© 2012 Andreas Stefan Manfred Hagemann
THREE ESSAYS IN ECONOMETRICS

BY

ANDREAS STEFAN MANFRED HAGEMANN

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Graduate College of the University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Roger Koenker, Chair
Professor Xiaofeng Shao
Professor Daniel P. McMillen
Professor Anil K. Bera
ABSTRACT

This dissertation consists of three essays. In the first essay, entitled “Robust Spectral Analysis,” I introduce quantile spectral densities that summarize the cyclical behavior of time series across their whole distribution by analyzing periodicities in quantile crossings. This approach can capture systematic changes in the impact of cycles on the distribution of a time series and allows robust spectral estimation and inference in situations where the dependence structure is not accurately captured by the auto-covariance function. I study the statistical properties of quantile spectral estimators in a large class of nonlinear time series models and discuss inference both at fixed and across all frequencies. Monte Carlo experiments and an empirical example illustrate the advantages of quantile spectral analysis over classical methods when standard assumptions are violated.

In the second essay, “Stochastic Equicontinuity in Nonlinear Time Series Models,” I provide simple and easily verifiable conditions under which a strong form of stochastic equicontinuity holds in a wide variety of modern time series models. In contrast to most results currently available in the literature, my methods avoid mixing conditions. I discuss two applications in detail.

In the third essay, “A Simple Test for Regression Specification with Non-Nested Alternatives,” I introduce a simple test for the presence of the data-generating process among several non-nested alternatives. The test is an extension of the classical J test for non-nested regression models. I also provide a bootstrap version of the test that avoids possible size distortions inherited from the J test.
Für meine Eltern
I will always be indebted to my thesis advisor Roger Koenker for his guidance, insight, and patience. Working with him has been a constant source of inspiration and motivation for me. Xiaofeng Shao taught me a lot about econometrics and statistics, and I benefited enormously from his comments on the first chapter of my dissertation. Dan Bernhardt tirelessly read through several drafts of my manuscripts and provided advice and criticism that helped me improve the presentation of my work. I also received helpful comments from Anil Bera, Dan McMillen, Jiaying Gu, Euler Mello, Tom Parker, and seminar participants at the University of Illinois.

Before coming to the University of Illinois, I was fortunate enough to meet Gabriel Lee and Rolf Tschernig at the Universität Regensburg. Gabe convinced me to go to graduate school, which I still think is one of the best decisions I have ever made. I wish to thank him for all of his advice over the years. Rolf encouraged me to pursue econometrics and gave me important early feedback on my projects. I am grateful for his continuing interest in my work. I would also like to thank Harry Haupt at the Universität Bielefeld for his help in getting me into graduate school.

Finally, my most heartfelt thanks to Sarah Miller and my parents for all the love and support they have given me throughout these years. This dissertation is dedicated to my parents.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER 1</th>
<th>ROBUST SPECTRAL ANALYSIS</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Quantile Spectra and Two Estimators</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Asymptotic Properties of Quantile and Smoothed Quantile Periodograms</td>
<td>9</td>
</tr>
<tr>
<td>1.4</td>
<td>Testing for Flatness of a Quantile Spectrum</td>
<td>17</td>
</tr>
<tr>
<td>1.5</td>
<td>Numerical Results</td>
<td>22</td>
</tr>
<tr>
<td>1.6</td>
<td>Conclusion</td>
<td>39</td>
</tr>
<tr>
<td>1.7</td>
<td>Proofs</td>
<td>39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 2</th>
<th>STOCHASTIC EQUICONTINUITY IN NONLINEAR TIME SERIES MODELS</th>
<th>55</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>55</td>
</tr>
<tr>
<td>2.2</td>
<td>Stochastic Equicontinuity in Nonlinear Time Series Models</td>
<td>56</td>
</tr>
<tr>
<td>2.3</td>
<td>Proofs</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 3</th>
<th>A SIMPLE TEST FOR REGRESSION SPECIFICATION WITH NON-NESTED ALTERNATIVES</th>
<th>63</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>63</td>
</tr>
<tr>
<td>3.2</td>
<td>The MJ Test for Non-Nested Regression Models</td>
<td>65</td>
</tr>
<tr>
<td>3.3</td>
<td>Bootstrapping the MJ Test Statistic</td>
<td>70</td>
</tr>
<tr>
<td>3.4</td>
<td>Simulation Study</td>
<td>72</td>
</tr>
<tr>
<td>3.5</td>
<td>Conclusion</td>
<td>76</td>
</tr>
<tr>
<td>3.6</td>
<td>Auxiliary Results and Definitions</td>
<td>78</td>
</tr>
<tr>
<td>3.7</td>
<td>Proofs</td>
<td>79</td>
</tr>
</tbody>
</table>

REFERENCES 83
CHAPTER 1

ROBUST SPECTRAL ANALYSIS

I introduce quantile spectral densities that summarize the cyclical behavior of time series across their whole distribution by analyzing periodicities in quantile crossings. This approach can capture systematic changes in the impact of cycles on the distribution of a time series and allows robust spectral estimation and inference in situations where the dependence structure is not accurately captured by the auto-covariance function. I study the statistical properties of quantile spectral estimators in a large class of nonlinear time series models and discuss inference both at fixed and across all frequencies. Monte Carlo experiments and an empirical example illustrate the advantages of quantile spectral analysis over classical methods when standard assumptions are violated.

1.1 Introduction

Classical spectral analysis uses estimates of the spectrum or spectral density, a weighted sum of auto-covariances, to quantify the relative magnitude and frequency of cycles present in a time series. However, if the dependence structure is not accurately captured by the auto-covariance function, for example, because the time series under consideration is uncorrelated or heavy-tailed, then spectral analysis can provide only uninformative or even misleading results. In this chapter I discuss estimation and inference for a new class of spectral densities that summarize the cyclical behavior across the whole distribution of a time series by analyzing how frequently a process crosses its marginal quantiles. Functions from this class, which I refer to as quantile spectra or quantile spectral densities, are similar to classical spectral densities in both shape and interpretation, but can capture systematic changes
in the impact of cycles on the distribution of a time series. Such changes arise naturally in a variety of modern time series models, including stochastic volatility and random coefficient autoregressive models, and cannot be identified through classical spectral analysis, where cycles are assumed to be global phenomena with a constant effect on the whole distribution. Quantile spectral analysis fundamentally changes this view because it distinguishes between the effects of cycles at different points of the distribution of a process and permits a local focus on the parts of the distribution that are most affected by the cyclical structure.

Spectral analysis has traditionally played an important role in the analysis of economic time series; see, among many others, Granger (1966), Sargent (1987, chap. 11), Diebold, Ohanian, and Berkowitz (1998), and Qu and Tkachenko (2011), where the shape of the sample spectral density is typically taken to be one of the “stylized facts” that the predictions of a model must match. For macroeconomic data, these stylized facts often refer to high-frequency (seasonal) and low-frequency (business cycle) peaks in the spectrum. However, both observed data and the posterior distributions of economic models can exhibit heavy tails (Cogley and Sargent, 2002) that can induce peaks at random in the sample spectra of the data and the model output, invalidating comparisons between the two. For financial data, the stylized facts include the absence of auto-correlation, i.e., peakless spectra, and heavy-tailed marginal distributions (Cont, 2001). Stochastic volatility models such as GARCH processes (Bollerslev, 1986) can cross almost every quantile of their distribution in a periodic manner and at the same time satisfy these and other stylized facts, leading the researcher to incorrectly conclude from the spectrum that no periodicity is present. Bispectra and higher-order spectra can possibly detect cycles in quantile crossings, but rely on the presumption of light tails since they require the existence of at least third moments to be well defined and sixth moments to be estimated reliably (see, e.g., Rosenblatt and Van Ness, 1965). Financial time series such as log-returns of foreign exchange rates or stock prices may lack finite fourth or even third moments (Loretan and Phillips, 1994; Longin, 1996).

My proposed approach is robust to each of these concerns: Quantile spectral methods consistently recover the spectral shape and detect periodicities even in uncorrelated or heavy-tailed data. Inference about quantile spectra both at fixed frequencies and across frequencies does not require assumptions
about the moments of the process. Although moment conditions can be used to verify some of the assumptions below, arbitrarily low fractional moments suffice. Because several common time series models can induce situations where cycles are present at some but not at all quantiles, I also provide a general Cramér-von Mises specification test for peakless quantile spectra. Under conditions that are routinely imposed in the literature when testing for the absence of peaks, these tests are distribution-free and, depending on the strength of the assumptions, sometimes even exact in finite samples. The test remains valid asymptotically under much weaker conditions when a bootstrap approximation is used.

Several recent papers apply quantiles in spectral or correlogram (autocorrelation) analysis. Li (2008, 2011) obtains robust spectral estimators via quantile regressions for harmonic regression models. Although his estimation method is quite different from that developed here, there is some overlap in our results. I provide a detailed discussion in section 1.3. Katkovnik (1998) relies on the same idea as Li (2008), but only works with sinusoidal models and iid noise. Linton and Whang (2007) introduce the “quantilogram,” a correlogram that is essentially the inverse Fourier transform of a quantile spectrum, but their focus is on testing for directional predictive ability of financial data in the time domain, rather than spectral analysis. Chung and Hong (2007) test for directional predictive ability with the generalized spectrum (Hong, 1999) by investigating the frequency domain behavior of processes around a given threshold. This approach is similar in spirit to quantile spectral analysis but, as Linton and Whang point out, Chung and Hong rescale their data with sample standard deviations but do not account for the randomness introduced by the rescaling in the derivation of their tests. In contrast, the scaling of the data for quantile spectral analysis is provided automatically through the marginal quantile function and all of my results are derived under the assumption that these quantiles are estimated.

Other robust spectral methods are discussed by Kleiner and Martin (1979) and Klüppelburg and Mikosch (1994): Kleiner and Martin focus on time series where the dependence structure is accurately captured by an autoregressive model of sufficiently high order. Quantile spectral analysis differs from these methods in that it is completely nonparametric and, most importantly, it robustly estimates cyclical components even when an autoregression is not an appropriate model for the data. Klüppelburg and Mikosch robustify clas-
sical spectral estimates by a self-normalization procedure to estimate normalized spectra under arbitrarily weak fractional moments conditions. However, their results are of limited use for applications because little is known about the asymptotic distribution of their procedures. In contrast, I show that quantile spectral estimates have relatively simple asymptotic distributions even when no moments exist.

After completing the first draft of this manuscript (Hagemann, 2011), the papers by Dette, Hallin, Kley, and Volgushev (2011) and Lee and Subba Rao (2011) became available. Both describe methods based on analyzing cross-covariances of quantile hits and copulas in the frequency domain that are similar to the quantile spectral estimators presented here. However, both Dette et al. and Lee and Subba Rao develop their methods as alternatives to the generalized spectrum to discover the presence of any type of dependence structure in time series data. My estimators are constructed to identify cyclical dependence.

The remainder of this chapter is organized as follows: Section 1.2 discusses quantile spectral analysis and introduces two classes of estimators. Section 1.3 establishes the asymptotic validity of the estimators under weak regularity conditions. In section 1.4, I show the consistency of Cramér-von Mises tests for peakless spectra. The Monte Carlo experiments and an empirical example in section 1.5 illustrate the finite sample properties of the estimators and tests. Section 1.6 concludes. Section 1.7 contains auxiliary results and proofs.

I use the following notation throughout this chapter: \(1\{\cdot\}\) and \(1_{\{\cdot\}}\) both represent the indicator function and \(\|X\|_p\) denotes \((E|X|^p)^{1/p}\). Limits are as \(n \to \infty\) unless otherwise noted and convergence in distribution is indicated by \(\sim\). The inner product \(\langle \cdot, \cdot \rangle_\Pi\) and norm \(\| \cdot \|_\Pi\) are defined at the beginning of section 1.4.

### 1.2 Quantile Spectra and Two Estimators

This section introduces quantile spectral density estimation as a robust complement to classical spectral methods. Spectral analysis aims to reveal periodic behavior in a stationary time series \(X_t\) with auto-covariance function \(\gamma_X(j) := E X_0 X_j - (E X_0)^2\) at lag \(j\) by estimating the spectrum or spectral
density at frequency $\lambda$, defined as

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{j \in \mathbb{Z}} \gamma_X(j) \cos(j\lambda), \quad \lambda \in (-\pi, \pi]. \quad (1.1)$$

The auto-covariance function is typically taken to be absolutely summable to ensure that $f_X$ is continuous and symmetric about 0; a stochastic process that does not possess at least finite second moments cannot be meaningfully analyzed by the spectrum. If $f_X$ has a peak at $\lambda$, then $X_t$ is expected to repeat itself on average after $2\pi/\lambda$ units of time; for example, a monthly time series with a peak in the spectrum at $2\pi/3$ has a three-month cycle, with a higher value of $f_X$ corresponding to a more pronounced cycle. The primary goal of this chapter is to develop spectral methods that go a step beyond summarizing the average impact of cycles by distinguishing between the effects of cycles at different points of the distribution of $X_t$.

The central idea is that if a stationary process $(X_t)_{t \in \mathbb{Z}}$ contains cycles, then its realizations will tend to stay above or below a given threshold in an approximately periodic manner. The pattern in which the process crosses a threshold at the center of its distribution reflects the most prominent cycles, but provides little information about their relative sizes. Patterns in threshold crossings near the extremes of the distribution help to identify amplitudes of these cycles and they also recover periodicities that are obscured at the center of the distribution. The quantiles of $X_t$, arising from the quantile function $\xi_0(\tau) := \inf \{x : P(X_0 \leq x) \geq \tau\}$, are natural choices for such thresholds because they give precise meaning to the notion of the center and extremes of a distribution. Spectral analysis of quantile crossing patterns can then discover cycles in the process and reveal the extent to which they are present at a given quantile without relying on moments.

To formalize this idea, pick probabilities $\tau \in (0, 1)$ corresponding to the marginal quantiles $\xi_0(\tau)$ of $X_t$. The variable of interest for the analysis is

$$V_t(\tau, \xi) = \tau - 1\{X_t < \xi\}, \quad (\tau, \xi) \in (0, 1) \times \mathbb{R},$$

such that $V_t(\tau) := V_t(\tau, \xi_0(\tau))$ takes on the value $\tau - 1$ if $X_t$ is below its $\tau$-th quantile at $t$, and $\tau$ otherwise. Here the quantiles are not assumed to be known, which enables the researcher to choose the values of $\tau$ according to the amount and nature of information that is needed about the cyclical structure
of the time series. For example, $\tau = 0.5$ only analyzes fluctuations about the median, whereas varying $\tau$ between 0.5 and 0.9 also provides information about the positive amplitudes by including values in the upper tail of the process.

If the distribution function of $X_t$ is continuous and increasing at $\xi_0(\tau)$, then the $\tau$-th quantile crossing indicator $V_t(\tau)$ is a bounded, stationary, mean-zero random variable with auto-covariance function $r_\tau(j) := \gamma_{V(\tau)}(j) = \mathbb{E}V_0(\tau)V_j(\tau)$. Periodicities in $V_t(\tau)$ are summarized by peaks in its spectral density

$$g_\tau(\lambda) := f_{V(\tau)}(\lambda) = \frac{1}{2\pi} \sum_{j \in \mathbb{Z}} r_\tau(j) \cos(j\lambda),$$

which I refer to as the $\tau$-th quantile spectrum or $\tau$-th quantile spectral density in the sequel. Analyzing $g_\tau$ across a grid of probabilities $\tau \in (0, 1)$ therefore reveals cycles in events of the form $\{X_t < \xi_0(\tau) \mid t \in \mathbb{Z}\}$, which in turn summarize $(X_t)_{t \in \mathbb{Z}}$ with arbitrary precision as long as the grid is fine enough.

As the next two Examples show, quantile spectral analysis can in fact yield additional insights beyond classical spectral analysis; Linton and Whang (2007) consider similar models. I discuss estimation of quantile spectra below.

**Example 1.1 (Stochastic volatility).** Let $(\varepsilon_t)_{t \in \mathbb{Z}}$ be a sequence of iid mean-zero random variables and suppose the data are generated by the stochastic volatility model

$$X_t = \xi_0(\tau_0) + \varepsilon_t v(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$$

for some $\tau_0 \in (0, 1)$, where $v > 0$ is a measurable function that drives the volatility of the process. If $X_t$ has finite second moments, then it is an uncorrelated time series and its spectrum contains no information about the dependence structure beyond that it is “flat,” i.e., $f_X(\lambda) = \gamma_X(0)/(2\pi)$ at all frequencies.

However, any stationary time series with a continuous and increasing distribution function at $\xi_0(\tau)$ satisfies $r_\tau(0) = \tau(1-\tau)$ and the stochastic volatility process also has the property that

$$r_{\tau_0}(j) = \mathbb{E}V_0(\tau_0)(\tau_0 - \mathbb{P}(X_j < \xi_0(\tau_0) \mid \varepsilon_{j-1}, \ldots))$$

$$= (\tau_0 - \mathbb{P}(\varepsilon_j < 0)) \mathbb{E}V_0(\tau_0) = 0 \quad \text{for all} \ j > 0.$$ 

Therefore its $\tau_0$-th quantile spectrum will also flat in the sense that $g_{\tau_0}(\lambda) \equiv \tau_0(1-\tau_0)/(2\pi)$, but the other quantile spectra of the stochastic process will
be informative because \( r_\tau(j) \) generally does not vanish for \( \tau \neq \tau_0 \).

□

**Example 1.2 (QAR).** Let \((\varepsilon_t)_{t \in \mathbb{Z}}\) be a sequence of independent Uniform(0, 1) variables and consider the second-order quantile autoregressive (QAR(2)) process of Koenker and Xiao (2006),

\[
X_t = \beta_0(\varepsilon_t) + \beta_1(\varepsilon_t)X_{t-1} + \beta_2(\varepsilon_t)X_{t-2}
= \mathbb{E}(\beta_1(\varepsilon_0))X_{t-1} + \mathbb{E}(\beta_2(\varepsilon_0))X_{t-2} + Y_t,
\]

where \( Y_t = \beta_0(\varepsilon_t) + [\beta_1(\varepsilon_t) - \mathbb{E}(\beta_1(\varepsilon_0))]X_{t-1} + [\beta_2(\varepsilon_t) - \mathbb{E}(\beta_2(\varepsilon_0))]X_{t-2} \). Here \( \beta_0, \beta_1, \) and \( \beta_2 \) are unknown functions that satisfy regularity conditions that ensure stationarity and \( X_t \) is assumed to be increasing in \( \varepsilon_t \) conditional on \( X_{t-1}, X_{t-2} \). Provided that its second moments exist, the sequence \((Y_t)_{t \in \mathbb{Z}}\) has no influence on the shape of the spectrum because it is an uncorrelated sequence that is also uncorrelated with the other variables on the right-hand side of the preceding display (Knight, 2006). Hence, if \( \mathbb{E} \beta_1(\varepsilon_0) = \mathbb{E} \beta_2(\varepsilon_0) = 0 \), the spectrum of \( X_t \) satisfies \( f_X(\lambda) \equiv \gamma_Y(0)/(2\pi) \) and classical spectral analysis cannot reveal anything about cycles in \( X_t \). If \( \mathbb{E} \beta_1(\varepsilon_0) \) and \( \mathbb{E} \beta_2(\varepsilon_0) \) are nonzero, then the spectrum of \( X_t \) is the same as that of an AR(2) process with the same mean and covariances as the QAR(2). If, instead, there is some \( \tau_0 \in (0, 1) \) such that \( \beta_1(\tau_0) = \beta_2(\tau_0) = 0 \), then the \( \tau_0 \)-th quantile spectrum is also flat (see Example 1.14 below), but cycles can be recovered at other quantiles. Further, the quantile spectra of the QAR(2) process and those of an AR(2) process with the same mean and covariance structure will generally be different. □

For a given sample \( S_n := \{X_t : t = 1, \ldots, n\} \), I consider two estimators of the quantile spectrum that correspond to the periodograms and smoothed periodograms used in classical spectral analysis. The key difference from the classical case is that the variable of interest \( V_t(\tau) \) is indexed by the unknown quantity \( \xi_0(\tau) \) and therefore itself has to be estimated. To this end, let \( \hat{\xi}_n(\tau) \) be the \( \tau \)-th sample quantile determined implicitly by solutions to the minimization problem

\[
\min_{x \in \mathbb{R}} \sum_{t=1}^n \rho_\tau(X_t - x),
\]

where \( \rho_\tau(x) := x(\tau - 1\{x < 0\}) \) is the Koenker and Bassett (1978) check
Let $\hat{V}_t(\tau) := V_t(\tau, \hat{\xi}_n(\tau))$ be the estimate of $V_t(\tau)$. The $\tau$-th quantile periodogram is then the "plug-in" estimator

$$Q_{n,\tau}(\lambda) := \frac{1}{2\pi} \left| \sum_{t=1}^{n} \hat{V}_t(\tau) e^{-it\lambda} \right|^2 = \frac{1}{2\pi} \sum_{|j|<n} \hat{r}_{n,\tau}(j) \cos(j\lambda), \quad (1.3)$$

where $i := \sqrt{-1}$ and $\hat{r}_{n,\tau}(j) := n^{-1} \sum_{t=|j|+1}^{n} \hat{V}_t(\tau) \hat{V}_{t-|j|}(\tau)$ for $|j| < n$. As I will show in the next section, the quantile periodogram inherits the properties of the classical periodogram in the sense that it allows the construction of valid confidence intervals, but does not provide consistent estimates for the spectrum of interest.

Consistent estimation of the quantile spectrum requires additional smoothing to assign less weight to the imprecisely estimated auto-covariances with lags $|j|$ near $n$. For this I apply the Parzen (1957) class of kernel spectral density estimators to the present framework. The estimators, which I refer to as smoothed $\tau$-th quantile periodograms, are given by

$$\hat{g}_{n,\tau}(\lambda) = \frac{1}{2\pi} \sum_{|j|<n} w(j/B_n) \hat{r}_{n,\tau}(j) \cos(j\lambda), \quad (1.4)$$

where $B_n$ is a scalar "bandwidth" parameter that grows with $n$ at a rate specified in Theorem 1.11 below, and $w$ is a real-valued smoothing weight function from the set

$$\mathcal{W} := \left\{ w \text{ is bounded and continuous, } w(x) = w(-x) \quad \forall x \in \mathbb{R}, \right.$$  

$$w(0) = 1, \quad \bar{w}(x) := \sup_{y \geq x} |w(y)| \text{ satisfies } \int_{0}^{\infty} \bar{w}(x) \, dx < \infty, \quad W(\lambda) := \frac{1}{2\pi} \int_{-\infty}^{\infty} w(x) e^{-ix\lambda} \, dx \text{ satisfies } \int_{-\infty}^{\infty} |W(\lambda)| \, d\lambda < \infty \right\}. \quad (1.5)$$

In the literature, $w$ and $W$ are usually called the lag window and spectral window, respectively. Both functions are also often referred to as kernels, although $w$ does not necessarily integrate to one.

