^*)\right). \tag{A.2}$$

The assumptions of Theorem 4.2 of Paparoditis and Politis (2002) are satisfied[1] under Assumption A.1. Applying this result in combination with (A.1) and (A.2) we obtain

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

[1]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.

a.s., where d_{KS} is the Kolmogorov-Smirnov metric on the space of probability distributions on $\mathbb{R}^s$. In view of Lemma A.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

$$\mathscr{B}_T^*(x, y) = \frac{\sqrt{T}}{T-1} \sum_{t=1}^{T-1} \left(f(Y_{t,T-1}^*) - E^* f(Y_{t,T-1}^*)\right), \tag{A.3}$$

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 (A.3), 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). $\square$

Proof of Theorem 3.2. We know from Lemma 3.1 that $\mathscr{L}_T^*(\mathscr{B}_T^*) \rightsquigarrow \mathscr{B}$ a.s. An application of the continuous mapping theorem yields

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

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

Proof of Theorem 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 (4.1) 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 4.2. In view of (4.1) 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 $\int P(X_{t+1} \leq x|X_t = x)\mathrm{d}F(x) = P(X_{t+1} \leq X_t)$. □

To prove Theorem 4.3, the following two preliminary results will be useful.

Lemma A.2. *Suppose Assumption 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 A.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,[2] 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

[2]Fermanian and Scaillet (2003, pp. 48–49) do not exploit the fact that k is even in the proof of their Theorem 7, thereby obtaining $E\hat{f}_T(x) = f(x) + O(h)$. This leads them to impose the condition $Th^3 \to 0$ in order to achieve asymptotically negligible bias, which is stronger than necessary.

we need only show that

$$(Th)^{1/2}\left(\sum_{i=1}^{2}\lambda_i(\partial_i\hat{H}_T(x,x)-E\partial_i\hat{H}_T(x,x))+\lambda_3(\hat{f}_T(x)-E\hat{f}_T(x))\right)\to_d N(0,\lambda^\top\Sigma\lambda). \tag{A.4}$$

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 (A.4) 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[3] 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) 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

$$h^{-1}EV_{1tT}V_{2tT}=\lambda_1\lambda_2 hE\left(K_h(x-X_t)k_h(x-X_{t+1})^2K_h(x-X_{t+2})\right)+O(h). \tag{A.5}$$

[3]Fermanian and Scaillet (2003) assume that $\alpha_T=O(T^{-2})$, but this is not quite enough to ensure that A3.1 of Robinson (1983), which requires $\sum_{j=T}^{\infty}\alpha_j=o(T^{-1})$, is satisfied. $\alpha_T=O(T^{-\eta})$ for some $\eta>2$ suffices. In fact, Lemmas A.2 and A.3 and Theorem 4.3 remain true if our Assumption 4.1(c) is replaced with A3.1 of Robinson (1983).

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

$$\begin{aligned} & E\left(K_h(x - X_t) K_h(x - X_{t+2}) | X_{t+1} = y\right) \qquad \text{(A.6)} \\ = \; & \left(\int K_h(x-w)\partial_2 C(F(\mathrm{d}w), F(y))\right)\left(\int K_h(x-z)\partial_1 C(F(y), F(\mathrm{d}z))\right). \end{aligned}$$

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 (A.6), 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

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

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 (A.5) 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

$$\begin{aligned} h^2Ek_h(x-X_{t+1})k_h(x-X_{t+s+1}) &= h^2\iint k_h(x-y)k_h(x-z)H^s(\mathrm{d}y,\mathrm{d}z) \\ &= \iint k'(v)k'(w)H^s(x-hv,x-hw)\mathrm{d}v\mathrm{d}w. \quad \text{(A.7)} \end{aligned}$$

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 (A.7) 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 (A.4) holds, with Σ having $(i,j)^{\text{th}}$ element σ_{ij}. This completes the proof. □

Lemma A.3. *Suppose Assumption 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 A.3. We begin by noting that Theorem 6 of Fermanian and Scaillet (2003) implies[4] that $\hat{x}_T=x+O_p(T^{-1/2})$. Next, using a third-order Taylor expansion,[5] we find that

$$\partial_1\hat{H}_T(\hat{x}_T,\hat{x}_T)-\partial_1\hat{H}_T(x,x)=\sum_{j=1}^{3}\frac{1}{j!}(\hat{x}_T-x)^j\frac{\mathrm{d}^j}{\mathrm{d}z^j}\partial_1\hat{H}_T(z,z)|_{z=x}+R_T, \tag{A.8}$$

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 (A.8) is $o_p(T^{-1/2}h^{-1/2})$, it now suffices for us to show that

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

[4]Strictly speaking, Theorem 6 of Fermanian and Scaillet (2003) requires that f is positive on the interior of its support. However, for our purposes, it suffices that f is positive at x.