**Remarks.** 1. The class $\mathcal{W}$ includes most of the kernels that are used in practice, for example the Bartlett (i.e., triangular), Parzen, Tukey-Hanning, Daniell, and quadratic-spectral windows. However, it excludes the truncated (also known as rectangular or Dirichlet) window. See Andrews (1991) and Brockwell and Davis (1991, pp. 359-362) for thorough descriptions of these
windows and their properties. I provide a brief discussion on how to choose $w$ and $B_n$ at the end of the next section.

2. The restriction $\int_0^\infty \tilde{w}(x) \, dx < \infty$ is not standard in the spectral density estimation literature. As pointed out by Jansson (2002), it is needed for asymptotic bounds on expressions such as $B_n^{-1} \sum_{|j|<n} |w(j/B_n)|$ that typically arise in consistency proofs of spectral density estimates indexed by estimated parameters; see also Robinson (1991).

3. Spectra are non-negative. It is therefore common practice to choose a window such that $W \geq 0$ to ensure non-negativity of the spectral density estimate; see, e.g., Andrews (1991) and Smith (2005). The condition $\int_{-\infty}^{\infty} |W(\lambda)| \, d\lambda < \infty$ is immediately satisfied for such windows in view of the inverse Fourier transform $w(x) = \int_{-\infty}^{\infty} e^{ix\lambda} W(\lambda) \, d\lambda$, which implies $\int_{-\infty}^{\infty} W(\lambda) \, d\lambda = w(0) = 1$ for $w \in W$. The Tukey-Hanning window is an example of a window that satisfies $\int_{-\infty}^{\infty} |W(\lambda)| \, d\lambda < \infty$, but not $W \geq 0$.

The next section characterizes the asymptotic properties of the quantile and smoothed quantile periodograms.

1.3 Asymptotic Properties of Quantile and Smoothed Quantile Periodograms

In this section I construct confidence intervals for the quantile spectrum and establish the consistency of the smoothed quantile periodogram under regularity conditions. I also compare the quantile periodogram to the periodograms of Li (2008, 2011).

Throughout the remainder of this chapter I assume that $(X_t)_{t \in \mathbb{Z}}$ is a non-linear process of the form

$$X_t = Y(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots),$$

(1.5)

where $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a sequence of iid copies of a random variable $\varepsilon$ and $Y$ is a measurable, possibly unknown function that transforms the input $\mathcal{F}_t := (\varepsilon_t, \varepsilon_{t-1}, \ldots)$ into the output $X_t$. The class (1.5) includes a large number of commonly-used stationary time series models. For instance, the processes in Examples 1.1 and 1.2 are of this form; I provide other examples below Proposition 1.4 in this section.
The essential conditions for the estimation of spectra are restrictions on the memory of the time series. As pointed out by Wu (2005), for time series of the form (1.5) such restrictions are most easily implemented by comparing $X_t$ to a slightly perturbed version of itself. Let $(\varepsilon_t^\ast)_t \in \mathbb{Z}$ be an iid copy of $(\varepsilon_t)_t \in \mathbb{Z}$, so that the difference between $X_t$ and $X'_t := Y(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0^\ast, \varepsilon_{-1}^\ast, \ldots)$ are the inputs before time $t = 1$. Define $X_\tau(\delta) := \{\xi \in \mathbb{R} : |\xi_0(\tau) - \xi| \leq \delta\}$ and assume the following:

**Assumption 1.3.** For a given $\tau \in (0, 1)$, there exist $\delta > 0$ and $\sigma \in (0, 1)$ such that

$$\sup_{\xi \in X_{\tau}(\delta)} \| \mathbb{1}\{X_n < \xi\} - \mathbb{1}\{X'_n < \xi\} \| = O(\sigma^n).$$

Intuitively, this condition requires the probability that $X_n$ is below but $X'_n$ is above a given threshold (or vice versa) to be sufficiently small for large $n$ as long as the threshold is near $\xi_0(\tau)$. It is the only dependence condition needed to construct asymptotically valid confidence intervals for the quantile spectrum. Assumption 1.3 avoids restrictions on the summability of the cumulants (Brillinger, 1975, pp. 19-21) of $X_t$ that are routinely imposed in the spectral estimation literature; see Andrews (1991) and the references therein. Cumulant conditions or “mixing” assumptions (Rosenblatt, 1984) that imply such conditions are sometimes difficult to establish for a given time series model and can easily fail or put unwanted restrictions on the parameter space when $X_t$ is, for example, generated by a standard GARCH process (Bollerslev, 1986).

Assumption 1.3 does not require the existence of any moments of $X_t$, but can be verified for most commonly used stationary time series models at the expense of an arbitrarily weak moment restriction via the geometric moment contracting (GMC) property introduced by Hsing and Wu (2004). A time series of the form (1.5) is said to be GMC for some $\alpha > 0$ if $\|X_n - X'_n\|_\alpha = O(\varrho^n)$ for some $\varrho \in (0, 1)$, where $\varrho$ may depend on $\alpha$.

**Proposition 1.4.** Assumption 1.3 is satisfied if $F_X(x) := \mathbb{P}(X_0 \leq x)$ is Lipschitz continuous in a neighborhood of $\xi_0(\tau)$ and $\|X_n - X'_n\|_\alpha = O(\varrho^n)$ for some $\alpha > 0$ and $\varrho \in (0, 1)$.

The GMC property is satisfied for stationary (causal) ARMA, ARCH (Engle, 1982), GARCH, ARMA-ARCH, ARMA-GARCH, asymmetric GARCH (Ding, Granger, and Engle, 1993; Ling and McAleer, 2002), generalized ran-
dom coefficient autoregressive (Bougerol and Picard, 1992), and QAR models; see Shao and Wu (2007) and Shao (2011b) for proofs and more examples. By Proposition 1.4, all of these models are included in the analysis if $F_X$ is Lipschitz near $\xi_0(\tau)$—a condition that is also needed for all of my results.

In addition to Lipschitz continuity, a restriction on $F_X$ is required to ensure both that $V_t(\tau)$ can be estimated consistently and that $\sqrt{n}(\hat{\xi}_n(\tau) - \xi_0(\tau))$ is bounded in probability:

**Assumption 1.5.** $F_X$ is Lipschitz continuous in a neighborhood of $\xi_0(\tau)$ and has a positive and continuous (Lebesgue) density at $\xi_0(\tau)$.

This assumption, or slight variations thereof, is standard in the quantile estimation and regression literature; see, e.g., Koenker (2005, p. 120) and Wu (2007).

As a preliminary step towards inference about quantile spectra, the following result establishes the joint asymptotic distribution of the quantile periodogram on a subset of the natural frequencies $\ldots, -4\pi/n, -2\pi/n, 0, 2\pi/n, 4\pi/n, \ldots \subset (-\pi, \pi]$. More precisely, Theorem 1.6 shows that the usual convergence of the periodogram at different frequencies to independent exponential variables is not affected by the presence of the estimated quantile $\hat{\xi}_n(\tau)$.

**Theorem 1.6.** Suppose Assumptions 1.3 and 1.5 hold for some $\tau \in (0, 1)$. Let $\lambda_n = 2\pi j_n/n$ with $j_n \in \mathbb{Z}$ be a sequence of natural frequencies such that $\lambda_n \to \lambda \in (0, \pi)$ with $g_\tau(\lambda) > 0$. Then, for any fixed $k \in \mathbb{Z}$, the collection of quantile periodograms

$$Q_{n,\tau}(\lambda_n - 2\pi k/n), Q_{n,\tau}(\lambda_n - 2\pi (k-1)/n), \ldots, Q_{n,\tau}(\lambda_n + 2\pi k/n)$$

converges jointly in distribution to independent exponential variables with mean $g_\tau(\lambda)$.

**Remarks.** 1. The natural frequencies induce invariance to centering in the quantity inside the modulus in (1.3) so we can write

$$n^{-1/2} \sum_{t=1}^n \hat{V}_t(\tau)e^{-it\lambda_n} = -n^{-1/2} \sum_{t=1}^n \left(1\{X_t < \hat{\xi}_n(\tau)\} - F_X(\hat{\xi}_n(\tau))\right)e^{-it\lambda_n}.$$  

Given the invariance, the strategy for the proof is to show that the empirical process on the right-hand side of the preceding display is stochastically
equicontinuous with respect to an appropriate semi-metric on bounded sets near $\xi_0(\tau)$. For this I extend Andrews and Pollard’s (1994) functional limit theorems to time series of the form (1.5) that satisfy Assumption 1.3. The equicontinuity property and a result of Shao and Wu (2007) on classical periodograms at natural frequencies then yield the desired results.

2. If a quantile of interest $\xi_0(\tau)$ is assumed to be known, for example $\xi_0(0.5) = 0$ as in Li (2008), then Theorem 1.6 remains valid when (i) $\xi_0(\tau)$ is used in $Q_{n,\tau}$ instead of $\hat{\xi}_n(\tau)$, (ii) Assumption 1.5 is replaced by the condition that $F_X$ is continuous and increasing at $\xi_0(\tau)$, and (iii) Assumption 1.3 is replaced by Assumption 1.9 below with $\delta = 0$. This is a direct consequence of Shao and Wu’s (2007) Corollary 2.1.

Theorem 1.6 yields a convenient way to construct point-wise confidence intervals for the quantile spectrum. The proof follows immediately from the properties of independent exponential variables. Example 1.8 provides an application.

**Corollary 1.7.** Suppose the conditions of Theorem 1.6 are satisfied. Define $\tilde{Q}_{n,\tau}(\lambda, k) = \sum_{|j| \leq k} Q_{n,\tau}(\lambda_n + 2\pi j/n)/(2k + 1)$, and let $\chi^2_{4k+2,\alpha}$ be the $\alpha$ quantile of a $\chi^2$ distribution with $4k + 2$ degrees of freedom. Then, for every fixed $k \in \mathbb{Z}$, the probability of the event

$$g_{\tau}(\lambda) \in \left( \frac{(4k + 2)\tilde{Q}_{n,\tau}(\lambda, k)}{\chi^2_{4k+2,1-\alpha/2}}, \frac{(4k + 2)\tilde{Q}_{n,\tau}(\lambda, k)}{\chi^2_{4k+2,\alpha/2}} \right)$$

converges to $1 - \alpha$.

**Example 1.8 (Testing for periodicities).** The processes in Examples 1.1 and 1.2 are instances where $V_t(\tau_0)$ is a white noise series for some $\tau_0 \in (0, 1)$. Then the $\tau_0$-th quantile spectrum of $X_t$ is $\tau_0(1 - \tau_0)/(2\pi)$ at all frequencies and therefore contains no periodicities at that quantile. Because a spike in the periodogram could either be evidence for a periodicity or an artifact generated by the sample, this leads to the problem of testing whether the $\tau_0$-th quantile spectrum behaves like a flat quantile spectrum at a given frequency. By Corollary 1.7, this hypothesis can be rejected at level $\alpha$ if the confidence interval in the Corollary does not contain $\tau_0(1 - \tau_0)/(2\pi)$. The same type of test is not as simple in classical spectral analysis because (1.1) reduces to the unknown quantity $\gamma_X(0)/(2\pi)$ if $X_t$ is white noise. I extend the idea of testing for flatness in section 1.4 to provide a test for
the more general hypothesis that \( g_\tau(\lambda) = \tau_0(1 - \tau_0)/(2\pi) \) jointly across all frequencies.

The results stated in Theorem 1.6 and its Corollary overlap to some extent with Theorem 2 of Li (2008). He uses the least absolute deviations estimator in the harmonic regression model

\[
\hat{\beta}_n(\lambda) = \arg\min_{(b_1, b_2) \in \mathbb{R}^2} \frac{1}{n} \sum_{t=1}^{n} \rho_{0.5}(X_t - \cos(t\lambda)b_1 - \sin(t\lambda)b_2),
\]

to define the Laplace periodogram \( L_n(\lambda) = n|\hat{\beta}_n(\lambda)|^2/4 \). In the special case that the time series of interest satisfies \( X_t = \cos(t\lambda_0)b_1 + \sin(t\lambda_0)b_2 + \varepsilon_t \), where \( \lambda_0, \beta_1, \) and \( \beta_2 \) are unknown constants, this approach has the advantage that the maximizer of \( L_n(\lambda) \) can be used as a robust estimator of \( \lambda_0 \), although Li provides only Monte Carlo evidence of this assertion. For general time series, he assumes that \( X_t \) has median zero and a density \( F_X \) with \( F_X'(0) > 0 \), and that certain short-range dependence conditions are satisfied. The proofs of his Theorems 1 and 2 then yield an asymptotically linear representation for \( \hat{\beta}_n(\lambda_n) \) that can be used to show

\[
L_n(\lambda_n) = F_X'(0)^{-2} \left| n^{-1/2} \sum_{t=1}^{n} \hat{V}_t(0.5)e^{-it\lambda_n} \right|^2 + o_p(1).
\]

The first term on the right-hand side is \( 2\pi/F_X'(0)^2 \) times the quantile periodogram evaluated at the median. Hence, if the median of \( X_t \) is indeed zero, the Laplace periodogram and the quantile periodogram at the median are asymptotically equivalent up to the unknown constant \( 2\pi/F_X'(0)^2 \). Li (2011) extends his idea of harmonic median regression to quantile regression.

Using Li’s (2008, 2011) periodograms instead of the quantile spectral methods introduced in this chapter has the following disadvantages: (i) All of Li’s asymptotic results depend on terms of the form \( \tau(1 - \tau)/F_X'(\xi(\tau))^2 \) that in his case must be estimated to make inference about the dimensionless quantity \( g_\tau(\lambda) \) even for simple tests such as in Example 1.8; my approach avoids this complication altogether. (ii) Li’s methods require quantile regression at every frequency, whereas the quantile periodogram (1.3) can be computed easily with the Fast Fourier Transform. (iii) Li does not provide consistent estimators. For example, \( L_n(\lambda) \) converges to a distribution with asymptotic mean \( [2\pi/(4F_X'(0)^2)] \times g_{0.5}(\lambda) \), but Li does not establish that a smoothed
version of $L_n(\lambda)$ converges in probability to this quantity. In contrast—as I will show now—the quantile periodogram can be smoothed by standard methods to provide uniformly consistent estimates of $g_r$.

Consistent estimation of the quantile spectrum requires weaker conditions than the construction of confidence intervals because much of the randomness introduced by replacing $r_r$ (as defined above (1.2)) with $\hat{r}_{n,r}$ is now controlled by the smoothing weight function $w$. Let $\varepsilon_0^*$ be an iid copy of $\varepsilon_0$ such that $X_t$ and $X_t^* := Y(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0^*, \varepsilon_{-1}, \ldots)$ differ only through the input at time $t = 0$. I assume $X_t$ satisfies the following:

**Assumption 1.9.** For a given $\tau \in (0, 1)$ and $\mathcal{X}_r(\delta)$ as in Assumption 1.3, there exists a $\delta > 0$ such that

$$\sum_{t=0}^{\infty} \sup_{\xi \in \mathcal{X}_r(\delta)} \|1\{X_t < \xi\} - 1\{X_t^* < \xi\}\| < \infty.$$ 

**Remarks.**

1. Assumption 1.3 implies Assumption 1.9 in view of the relation $\|1\{X_n < \xi\} - 1\{X_n^* < \xi\}\| = \|1\{X_{n+1} < \xi\} - 1\{X_{n+1}^* < \xi\}\|$; see the discussion below equation [13] of Wu (2005). For $\xi$ near $\xi_0(\tau)$, adding and subtracting $1\{X_n^* < \xi\}$ and the triangle inequality then yield $\|1\{X_n < \xi\} - 1\{X_n^* < \xi\}\| = O(\sigma^n)$, which remains valid after taking suprema over $\mathcal{X}_r(\delta)$.

2. A stationary stochastic process is usually called short-range dependent if its auto-covariance function is summable. Since $X_t$ can have heavy tails, this definition no longer has the desired meaning. However, Remark 2.1 of Shao (2011a) can be used to show that $V_t(\tau)$ is short-range dependent because

$$\sum_{j \in \mathbb{Z}} |r_r(j)| \leq \left( \sum_{t=0}^{\infty} \|1\{X_t < \xi_0(\tau)\} - 1\{X_t^* < \xi_0(\tau)\}\| \right)^2 < \infty,$$

provided Assumptions 1.3 or 1.9 hold. This suggests that these assumptions should still be regarded as short-range dependence conditions on $X_t$. Heyde (2002) argues similarly to quantify the dependence of the increments of certain Gaussian processes.

Assumption 1.9 is easily verified in most cases via Proposition 1.4. However, more direct arguments can also be useful:

**Example 1.10 (Linear processes with Cauchy innovations).** Consider the linear process $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, where $(a_j)_{j \in \mathbb{N}}$ is a sequence of constants and
$$(\varepsilon_t)_{t \in \mathbb{Z}}$$ is an sequence of iid copies of a standard Cauchy random variable. Without loss of generality, let $a_0 = 1$. Write $F_\varepsilon$ for the distribution function of $\varepsilon$; then $X_t$ has distribution function $F_X(x) = \mathbb{E} F_\varepsilon(x - \sum_{j=1}^{\infty} a_j \varepsilon_{t-j})$ and therefore also possesses a bounded density $F_X'$ by the Lebesgue Dominated Convergence Theorem. Furthermore, apply the point-wise inequality

$$|1\{X_n < \xi\} - 1\{X_n^* < \xi\}| \leq 1\{|X_n - \xi| < |X_n - X_n^*|\},$$

then the Mean Value Theorem and $\mathbb{P}(|\varepsilon_0| + |\varepsilon_0^*| \geq x) \leq \mathbb{P}(|\varepsilon_0| \geq x/2) + \mathbb{P}(|\varepsilon_0^*| \geq x/2)$ for any fixed $x$ to see that

$$|1\{X_n < \xi\} - 1\{X_n^* < \xi\}|^2 \leq \mathbb{P}(|X_n - \xi| < |a_n||\varepsilon_0 - \varepsilon_0^*|) \leq \mathbb{P}(|X_n| \leq |a_n|^{1/2}) + \mathbb{P}(|a_n||\varepsilon_0 - \varepsilon_0^*| \geq |a_n|^{1/2}) \leq 2|a_n|^{1/2} \sup_{x \in \mathbb{R}} F_X'(x) + 2\mathbb{P}(|\varepsilon_0| \geq 4|a_n|^{-1/2}),$$

which is $O(|a_n|^{1/2})$ because the tail probability $\mathbb{P}(|\varepsilon_0| > x)$ of a Cauchy random variable is proportional to $x^{-1}$ as $x \to \infty$. Because these bounds hold uniformly in $\xi$, take square roots in the preceding display to conclude that Assumption 1.9 is satisfied if $\sum_{j=0}^{\infty} |a_j|^{1/4} < \infty$. The same type of reasoning can be used more generally when the innovations come from a smooth distribution whose tails behave algebraically. Proposition 1.4 does not apply here because $(a_n)_{n \in \mathbb{N}}$ does not necessarily vanish at a geometric rate.

Theorem 1.11 below establishes uniform consistency of the smoothed quantile periodogram under the condition that the bandwidth $B_n$ grows at a sufficiently slow rate. In particular, due to the uniformity Theorem 1.11 allows for both fixed frequencies and sequences of frequencies such as the natural frequencies above.

**Theorem 1.11.** If Assumptions 1.5 and 1.9 hold for some $\tau \in (0,1)$, $w \in \mathcal{W}$, $B_n \to \infty$, and $B_n = o(\sqrt{n})$, then

$$\hat{g}_{n,\tau}(\lambda) \overset{p}{\to} g_\tau(\lambda)$$

uniformly in $\lambda \in (-\pi, \pi]$.

**Remarks.** 1. The proof of Theorem 1.11 relies in part on recent results for classical spectral density estimates obtained by Liu and Wu (2010).
2. At fixed frequencies, kernel spectral density estimates of differentiable functions are often valid for bandwidths up to order $B_n = o(n)$; see, e.g., Andrews (1991) and Davidson and de Jong (2000). The stronger requirement $B_n = o(\sqrt{n})$ reflects that $\hat{V}_i(\tau)$ is not a smooth function of $\hat{\xi}_n(\tau)$. However, this requirement is not much of a restriction because, as Andrews notes, optimal bandwidths are typically of order less than $\sqrt{n}$.

3. If the quantile of interest $\xi_0(\tau)$ is assumed to be known, then Theorem 1 of Liu and Wu (2010) implies that Theorem 1.11 continues to hold when (i) $\xi_0(\tau)$ is used in the definition of $\hat{g}_{n,\tau}$ instead of $\hat{\xi}_n(\tau)$, (ii) Assumption 1.5 is replaced the condition that $F_X$ is continuous and increasing at $\xi_0(\tau)$, (iii) $\delta = 0$ in Assumption 1.9, and (iv) $B_n = o(n)$.

4. The smoothed quantile periodogram at a known quantile $\xi_0(\tau)$ is just an ordinary smoothed periodogram of $V_i(\tau)$ and therefore optimality results from classical spectral analysis apply. In particular, the optimal lag window among the kernels in $\mathcal{W} \cap \{W \geq 0\}$ with respect to the relative mean-square error (MSE) criterion of Priestley (1962) is the quadratic-spectral (QS) window

$$w_{QS}(x) = \frac{25}{12\pi^2x^2}\left(\sin\left(\frac{6\pi x}{5}\right) - \cos\left(\frac{6\pi x}{5}\right)\right).$$

The mean-square optimal bandwidth for the QS kernel is $B_n = O(n^{1/5})$, which can be established under additional dependence conditions; for example, Assumption 1.3 with $\delta = 0$ suffices. In the general case where $\xi_0(\tau)$ is estimated, a truncated MSE criterion as in Andrews (1991) could be used to limit the influence of $\hat{\xi}_n(\tau)$. However, his results rely crucially on second-order differentiability of the smoothed periodogram with respect to the estimated parameter. A fundamentally different approach is therefore likely to be needed, which I leave for future research.

I investigate the finite sample properties of the smoothed quantile periodogram and confidence intervals based on the quantile periodogram in a small simulation study in section 1.5. The next section discusses the use of integrated quantile periodograms to test for uninformative quantile spectra.
1.4 Testing for Flatness of a Quantile Spectrum

In this section I provide two Cramér-von Mises tests (Procedures 1.15 and 1.17 below) for the null hypothesis that the \( \tau \)-th quantile spectrum is flat, i.e., \( g_\tau(\lambda) \equiv \tau(1-\tau)/(2\pi) \), against the alternative that the \( \tau \)-th quantile spectrum is informative.

If the distribution function of \( X_t \) is continuous and increasing at \( \xi_0(\tau) \), then \( r_\tau(0) = \tau(1-\tau) \) and the null and alternative hypotheses can be stated more precisely as

\[
H_0: r_\tau(j) = 0 \text{ for all } j > 0 \quad \text{and} \quad H_1: r_\tau(j) \neq 0 \text{ for some } j > 0.
\]

Provided that \( \sum_{j\in\mathbb{Z}} r_\tau(j) \) converges absolutely, the \( \tau \)-th quantile spectrum is symmetric about zero. One way to test for the null hypothesis is therefore to check if the sample equivalent of

\[
\int_0^\lambda g_\tau(u) \, du - \int_0^\lambda \frac{r_\tau(0)}{2\pi} \, du = \sum_{j>0} r_\tau(j) \psi_j(\lambda),
\]

where \( \psi_j(\lambda) := \sin(j\lambda)/(\pi\lambda) \), is near zero for all \( \lambda \in \Pi := [0,\pi] \).

The quantity in the preceding display is best understood as a function in \( L_2(\Pi) \), the set of Lebesgue-measurable functions \( f: \Pi \to \mathbb{R} \) with \( \int_{\Pi} f(\lambda)^2 \, d\lambda < \infty \). Under the equivalence relation “\( f \equiv g \) if and only if \( f = g \) Lebesgue-almost everywhere,” \( L_2(\Pi) \) is a proper Hilbert space with inner product \( \langle f, g \rangle_\Pi := \int_{\Pi} f(\lambda)g(\lambda) \, d\lambda \) for \( f, g \in L_2(\Pi) \) and norm \( \| f \|_\Pi := \sqrt{\langle f, f \rangle_\Pi} \). Since \( \| \psi_j \|_\Pi = 1/(2\pi j^2) \) for all \( j \in \mathbb{Z} \setminus \{0\} \), (1.6) indeed lies in \( L_2(\Pi) \) and satisfies

\[
\left\| \sum_{j>0} r_\tau(j) \psi_j \right\|_\Pi^2 = \sum_{j>0} r_\tau(j)^2 \| \psi_j \|_\Pi^2.
\]

Here we need the fact that \( \langle \psi_j, \psi_k \rangle_\Pi = 0 \) for all \( j \neq k \). Now replace \( r_\tau(j) \) by \( \hat{r}_{n,\tau}(j) \) and rescale to obtain the Cramér-von Mises statistic

\[
CM_{n,\tau} := \left\| \sqrt{n} \sum_{j=1}^{n-1} \hat{r}_{n,\tau}(j) \psi_j \right\|^2_\Pi = \frac{n}{2\pi} \sum_{j=1}^{n-1} \left( \frac{\hat{r}_{n,\tau}(j)}{j} \right)^2
\]

based on the random process \( S_{n,\tau}(\lambda) := \sqrt{n} \sum_{j=1}^{n-1} \hat{r}_{n,\tau}(j) \psi_j(\lambda) \) in \( L_2(\Pi) \). No
smoothing weight function and bandwidth is needed because the integral in (1.6) already acts as a smoothing operator. The scaling factor $\sqrt{n}$ in $S_{n,\tau}$ is included because $\sqrt{n} \hat{r}_{n,\tau}(j)$ can be expected to have an asymptotically normal distribution for each $j > 0$ under the null hypothesis. When viewed as a random function on $L_2(\Pi)$, the process $S_{n,\tau}(\lambda)$ should then converge in distribution to a mean-zero Gaussian process $S_\tau(\lambda)$ with covariances

$$
\mathbb{E} S_\tau(\lambda) S_\tau(\lambda') = \sum_{j>0} \sum_{k>0} \sum_{l \in \mathbb{Z}} \left( V_0(\tau) V_j(\tau), V_{j-l}(\tau) V_{j-l-k}(\tau) \right) \psi_j(\lambda) \psi_k(\lambda'),
$$

(1.7)

$\lambda, \lambda' \in \Pi$, so that $CM_{n,\tau} \rightsquigarrow \|S_\tau\|_{\Pi}^2$ by the Continuous Mapping Theorem (see, e.g., Theorem 18.11 of van der Vaart, 1998, p. 259).

As the following Theorem shows, this convergence indeed occurs if the conditions of the null hypothesis are strengthened slightly: Suppose that under $H_0$, for a given $\tau \in (0,1)$ there is a $\delta > 0$ such that

$$
P(0 < \xi, X_j < \xi') = P(0 < \xi) P(0 < \xi')
$$

(1.8)

for all $j > 0$ and all $\xi, \xi' \in X_\tau(\delta)$, where $X_\tau(\delta) = \{\xi \in \mathbb{R} : |\xi_0(\tau) - \xi| \leq \delta\}$ as before. The role of this condition is discussed in detail below.

**Theorem 1.12.** Suppose Assumptions 1.3 and 1.5 hold for some $\tau \in (0,1)$.

(i) If $H_0$ is satisfied in the sense of (1.8), then $CM_{n,\tau} \rightsquigarrow \|S_\tau\|_{\Pi}^2$, and

(ii) if $H_1$ is satisfied, then $P(CM_{n,\tau} > B) \to 1$ for every $B \in \mathbb{R}$.

**Remarks.** 1. For the proof of the Theorem, I show stochastic equicontinuity of the empirical process $(n-j)^{-1/2} \sum_{j=0}^{n-j} [V_\tau(\xi) V_{\tau+j}(\xi) - \mathbb{E} V_\tau(\xi) V_j(\xi)]$ indexed by $\xi$ under Assumptions 1.3 and 1.5 for each fixed $j$. Condition (1.8) is used to control the behavior of $\hat{r}_{n,\tau}(j)$ for large $j$ and $n$. These two results then allow me to apply a general result on Cramér-von Mises tests for spectral densities given in Shao (2011a).

2. Condition (1.8) imposes slightly more on the dependence structure of $V_\tau(\tau)$ than the white noise assumption $H_0$ (i.e., $\delta = 0$). However, since $\delta$ can be chosen to be as small as desired, it is much less restrictive than requiring that $(X_t)_{t \in \mathbb{Z}}$ be pairwise independent ($\delta = \infty$) or even iid, which is frequently imposed when testing for white noise; see, e.g., Milhøj (1981) and Hong (1996).
Example 1.13 (Stochastic volatility, continued). Recall that $F_\varepsilon$ is the distribution function of $\varepsilon$. The stochastic volatility process in Example 1.1 has a flat $\tau_0$-th quantile spectrum but fails to satisfy (1.8) because

$$
\mathbb{P}(X_0 < \xi, X_j < \xi') = \mathbb{E} 1 \left\{ \varepsilon_t < \frac{\xi - \xi_0(\tau_0)}{v(\varepsilon_{t-1}, \ldots)} \right\} F_\varepsilon \left( \frac{\xi' - \xi_0(\tau_0)}{v(\varepsilon_{j-1}, \ldots)} \right)
$$

can generally not be simplified further due to the lagged innovations contained in the volatility process $v$. Thus, Theorem 1.12 does not apply. However, the test procedure from Example 1.8 can still be used in this case to test for flatness of the $\tau_0$-th quantile spectrum, for if $g_{\tau_0}(\lambda_0) = \tau_0(1 - \tau_0)/(2\pi)$ is rejected at some frequency $\lambda_0$, then $H_1$ must be true. Linton and Whang (2007) investigate the stochastic volatility model of Example 1.1 with the sample quantilogram, defined as $\hat{r}_{n,\tau}(j)/\hat{r}_{n,\tau}(0)$, for a fixed, finite number of lags $j = 1, 2, \ldots$. From their results it can be seen that the failure of (1.8) for the stochastic volatility model manifests itself in terms of a non-vanishing drift term in $\sqrt{n}\hat{r}_{n,\tau}(j)$ due to the estimation of $\xi_0(\tau)$. A Cramér-von Mises test requires control of these drifts for large $j$ and $n$; this is nontrivial and left for future work. □

Example 1.14 (QAR, continued). The QAR process in Example 1.2 possesses a flat $\tau_0$-th quantile spectrum and has the property (1.8) if there exists a neighborhood $T$ of $\tau_0$ such that $\beta_1(\tau) = \beta_2(\tau) = 0$ for all $\tau \in T$: In this case, the conditional quantile function, defined as the solution $\xi(\tau | F_{t-1})$ of $\mathbb{P}(X_t \leq \xi | F_{t-1}) = \tau$, is given by $\xi(\tau | F_{t-1}) = \beta_0(\tau) + \beta_1(\tau)X_{t-1} + \beta_2(\tau)X_{t-2} = \beta_0(\tau)$ almost surely for all $\tau \in T$ by monotonicity. Take expectations to deduce that

$$
\tau = \mathbb{P}(X_t \leq \xi(\tau | F_{t-1}) | F_{t-1}) = \mathbb{P}(X_t \leq \beta_0(\tau)) = \mathbb{P}(X_0 \leq \xi_0(\tau)),
$$

almost surely for all $\tau \in T$ and therefore $\xi(\tau | F_t) = \xi_0(\tau)$ almost surely on $\tau \in T$. Conclude that for any $\tau, \tau' \in T$,

$$
\mathbb{P}(X_0 < \xi_0(\tau), X_j < \xi_0(\tau')) = \mathbb{E} 1 \{X_0 < \xi_0(\tau)\} \mathbb{P}(X_j < \xi_0(\tau') | F_{j-1})
\quad = \mathbb{P}(X_0 < \xi_0(\tau)) \mathbb{P}(X_0 < \xi_0(\tau')).
$$

Now (1.8) follows because as long as $F_X$ is continuous and increasing in a neighborhood of $\xi_0(\tau_0)$, there is a $\delta > 0$ such that for every $\xi, \xi' \in X_{\tau_0}(\delta)$,
there are $\tau, \tau' \in T$ such that $\xi = \xi_0(\tau)$ and $\xi' = \xi_0(\tau')$. The assertion in Example 1.2 about the flatness of the $\tau_0$-th quantile spectrum is obtained by letting $T = \{\tau_0\}$. □

The main difficulty with applying Theorem 1.12 in practice is the unknown covariance function (1.7) of the limiting process $S_{\tau}$. In standard spectral analysis, this has led researchers to assume that $X_t$ is iid normal under the null hypotheses of white noise (Durbin, 1967, is an important early reference) in order to avoid having to estimate the covariance function of a Gaussian process. In sharp contrast, in quantile spectral analysis the assumption that $X_t$ is iid is already enough to construct a test for flatness without imposing a distributional assumption: In large samples $\hat{V}_{\tau}(\tau)$ is close to $V_{\tau}(\tau) = \tau - 1\{X_t < \xi_0(\tau)\}$ in probability, but $1\{X_t < \xi_0(\tau)\}$ is a Bernoulli random variable with success probability $\tau$ as long as $F_{X}$ is continuous and increasing at $\xi_0(\tau)$. Hence, if $X_t$ is indeed iid and $J_1, J_2, \ldots, J_n$ are independent Bernoulli($\tau$) variables, then

$$C_{M_{n,\tau}} := \frac{1}{2\pi n} \sum_{j=1}^{n-1} j^{-2} \left( \sum_{t=1+j}^{n} V_t(\tau) V_{t-j}(\tau) \right)^2$$

and

$$CM'_{n,\tau} := \frac{1}{2\pi n} \sum_{j=1}^{n-1} j^{-2} \left( \sum_{t=1+j}^{n} (\tau - J_t)(\tau - J_{t-j}) \right)^2$$

have the same distribution. Because $CM_{n,\tau} = C_{M_{n,\tau}} + o_p(1)$ under the conditions of Theorem 1.12(i), this distributional equivalence leads to a simple, distribution-free Monte Carlo test. I prove its consistency in Corollary 1.16 below.

**Procedure 1.15 (Monte Carlo test for flatness).**

1. Draw $n$ iid copies $J_1, J_2, \ldots, J_n$ of a Bernoulli($\tau$) random variable.
2. Compute $CM'_{n,\tau}$ with the variables from step 1.
3. Repeat steps 1 and 2 $R$ times. Reject $H_0$ in favor of $H_1$ if $CM_{n,\tau}$ is larger than $c_{n,\tau}(1 - \alpha)$, the $1 - \alpha$ empirical quantile of the $R$ realizations of $CM'_{n,\tau}$.

**Remark.** Exploiting the distribution-free character of sign or quantile crossing indicators has a long history in statistics and econometrics; see, e.g., Walsh (1960). More recently, Chernozhukov, Hansen, and Jansson (2009) use it to construct finite sample confidence intervals for quantile regression estimators.
By choosing the number of Monte Carlo repetitions $R$ large enough, the quantiles of the null distribution of $\tilde{CM}_{n,\tau}$ can be approximated with arbitrary precision. I therefore let $R \to \infty$ and define the quantiles of the simulated distribution directly as $c_{n,\tau}(1-\alpha) := \inf\{x \in \mathbb{R} : \mathbb{P}(\tilde{CM}_{n,\tau} > x) \leq \alpha\}$.

The large sample properties of Procedure 1.15 can now be stated as follows:

**Corollary 1.16.** Suppose Assumption 1.5 is satisfied for some $\tau \in (0,1)$ and let $\alpha \in (0,1)$.

(i) If $(X_t)_{t \in \mathbb{Z}}$ is an iid sequence, then $\mathbb{P}(CM_{n,\tau} > c_{n,\tau}(1-\alpha)) \to \alpha$, and

(ii) if Assumption 1.3 and $H_1$ hold, then $\mathbb{P}(CM_{n,\tau} > c_{n,\tau}(1-\alpha)) \to 1$.

**Remark.** If $\xi_0(\tau)$ is known, then the test in Procedure 1.15 has level $\alpha$ even in finite samples provided that $\tilde{CM}_{n,\tau}$ is used in step 3 instead of $CM_{n,\tau}$.

In cases where it does not seem reasonable to assume that $X_t$ is iid under the null hypothesis, the block-wise wild bootstrap of Shao (2011a) should be used instead. This bootstrap is a modification of the standard wild bootstrap (Liu, 1988; Mammen, 1992). It perturbs whole blocks of observations with iid copies of a random variable $\eta$ that is independent of the data and satisfies $\mathbb{E}\eta = 0$, $\mathbb{E}\eta^2 = 1$, and $\mathbb{E}\eta^4 < \infty$. Since the blocks grow with the sample size, this eventually captures enough of the dependence structure to provide critical values for the null distribution under the more general condition (1.8).

**Procedure 1.17 (Shao’s block-wise wild bootstrap).**

1. Choose a block length $b_n \leq n$ and the corresponding number of blocks $L_n = n/b_n$, taken to be an integer for convenience. For each $s = 1, \ldots, L_n$ define a block $B_s = \{(s-1)b_n + 1, \ldots, sb_n\}$.

2. Draw $L_n$ iid copies $\eta_1, \eta_2, \ldots, \eta_{L_n}$ of $\eta$. For each $t = 1, \ldots, n$, define $\omega_t = \sum_{s=1}^{L_n} \eta_s 1\{t \in B_s\}$ so that $\omega_t$ takes on the value $\eta_s$ if $t$ lies in the $s$-th block.

3. Compute $\hat{r}_{n,\tau}(j) := n^{-1} \sum_{t=j+1}^{n} [\hat{V}_t(\tau)\hat{V}_{t-j}(\tau) - \hat{r}_{n,\tau}(j)]\omega_t$ and calculate the bootstrap statistic

$$CM^*_n,\tau := \frac{n}{2\pi} \sum_{j=1}^{n-1} \left(\frac{\hat{r}_{n,\tau}^*(j)}{j}\right)^2.$$ 

4. Repeat steps 2 and 3 $R$ times. Reject $H_0$ in favor of $H_1$ if $CM_{n,\tau}$ is larger than $c_{n,\tau}^*(1-\alpha)$, the $1-\alpha$ empirical quantile of the $R$ realizations of $CM^*_n,\tau$.  

21
Remark. The recommended choice for $\eta$ in practice is a Rademacher variable that takes on the value 1 with probability $1/2$ and the value $-1$ with probability $1/2$. Distributions other than the Rademacher distribution can be used for $\eta$, in particular if $\hat{V}_t(\tau)\hat{V}_{t-j}(\tau)$ has a skewed distribution, but there is no evidence that they would lead to better inference; see Davidson, Monticini, and Peel (2007) for a discussion of this point for the standard wild bootstrap.

As before, I take $R$ to be large and define the quantiles of the bootstrap distribution conditional on the sample $S_n$ as $c_{n,\tau}^*(1-\alpha) = \inf \{ x \in \mathbb{R} : \mathbb{P}(CM_{n,\tau}^* \leq x \mid S_n) \geq 1-\alpha \}$. Procedure 1.17 then has the following asymptotic properties:

**Theorem 1.18.** Suppose Assumptions 1.3 and 1.5 hold for some $\tau \in (0,1)$. Let $\alpha \in (0,1)$, $b_n \to \infty$ and $b_n/n \to 0$.

(i) If $H_0$ is satisfied in the sense of (1.8), then $\mathbb{P}(CM_{n,\tau} > c_{n,\tau}^*(1-\alpha)) \to \alpha$, and

(ii) if $H_1$ is satisfied, then $\mathbb{P}(CM_{n,\tau} > c_{n,\tau}^*(1-\alpha)) \to 1$.

Remark. If $\xi_0(\tau)$ is known, then Theorems 1.12 and 1.18 remain valid without condition (1.8) as long as $\tilde{CM}_{n,\tau}$ is used in place of $CM_{n,\tau}$.

The next section investigates the finite sample properties of the two Cramér-von Mises tests, the quantile periodogram, and the smoothed quantile periodogram in a Monte Carlo study and provides an empirical application.

### 1.5 Numerical Results

In this section I present a sequence of examples to illustrate quantile spectral methods in the context of some familiar time series models and macroeconomic data, and compare the results to those obtained from traditional spectral analysis.

**Example 1.19 (AR(2) with spectral peak).** Let $(\varepsilon_t)_{t \in \mathbb{Z}}$ be iid copies of an $N(0,1)$ variable with distribution function $\Phi$. Li (2008) investigates the frequency domain properties of a stationary AR(2) process of the form

$$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \varepsilon_t, \quad \beta_1 = 2 \times 0.95 \cos(2\pi \times 0.22), \beta_2 = -0.95^2. \tag{1.9}$$

Shao and Wu’s (2007) Theorem 5.2 implies that $X_t$ is GMC for all $\alpha > 0$.  

22
Figure 1.1: Left panel: spectral density (dotted line) of $X_t$ as in (1.9), QS-smoothed periodogram (solid) of a realization with $n = 300$ and $B_n = 13n^{1/5} \approx 40.68$, chi-squared point-wise 95% confidence bands (shaded grey) with $k = 4$, and $\hat{\gamma}_{n,X}(0)/(2\pi)$ (dashed). Right panel: median spectrum (dotted) of $X_t$, QS-smoothed median periodogram (solid), chi-squared point-wise 95% confidence bands (shaded grey), and $0.5(1 - 0.5)/(2\pi)$ (dashed). Both panels use the same data, $B_n$, and $k$, and are normalized by $\hat{\gamma}_{n,X}(0)/(2\pi)$ (left) and $0.5(1 - 0.5)/(2\pi)$ (right).

Since $X_t$ is also normally distributed, Proposition 1.4 applies and consequently Assumptions 1.3 and 1.5 hold. To study the finite sample properties of classical and quantile spectral estimates for sample sizes $n \in \{300, 600, 900\}$ in this model, I generated 10,000 realizations of the process of size $400 + n$ for each $n$ and then discarded the first 400 observations. Each realization was initialized by independent standard normal random variables. The solid black line in the left panel of Figure 1.1 plots a QS-smoothed periodogram of $X_t$, i.e.,

$$\hat{f}_{n,X}(\lambda) = \frac{1}{2\pi} \sum_{|j| < n} w_{QS}(j/B_n) \hat{\gamma}_{n,X}(j) \cos(j\lambda),$$

where $\hat{\gamma}_{n,X}(j) := n^{-1} \sum_{t=|j|+1}^{n} (X_t - X_n)(X_{t-|j|} - X_n)$ and $X_n := n^{-1} \sum_{t=1}^{n} X_t$, of one such realization with $n = 300$ and $B_n = 13n^{1/5} \approx 40.68$. The process (1.9) has little noise and a single pronounced peak at $2\pi \times 0.22$ in its spectral density, shown as the dotted line in the left panel of Figure 1.1. The smoothed periodogram therefore does not have much difficulty identifying the peak,
although its size is underestimated slightly due to the smoothing. The shaded area in the left panel shows 95% asymptotic point-wise confidence bands based on the periodogram of $X_t$, defined as

$$I_{n,X} (\lambda) = \frac{1}{2\pi} \sum_{|j| < n} \dot{\gamma}_{n,X}(j) \cos(j\lambda).$$

The point-wise confidence bands were computed by averaging over $2k + 1$ periodogram coordinates at natural frequencies in the same way as in Corollary 1.7, but with $Q_{n,\tau}$ replaced by $I_{n,X}$. Here and in all plots below, I used $k = 4$. The dashed line in the left panel plots $\dot{\gamma}_{n,X}(0)/(2\pi)$, i.e., the usual estimate of $f_X$ if the spectrum were known to be flat. It provides a natural point of comparison for the other quantities; in particular, it can be seen from the left panel that the peak at $2\pi \times 0.22$ is significantly different from a flat spectrum at the 5% level.

The right panel of Figure 1.1 analyzes the same data with quantile spectral methods. The black line is the QS-smoothed median (i.e., 0.5-th quantile) periodogram and the shaded area graphs 95% point-wise confidence bands computed as described in Corollary 1.7. Here I used the same values for $B_n$ and $k$ as in the left panel. The dashed line is $0.5(1 - 0.5)/(2\pi)$, i.e., the median spectrum under the hypothesis that it is flat. The dotted line shows the median spectrum $g_{0.5}$, which can be calculated exactly from equation (6) in Li (2008). The smoothed median periodogram clearly identifies the peak, although the estimate of the actual size of the peak is slightly worse than the one obtained in the left panel. However, the median spectrum is completely contained inside the confidence bands and the peak at $2\pi \times 0.22$ differs significantly from a flat median spectrum at the 5% level.

For both panels the choice of $B_n$ and $k$ matters, with lower values of $B_n$ and higher values of $k$ leading to smoother—but not necessarily better—estimates: Figure 1.2 shows the mean integrated square error (MISE) of the QS-smoothed periodogram (left panel) and the QS-smoothed median periodogram (right) estimated from the 10,000 realizations as a function of $B_n/n^{1/5}$. Here the behavior of both methods is quite similar and the MISEs attain their minimum at $B_n/n^{1/5} \approx 13$ for each $n \in \{300, 600, 900\}$, which provides evidence that the optimal growth rate $B_n = O(n^{1/5})$ for the QS-smoothed periodogram is also a good choice for QS-smoothed quantile

24
periodograms. Further, Table 1.1 shows the empirical frequency of the event that the 95% confidence interval at $\lambda \in \{\pi \times 0.22, 2\pi \times 0.22, 3\pi \times 0.22\}$ covered the spectrum and median spectrum, respectively, in the experiments for $k \in \{2, 4, 6\}$ and $n$ as before. The confidence intervals constructed from the periodogram and the median periodogram behaved very similar at the three frequencies and covered the population value in nearly 95% of all cases unless $n$ was small and $k$ was large. For these values both methods had low coverage frequencies.