[5]Fermanian and Scaillet (2003, p. 47) seek to establish a result comparable to our Lemma A.3 by employing a first-order Taylor expansion where we have employed a third-order expansion. They obtain an approximation error of order $O_p(T^{-1/2}h^{-5/2})$, which is claimed to be $o_p(1)$, but which in fact diverges under their assumptions. Here we avoid this difficulty by using a higher order Taylor expansion. Consequently, our bandwidth and kernel conditions differ from theirs.

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

$$\frac{1}{T}\sum_{t=1}^{T} k_h^{(i)}(x - X_t)K_h^{(j-i)}(x - X_{t+1}) = o_p(T^{(j-1)/2}h^{-1/2}) \tag{A.9}$$

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

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

It follows that

$$\frac{1}{T}\sum_{t=1}^{T} k_h^{(i)}(x - X_t)K_h^{(j-i)}(x - X_{t+1}) = T^{(j-1)/2}h^{-1/2}S_T + O(1), \tag{A.10}$$

where, suppressing the dependence of S_T and V_{tT} on i and j in our notation[6], 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 (A.10) and the fact that $T^{(j-1)/2}h^{-1/2} \to \infty$, we may verify (A.9) 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

$$EV_{tT}^2 \leq 2T^{-j/2}h^{2j+1}Ek_h^{(i)}(x - X_t)^2K_h^{(j-i)}(x - X_{t+1})^2 = O(T^{-j/2}h) = o(T^{j/2}h^{2j}), \tag{A.11}$$

[6]As defined here, S_T and V_{tT} differ from S_T and V_{itT} as defined in the proof of Lemma A.2, but play the same role in the application of Lemma 7.1 of Robinson (1983).

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) $\sup_{1\le t\le T}|V_{tT}| = O(1)$, (b) $E|V_{tT}V_{t+1,T}| = o(T^{j/2}h^{2j})$, and (c) $E|V_{tT}V_{t+s,T}| = O(T^j h^{4j})$ for $s \ge 2$. Boundedness of $k^{(i)}$ and $K^{(j-i)}$ may be used to show that $\sup_{1\le t\le T}|V_{tT}| = O(T^{-j/4}h^{-1/2}) = O(1)$, yielding (a). Parts (b) and (c) follow from (A.11) using the Cauchy-Schwarz inequality. We have now verified all assumptions of Lemma 7.1 of Robinson (1983), which allows us to conclude that $S_T = o_p(1)$. Thus, (A.9) holds for any $j = 1, \ldots, 3$ and $i = 0, \ldots, j$, and so the right-hand side of (A.8) is $o_p(T^{-1/2}h^{-1/2})$. This proves part (i) of Lemma A.3. Parts (ii) and (iii) may be proved using the same approach. □

Proof of Theorem 4.3. Lemma A.2 and Lemma A.3 jointly imply that

$$(Th)^{1/2} \cdot \begin{bmatrix} \partial_1 \hat{H}_T(\hat{x}_T, \hat{x}_T) - \partial_1 H(x,x) \\ \partial_2 \hat{H}_T(\hat{x}_T, \hat{x}_T) - \partial_2 H(x,x) \\ \hat{f}_T(\hat{x}_T) - f(x) \end{bmatrix} \to_d N(0, \Sigma), \tag{A.12}$$

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

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

That $\hat{\sigma}^2_T(u) \to_p \sigma^2(u)$ follows easily from (A.12). □

References

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.

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

BEARE, B. K. (2012). Archimedean copulas and temporal dependence. To appear in *Econometric Theory.*

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31** 307–327.

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.

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

CASON, T. N., FRIEDMAN, D. AND WAGENER, F. (2005). The dynamics of price dispersion, or Edgeworth variations. *Journal of Economic Dynamics and Control* **29** 801-822.

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., Mulinacci, S., Gobbi, F. and Romagnoli, S. (2011). *Dynamic Copula Methods in Finance*. Wiley Finance Series.

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.

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.

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.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticty with estimates of the variance of U.K. inflation. *Econometrica* **50** 987–1008.

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.

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. (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.

Hinich, M. J. and Rothman, P. (1998). Frequency-domain test of time reversibility. *Macroeconomic Dynamics* **2** 72–88.

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

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

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

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.

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.

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.

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.

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.

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.

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.

Tappata, M. (2009). Rockets and feathers: understanding asymmetric pricing. *RAND Journal of Economics* **40** 673–687.

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

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.