Robust estimators (in the sense of Huber and Ronchetti, 2009, p. 5) exhibit stability, i.e., small deviations from the model assumptions should have small effects on the performance of the estimator, and high breakdown resistance, i.e., larger deviations should not cause catastrophic results. The following two examples illustrate that classical spectral estimates are not robust to outliers in the data, whereas quantile spectral estimators provide reliable results in such situations.

Example 1.20 (Stability of quantile spectral estimators). Suppose that each observation in a realization of the AR(2) process from Example 1.19 has a probability $p$ of being contaminated by an additional additive error component. For this I drew iid Bernoulli($p$) variables $J_1, \ldots, J_n$ and iid cen-
Table 1.1: Finite-sample coverage frequencies of an asymptotic 95% confidence interval (CI) for the spectrum and median spectrum of the process in Example 1.19 at $\lambda \in \{\pi \times 0.22, 2\pi \times 0.22, 3\pi \times 0.22\}$ as a function of $n$ and $k$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>Periodogram CI</th>
<th>Median Periodogram CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\pi \times .22$</td>
<td>$2\pi \times .22$</td>
</tr>
<tr>
<td>300</td>
<td>2</td>
<td>0.940</td>
<td>0.931</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.936</td>
<td>0.676</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.921</td>
<td>0.249</td>
</tr>
<tr>
<td>600</td>
<td>2</td>
<td>0.943</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.946</td>
<td>0.915</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.944</td>
<td>0.774</td>
</tr>
<tr>
<td>900</td>
<td>2</td>
<td>0.950</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.949</td>
<td>0.941</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.948</td>
<td>0.904</td>
</tr>
</tbody>
</table>

Central Student $t(\nu)$ variables $T_1, \ldots, T_n$ to generate the observed samples as $S_n = \{X_t + J_t T_t : t = 1, \ldots, n\}$, where the $X_1, \ldots, X_n$ were taken from Example 1.19. The spectral density of the corresponding process $(X_t + J_t T_t)_{t \in \mathbb{Z}}$ is

$$f_{X+JT}(\lambda) = f_X(\lambda) + \frac{p}{2\pi \nu - 2},$$

which, for any given $p$, can be made as large as desired by choosing $\nu > 2$ sufficiently close to 2 without violating the assumptions of classical spectral theory. Figure 1.3 plots $f_{X+JT}(\lambda)$ for $p = 0.15$ and $\nu = 2.001$ as a dotted line in the left panel; the median spectrum (dotted, right) needed no adjustment because it is invariant under such contamination. The other quantities are the same as in Figure 1.1 and the same 300 observations were used, but 46 of these were contaminated. The smoothed periodogram retains the spectral shape and has a significant spike at $2\pi \times 0.22$, but grossly underestimates the location of the spectrum. Moreover, the confidence bands no longer contain the spectrum at any frequency. In sharp contrast, the smoothed median periodogram is barely affected by the contamination and the confidence bands cover the median spectrum at almost all frequencies. The hypothesis that $g_{0.5}(2\pi \times 0.22) = 0.5(1 - 0.5)/(2\pi)$ can also be clearly rejected.

I repeated the experiment from Table 1.1 with the contaminated data. The estimated coverage probabilities for the confidence intervals constructed
Figure 1.3: Left panel: spectral density (dotted line) of the process in Example 1.20, QS-smoothed periodogram (solid) of a realization with $n = 300$ and $B_n = 13n^{1/5} \approx 40.68$, chi-squared point-wise 95% confidence bands (shaded grey) with $k = 4$, and $\hat{\gamma}_{n,X}(0)/(2\pi)$ (dashed). Right panel: median spectrum (dotted), QS-smoothed median periodogram (solid), chi-squared point-wise 95% confidence bands (shaded grey), and $0.5(1 - 0.5)/(2\pi)$ (dashed). Both panels use the same data, $B_n$, and $k$, and are normalized by $\hat{\gamma}_{n,X}(0)/(2\pi)$ (left) and $0.5(1 - 0.5)/(2\pi)$ (right).

from the periodogram and the median periodogram are shown in Table 1.2. As can be seen, the presence of outliers had little effect on the performance of the quantile spectral estimates. In sharp contrast, the coverage probability for the classical spectrum was almost zero in most cases and 0.245 in the best scenario ($k = 2, n = 900$).

The odd behavior of the classical spectral density estimates in this example is likely due to the imprecisely estimated auto-covariances of the contaminated process. As Basraka, Davis, and Mikosch (2002) point out, for near-infinite variance time series the convergence rate of sample auto-covariances to their population equivalent is much slower than $n^{-1/2}$. Since periodograms are weighted sums of sample auto-covariances, they can be expected to inherit this lack of precision. In contrast, the sample auto-covariances of $\hat{V}_t(\tau)$ can be shown to converge at rate $n^{-1/2}$ as long as Assumption 1.3 and a slightly strengthened version of Assumption 1.5 hold.

**Example 1.21** (Breakdown resistance of quantile spectral estimators). Now suppose instead that each observation from Example 1.19 has a 15 per-
Table 1.2: Finite-sample coverage frequencies of an asymptotic 95% confidence interval (CI) for the spectrum and the median spectrum of the process in Examples 1.20 as a function of $n$ and $k$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>Periodogram CI</th>
<th>Median Periodogram CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\pi \times .22$</td>
<td>$2\pi \times .22$</td>
</tr>
<tr>
<td>300</td>
<td>2</td>
<td>0.001</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>600</td>
<td>2</td>
<td>0.001</td>
<td>0.208</td>
</tr>
<tr>
<td>600</td>
<td>4</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>900</td>
<td>2</td>
<td>0.001</td>
<td>0.245</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.001</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.001</td>
<td>0.002</td>
</tr>
</tbody>
</table>

The observed samples then were $S_n = \{X_t + J_t C_t : t = 1, \ldots, n\}$ with the $X_1, \ldots, X_n$ as before. Since these outliers do not have a well defined mean, the spectral density of the corresponding contaminated process no longer exists. Spectral analysis by ordinary methods broke down completely when 46 of the 300 observations used for Figure 1.1 were contaminated: The smoothed periodogram in Figure 1.4 no longer has the expected spectral shape and fails to give any indication of a periodicity present in the data. A comparison of the confidence bands to the estimate of $\gamma_X(0)/(2\pi)$ now provides overwhelming evidence for the false hypothesis that the process is white noise. In sharp contrast, the median spectrum is unaffected by the contamination and the smoothed median periodogram significantly identifies the periodicity. In addition, the confidence bands remain essentially unchanged from Example 1.20, which is also confirmed by the coverage probability estimates of the confidence intervals constructed from median periodograms provided in Table 1.3. Here the estimates were nearly identical to the ones presented in Table 1.2 for the median spectrum. Corresponding estimates for the classical spectrum cannot be computed because it is unbounded at all frequencies.

For the next Monte Carlo exercise, I return to the stochastic volatility model from Example 1.1 to illustrate that even if the classical spectrum
Figure 1.4: Left panel: QS-smoothed periodogram (solid black) of a realization of the process in Example 1.21 with $n = 300$ and $B_n = 13n^{1/5} \approx 40.68$, chi-squared point-wise 95% confidence bands (shaded grey) with $k = 4$, and $\hat{\gamma}_{n,X}(0)/(2\pi)$ (dashed). The spectral density does not exist. Right panel: median spectrum (dotted), QS-smoothed median periodogram (solid), chi-squared point-wise 95% confidence bands (shaded grey), and $0.5(1 - 0.5)/(2\pi)$ (dashed). Both panels use the same data, $B_n$, and $k$, and are normalized by $\hat{\gamma}_{n,X}(0)/(2\pi)$ (left) and $0.5(1 - 0.5)/(2\pi)$ (right).

Table 1.3: Finite-sample coverage frequencies of an asymptotic 95% confidence interval for the median spectrum of the process in Example 1.21 as a function of $n$ and $k$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>$\pi \times .22$</th>
<th>$2\pi \times .22$</th>
<th>$3\pi \times .22$</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>2</td>
<td>0.904</td>
<td>0.971</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.889</td>
<td>0.827</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.861</td>
<td>0.059</td>
<td>0.966</td>
</tr>
<tr>
<td>600</td>
<td>2</td>
<td>0.915</td>
<td>0.973</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.901</td>
<td>0.961</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.884</td>
<td>0.872</td>
<td>0.955</td>
</tr>
<tr>
<td>900</td>
<td>2</td>
<td>0.918</td>
<td>0.970</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.905</td>
<td>0.966</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.883</td>
<td>0.940</td>
<td>0.951</td>
</tr>
</tbody>
</table>
shows no sign of periodicity, almost all quantiles of the distribution can be crossed in a periodic manner.

**Example 1.22** (Stochastic volatility, continued). Take \((\varepsilon_t)_{t\in\mathbb{Z}}\) to be iid copies of a \(N(0, \theta^2)\) variable and let \(u_t = \log v(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)\) be the stationary solution of the process \(u_t = \beta_1 u_{t-1} + \beta_2 u_{t-2} + \varepsilon_{t-1}\) with \(\beta_1, \beta_2\) as in (1.9). Then \(e^{u_t}\) is log-normally distributed and \(X_t = \varepsilon_t v(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = \varepsilon_t e^{u_t}\) has median zero. To show that \(X_t\) is GMC, apply the Mean Value Theorem and the Cauchy-Schwarz inequality to obtain the bound 
\[
\|X_n - X'_n\|_\alpha \leq \|\varepsilon_n\|_\alpha \|e^{u_n}\|_{2\alpha} \|u_n - u'_n\|_{2\alpha},
\]
where \(u'_n\) is \(u_n\) with \((\varepsilon_0, \varepsilon_{-1}, \ldots)\) replaced by \((\varepsilon^*_{0}, \varepsilon^*_{-1}, \ldots)\) and \(\bar{u}_n\) lies on the line segment joining \(u_n\) and \(u'_n\). By monotonicity of the exponential function and the Minkowski inequality, we have
\[
\|e^{\bar{u}_n}\|_{\min\{1, 2\alpha\}} \leq \|\max\{e^{u_n}, e^{-u_n}, e^{u'_n}, e^{-u'_n}\}\|_{\min\{1, 2\alpha\}} \leq 4\|e^{u_n}\|_{\min\{1, 2\alpha\}} < \infty
\]
because the four terms inside the maximum have the same log-normal distribution. If needed, the Loève c.r inequality provides a similar bound for the case \(0 < 2\alpha < 1\). The GMC property then follows since \(u_t\) is GMC by Theorem 5.2 of Shao and Wu (2007). The distribution function of \(X_t\) is given by \(F_X(x) = \mathbb{E} \Phi(x/(e^{u}\theta))\), which can be seen to have a bounded density with the help of the Lebesgue Dominated Convergence Theorem. Therefore, Assumptions 1.3 and 1.5 again hold.

The top two panels of Figure 1.5 graph the same spectral estimates as in Figures 1.1 for \(n = 600\) observations of the stochastic volatility model with \(\theta = 1\). The spectrum (not shown to prevent clutter) and the median spectrum (identical to the dashed line in the top right panel) of the model are flat, which is also correctly identified at almost all frequencies by both point-wise confidence bands. The bottom two panels show the smoothed quantile periodograms (black lines) and point-wise confidence bands (shaded grey) at \(\tau = 0.25\) (left) and \(\tau = 0.75\) (right) computed from the same data. In both panels, the estimated quantile spectra show a considerable spike that is significantly different from a flat \(\tau\)-th quantile spectrum at frequency \(2\pi \times 0.22\), thereby providing evidence of a dependence structure that is not present in the mean and auto-covariance of the process. Since the quantile spectra of the process do not possess a closed-form expression for \(\tau \neq 0.5\), I instead also plot smoothed quantile periodograms of \(n = 10^6\) observations at \(\tau = 0.25\) (left) and \(\tau = 0.75\) (right) as dotted lines in the bottom panels to illustrate
Figure 1.5: Top left panel: QS-smoothed periodogram (solid) of a realization of the process in Example 1.22 with \( n = 600 \) and \( B_n = 13n^{1/5} \approx 46.73 \), chi-squared point-wise 95% confidence bands (shaded grey) with \( k = 4 \), and \( \hat{\gamma}_{n,X}(0)/(2\pi) \) (dashed). Other panels: QS-smoothed \( \tau \)-th quantile periodogram (solid), chi-squared point-wise 95% confidence bands (shaded grey), and \( \tau(1-\tau)/(2\pi) \) (dashed) for \( \tau = 0.5 \) (top right), 0.25 (bottom left), and 0.75 (bottom right). All panels use the same data, \( B_n \), and \( k \). The top left panel is normalized by \( \hat{\gamma}_{n,X}(0)/(2\pi) \). The other panels are normalized by \( \tau(1-\tau)/(2\pi) \). The bottom two panels also show QS-smoothed \( \tau \)-th quantile periodograms (dotted) with \( n = 10^6 \) for \( \tau = 0.25 \) (left) and 0.75 (right). Frequencies near zero are not shown to enhance readability.
how much of the spectral shape is already recovered in a sample with 600 observations. Indeed, although the estimates from the smaller sample are more volatile, the size and shape of the peaks at $2\pi \times 0.22$ are nearly identical for the two sample sizes.

To evaluate how reliably the quantile spectral estimates discover the cycle at frequency $2\pi \times 0.22$, I recorded the relative number of the test decisions in favor of the hypothesis $H_0$: $g_\tau(2\pi \times 0.22) = \tau(1 - \tau)/(2\pi)$ in 10,000 realizations of the stochastic volatility model using a 95% confidence interval with $k = 4$. The results are shown in Figure 1.6 for different sample sizes as a function of $\tau \in (0, 1)$. At $\tau = 0.5$, the null hypothesis is true and the tests almost attained the 5% level (lower grey line) for the three sample sizes. At the other quantiles, the null hypothesis is false, which was also correctly recognized at all sample sizes as long as a quantile not too close to $\tau = 0.5$ was chosen.

The additional information obtained from quantile spectral analysis can also be seen in Figure 1.7, where I graph the QS-smoothed quantile periodogram as a function of both $\lambda$ and $\tau$. Here I chose $n = 900$ and $B_n = 8n^{1/5} \approx 31.18$ for a smoother appearance of the plot. The two humps in the figure make it clear that most of the dependence structure is in fact present near the lower and upper quartiles of the process, whereas working with the mean or median provides no insight in this case.
Figure 1.7: QS-smoothed quantile periodogram across all quantiles of a realization of the process in Example 1.22 with $n = 900$ and $B_n = 8n^{1/5} \approx 31.18$, normalized by the joint maximum of all quantile periodograms. Frequencies near zero are not shown to enhance readability.

The following examples illustrate the size and power of the two Cramér-von Mises tests introduced in section 1.4.

**Example 1.23 (QAR(2) and Procedure 1.15).** Table 1.4 shows the empirical rejection frequency of the null hypothesis of a flat $\tau$-th quantile spectrum as a function of $n \in \{100, 200, 300\}$ and $\tau \in \{0.1, 0.5, 0.9\}$ in a variety of settings.

For each entry, I recorded the test decision of Procedure 1.15 in 10,000 realizations by comparing the test statistics to 5% critical values obtained from $10^6$ simulations each. The first column of the “Size” portion provides the rejection frequencies when the data were iid $\chi^2_3$ variables. In this case, the null hypothesis is true at all quantiles. The test behaved mildly conservatively for $\tau = 0.1$ in smaller samples, but was close to the level of the test at other quantiles and samples sizes. In samples larger than 300 (not reported), the test was essentially exact at all quantiles. I also experimented with other distributions, including normal, Student $t(2)$, and standard Cauchy variables, but found that they had little impact on the results.

The first column of the “Power” portion shows the relative number of rejections when the data-generating process was the AR(2) from Example 1.19. Here the null hypothesis is false at all quantiles, which was also reliably
Table 1.4: Rejection frequencies of the null hypothesis for the Monte Carlo Cramér-von Mises test (Procedure 1.15) at the 5% level.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\tau$</th>
<th>Size $\chi^2_3$</th>
<th>Power Ex. 1.22</th>
<th>QAR</th>
<th>Power Ex. 1.19</th>
<th>Power Ex. 1.22</th>
<th>QAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.1</td>
<td>0.022</td>
<td>0.024</td>
<td>0.093</td>
<td>0.007</td>
<td>0.993</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.053</td>
<td>0.068</td>
<td>0.999</td>
<td></td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.037</td>
<td></td>
<td>0.169</td>
<td>0.332</td>
<td>0.993</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>0.019</td>
<td>0.021</td>
<td>0.405</td>
<td>0.043</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.052</td>
<td>0.076</td>
<td>1.000</td>
<td></td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.046</td>
<td></td>
<td>0.504</td>
<td>0.468</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>0.1</td>
<td>0.048</td>
<td>0.029</td>
<td>0.795</td>
<td>0.188</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.052</td>
<td>0.080</td>
<td>1.000</td>
<td></td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.050</td>
<td></td>
<td>0.875</td>
<td>0.724</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

identified at the median at all samples. However, at the outer quantiles the spectral peak is smaller and therefore larger samples were needed to detect its presence. The results for the contaminated processes from Examples 1.20 and 1.21 are not shown because they were virtually identical.

The second “Size” and “Power” columns give the rejection frequencies for the stochastic volatility model from Example 1.22. The null hypothesis is true at $\tau = 0.5$, but the process is not covered by the assumptions underlying the Monte Carlo test because the stochastic volatility model is not iid, which resulted in a mild over-rejection at all sample sizes. At the other quantiles, the process satisfies $H_1$ and the test has power against this alternative by Corollary 1.16(ii). The power of the test increased sharply with the sample size for $\tau = 0.9$, whereas for $\tau = 0.1$ the increase was considerably slower. Some intuition for this result can be gathered from Figure 1.5, where the estimated quantile spectrum in the lower quantiles can be seen to have a long stretch on which it is close to the hypothetical quantile spectrum implied by the null hypothesis. In contrast, this stretch is somewhat shorter in the upper quantiles. Moreover, as shown in Figure 1.6, the setup for the test is quite demanding because the spectral peak near the extremes of the distribution is small. Larger samples (not reported) yielded better results, with the power being nearly one at all quantiles for $n = 600$.

The third columns of the “Size” and “Power” portions show the relative number of rejections of the hypothesis of a flat $\tau$-th quantile spectrum for
realizations of the QAR(2) process (see Example 1.2)

\[ X_t = 4 + \Phi^{-1}(\varepsilon_t) + 0.8 \times 1\{\varepsilon_t > 0.2\} X_{t-1} + 0.6 \times 1\{\varepsilon_t > 0.6\} X_{t-2} \quad (1.10) \]

where, as before, \((\varepsilon_t)_{t\in\mathbb{Z}}\) is a sequence of iid copies of a Uniform(0, 1) variable. By Theorem 5.1 of Shao and Wu (2007), this recursion admits a stationary solution of the form (1.5) and satisfies the GMC property. Further, the marginal distribution function of \(X_t\) can be seen to possess a bounded Lebesgue density from the properties of truncated normal variables and dominated convergence. If \((X_t)_{t\in\mathbb{Z}}\) is positive, the right-hand side of (1.10) is guaranteed to be increasing in \(\varepsilon_t\) conditional on \(X_{t-1}, X_{t-2}\) and the model in the preceding display is indeed a proper QAR model. Since the process has a very small probability of generating a negative observation, I therefore considered only positive realizations of (1.10) in order to enforce well-behaved sample paths.

The QAR process satisfies the null hypothesis of a flat quantile spectrum for \(\tau \in (0, 0.2]\) and the alternative at the other quantiles. In particular, it behaves like a stationary QAR(1) on \(\tau \in (0.2, 0.6]\) that exhibits enough mean reversion to regulate the explosive behavior of the process on \(\tau \in (0.6, 1]\). This dependence structure induces an asymmetric spectral shape across quantiles, with spectral peaks of different sizes at frequency zero in the middle to upper quantiles. The QS-smoothed quantile periodogram of a realization with \(n = 900\) plotted in Figure 1.8 illustrates this shape. As can be seen from Table 1.4, the Monte Carlo Cramér-von Mises test very reliably detected the presence of the alternative hypothesis at \(\tau = 0.5\) and 0.9 even for \(n = 100\). At \(\tau = 0.1\) the null hypothesis is true and, although Procedure 1.15 does not apply because the observations are not iid, the test was only mildly conservative.

\textbf{Example 1.24 (QAR(2) and Procedure 1.17):} I repeated the experiments outlined in the previous example with the wild bootstrap test described in Procedure 1.17. I experimented with the block size \(b_n\), but found that the results were not overly sensitive to this choice as long as the blocks were not too large. I therefore settled for block sizes near \(\sqrt{n}/2\) and used \(b_n = 5, 8,\) and 10 for \(n = 100, 200,\) and 300, respectively, although other choices are clearly possible; see Shao (2011a) for a thorough discussion.
The results are shown in Table 1.5. The important difference to the preceding example is that the QAR(2) model (1.10) is now fully covered by the assumptions of the test; see Theorem 1.18. This is also reflected in the test for a flat quantile spectrum of the QAR process at $\tau = 0.1$, which was nearly exact for $n = 300$. The other results in the “Size” portion of the table were similar to the ones given in Table 1.4 for the Monte Carlo test. The power of the bootstrap test was also comparable to the other test, but neither of the tests dominated the other: For the AR model both test behaved similarly, for the stochastic volatility model the bootstrap test showed a more balanced performance, and for the QAR model the Monte Carlo test was more powerful at the outer quantiles.

Example 1.25 (Building permits data). Finally, to illustrate what kind of insights quantile spectral analysis of actual economic data can provide, I consider the series “New Privately Owned Housing Units Authorized by Building Permits in Permit-Issuing Places” from the US Census. The data consist of 634 (seasonally unadjusted) monthly observations from January 1959 to October 2011 of the total number of permits from permit-issuing places in

---

1I downloaded the data from http://www.census.gov/const/permits_cust.xls on November 30, 2011.
Table 1.5: Rejection frequencies of the null hypothesis for the bootstrap Cramér-von Mises test (Procedure 1.17) at the 5% level. The block sizes are $b_n = 5, 8, \text{ and } 10$ for $n = 100, 200, \text{ and } 300$, respectively. I used the warp-speed method of Giacomini et al. (2007) to estimate size and power of the bootstrap test; this method considerably sped up the simulations because only one bootstrap replication per Monte Carlo replication was needed.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\tau$</th>
<th>$\chi^2_3$</th>
<th>Ex. 1.22</th>
<th>QAR</th>
<th>Ex. 1.19</th>
<th>Ex. 1.22</th>
<th>QAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.1</td>
<td>0.027</td>
<td>–</td>
<td>0.026</td>
<td>0.113</td>
<td>0.056</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.055</td>
<td>0.094</td>
<td>–</td>
<td>1.000</td>
<td>–</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.029</td>
<td>–</td>
<td>–</td>
<td>0.170</td>
<td>0.110</td>
<td>0.374</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>0.031</td>
<td>–</td>
<td>0.030</td>
<td>0.430</td>
<td>0.339</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.058</td>
<td>0.083</td>
<td>–</td>
<td>1.000</td>
<td>–</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.049</td>
<td>–</td>
<td>–</td>
<td>0.486</td>
<td>0.422</td>
<td>0.635</td>
</tr>
<tr>
<td>300</td>
<td>0.1</td>
<td>0.050</td>
<td>–</td>
<td>0.051</td>
<td>0.754</td>
<td>0.550</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.056</td>
<td>0.090</td>
<td>–</td>
<td>1.000</td>
<td>–</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.052</td>
<td>–</td>
<td>–</td>
<td>0.780</td>
<td>0.567</td>
<td>0.820</td>
</tr>
</tbody>
</table>

the United States that report to the Census. Such a permit is typically issued by a town or a county and enables an individual to begin construction on a new housing unit.

Figure 1.9 graphs this time series in the frequency domain: The smoothed periodogram (solid line, top left) and smoothed median periodogram (solid, top right) behave similarly and have their largest peaks at frequencies 0.045 and 0.039, respectively, which translates into an estimated business cycle length of 11.58 years when measured by the smoothed periodogram and 13.39 years when measured by the smoothed median periodogram. Both lines also have peaks of similar size at the yearly ($2\pi/12 \approx 0.52$) and half-yearly ($2\pi/6 \approx 1.05$) frequencies, which provides evidence of considerable seasonality in the data. However, as illustrated by the smoothed 0.10-th quantile periodogram (solid, bottom left) and 0.90-th quantile periodogram (solid, bottom right), these seasonal cycles do not appear uniformly across the distribution of the data. At the 0.90-th quantile, the yearly and—to some extent—the half-yearly cycles are still present, but at the 0.10-th quantile, this seasonality vanishes completely; the smoothed 0.10-th quantile periodogram also has some smaller peaks between 0.2 and 0.5, but comparison
Figure 1.9: Top left panel: QS-smoothed periodogram (solid) of building permit data with $B_n = 28n^{1/5} \approx 101.76$ to prevent smaller peaks from being smoothed out, chi-squared point-wise 95% confidence bands (shaded grey, not shown completely to prevent clutter) with $k = 4$, and $\hat{\gamma}_{n,X}(0)/(2\pi)$ (dashed). Other panels: QS-smoothed $\tau$-th quantile periodogram (solid), chi-squared point-wise 95% confidence bands (shaded grey), and $\tau(1 - \tau)/(2\pi)$ (dashed) for $\tau = 0.5$ (top right), 0.10 (bottom left), and 0.90 (bottom right). All panels use the same data, $B_n$, and $k$. The top left panel is normalized by $\hat{\gamma}_{n,X}(0)/(2\pi)$. The other panels are normalized by $\tau(1 - \tau)/(2\pi)$.
to the confidence intervals (shaded grey) shows that these peaks are not significantly different from a straight line. All graphs have in common, however, that the business cycle explains most of the cyclical variation, which indicates that seasonal patterns disappear during economic troughs.

1.6 Conclusion

In this chapter I introduced quantile spectral densities that summarize the cyclical behavior of time series across their whole distribution by analyzing periodicities in quantile crossings. I discussed robust spectral estimation and inference in situations where the dependence structure of a time series is not accurately captured by the auto-covariance function, in particular when the time series under consideration is uncorrelated or heavy-tailed. I established the statistical properties of quantile spectral estimators in a large class of nonlinear time series models and discussed inference both at fixed and across all frequencies. Monte Carlo experiments and an empirical example showed that quantile spectral estimates are similar to regular spectral density estimates in both shape and interpretation when standard conditions are satisfied, but can still reliably identify dependence structures when these conditions fail to hold.

1.7 Proofs

Throughout this section, \( \mathbb{P}^* \) and \( \mathbb{E}^* \) respectively denote outer probability and outer expectation (see, e.g., van der Vaart, 1998, p. 258). Probability and expectation conditional on the observed sample \( S_n \) is abbreviated by \( \hat{\mathbb{P}}(\cdot) := \mathbb{P}(\cdot \mid S_n) \) and \( \hat{\mathbb{E}}(\cdot) := \mathbb{E}(\cdot \mid S_n) \).

Proof of Proposition 1.4. By assumption, we can find a \( \delta' > 0 \) such that \( F_X \) is Lipschitz on \( (\xi_0(\tau) - \delta', \xi_0(\tau) + \delta') \). Choose a large enough \( N \in \mathbb{N} \) such that \( \delta := \delta' - \theta^{N/(1+\alpha)} > 0 \) and pick any \( \xi \in X_r(\delta) \); then, for all \( n \geq N \), apply the pointwise bound

\[
|1\{X_n < \xi\} - 1\{X'_n < \xi\}| \leq 1\{|X_n - \xi| < |X_n - X'_n|\},
\]

39
the Markov inequality, and the GMC property to see that

\[
\|1\{X_n < \xi\} - 1\{X'_n < \xi\}\|^2 \\
\leq \mathbb{P}(|X_n - \xi| < |X_n - X'_n|) \\
\leq \mathbb{P}(|X_n - \xi| < \varrho^{na/(1+\alpha)}) + \mathbb{E} |X_n - X'_n|^\alpha \varrho^{-na^2/(1+\alpha)} \\
\leq M \varrho^{na/(1+\alpha)}
\]

for a large enough absolute constant \(M\). This constant can be enlarged slightly to ensure that the inequality also holds for the remaining \(n < N\).

With \(\sigma := \varrho^{\alpha/(2+2\alpha)}\), take square-roots on both sides and suprema over \(X_\tau(\delta)\) to establish the desired result.

**Proof of Theorem 1.6.** Let \(Z_{t,n} = (X_t, t\lambda_n)\) and define the maps \(h_{\xi}(Z_{t,n}) = 1\{X_t < \xi\} \times \cos(t\lambda_n)\) and \(h^*_\xi(Z_{t,n}) = 1\{X_t < \xi\} \times \sin(t\lambda_n)\). The empirical process evaluated at some function \(h\) is denoted by

\[
\nu_n h := n^{-1/2} n \sum_{t=1}^n (h(Z_{t,n}) - \mathbb{E} h(Z_{t,n})).
\]

The finite Fourier transform at nonzero natural frequencies is invariant to centering. Hence, from (1.3), we can decompose \(Q_{n,\tau}(\lambda_n)\) into

\[
\frac{1}{2\pi} |\nu_n (h_{\xi_n}(\tau) - h_{\xi_0}(\tau)) + i \nu_n (h^*_\xi(\tau) - h^*_0(\tau)) - n^{-1/2} \sum_{t=1}^n V_t(\tau) e^{-it\lambda_n}|^2
\]

(1.11)

For the proof of the theorem, I proceed in three steps: I show that (i) the first term and (ii) the second term inside the modulus in the display are small in probability and that (iii) the remainder of (1.11) has the desired asymptotic distribution jointly for frequencies \(\lambda_n + 2\pi j/n\) with \(|j| \leq k\).

**Step (i):** Define a norm \(\rho(h_\xi) = \sup_{t,n \in \mathbb{N}} \|h_\xi(Z_{t,n})\|\). Take a grid of points \(\xi_0(\tau) - \delta =: \xi_0 < \xi_1 < \cdots < \xi_N := \xi_0(\tau) + \delta\) and let \(b_k(Z_{t,n}) := (h_{\xi_k}(Z_{t,n}) - h_{\xi_{k-1}}(Z_{t,n}))/\cos(t\lambda_n)\). Given a \(\xi \in X_\tau(\delta)\), we can then find an index \(k\) such that \(|h_\xi - h_{\xi_{k-1}}| \leq b_k\). In addition, we have

\[
\rho(b_k) = \|1\{X_0 < \xi_k\} - 1\{X_0 < \xi_{k-1}\}\| \leq \sqrt{F_X(\xi_k) - F_X(\xi_{k-1})},
\]

which is bounded above by a constant multiple of \(\sqrt{\xi_k - \xi_{k-1}}\) due to Lipschitz
continuity. Hence, if we choose the grid such that \( \rho(b_k) \leq \epsilon \) for all \( k = 1, \ldots, N \), the parametric class \( H := \{ h_\xi : \xi \in \mathcal{X}(\delta) \} \) has bracketing numbers (see Andrews and Pollard, 1994; van der Vaart, 1998, pp. 270-271) with respect to \( \rho \) of order \( N(\epsilon, H) = O(\epsilon^{-2}) \) as \( \epsilon \to 0 \).

By the same calculations as in the preceding display, there is some \( M > 0 \) such that all \( \xi, \xi' \in \mathcal{X}(\delta) \) satisfy \( \rho(h_\xi - h_{\xi'}) \leq M|\xi - \xi'|^{1/2} \) and therefore \( \rho(h_{\hat{\xi}_n(\tau)} - h_{\xi_0(\tau)}) \to_p 0 \) in view of Lemma 1.26 below. For \( \epsilon, \eta > 0 \), the limit superior of \( P(|\nu_n(h_{\hat{\xi}_n(\tau)} - h_{\xi_0(\tau)})| \geq \epsilon, \rho(h_{\hat{\xi}_n(\tau)} - h_{\xi_0(\tau)}) \leq \eta) \) is then at most

\[
\limsup_{n \to \infty} P\left( |\nu_n(h_{\hat{\xi}_n(\tau)} - h_{\xi_0(\tau)})| \geq \epsilon, \rho(h_{\hat{\xi}_n(\tau)} - h_{\xi_0(\tau)}) \leq \eta \right) \leq \limsup_{n \to \infty} \mathbb{P}\left( \sup_{\xi \in \mathcal{X}(\delta) : \rho(h_\xi - h_{\xi_0(\tau)}) \leq \eta} |\nu_n(h_\xi - h_{\xi_0(\tau)})| \geq \epsilon \right) \tag{1.12}
\]

The Markov inequality and Lemma 1.28 below imply that the term on the right can be made as small as desired by choosing \( \eta \) small enough. This is also true for the frequencies \( \lambda_n + 2\pi j/n \) with \( |j| \leq k \).

**Step (ii):** Replace cosines with sines in the proofs of Lemmas 1.27 and 1.28 (with the same bounding functions \( b_k \) as above) to reach the same conclusion for \( \nu_n(h_{\hat{\xi}_n(\tau)} - h_{\xi_0(\tau)}) \).

**Step (iii):** In view of (i), (ii), and continuity of the modulus, I only have to show that the remainder of (1.11) converges jointly at each \( \lambda_n + 2\pi j/n, |j| \leq k \), in distribution to independent exponential variables with mean \( g_\tau(\lambda) \). For this I apply Corollary 2.1 of Shao and Wu (2007). Because \( V_t(\tau) \) is a bounded mean-zero variable, the only condition that has to be checked is

\[
\sum_{t=0}^{\infty} \| \mathbb{E}(V_t(\tau) \mid \mathcal{F}_0) - \mathbb{E}(V_t(\tau) \mid \mathcal{F}_{-1}) \| < \infty.
\]

By the conditional Jensen inequality, the law of iterated expectations, and Assumption 1.9, this summability condition is satisfied because

\[
\| \mathbb{E}(V_t(\tau) \mid \mathcal{F}_0) - \mathbb{E}(V_t(\tau) \mid \mathcal{F}_{-1}) \| \\
= \| \mathbb{E}(1\{X_t < \xi_0(\tau)\} - 1\{X_t^* < \xi_0(\tau)\} \mid \mathcal{F}_0) \| \\
\leq \| 1\{X_t < \xi_0(\tau)\} - 1\{X_t^* < \xi_0(\tau)\} \|.
\]

Assumption 1.3 implies Assumption 1.9, and so the joint convergence asserted.
in Theorem 1.6 follows.

Lemma 1.26. Suppose Assumptions 1.5 and 1.9 hold; then \( \hat{\xi}_n(\tau) - \xi_0(\tau) = O_p(n^{-1/2}) \).

Proof of Lemma 1.26. Arguing as in the proof of Theorem 1 of Wu (2005), use the conditional Jensen inequality and the law of iterated expectations to deduce that

\[
\| \mathbb{E}(1\{X_t < \xi\} | \mathcal{F}_0) - \mathbb{E}(1\{X^*_t < \xi\} | \mathcal{F}_{-1}, \varepsilon^*_0) \| \\
= \| \mathbb{E}(1\{X_t < \xi\} - 1\{X^*_t < \xi\} | \mathcal{F}_0, \varepsilon^*_0) \| \\
\leq \| 1\{X_t < \xi\} - 1\{X^*_t < \xi\} \|.
\]

Taking suprema over \( X_t(\delta) \) shows that Assumption 1.9 implies condition (7) of Wu (2007) and his Theorem 1 then yields the desired result.

For Lemmas 1.27 and 1.28, I mimic the proofs of Andrews and Pollard’s (1994) Theorem 2.2 and Lemma 3.1; their arguments do not apply directly since Andrews and Pollard work with strongly mixing arrays.

Lemma 1.27. Let \( \phi(h_\xi - h_{\xi'}) := \rho(h_\xi - h_{\xi'})^{2/(2+\gamma)} \) for some \( \gamma > 0 \) and suppose that Assumption 1.3 holds. Then, for all \( n \in \mathbb{N} \), all \( \xi, \xi' \in X_\tau(\delta) \), and every even integer \( Q \geq 2 \),

\[
\mathbb{E} | \nu_n(h_\xi - h_{\xi'}) |^Q \leq n^{-Q/2} C((\phi(h_\xi - h_{\xi'})^2n) + \cdots + (\phi(h_\xi - h_{\xi'})^2n)^{Q/2}),
\]

where \( C \) depends only on \( Q, \gamma, \) and \( \sigma \). The inequality remains valid when \( h_\xi - h_{\xi'} \) is replaced by \( b_k \) for any given \( k \geq 1 \).

Proof of Lemma 1.27. It suffices to show the inequality given in the lemma after dividing both sides by \( 4^Q \) to ensure that the absolute value of

\[
H_t := (h_\xi(Z_{t,n}) - h_{\xi'}(Z_{t,n}) - (\mathbb{E} h_\xi(Z_{t,n}) - \mathbb{E} h_{\xi'}(Z_{t,n}))) / 4
\]

is bounded by 1. The \( 4^{-Q} \) on the right hand can be absorbed into \( C \). Define \( H'_t \) in the same way as \( H_t \) but replace \( X_t \) with \( X'_t \). Here I suppress the dependence of \( H_t \) and \( H'_t \) on \( n, \xi, \) and \( \xi' \) because they are irrelevant in the following. Also note that \( \mathbb{E} H_t = \mathbb{E} H'_t = 0 \) for all \( t, n \in \mathbb{N} \) and all \( \xi, \xi' \in X_\tau(\delta) \) because \( X_t \) and \( X'_t \) are identically distributed.
For fixed $k \geq 2$, $d \geq 1$, and $1 \leq m < k$, consider integers $t_1 \leq \cdots \leq t_m \leq t_{m+1} \leq \cdots \leq t_k$ so that $t_{m+1} - t_m = d$ and define $a_k(\lambda_n) = |2^{-k} \prod_{i=1}^{k} \cos(t_i \lambda_n)|$. Since $U_t := H_t/\cos(t \lambda_n)$ and $U'_t := H'_t/\cos(t \lambda_n)$ are stationary, repeatedly add and subtract to see that

$$\left| \mathbb{E} H_{t_1} H_{t_2} \cdots H_{t_k} - \mathbb{E} H_{t_1} H_{t_2} \cdots H_{t_m} \mathbb{E} H_{t_{m+1}} \cdots H_{t_k} \right|$$

$$= a_k(\lambda_n) \left| \mathbb{E} U_{t_1} - t_m U_{t_2} - t_m \cdots U_{t_k} - t_m - \mathbb{E} U_{t_1} - t_m U_{t_2} - t_m \cdots U_{t_m} - t_m \right|$$

$$\leq a_k(\lambda_n) \left| \mathbb{E} U_{t_1 - t_m} \cdots U_{t_m} - U_0 \mathbb{E} U_{t_m+2} - t_m \cdots U_{t_k - t_m} \right|$$

$$+ \sum_{i=2}^{k-m-1} a_k(\lambda_n) \left| \mathbb{E} U_{t_1 - t_m} \cdots U_0 U'_{t_m} \cdots (U_{t_{m+i} - t_m} - U'_{t_{m+i} - t_m}) \cdots U_{t_k - t_m} \right|$$

$$+ a_k(\lambda_n) \left| \mathbb{E} U_{t_1 - t_m} \cdots U_0 U'_{t_m} \cdots U'_{t_k - t_m} - \mathbb{E} U_{t_1 - t_m} \cdots U_0 \mathbb{E} U_{t_m} \cdots U_{t_k - t_m} \right|$$

(1.13)

In particular, the last term on the right-hand side is zero because $U_{t_1 - t_m} \cdots U_0$ and $U'_{t_m} \cdots U'_{t_k - t_m}$ are independent and $U'_{t_m} \cdots U'_{t_k - t_m}$ and $U_0 \cdots U_{t_k - t_m}$ are identically distributed.

By Assumption 1.3, $\|U_d - U'_d\|_s \leq \|1\{X_d < \xi\} - 1\{X'_d < \xi\}\|_s + \|1\{X_d < \xi'\} - 1\{X'_d < \xi'\}\|_s \leq 2 \sup_{\xi \in X_d(\delta)} \|1\{X_d < \xi\} - 1\{X'_d < \xi\}\|_s \leq C' \sigma^d$ for some $C' > 0$ and $s \geq 1$. Here the choice of $s$ does not matter because Assumption 1.3 still applies when $\| \cdot \|$ is replaced by $\| \cdot \|_s$ for any $s > 0$; see Lemma 2 of Wu and Min (2005). Hölder’s inequality then bounds the first term on the right-hand side of the preceding display by

$$\|H_{t_1} \cdots H_{t_m}\|_p \|H_{t_{m+2}} \cdots H_{t_k}\|_q C' \sigma^d,$$

(1.14)

where the reciprocals of $p$, $q$, and $s$ sum to 1. Proceeding as in Andrews and Pollard (1994), another application of the Hölder inequality yields

$$\|H_{t_1} \cdots H_{t_m}\|_p \leq \left( \prod_{i=1}^{m} \mathbb{E} |H_{t_i}|^{mp} \right)^{1/(mp)} \leq \phi(h_\xi - h_\xi')(2+\gamma)/p$$

whenever $mp \geq 2$ and similarly $\|H_{t_{m+2}} \cdots H_{t_k}\|_q \leq \phi(h_\xi - h_\xi')(2+\gamma)/q$ whenever $(k - m - 1)q \geq 2$. Suppose for now that $k \geq 3$. If $k > m + 1$, take $s = (\gamma + Q)/\gamma$ and $mp = (k - m - 1)q = (k - 1)/(1 - 1/s)$. Decrease the resulting exponent of $\phi(h_\xi - h_\xi')$ from $Q(2 + \gamma)/(Q + \gamma)$ to 2 to see that (1.14) is bounded by $C' \sigma^d \phi(h_\xi - h_\xi')^2$. If $k \geq 2$ and $k = m + 1$,
the factor $\|H_{m+1} \cdots H_k\|_q$ is not present in (1.14), but we can still choose $s = (\gamma + Q)/\gamma$ and $mp = (k - 1)/(1 - 1/s)$ to obtain the same bound. Since the same argument also applies to each of the other summands in (1.13), we can find a constant $M > 0$ such that

$$|\mathbb{E} H_{t_1} H_{t_2} \cdots H_{t_k}| \leq |\mathbb{E} H_{t_1} H_{t_2} \cdots H_{t_n} \mathbb{E} H_{t_{m+1}} \cdots H_{t_k}| + M\sigma^d \phi(h_{\xi} - h_{\xi'})^2.$$ 

Here $M$ in fact depends on $k$, but this does not disturb any of the subsequent steps.

Now replace (A.2) in Andrews and Pollard (1994) by the inequality in the preceding display. In particular, replace their $8\alpha(d)^{1/s}$ with $M\sigma^d$ and their $\tau^2$ with $\phi(h_{\xi} - h_{\xi'})^2$. The rest of their arguments now go through without changes.

The inequality for $b_k$ follows by letting $\lambda_n \equiv 0$; this is not a contradiction to the assumptions of Theorem 1.6 because this proof is valid for any sequence $(\lambda_n)_{n \in \mathbb{N}}$.

**Lemma 1.28.** Suppose the assumptions of Theorem 1.6 hold. For every $\epsilon > 0$ and every even integer $Q \geq 4$, there is an $\eta > 0$ such that

$$\lim_{n \to \infty} \mathbb{E}^* \left( \sup_{\xi, \xi' \in \mathcal{X} \cap (\delta); \rho(h_{\xi} - h_{\xi'}) \leq \eta} |\nu_n(h_{\xi} - h_{\xi'})| \right)^Q \leq \epsilon.$$ 

**Proof of Lemma 1.28.** I follow Andrews and Pollard’s (1994) proof of their Theorem 2.1. It requires three steps: (i) Their “Proof of inequality (3.2),” (ii) their “Proof of inequality (3.3),” and (iii) their “Comparison of pairs” argument. Replace their $i$ with $k$ and their $\tau(h_i)$ with $\phi(b_k)$; then apply Lemma 1.27 above instead of Andrews and Pollard’s (1994) Lemma 3.1 in the derivation of their inequality (3.5) to deduce

$$\left\| \max_{1 \leq k \leq N} |\nu_n b_k| \right\|_Q \leq C'' N^{1/Q} \max \left\{ n^{-1/2}, \max_{1 \leq k \leq N} \phi(b_k) \right\}$$

and use this inequality in (i) instead of their inequality (3.5). In (i) Andrews and Pollard also require the finiteness of the bracketing integral

$$\int_0^1 x^{-\gamma/(2+\gamma)} N(x, \mathcal{H})^{1/Q} dx,$$
which follows immediately by choosing $\gamma = Q - 4$. Another application of Lemma 1.27 establishes the required analogue of Andrews and Pollard’s inequality (3.5) used in (ii). The same inequality can also be applied in (iii). The other arguments remain valid without changes. \hfill \Box

**Proof of Theorem 1.11.** Denote by $\tilde{r}_{n,\tau}(\cdot) = n^{-1} \sum_{t=|j|-1}^n V^\tau_t V_{t-[j]}(\tau)$ and $$\tilde{g}_{n,\tau}(\lambda) := \frac{1}{2\pi} \sum_{|j|<n} w(j/B_n) \tilde{r}_{n,\tau}(j)e^{-ij\lambda}$$

the infeasible sample auto-covariance and smoothed quantile spectrum, respectively, based on the unknown quantile $\xi_0(\tau)$. The triangle inequality and $|V^\tau_t(\cdot, \cdot)| < 1$ yield

$$2\pi \sup_{\lambda \in (-\pi,\pi]} |\tilde{g}_{n,\tau}(\lambda) - \tilde{g}_{n,\tau}(\lambda)|$$

$$\leq \frac{1}{n} \sum_{|j|<n} |w(j/B_n)| \sum_{t=|j|-1}^n |\hat{V}^\tau_t - V^\tau_t| |V_{t-[j]}(\tau)|$$

$$\leq \frac{1}{n} \sum_{|j|<n} |w(j/B_n)| \sum_{t=|j|-1}^n \left( |\hat{V}^\tau_t - V^\tau_t| + |\hat{V}_{t-[j]}(\tau) - V_{t-[j]}(\tau)| \right)$$

$$\leq \frac{1}{n} \sum_{|j|<n} |w(j/B_n)| \sum_{t=|j|-1}^n \left( 1_{\{X_t - \xi_0(\tau) < |\xi_0(\tau) - \xi_0(\tau)|\}} + 1_{\{X_{t-[j]} - \xi_0(\tau) < |\xi_0(\tau) - \xi_0(\tau)|\}} \right).$$

Consider the first indicator function on the right-hand side of the preceding display and recall that $\sqrt{n}(|\hat{\xi}_n(\tau) - \xi_0(\tau)|)$ is uniformly tight by Lemma 1.26. For any given $\epsilon > 0$ and $\eta > 0$, the Markov inequality implies for large enough $M > 0$

$$\limsup_{n \to \infty} \mathbb{P}\left( \frac{1}{n} \sum_{|j|<n} |w(j/B_n)| \sum_{t=|j|-1}^n 1_{\{X_t - \xi_0(\tau) < |\xi_0(\tau) - \xi_0(\tau)|\}} \geq \eta \right)$$

$$\leq \limsup_{n \to \infty} \mathbb{P}\left( \frac{1}{n} \sum_{|j|<n} |w(j/B_n)| \sum_{t=|j|-1}^n 1_{\{X_t - \xi_0(\tau) < M^{-1/2}\}} \geq \eta \right)$$

$$+ \sup_{n \in \mathbb{N}} \mathbb{P}(|\hat{\xi}_n(\tau) - \xi_0(\tau)| \geq M^{-1/2})$$

$$\leq \limsup_{n \to \infty} \frac{1}{\eta} \sum_{|j|<n} |w(j/B_n)| \mathbb{P}(|X_0 - \xi_0(\tau)| < M^{-1/2}) + \epsilon.$$
By Lemma 1 of Jansson (2002), the limit superior of $B_n^{-1} \sum_{|j| < n} |w(j/B_n)|$ is finite, and in view of the assumed Lipschitz continuity, the first term on the right-hand side of the preceding display then vanishes because $B_n n^{-1/2} \to 0$. The same argument applies to the second indicator function above due to stationarity. Together this yields $\sup_{\lambda} |\hat{g}_{n,\tau}(\lambda) - \tilde{g}_{n,\tau}(\lambda)| \to_p 0$.

To show $\tilde{g}_{n,\tau}(\lambda) \to_p g_{\tau}(\lambda)$ uniformly in $\lambda$, I use Liu and Wu’s (2010) Theorem 1, which applies whenever the windows $w \in \mathcal{W}$ satisfy their Condition 1. The only two conditions that need to be established are the absolute integrability of $w$, which is immediate from $\int_{-\infty}^{\infty} |w(x)| \, dx \leq 2 \int_0^{\infty} \bar{w}(x) \, dx < \infty$, and

$$\limsup_{n \to \infty} B_n^{-1} \sum_{j \in \mathbb{Z}} w(j/B_n)^2 < \infty.$$  

Although Liu and Wu (2010) provide a specific value for the limit in the preceding display, its boundedness is in fact all that is needed for the proof of their Theorem 1. To this end, take $M \geq \sup_{x \in \mathbb{R}} |w(x)|$ such that for $j \geq 1$

$$w(j/B_n)^2 \leq M |w(j/B_n)| \leq M \bar{w}(j/B_n) \leq M B_n \int_{(j-1)/B_n}^{j/B_n} \bar{w}(x) \, dx$$

by monotonicity, and therefore symmetry implies

$$B_n^{-1} \sum_{j \in \mathbb{Z}} w(j/B_n)^2 \leq B_n^{-1} + 2M \sum_{j=1}^{\infty} \int_{(j-1)/B_n}^{j/B_n} \bar{w}(x) \, dx$$

$$= B_n^{-1} + 2M \int_0^{\infty} \bar{w}(x) \, dx,$$

which is finite by assumption. This is also true for its limit superior as $n \to \infty$. The triangle inequality completes the proof.

Proof of Theorem 1.12. (i) The process $S_{n,\tau}(\lambda)$ can be decomposed into

$$\sqrt{n} \sum_{j=1}^{n-1} (\hat{r}_{n,\tau}(j) - \tilde{r}_{n,\tau}(j)) \psi_j(\lambda) + \sqrt{n} \sum_{j=1}^{n-1} \hat{r}_{n,\tau}(j) \psi_j(\lambda).$$

The second term side converges weakly in $L_2(\Pi)$ to $S_{\tau}(\lambda)$ by the proof of Theorem 2.1 of Shao (2011a). The Continuous Mapping Theorem then yields $CM_{n,\tau} \to \|S_{\tau}\|_{\Pi}^2$ as long as the $L_2(\Pi)$-norm of the first term of the display is eventually small in probability.
To this end, define $\varphi_j = \|\psi_j\|_\Pi$ and note that $\langle \psi_j, \psi_k \rangle = 0$ for $j \neq k$. Use this orthogonality to write
\[
\left\| \sqrt{n} \sum_{j=1}^{n-1} (\hat{r}_{n,\tau}(j) - \tilde{r}_{n,\tau}(j)) \psi_j \right\|_\Pi^2 = n \sum_{j=1}^{n-1} (\hat{r}_{n,\tau}(j) - \tilde{r}_{n,\tau}(j))^2 \varphi_j^2.
\]

Let $r_{n,\tau}(j, \xi) = (\tau - F_X(\xi))^2/(n-j)/n$ for $j > 0$ and $\xi \in \mathbb{R}$. Under the null hypothesis, we have $r_\tau(j) = 0 = r_{n,\tau}(j, \xi_0(\tau))$ for all $j > 0$ and, by Lipschitz continuity, there exists $M > 0$ such that
\[
\hat{r}_{n,\tau}(j) := r_{n,\tau}(j, \hat{\xi}_n(\tau)) \leq |F_X(\hat{\xi}_n(\tau)) - \tau|^2 \leq M|\hat{\xi}_n(\tau) - \xi_0(\tau)|^2
\]
for all $j > 0$. In view of these properties, fix some $K \leq n$ and apply the Loève $c_r$ inequality to bound the quantity in the preceding display by
\[
2n \sum_{j=1}^{K-1} \left( \hat{r}_{n,\tau}(j) - \tilde{r}_{n,\tau}(j) - (\bar{r}_{n,\tau}(j) - r_\tau(j)) \right)^2 \varphi_j^2 + 2n \sum_{j=1}^{K-1} \tilde{r}_{n,\tau}(j)^2 \varphi_j^2 + n \sum_{j=K}^{n-1} (\hat{r}_{n,\tau}(j) - \tilde{r}_{n,\tau}(j))^2 \varphi_j^2.	ag{1.15}
\]

Fix $\epsilon, \epsilon' > 0$ and let $e_{\xi, j}(X_t, X_{t+j}) := V_\tau(\tau, \xi) V_{t+j}(\tau, \xi)$. For given $j > 0$, take $\rho(e_{\xi, j} - e_{\xi', j})$ as the distance of $\xi$ and $\xi'$ on $\mathcal{X}_\tau(\delta)$, where $\rho$ is as in the proof of Theorem 1.6. The distance also depends on $\tau$, but this is irrelevant in the following. Note that $\rho(e_{\xi, j} - e_{\xi', j}) \leq 2\|1\{X_0 < \xi\} - 1\{X_0 < \xi'\}\| \leq M'|\xi - \xi'|$ uniformly in $j > 0$ for some $M' > 0$ by stationarity. Hence, for any $\eta_j > 0$, $1 \leq j < K$, we have
\[
\mathbb{P} \left\{ \rho(e_{\xi_n(\tau), j} - e_{\xi_0(\tau), j}) > \eta_j \right\} \leq \mathbb{P} \left( |\hat{\xi}_n(\tau) - \xi_0(\tau)|^{1/2} > \min_{1 \leq j < K} \eta_j / M' \right) = o(1).
\]

Under the null hypothesis, we can write $\sqrt{n}(\hat{r}_{n,\tau}(j) - r_{\tau}(j)) = (n-j)/n \times \nu_{n-j} e_{\xi_0(\tau)}$ and, by (1.8), $\sqrt{n}(\hat{r}_{n,\tau}(j) - \tilde{r}_{n,\tau}(j)) = (n-j)/n \nu_{n-j} e_{\hat{\xi}_n(\tau)}$ as long as $\hat{\xi}_n(\tau) \in \mathcal{X}_\tau(\delta)$, where I use the notation from the proof of Theorem
1.6. In view of the preceding display, (1.15) then satisfies

\[
\limsup_{n \to \infty} \mathbb{P}\left( n \sum_{j=1}^{K-1} \left( \hat{r}_{n,\tau}(j) - \tilde{r}_{n,\tau}(j) - (\hat{r}_{n,\tau}(j) - r_{\tau}(j)) \right)^2 \varphi_j^2 \geq \epsilon/2 \right) \\
\leq \limsup_{n \to \infty} \mathbb{P}\left( \sum_{j=1}^{K-1} \left( \sup_{\xi \in \mathcal{X}_{\tau}(\delta)} |\nu_{n-j}(e_{\xi,j} - e_{\xi'(\tau),j})| \right)^2 \varphi_j^2 \geq \epsilon/2 \right) \\
\leq \frac{2}{\epsilon} \sum_{j=1}^{K-1} \limsup_{n \to \infty} \mathbb{E}\left( \sup_{\xi \in \mathcal{X}_{\tau}(\delta)} |\nu_{n-j}(e_{\xi,j} - e_{\xi'(\tau),j})| \right)^2 \varphi_j^2 \\
\leq \frac{2}{\epsilon} \sum_{j=1}^{K-1} \left( \limsup_{n \to \infty} \mathbb{E}\left( \sup_{\xi \in \mathcal{X}_{\tau}(\delta)} \left| \nu_{n-j}(e_{\xi,j} - e_{\xi'(\tau),j}) \right| \right)^Q \right) \varphi_j^2 \\
\leq \frac{2}{\epsilon} \sum_{j=1}^{K-1} \varphi_j^2 \leq \frac{2\epsilon'}{\epsilon} \sum_{j>0} \varphi_j^2 = \left( \frac{\pi}{6\epsilon} \right)^2 \epsilon',
\]

where the first inequality is the Markov inequality, the second follows from Lemma 1.29 below, the third is Jensen’s, and the equality uses \( \varphi_j^2 = 1/(2\pi j^2) \) for \( j > 0 \).

Now consider (1.16). The first term can be bounded by

\[
2nM^2 |\hat{\xi}_n(\tau) - \xi_0(\tau)|^4 \sum_{j>0} \varphi_j^2 = O_p(n^{-1}) = o_p(1).
\]

By arguments as in the proof of Theorem 1.11, for a large enough \( M' \) the probability that the second term of (1.16) exceeds \( \epsilon \) is at most

\[
\mathbb{P}\left( n \sum_{j=K}^{n-1} \left( \hat{r}_{n,\tau}(j) - \tilde{r}_{n,\tau}(j) \right)^2 \varphi_j^2 \geq \epsilon, |\hat{\xi}_n(\tau) - \xi_0(\tau)| \leq M'n^{-1/2} \right) + \epsilon' \\
\leq \mathbb{P}\left( n^{-1} \sum_{j=K}^{n-1} \sum_{t=j+1}^{n} 1_{\{|X_t-\xi_0(\tau)|>|\hat{\xi}_n(\tau)-\xi_0(\tau)|\}} \right) \varphi_j^2 \geq \epsilon, |\hat{\xi}_n(\tau) - \xi_0(\tau)| \leq M'n^{-1/2} \right) + \epsilon' \\
\leq \mathbb{P}\left( n^{-1} \sum_{j=K}^{n-1} \sum_{t=j+1}^{n} 1_{\{|X_t-\xi_0(\tau)|\leq M'n^{-1/2}\}} \right) \varphi_j^2 \geq \epsilon/2 \right) + \epsilon'
\]
\[
\leq \frac{4}{\epsilon} \mathbb{E} \left( n^{-1/2} \sum_{t=1}^{\infty} \mathbb{1}_{[\|X_t - \xi_0(\tau)\| \leq M' n^{-1/2}]} \right)^2 \sum_{j \geq K} \varphi_j^2 + \epsilon' \\
\leq \frac{4}{\epsilon} \left( \mathbb{P}(\|X_0 - \xi_0(\tau)\| \leq M' n^{-1/2}) + n \mathbb{P}(\|X_0 - \xi_0(\tau)\| \leq M' n^{-1/2})^2 \right) \sum_{j \geq K} \varphi_j^2 + \epsilon' \\
= O(1) \sum_{j \geq K} \varphi_j^2 + \epsilon',
\]

which can be made smaller than \(2\epsilon'\) by choosing \(K\) large enough. This does not affect any of the other bounds. Since \(\epsilon'\) was arbitrary, we have \(CM_{n,\tau} = \tilde{CM}_{n,\tau} + o_p(1)\) and \(CM_{n,\tau} \sim \|S_\tau\|_1^2\), which proves the first result.

(ii) Fix some \(K \leq n\) and decompose the statistic into

\[
CM_n(\tau)/n = \sum_{j=1}^{K-1} (\hat{r}_{n,\tau}(j) - \bar{r}_{n,\tau}(j))^2 \varphi_j^2 + \sum_{j=1}^{K-1} (\hat{r}_{n,\tau}(j) - \bar{r}_{n,\tau}(j)) \bar{r}_{n,\tau}(j) \varphi_j^2 \\
+ \sum_{j=1}^{K-1} \bar{r}_{n,\tau}(j)^2 \varphi_j^2 + \sum_{j=K}^{n-1} \bar{r}_{n,\tau}(j)^2 \varphi_j^2.
\]

The first and second terms on the right-hand side of the displayed equation converge to zero in probability as \(n \to \infty\) because \(\hat{r}_{n,\tau}(j) - \bar{r}_{n,\tau}(j) = o_p(1)\) for each \(j\) under the assumptions of the theorem. The third term converges in probability to \(\sum_{j=1}^{K-1} r_\tau(j)^2 \varphi_j^2\) as \(n \to \infty\) by Wu’s (2005) Theorem 2(i). The absolute value of the last term is bounded by \(\sum_{j \geq K} \varphi_j^2\), where I have used the fact that \(|\hat{r}_{n,\tau}| \leq 1\). Hence, let \(K \to \infty\) to conclude \(CM_n(\tau)/n \to_p \sum_{j \geq 0} r_\tau(j)^2 \varphi_j^2 > 0\). The desired result now follows from a routine argument.

\[
\square
\]

**Lemma 1.29.** Suppose the assumptions of Theorem 1.12 hold. For every \(j > 0\), every \(Q\), and every \(\epsilon > 0\), there is an \(\eta > 0\) such that

\[
\limsup_{n \to \infty} \mathbb{E}^*(\sup_{\xi, \xi' \in \mathcal{X}_{\tau}(\delta) : \xi \neq \xi'} \sup_{\nu_\tau \leq \eta} \nu_{n-j}(\xi, \xi - \xi'))^Q \leq \epsilon.
\]

**Proof of Lemma 1.29.** Take a grid of points \(\xi_0(\tau) - \delta = \xi_0 < \xi_1 < \cdots < \xi_N = \xi_0(\tau) + \delta\) and let \(b_{k,j}(X_t, X_{t+j}) := 1\{X_t < \xi_k\} - 1\{X_t < \xi_{k-1}\} + 1\{X_{t+j} < \xi_k\} - 1\{X_{t+j} < \xi_{k-1}\}\). Given a \(\xi \in \mathcal{X}_{\tau}(\delta)\), we can then find an index \(k\) such
that \(|e_{\xi,j} - e_{\xi_{k-1},j}| \leq b_{k,j}\). Further,

\[
\rho(b_{k,j}) \leq 2\|1\{X_0 < \xi_k\} - 1\{X_0 < \xi_{k-1}\}\| \leq 2\sqrt{F_X(\xi_k) - F_X(\xi_{k-1})},
\]

which is proportional to \(\sqrt{\xi_k - \xi_{k-1}}\) due to Lipschitz continuity. If \(\rho(b_{k,j}) \leq \epsilon\) for all \(k = 1, \ldots, N\) then, as above, for each \(j\) the parametric class \(\mathcal{E}_j := \{e_{\xi,j} : \xi \in \mathcal{X}_\tau(\delta)\}\) has bracketing numbers with respect to \(\rho\) of order \(N(\epsilon, \mathcal{E}_j) = O(\epsilon^{-2})\) as \(\epsilon \to 0\). Hence, bracketing integrals of the class \(\mathcal{H}\) above and the classes \(\mathcal{E}_j\) have the same behavior. The proof of Lemma 1.28 therefore also applies to this lemma as long as the reference to Lemma 1.27 is replaced by Lemma 1.30 below.

\[\boxdot\]

**Lemma 1.30.** Fix some \(\gamma > 0\) and suppose that Assumption 1.3 holds. For all \(n \in \mathbb{N}\), all \(j < n\), all \(\xi, \xi' \in \mathcal{X}_\tau(\delta)\), and every even integer \(Q \geq 2\) we have

\[
\mathbb{E}|\nu_{n-j}(e_{\xi,j} - e_{\xi',j})|^Q \leq (n-j)^{-Q/2}C((\phi(e_{\xi,j} - e_{\xi',j})^2(n-j)) + \cdots + (\phi(e_{\xi,j} - e_{\xi',j})^2(n-j))^{Q/2}),
\]

where \(C\) depends only on \(j, Q, \gamma, \) and \(\sigma\). The inequality remains valid when \(e_{\xi,j} - e_{\xi',j}\) is replaced by \(b_{k,j}\) for any given \(k \geq 1\).

**Proof of Lemma 1.30.** As in the proof of Lemma 1.27, it suffices to show the inequality given in the Lemma after dividing both sides by \(4^Q\) to ensure that the absolute value of

\[E_{t,t+j} := (e_{\xi,j}(X_t, X_{t+j}) - e_{\xi',j}(X_t, X_{t+j}) - (\mathbb{E}e_{\xi,j}(X_0, X_j) - \mathbb{E}e_{\xi',j}(X_0, X_j)))/4\]

is bounded by 1. Define \(E'_{t,t+j}\) in the same way as \(E_{t,t+j}\) but replace \(X_t\) with \(X_t'\) and \(X_{t+j}\) with \(X_{t+j}'\). For fixed \(k \geq 2, d \geq 1,\) and \(1 \leq m < k\), consider integers \(t_1 \leq \cdots \leq t_m \leq t_{m+1} \leq \cdots \leq t_k\) so that \(t_{m+1} - t_m = d\). Repeatedly add and subtract to see that

\[
\left|\mathbb{E} E_{t_1,t_{k+1}} \cdots E_{t_k,t_{k+1}} - \mathbb{E} E_{t_1,t_{k+1}} \cdots E_{t_m,t_{m+1}} + \mathbb{E} E_{t_{m+1},t_{m+1}+j} \cdots E_{t_k,t_{k+1}}\right|
\]

\[
= \left|\mathbb{E} E_{t_1-t_m-j,t_{k+1}-t_m} \cdots E_{t_k-t_m-j,t_{k+1}-t_m}
\right|
\]

\[
\leq \left|\mathbb{E} E_{t_1-t_m-j,t_{k+1}-t_m} \cdots E_{t_k-t_m-j,t_{k+1}-t_m}\right|
\]

\[
\times E_{t_{m+2}-t_m-j,t_{m+2}-t_m} \cdots E_{t_k-t_m-j,t_{k+1}-t_m}
\]

\[50\]
where the last term on the right-hand side can again been seen to be zero.

Since $j$ is fixed, it is possible to write $\|E_{d-j,d} - E'_{d-j,d}\|_s \leq \|1\{X_d < \xi\} - 1\{X_d'< \xi\}\|_s + \|1\{X_{d-j} < \xi'\} - 1\{X_{d-j} < \xi'\}\|_s \leq C'\sigma^d(1 + \sigma^{-j})$ for some $C' > 0$, where the cases where $d \leq j$ were absorbed into $C'$. The same can then be done for $(1 + \sigma^{-j})$. Hence, proceed exactly as above to find a constant $M > 0$ such that

$$\|\mathbb{E} E_{t_1,t_1+j} E_{t_2,t_2+j} \cdots E_{t_k,t_k+j}\|_s \leq \|\mathbb{E} E_{t_1,t_1+j} E_{t_2,t_2+j} \cdots E_{t_{m+1},t_{m+1}+j} E_{t_{m+1},t_{m+1}+j} \cdots E_{t_k,t_k+j}\|_s + M\sigma^d(e_\xi,j - e_{\xi,j})^2.$$

The rest of the arguments in the proof of Lemma 1.27 now go through without changes. The proof for the bounding functions $b_k$ is almost identical and therefore omitted. \hfill \square

**Proof of Corollary 1.16.** (i) Theorem 1 of Lifshits (1982) guarantees that $\|S_\tau\|_H$ has a continuous distribution function, and therefore $c_{n,\tau}(1 - \alpha) \rightarrow c_{\infty,\tau}(1 - \alpha)$ by Lemma 21.2 of van der Vaart (1998), where $c_{\infty,\tau}$ is the quantile function of $\|S_\tau\|_H$. Hence, $CM_{n,\tau} - c_{n,\tau}(1 - \alpha) \sim \|S_\tau\|_H^2 - c_{\infty,\tau}(1 - \alpha)$ and, in particular, $\|S_\tau\|_H^2 - c_{\infty}(1 - \alpha)$ also has a continuous distribution function. This in turn implies

$$|\mathbb{P}(CM_{n,\tau} > c_{n,\tau}(1 - \alpha)) - \alpha| = |\mathbb{P}(CM_{n,\tau} \leq c_{n,\tau}(1 - \alpha)) - \mathbb{P}(\|S_\tau\|_H^2 \leq c_{\infty,\tau}(1 - \alpha))| \rightarrow 0.$$

(ii) Let $CM_{\infty,\tau} := \sum_{j>0} r_\tau(j)^2 \varphi_j^2$ and pick an $\epsilon > 0$ such that $CM_{\infty,\tau} - \epsilon > 0$. By Theorem 1.12(ii) and the properties of quantile functions,

$$\mathbb{P}(CM_{n,\tau} \leq c_{n,\tau}(1 - \alpha)) \leq 1\{c_{n,\tau}(1 - \alpha) > n(CM_{\infty,\tau} - \epsilon)\} + \mathbb{P}(|CM_{n,\tau}/n - CM_{\infty,\tau}| \geq \epsilon)$$

51
\[ 1 - \alpha > \mathbb{P}(CM'_{n,\tau}/n \leq CM_{\infty,\tau} - \epsilon) + o(1). \]

It therefore suffices to show that \( CM'_{n,\tau}/n \to_p 0 \), which follows from an application of Birkhoff’s Ergodic Theorem to the first term on the right-hand side of

\[
n^{-1} CM'_{n,\tau} \leq \sum_{j=1}^{K-1} \left( n^{-1} \sum_{t=1+j}^{n} (\tau - J_t) (\tau - J_{t-j}) \right)^2 \phi_j^2 + \sum_{j \geq K} \phi_j^2
\]

and then letting \( K \to \infty \).

**Proof of Theorem 1.18.** (i) Recall that \( r_\tau(j) = 0 \) for all \( j > 0 \) under the null hypothesis and let \( \tilde{r}^*_{n,\tau}(j) := n^{-1} \sum_{t=j+1}^{n} V_t(\tau) V_{t-j}(\tau) \omega_t \). We can write

\[ CM^*_{n,\tau} = \|S^*_{n,\tau}\|^2_{\Pi}, \]

where \( S^*_{n,\tau} \) is

\[
n^{-1/2} \sum_{j=1}^{n-1} \left( \sum_{t=j+1}^{n} \left( \hat{V}_t(\tau) \hat{V}_{t-j}(\tau) - V_t(\tau) V_{t-j}(\tau) \omega_t \right) \right) \psi_j(\lambda) \tag{1.17}
\]

\[
- n^{-1/2} \sum_{j=1}^{n} \tilde{r}^*_{n,\tau}(j) \psi_j(\lambda) \left( \sum_{t=j+1}^{n} \omega_t \right) + \sqrt{n} \sum_{j=1}^{n} \tilde{r}^*_{n,\tau}(j) \psi_j(\lambda). \tag{1.18}
\]

As a preliminary step, I show that the \( L^2(\Pi) \)-norms the first two terms have a \( \hat{P} \)-probability limit of zero with high \( P \)-probability; the \( L^2(\Pi) \)-norm of the third term converges \( \hat{P} \)-weakly in \( P \)-probability to \( \|S^*\|^2_{\Pi} \) by Shao’s (2011a) Theorem 3.1. I then use these results below to prove that the bootstrap test has asymptotic size \( \alpha \).

The \( \hat{P} \)-expectation of the square of the \( L^2(\Pi) \)-norm of (1.17) can be written as

\[
\hat{E} \left\| n^{-1/2} \sum_{j=1}^{n-1} \left( \sum_{t=j+1}^{n} \left( \hat{V}_t(\tau) \hat{V}_{t-j}(\tau) - V_t(\tau) V_{t-j}(\tau) \omega_t \right) \right) \psi_j(\lambda) \right\|^2_{\Pi}
\]

\[ = n^{-1} \sum_{j=1}^{n-1} \phi_j^2 \sum_{s=1}^{L_n} \left( \sum_{t \in B_s \cap [j+1,n]} \left( \hat{V}_t(\tau) \hat{V}_{t-j}(\tau) - V_t(\tau) V_{t-j}(\tau) \right) \right)^2. \]

Fix \( \epsilon, \epsilon' > 0 \) and pick a large enough \( M > 0 \) such that \( \sup_{n \in \mathbb{N}} \mathbb{P}(|X_0 - \xi_0(\tau)| > Mn^{-1/2}) < \epsilon' \). As in the proof of Theorem 1.12(i), the probability that the
term on the right is larger than \( \epsilon \) is at most \( \epsilon' \) plus

\[
\frac{2}{\epsilon} \sum_{j=1}^{n-1} \varphi_j^2 n^{-1} \sum_{s=1}^{L_n} \mathbb{E} \left( \sum_{t \in B_s \cap [j+1,n]} 1_{\{|X_t - \xi_0(\tau)| \leq Mn^{-1/2}\}} \right)^2
\]

\[
\leq \frac{2}{\epsilon} \sum_{j=1}^{n-1} \varphi_j^2 b_n^{-1} \left( b_n \mathbb{P}(\{|X_0 - \xi_0(\tau)| \leq Mn^{-1/2}\})
\right.
\]
\[
+ b_n^2 \mathbb{P}(\{|X_0 - \xi_0(\tau)| \leq Mn^{-1/2}\})^2
\]
\[
\leq \frac{2}{\epsilon} \left( O(n^{-1/2}) + O(b_n/n) \right) \sum_{j>0} \varphi_j^2,
\]

which can be made arbitrarily small by first letting \( n \to \infty \) and then \( M \to \infty \).

Now consider the \( \hat{\mathbb{P}} \)-expectation of the square of the \( L_2(\Pi) \)-norm of (1.18), which can be written as

\[
\mathbb{E} \left\| n^{-1/2} \sum_{j=1}^{n-1} \hat{r}_{n,\tau}(j) \left( \sum_{t=j+1}^{n} \omega_t \right) \psi_j \right\|_{\Pi}^2
\]

\[
= n^{-1} \sum_{j=1}^{n-1} \hat{r}_{n,\tau}(j)^2 \varphi_j^2 \sum_{s=1}^{L_n} \mathbb{E} \left( \sum_{t \in B_s \cap [j+1,n]} \omega_t \right)^2
\]

\[
\leq b_n \sum_{j=1}^{n-1} \hat{r}_{n,\tau}(j)^2 \varphi_j^2
\]

\[
\leq 2b_n \sum_{j=1}^{n-1} \left( \hat{r}_{n,\tau}(j) - \hat{r}_{n,\tau}(j) \right)^2 \varphi_j^2 + 2b_n \sum_{j=1}^{n-1} \hat{r}_{n,\tau}(j)^2 \varphi_j^2
\]

by the Loève \( c_r \) inequality. The first term on the right-hand side of the display converges to zero in probability by arguments similar to those given in the proof of Theorem 1.12(i) provided that \( b_n/n \to 0 \). The second term is \( O_p(b_n/n) \) by Corollary 2.1 of Shao (2011a). It follows that \( CM_{n,\tau}^* \sim \| S_\tau \|_{\Pi}^2 \) in probability.

Theorem 1 of Lifshits (1982) and Lemma 21.2 of van der Vaart (1998) then give \( c_{n,\tau}(1 - \alpha) \to_p c_\infty(1 - \alpha) \). Thus, \( CM_{n,\tau} - c_{n,\tau}(1 - \alpha) \sim \| S_\tau \|_{\Pi}^2 - c_{\infty,\tau}(1 - \alpha) \), which yields

\[
\left| \mathbb{P}(\{CM_{n,\tau} > c_{n,\tau}^*(1 - \alpha)\}) - \alpha \right|
\]
\[
= \left| \mathbb{P}(\{CM_{n,\tau} \leq c_{n,\tau}^*(1 - \alpha)\}) - \mathbb{P}(\| S_\tau \|_{\Pi}^2 \leq c_{\infty,\tau}(1 - \alpha)) \right| \to 0.
\]
(ii) Recall that \( CM_{\infty, \tau} = \sum_{j>0} r_\tau(j)^2 \varphi_j^2 \). Pick an \( \epsilon > 0 \) such that \( CM_{\infty, \tau} > \epsilon \) and, as in the proof of Corollary 1.16(ii), the properties of quantile functions and Theorem 1.12(ii) imply

\[
\mathbb{P}(CM_{n, \tau} \leq c_{n, \tau}^*(1 - \alpha)) \leq \mathbb{P}(c_{n, \tau}^*(1 - \alpha) > n(CM_{\infty, \tau} - \epsilon)) \\
+ \mathbb{P}(|CM_{n, \tau}/n - CM_{\infty, \tau}| \geq \epsilon) \\
= \mathbb{P}(1 - \alpha > \hat{P}(CM_{n, \tau}/n \leq CM_{\infty, \tau} - \epsilon)) + o(1).
\]

Hence it suffices to show that \( \hat{E}CM_{n, \tau}^*/n \to_p 0 \), which is seen from

\[
n^{-1} \hat{E}CM_{n, \tau}^* = n^{-2} \sum_{j=1}^{n-1} \varphi_j^2 \sum_{s=1}^{L_n} \left( \sum_{t \in B_s \cap [j+1, n]} (\hat{V}_t(\tau)\hat{V}_{t+j}(\tau) - \hat{r}_{n, \tau}(j)) \right)^2 \\
\leq 4 \frac{b_n}{n} \sum_{j>0} \varphi_j^2
\]

almost surely and \( b_n/n \to 0 \). \qed
I provide simple and easily verifiable conditions under which a strong form of stochastic equicontinuity holds in a wide variety of modern time series models. In contrast to most results currently available in the literature, my methods avoid mixing conditions.

2.1 Introduction

Stochastic equicontinuity typically captures the key difficulty in weak convergence proofs of estimators with non-differentiable objective functions. Precise and elegant methods have been found to deal with cases where the data dependence structure can be described by mixing conditions; see Dedecker, Doukhan, Lang, León, Louhichi, and Prieur (2007) for an excellent summary. Mixing assumptions are convenient in this context because they measure how events generated by time series observations—rather than the observations themselves—relate to one another and therefore also measure dependence of functions of such time series. The downside to these assumptions is that they can be hard to verify for a given application. Hansen (1996) describes alternatives and considers parametric classes of functions that behave like mixingales, but his results come at the expense of Lipschitz continuity conditions on these functions and rule out many applications of interest.

In this chapter I give simple and easily verifiable conditions under which objective functions of econometric estimators are stochastically equicontinuous when the underlying process is a stationary time series of the form

\[ \xi_i = \xi(\varepsilon_i, \varepsilon_{i-1}, \varepsilon_{i-2}, \ldots). \] (2.1)

Here \((\varepsilon_i)_{i \in \mathbb{Z}}\) is a sequence of iid copies of a random variable \(\varepsilon\) and \(\xi\) is a mea-
surable, possibly unknown function that transforms the input \((\varepsilon_i, \varepsilon_{i-1}, \ldots)\) into the output \(\xi_i\). The stochastic equicontinuity problem does not have to be parametric and no continuity conditions are needed. The class (2.1) allows for the construction of dependence measures that are directly related to the stochastic process and includes a large number of commonly-used stationary time series models. The next section provides several specific examples.

In the following, \(\|X\|_p\) denotes \((\mathbb{E}|X|^p)^{1/p}\) and \(\mathbb{P}^*\) and \(\mathbb{E}^*\) are outer probability and expectation, respectively (see van der Vaart, 1998, p. 258). Limits are as \(n \to \infty\).

\[ \text{2.2 Stochastic Equicontinuity in Nonlinear Time Series Models} \]

Let \(\nu_n f := n^{-1/2} \sum_{i=1}^{n} (f(\xi_i) - \mathbb{E} f(\xi_0))\) be the empirical process evaluated at some function \(f\). Here \(f\) is a member of a class of real-valued functions \(\mathcal{F}\). In econometric applications, \(\mathcal{F}\) is typically a parametric class \(\{f_\theta : \theta \in \Theta\}\), where \(\Theta\) is a bounded subset of \(\mathbb{R}^k\), although no parametric restriction on \(\mathcal{F}\) is necessary in the following. Define a norm by \(\rho(f) = \|f(\xi_0)\|_2\). An empirical process is said to be \textit{stochastically equicontinuous} (see, e.g., Pollard, 1985, p. 139) on \(\mathcal{F}\) if for all \(\epsilon > 0\) and \(\eta > 0\), there is a \(\delta > 0\) such that

\[ \limsup_{n \to \infty} \mathbb{P}^* \left( \sup_{f,g \in \mathcal{F}, \rho(f-g)<\delta} \left| \nu_n(f-g) \right| > \eta \right) < \epsilon. \]  

(2.2)

As mentioned above, proving stochastic equicontinuity is often the key difficulty in weak convergence proofs. The next two examples illustrate typical applications.

\[ \text{Example 2.1 (Quantilograms).} \] Linton and Whang (2007) measure the directional predictive ability of stationary time series \((X_i)_{i \in \mathbb{Z}}\) with the quantilogram, a normalized version of \(\mathbb{E}(\alpha - 1\{X_0 < \theta_\alpha\})(\alpha - 1\{X_h < \theta_\alpha\})\) with \(\alpha \in (0, 1)\) and \(h = 1, 2, \ldots\), where \(\theta_\alpha\) is the \(\alpha\)-quantile of the marginal distribution of \((X_i)_{i \in \mathbb{Z}}\). Let \(\xi_i = (X_{i-h}, X_i)\) and \(f_\theta(\xi_i) = (\alpha - 1\{X_{i-h} < \theta\})(\alpha - 1\{X_i < \theta\})\). Under the null hypothesis of no directional predictability, we have \(\mathbb{E} f_{\theta_\alpha}(\xi_0) = 0\) for all \(h = 1, 2, \ldots\). Let \(\hat{\theta}_{n,\alpha}\) be the sample \(\alpha\)-quantile and replace population moments by sample moments to obtain
\[ (n - h)^{-1} \sum_{i=1}^{n} f_{\hat{\theta}_{n,\alpha}}(\xi_i), \] the sample version of \( \mathbb{E} f_{\theta_0}(\xi_0) \). Apart from a scaling factor, the asymptotic null distribution of the sample quantilogram can be determined through the decomposition

\[
(n - h)^{-1/2} \sum_{i=1}^{n} f_{\hat{\theta}_{n,\alpha}}(\xi_i) = \sqrt{n - h} \mathbb{E} f_{\hat{\theta}_{n,\alpha}}(\xi_0) + \nu_{n-h} f_{\theta_0} + \nu_{n-h}(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_0}).
\]

If the distribution of \( X_i \) is smooth, the delta method can be used to control the first term on the right and, under dependence conditions, an ordinary central limit theorem applies to the second term. Further, we have \( \rho(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_0}) \to_p 0 \) whenever \( \hat{\theta}_{n,\alpha} \to_p \theta_0 \) (see Example 3 below). Hence, we can take \( F = \{ f_{\theta} : \theta \in \Theta \} \), where \( \Theta \) is a compact neighborhood of \( \theta_0 \), and as long as (2.2) holds, the third term on the right-hand side of the preceding display converges to zero in probability because in large samples

\[
\mathbb{P} \left( \nu_{n-h}(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_0}) > \eta, \rho(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_0}) < \delta \right) \leq \mathbb{P}^* \left( \sup_{f_{\theta} \in F, \rho(f_{\theta} - f_{\theta_0}) < \delta} |\nu_{n-h}(f_{\theta} - f_{\theta_0})| > \eta \right). \quad \square
\]

**Example 2.2 (Robust M-estimators of location).** Robust location estimators can often be defined implicitly as an M-estimator \( \hat{\theta}_n \) that nearly solves \( n^{-1} \sum_{i=1}^{n} f_{\theta}(\xi_i) = 0 \) in the sense that \( \sum_{i=1}^{n} f_{\hat{\theta}_n}(\xi_i) = o_p(\sqrt{n}) \). Popular examples include the median with \( f_{\theta}(x) = \text{sign}(x - \theta) \) and Huber estimators with \( f_{\theta}(x) = -\Delta 1\{x - \theta < -\Delta\} + (x - \theta) 1\{|x - \theta| \leq \Delta\} + \Delta 1\{x - \theta > \Delta\} \) for some \( \Delta > 0 \). Add and subtract in \( \sum_{i=1}^{n} f_{\hat{\theta}_n}(\xi_i) = o_p(\sqrt{n}) \) to see that stochastic equicontinuity implies \( \sqrt{n} \mathbb{E} f_{\hat{\theta}_n}(\xi_0) + \nu_n f_{\theta_0} = o_p(1) \). The limiting behavior of \( \sqrt{n}(\hat{\theta}_n - \theta_0) \) can then again be determined through the delta method and a central limit theorem. \( \square \)

Stochastic equicontinuity cannot hold without restrictions on the complexity of the set \( F \); see, e.g., Andrews (1994, pp. 2252–2253). Here, complexity of \( F \) is measured via its bracketing number \( N = N(\delta, F) \), the smallest number for which there are functions \( f_1, \ldots, f_N \in F \) and functions \( b_1, \ldots, b_N \) (not necessarily in \( F \)) such that \( \rho(b_k) \leq \delta \) and \( |f - f_k| \leq b_k \) for all \( 1 \leq k \leq N \). In addition, some restrictions are required on the memory of the time series. For processes of the form (2.1), the memory is most easily controlled

57
by comparing $\xi_i$ to a slightly perturbed version of itself (see Wu, 2005).

Let $(\varepsilon^*_i)_{i \in \mathbb{Z}}$ be an iid copy of $(\varepsilon_i)_{i \in \mathbb{Z}}$, so that the difference between $\xi_i$ and $\xi'_i := \xi(\varepsilon_i, \ldots, \varepsilon_1, \varepsilon^*_0, \varepsilon^*_{-1}, \ldots)$ are the inputs prior to period 1. Assume the following:

**Assumption 2.3.** Let $\mathcal{F}$ be a uniformly bounded class of real-valued functions with bracketing numbers $N(\delta, \mathcal{F}) < \infty$. Then there exists some $\alpha \in (0, 1)$ and $p > 0$ such that

(i) $\sup_{f \in \mathcal{F}} \|f(\xi_n) - f(\xi'_n)\|_p = O(\alpha^n)$ and

(ii) $\max_{1 \leq k \leq N(\delta, \mathcal{F})} \|b_k(\xi_n) - b_k(\xi'_n)\|_p = O(\alpha^n)$ for any given $\delta > 0$.

**Remarks.** (i) The examples at the end of this section show that Assumption 2.3 often represents only a mild restriction on the dependence structure.

(ii) Because $\mathcal{F}$ is assumed to be uniformly bounded, the bounding functions $b_k$ can be chosen to be bounded as well. Hence, in view of Lemma 2 of Wu and Min (2005), the exact choice of $p$ in Assumption 2.3 is irrelevant, for if the assumption holds for some $p$, then it holds for all $p > 0$.

Assumption 2.3 and a complexity requirement on $\mathcal{F}$ given by a bracketing integral imply a strong form of stochastic equicontinuity. The following theorem (the proof of which is found in the next section) is similar to Andrews and Pollard’s (1994) Theorem 2.2 with their mixing condition replaced by Assumption 2.3. It implies (2.2) via the Markov inequality.

**Theorem 2.4.** Suppose that Assumption 2.3 holds and

$$\int_0^1 x^{-\gamma/(2+\gamma)} N(x, \mathcal{F})^{1/Q} dx < \infty$$

for some $\gamma > 0$ and an even integer $Q \geq 2$. Then for every $\epsilon > 0$, there is a $\delta > 0$ such that

$$\limsup_{n \to \infty} \mathbb{E}^* \left( \sup_{f, g \in \mathcal{F}, \rho(f-g) < \delta} |\nu_n(f - g)| \right)^Q < \epsilon.$$

A useful feature of this theorem is that the constants $\gamma$ and $Q$ are not connected to the dependence measures as in Andrews and Pollard (1994). In contrast to their result for mixing arrays, $\gamma$ and $Q$ can therefore be chosen to be as small and large, respectively, as desired to make the bracketing integral converge without restricting the set of time series under consideration.

Before concluding this section, the next two examples illustrate how to
apply the theorem and how to verify Assumption 2.3 in practice.

**Example 2.5** (Quantilograms, continued). Take a grid of points \(\min \Theta := \theta_0 < \theta_1 < \cdots < \theta_N := \max \Theta\) and let \(b_k(\xi_i) = 1\{X_i < \theta_k\} - 1\{X_i < \theta_{k-1}\} + 1\{X_i < \theta_k\} - 1\{X_i < \theta_{k-1}\}\). Suppose for simplicity that \(F_X(\theta) := \mathbb{P}(X_0 \leq \theta)\) is Lipschitz on \(\Theta\). Given a \(\theta \in \Theta\), we can then find an index \(k\) such that \(|f_\theta - f_{\theta_k}| \leq b_k\), where I used the fact that \(|\alpha - 1\{\cdot\}| \leq \max\{\alpha, 1 - \alpha\} < 1\). Moreover, by stationarity

\[
\rho(b_k) \leq 2\|1\{X_0 < \theta_k\} - 1\{X_0 < \theta_{k-1}\}\|_2 \leq 2\sqrt{F_X(\theta_k) - F_X(\theta_{k-1})},
\]

which is bounded above by a constant multiple of \(\sqrt{\theta_k - \theta_{k-1}}\) due to Lipschitz continuity. Hence, if \(\rho(b_k) \leq \delta\) for all \(k = 1, \ldots, N\), we can choose bracketing numbers with respect to \(\rho\) of order \(N(\delta, \mathcal{F}) = O(\delta^{-2})\) as \(\delta \to 0\) (see Andrews and Pollard, 1994; van der Vaart, 1998, pp. 270–272) and the bracketing integral converges, e.g., for \(\gamma = 1\) and \(Q = 4\). By the same calculations as in the preceding display, all \(\theta, \theta' \in \Theta\) satisfy \(\rho(f_\theta - f_{\theta'}) = O(|\theta - \theta'|^{1/2})\) as \(\theta \to \theta'\) and therefore \(\rho(f_{\theta_n,\alpha} - f_{\theta_a}) \to_p 0\) if \(\theta_n,\alpha \to_p \theta_a\). In addition, suppose that the geometric contraction (GMC) property of Wu and Min (2005) holds, i.e., there is some \(\beta \in (0, 1)\) and \(p > 0\) such that \(\|\xi_n - \xi_n'\|_p = O(\beta^n)\). Then Assumption 2.3(i) is also satisfied because

\[
\|f_\theta(\xi_n) - f_\theta(\xi_n')\|_p \leq 2\|1\{X_n < \theta\} - 1\{X_n' < \theta\}\|_p + 2\|1\{X_{n-h} < \theta\} - 1\{X_{n-h}' < \theta\}\|_p = O(\alpha^n) \text{ uniformly in } \theta \text{ for some } \alpha \in (0, 1) \text{ by Proposition 3.1 of Hagemann (2011). The GMC property holds, e.g., for stationary (causal) ARMA, ARCH, GARCH, ARMA-ARCH, ARMA-GARCH, asymmetric GARCH, generalized random coefficient autoregressive, and quantile autoregressive models; see Shao and Wu (2007) and Shao (2011b) for proofs and more examples. All of these models therefore also satisfy Assumption 2.3(i). The same reasoning applies to } b_k. \]

**Example 2.6** (Robust \(M\)-estimators of location, continued). Nearly identical arguments as in the preceding example yield stochastic equicontinuity for the median. For the Huber estimator, take the grid from before and note that we can find a \(k\) such that \(|f_\theta - f_{\theta_k}| \leq \min\{\theta_k - \theta_{k-1}, 2\Delta\} =: b_k\). A routine argument (Andrews and Pollard, 1994; van der Vaart, 1998, Example 19.7, pp. 270–271) yields bracketing numbers of order \(N(\delta, \mathcal{F}) = O(\delta^{-1})\) as \(\delta \to 0\); the bracketing integral is finite, e.g., for \(\gamma = 1\) and \(Q = 2\). Assumption 2.3(i) can be verified via the bound \(\sup_{\theta \in \Theta} \|f_\theta(\xi_n) - f_\theta(\xi'_n)\|_p \leq \|\xi_n - \xi'_n\|_p\) and (ii) holds trivially. \(\square\)
2.3 Proofs

Proof of the Theorem. This follows from a simple modification of Andrews and Pollard’s (1994) proof of their Theorem 2.1. The proof requires three steps: (i) Their “Proof of inequality (3.2),” (ii) their “Proof of inequality (3.3),” and (iii) their “Comparison of pairs” argument. Replace their $i$ with $k$ and their $\tau(h)$ with $\tau(b_k)$; then apply the Lemma below instead of Andrews and Pollard’s (1994) Lemma 3.1 in the derivation of their inequality (3.5) to deduce $\max_{1 \leq k \leq N} |\nu_n b_k|Q \leq C'N^{1/Q} \max\{n^{-1/2}, \max_{1 \leq k \leq N} \tau(b_k)\}$ and use this in (i) instead of their inequality (3.5). Another application of the Lemma establishes the required analogue of their inequality (3.5) used in (ii). The same inequality can also be applied in (iii). The other arguments remain valid without changes.

Lemma 2.7. Let $\tau(f) := \rho(f)^{2/(2+\gamma)}$ for some $\gamma > 0$ and suppose that Assumption 2.3 holds. For all $n \in \mathbb{N}$, all $f, g \in F$, and every even integer $Q \geq 2$ we have

$$\mathbb{E} |\nu_n(f-g)|^Q \leq n^{-Q/2}C\left((\tau(f-g)^2)^n + \cdots + (\tau(f-g)^2)^n\right)^{Q/2},$$

where $C$ depends only on $Q$, $\gamma$, and $\alpha$. The inequality remains valid when $f - g$ is replaced by $b_k$ for any given $k \geq 1$.

Proof of the Lemma. Let $Z(i) := f(\xi_i) - \mathbb{E} f(\xi_0) - (g(\xi_i) - \mathbb{E} g(\xi_0))$. Assume without loss of generality that $|Z(i)| \leq 1$ for all $i \geq 1$; otherwise rescale and redefine $C$. Define $Z'(i) = f(\xi'_i) - \mathbb{E} f(\xi_0) - (g(\xi'_i) - \mathbb{E} g(\xi_0))$ and note that $\mathbb{E} Z(i) = \mathbb{E} Z'(i) = 0$ for all $i \in \mathbb{Z}$ and all $f, g \in F$ because $\xi_i$ and $\xi'_i$ are identically distributed. For fixed $k \geq 2$, $d \geq 1$, and $1 \leq m < k$, consider integers $i_1 \leq \cdots \leq i_m \leq i_{m+1} \leq \cdots \leq i_k$ so that $i_{m+1} - i_m = d$. Since $Z(i)$ and $Z'(i)$ are stationary, repeatedly add and subtract to see that

$$\left|\mathbb{E} Z(i_1)Z(i_2)\cdots Z(i_k) - \mathbb{E} Z(i_1)Z(i_2)\cdots Z(i_m)\mathbb{E} Z(i_{m+1})\cdots Z(i_k)\right|\leq \left|\mathbb{E} Z(i_1 - i_m)Z(i_2 - i_m)\cdots Z(i_k - i_m)\right|$$

$$- \left|\mathbb{E} Z(i_1 - i_m)Z(i_2 - i_m)\cdots Z(0)\mathbb{E} Z(d)\cdots Z(i_k - i_m)\right|\leq \left|\mathbb{E} Z(i_1 - i_m)\cdots Z(0)(Z(d) - Z'(d))Z(i_{m+2} - i_m)\cdots Z(i_k - i_m)\right|$$

60
\[ + \sum_{j=2}^{k-m-1} \left| \mathbb{E} Z(i_1 - i_m) \cdots Z(0) Z'(d) \times \cdots \times (Z(i_{m+j} - i_m) - Z'(i_{m+j} - i_m)) \cdots Z(i_k - i_m) \right| 
\]

\[ + \left| \mathbb{E} Z(i_1 - i_m) \cdots Z(0) Z'(d) \cdots Z'(i_k - i_m) \right| 
\]

- \left| \mathbb{E} Z(i_1 - i_m) \cdots Z(0) \mathbb{E} Z(d) \ldots Z(i_k - i_m) \right| \]

(2.3)

In particular, the last summand on the right-hand side is zero because \(Z(i_1 - i_m) \cdots Z(0)\) and \(Z'(d) \cdots Z'(i_k - i_m)\) are independent and \(Z(d) \times \cdots \times Z(i_k - i_m)\) and \(Z'(d) \cdots Z'(i_k - i_m)\) are identically distributed. For a large enough \(M > 0\) and some \(s > 1\), Assumption 2.3(i) and distributional equivalence of \(Z(d)\) and \(Z'(d)\) imply \(\|Z(d) - Z'(d)\|_s \leq \|f(\xi_d) - f(\xi'_d)\|_s + \|g(\xi_d) - g(\xi'_d)\|_s \leq 2 \sup_{f \in \mathcal{F}} \|f(\xi_d) - f(\xi'_d)\|_s \leq M \alpha^d\). Hölder’s inequality then bounds the first term on the right-hand side of the preceding display by

\[ \|Z(i_1) \cdots Z(i_m)\|_p \|Z(i_{m+2}) \cdots Z(i_k)\|_q M \alpha^d, \]

(2.4)

where the reciprocals of \(p, q\), and \(s\) sum to 1. Proceeding similarly to Andrews and Pollard (1994), another application of the Hölder inequality yields

\[ \|Z(i_1) \cdots Z(i_m)\|_p \leq \left( \prod_{j=1}^{m} \mathbb{E} |Z(i_j)|^{mp} \right)^{1/(mp)} \leq \tau(f - g)^{(2+\gamma)/p} \]

whenever \(mp \geq 2\) and similarly \(\|Z(i_{m+2}) \cdots Z(i_k)\|_q \leq \tau(f - g)^{(2+\gamma)/q}\) whenever \((k - m - 1)q \geq 2\). Suppose for now that \(k \geq 3\). If \(k > m + 1\), take \(s = (\gamma + Q)/\gamma\) and \(mp = (k - m - 1)q = (k - 1)/(1 - 1/s)\). Decrease the resulting exponent of \(\tau(f - g)\) from \(Q(2 + \gamma)/(Q + \gamma)\) to 2 so (2.4) is bounded by \(M \alpha^d \tau(f - g)^2\). If \(k \geq 2\) and \(k = m + 1\), the factor \(\|Z(i_{m+2}) \cdots Z(i_k)\|_q\) is not present in (2.4), but we can still choose \(s = (\gamma + Q)/\gamma\) and \(mp = (k - 1)/(1 - 1/s)\) to obtain the same bound. Identical arguments also apply to each of the other summands in (2.3). Hence, we can find some \(M' > 0\) so that

\[ |\mathbb{E} Z(i_1)Z(i_2) \cdots Z(i_k)| \]

\[ \leq |\mathbb{E} Z(i_1)Z(i_2) \cdots Z(i_m)\mathbb{E} Z(i_{m+1}) \cdots Z(i_k)| + M' \alpha^d \tau(f - g)^2. \]
Here $M'$ in fact depends on $k$, but this does not disturb any of the subsequent steps.

Now replace (A.2) in Andrews and Pollard (1994) by the inequality in the preceding display. In particular, replace their $8\alpha(d)^{1/s}$ with $M'\alpha^d$ and their $\tau^2$ with $\tau(f-g)^2$. The rest of their arguments now go through without changes. The desired result for $b_k$ follows mutatis mutandis: Simply define $Z(i) = b_k(\xi_i)$, repeat the above steps, and invoke Assumption 2.3(ii) in place of Assumption 2.3(i). \qed
CHAPTER 3

A SIMPLE TEST FOR REGRESSION SPECIFICATION WITH NON-NESTED ALTERNATIVES\textsuperscript{1}

I introduce a simple test for the presence of the data-generating process among several non-nested alternatives. The test is an extension of the classical $J$ test for non-nested regression models. I also provide a bootstrap version of the test that avoids possible size distortions inherited from the $J$ test.

3.1 Introduction

Non-nested testing problems typically do not have a natural null hypothesis. For example, it is a priori not clear what should be the null hypothesis when testing whether a specific covariate enters the regression equation in level or in log form. For the Davidson and MacKinnon (1981) $J$ test and the Cox (1961, 1962) test, the literature therefore usually suggests a sequence of tests where each possible null hypothesis is considered; see, among others, Fisher and McAleer (1979), Dastoor (1981), and Pesaran and Weeks (2003). In this chapter, I introduce a simple test for the presence of the correct model among several non-nested specifications that avoids sequential testing. The test, which I refer to as the $MJ$ (minimum $J$) test, is an extension of the $J$ test and bases its decision on the model with the least significant $J$ statistic.

Non-nested hypothesis tests such as the $J$ or the Cox tests rely heavily on the assumption that one of the models under consideration is correct, and therefore all other non-nested specifications must be wrong. However, it may well happen that a non-nested hypothesis test does not reject a model in the presence of an alternative model, but also does not reject the alternative in the presence of the original model when the hypotheses are reversed. This

leaves the researcher in the unfortunate situation of having to conclude that both specifications “explain the data equally well” even though at most one of them can be correct. Similar problems arise when all models are rejected. A further issue is that the sequential testing is typically conducted without regard to overall size, and thus two researchers working with the same data can arrive at different specifications simply because they used different levels of significance. Non-nested testing procedures have been subject to substantial criticism because of these features; see, e.g., Granger, King, and White (1995) and Shi (2010).

I show that the $MJ$ test is robust to each of these concerns: It does not require the correct model to be among the considered specifications and avoids ambiguous test outcomes. The $MJ$ test determines with asymptotically correct size if the correct model is among the specifications under consideration. If the correct model is present, it is chosen with probability approaching one as the sample size becomes large. I also provide a bootstrap version of the $MJ$ test that possesses all of these properties.

As the Monte Carlo study in this chapter shows, the bootstrap is crucial here because the $MJ$ test can over-reject in empirically relevant cases, but a careful application of the bootstrap transforms it into an almost exact test even in quite small samples. This feature of the $MJ$ test is inherited from the traditional $J$ test, which is known to over-reject severely (Godfrey and Pesaran, 1983) unless the bootstrap is used; see Fan and Li (1995), Godfrey (1998), and Davidson and MacKinnon (2002). My simulation study also suggests that the bootstrap test has good power except when the researcher makes a particularly bad guess about the correct model, i.e., when there is little correlation between the designs of the true model and the models under consideration.

The chapter is organized as follows: Section 3.2 establishes the large sample properties of the $MJ$ test. Section 3.3 discusses bootstrap methods. Section 3.4 contains the simulation study. Section 3.5 concludes. Sections 3.6 and 3.7 present auxiliary results and proofs.

I will use the following notation throughout the chapter: For an index set of increasing integers $\mathcal{I} = \{1, \ldots, I\}$, $(a_i)_{i \in \mathcal{I}}$ denotes the column vector $(a_1, a_2, \ldots, a_I)^\top \in \mathbb{R}^I$, and an $I \times I$ matrix with generic element $a_{i,i'}$ in its $i^{th}$ row and $i'^{th}$ column is denoted by $(a_{i,i'})_{i,i' \in \mathcal{I}}$. If the integers in $\mathcal{I}$ are not consecutive, the notation is meant to indicate that the $a_i$ and $a_{i,i'}$ enter $(a_i)_{i \in \mathcal{I}}$.
and \((a_{i,i'})_{i,i' \in I}\) sequentially from smallest to largest index. Convergence in distribution as \(n \to \infty\) is denoted by \(\rightsquigarrow\); \(|\cdot|\) is the Euclidean norm.

### 3.2 The MJ Test for Non-Nested Regression Models

This section extends the \(J\) test for non-nested linear regression models to handle heteroscedastic errors. I then introduce the \(MJ\) test (Procedure 3.6 below).

Suppose we observe covariates \(\{(x_{i,1}^\top, \ldots, x_{i,M}^\top) \in \mathbb{R}^{d_1 + \cdots + d_M} : i = 1, \ldots, n\}\) that give rise to \(M \geq 2\) different possible linear regression models for \(y := (y_1, \ldots, y_n)^\top \in \mathbb{R}^n\), i.e.,

\[ y = X_m \beta_m + u_m, \quad m \in \mathcal{M} := \{1, \ldots, M\}, \tag{3.1} \]

where \(X_m := (x_{1,m}, \ldots, x_{n,m})^\top \in \mathbb{R}^{n \times d_m}\) is the design matrix of model \(m\). The matrices \(X_1, \ldots, X_M\) are assumed to be non-nested, i.e., for any two matrices with index \(m \neq l\) in \(\mathcal{M}\), no matrix can be obtained from another by a linear transformation. This does not rule out the possibility that some of the columns of \(X_m\) and \(X_l\) are identical or that they may be nonlinear transformations of one another. In addition, there is an observed or unobserved design matrix \(X_m^* := (x_{1,m^*}, \ldots, x_{n,m^*})^\top\) associated with the correct model \(m^*\). Suppose that \(\mathcal{F} := \{X_m : m \in \mathcal{M} \cup \{m^*\}\}\) has the following properties:

**Assumption 3.1.** \(\{(y_i, (x_{i,m})_{m \in \mathcal{M} \cup \{m^*\}})^\top : i \geq 1\}\) is a sequence of iid random vectors. We have \(\mathbb{E}|x_{i,m}|^4 < \infty\) for all \(m \in \mathcal{M} \cup \{m^*\}\), where the number of elements of \(\mathcal{M}\) does not depend on \(n\). For all \(m \in \mathcal{M}\), the matrices \(\mathbb{E}x_{i,m}x_{i,m}^\top\) are positive definite.

Tests constructed for non-nested environments such as the Davidson and MacKinnon (1981) \(J\) test typically assume that there is an \(m^* \in \mathcal{M}\) such that the conditional mean of \(y\) can be written as \(\mathbb{E}(y \mid \mathcal{F}) = X_{m^*} \beta_{m^*}\). In the following, I depart from this condition and only assume that a correct specification exists.

**Assumption 3.2.** Model \(m^*\) satisfies \(\mathbb{E}(y \mid \mathcal{F}) = X_{m^*} \beta_{m^*}\). Let \(u_{i,m^*} := y_i - x_{i,m^*}^\top \beta_{m^*}\) for all \(i \geq 1\) and \(\mathbb{E}u_{i,m^*}^4 < \infty\).

**Remarks.** 1. Although I only consider linear regression models with independent data, the results of the chapter can be extended to nonlinear (para-
metric) models with weakly dependent data; see also the discussion below Theorem 3.7.

2. Davidson and MacKinnon (1981) point out that the assumption \( m^* \in \mathcal{M} \) is not crucial since the \( J \) test is capable of rejecting all models in \( \mathcal{M} \).

The setup of the \( J \) test presumes that for some predetermined \( m \in \mathcal{M} \), the researcher wants to test the null hypothesis \( m = m^* \) against \( m \neq m^* \) in the presence of the non-nested alternatives \( l \in \mathcal{M} \setminus \{m\} \). This can be done by artificially nesting the models in (3.1) via an additional parameter vector \( \alpha_m := (\alpha_{l,m})_{l \in \mathcal{M} \setminus \{m\}} \in \mathbb{R}^{M-1} \) such that

\[
y = X_m b_m + \sum_{l \in \mathcal{M} \setminus \{m\}} \alpha_{l,m} X_l \beta_l + u, \quad \text{where } b_m := \left( 1 - \sum_{l \in \mathcal{M} \setminus \{m\}} \alpha_{l,m} \right) \beta_m.
\]

Since the vectors \((\alpha_{l,m}, \beta_l)_{l \in \mathcal{M} \setminus \{m\}}\) of the nesting model may not be identified, Davidson and MacKinnon replace the \( \beta_l \) in the preceding display by the least squares (LS) estimates \( \hat{\beta}_l = (X_l^\top X_l)^{-1} X_l^\top y \). After redefining the error term appropriately, this yields

\[
y = X_m b_m + \sum_{l \in \mathcal{M} \setminus \{m\}} \alpha_{l,m} X_l \hat{\beta}_l + u, \quad \text{(3.2)}
\]

which can be estimated by LS. A Wald test for \( \alpha_m = 0 \) is a \( J \) test for the validity of model \( m \) in the presence of the alternatives \( \mathcal{M} \setminus \{m\} \). To construct the test statistic, let

\[
\lambda_{n,m} := n^{-1/2} \left( y^\top P_l M_m y \right)_{l \in \mathcal{M} \setminus \{m\}} \quad \text{and}
\]

\[
\hat{\Sigma}_{n,m} := n^{-1} \left( y^\top P_l M_m \hat{\Omega}_{n,m} M_m P_l^\top y \right)_{l,l' \in \mathcal{M} \setminus \{m\}},
\]

where \( P_m := X_m (X_m^\top X_m)^{-1} X_m^\top \) and \( M_m := I_n - P_m \) are the usual projection matrices and \( \hat{\Omega}_{n,m} \) is an “estimate” of \( \mathbb{E}(u_m u_m^\top | \mathcal{F}) \). The \( J \) test statistic for model \( m \) is then

\[
J_{n,m} := \lambda_{n,m}^\top \hat{\Sigma}_{n,m}^{-1} \lambda_{n,m},
\]

and the hypothesis that model \( m \) is the true model is rejected for large values of \( J_{n,m} \). In practice, the \( J \) statistic is easily computed by running regression (3.2) and performing a Wald test for \( \alpha_m = 0 \) with weighting matrix \( \hat{\Omega}_{n,m} \).

Given the independence of the observations, \( \hat{\Omega}_{n,m} \) should be a diagonal
matrix of squared residuals. For heteroscedasticity-robust testing, the literature frequently recommends using the residuals of the model under the null hypothesis; see, e.g., Davidson and MacKinnon (1985). Hence, I use the residuals $\hat{u}_{i,m} := y_i - x_{i,m}^\top \hat{\beta}_m$ and define

$$\hat{\Omega}_{n,m} := \text{diag}(\hat{u}_{1,m}^2, \ldots, \hat{u}_{n,m}^2),$$

although other approaches are clearly possible. If Assumptions 3.1 and 3.2 hold, I show in Section 3.6 that we can then write $\hat{\Sigma}_m = \Sigma_m + \Pi_m + o_p(1)$, where $\Sigma_m$ is $\hat{\Sigma}_{n,m}$ with products of the form $n^{-1}X_m^\top X_l$ replaced by $\mathbb{E} x_{i,m} x_{i,l}^\top$ and products of the form $n^{-1}X_l^\top \hat{\Omega}_{n,m} X_l$ replaced by $\mathbb{E}(u_{i,m}^2 x_{i,l} x_{i,l}^\top)$, and $\Pi_m$ is a positive semi-definite matrix; in particular, $\Pi_{m^*}$ is equal to zero. To ensure that the $J$ statistics are well behaved in large samples, I make the following assumption:

**Assumption 3.3.** For all $m \in \mathcal{M}$, $\Sigma_m$ is positive definite.

**Remarks.** Assumption 3.3 is needed to identify the artificial parameters $\alpha_m$, $m \in \mathcal{M}$. This condition fails when the covariates of $m^*$ are completely uncorrelated with the covariates of one of the models in $\mathcal{M}$ or when the design matrices of any two models in $\mathcal{M}$ are nested, i.e., one matrix is a linear transformation of the other. Michelis (1999) discusses the consequences of near-singularity of $\Sigma_m$ for the case $M = 2$.

For the $J$ test to have power against the alternative $m^* \not\in \mathcal{M}$, I also assume that the covariates of $m^*$ have enough idiosyncratic variation so that the following condition holds:

**Assumption 3.4.** For all $m \in \mathcal{M} \setminus \{m^*\}$, there exists $l \in \mathcal{M}$ such that

$$|\beta_m^\top \mathbb{E} x_{i,m^*} x_{i,l}^\top (\mathbb{E} x_{i,l} x_{i,l}^\top)^{-1} [\mathbb{E} x_{i,l} x_{i,m^*}^\top - \mathbb{E} x_{i,l} x_{i,l}^\top (\mathbb{E} x_{i,m^*} x_{i,m^*}^\top)^{-1} \mathbb{E} x_{i,m^*} x_{i,m^*}^\top] \beta_m^*| > 0.$$  \hfill (3.4)

**Remarks.** In addition to the restrictions imposed via Assumption 3.3, this property rules out that the covariates of $m^*$ are fully correlated with (but not identical to) the covariates of any model $m$ in $\mathcal{M}$, e.g., if $X_m = X_{m^*} + Z$, where $Z$ is an independent measurement error with mean zero. In such a case, the $J$ test would mistake $X_m$ for $X_{m^*}$.

The following Lemma summarizes the large sample properties of the $J_{n,m}$ statistics under Assumptions 3.1–3.4.
Lemma 3.5. Suppose Assumptions 3.1–3.4 are satisfied.

(i) If $m^* \in \mathcal{M}$, then $J_{n,m^*} \rightsquigarrow \chi^2_{M-1}$.
(ii) For every $m \in \mathcal{M} \setminus \{m^*\}$ and every $B \in \mathbb{R}$, we have $\lim_{n \to \infty} \mathbb{P}(J_{n,m} > B) = 1$.

The $J$ test does not have a natural null hypothesis, and therefore the researcher is expected to have a preferred model $m$ to test the null hypothesis $m = m^*$. Because there is usually little guidance in applied work about what $m$ could be, this makes a seemingly simple task such as testing whether to include a specific covariate in level or in log form a surprisingly difficult problem. However, a closer look at Lemma 3.5 reveals an easy way out of this dilemma: If one of the models under consideration is the correct model, then its $J$ statistic has an asymptotic distribution and the statistics of the other models diverge; if, instead, the correct model is not among the $M$ models, then all statistics will diverge. Thus, only the model with the smallest $J$ statistic can possibly be the correct model and we can reject the hypothesis $m^* \in \mathcal{M}$ when the smallest $J$ statistic is large. This motivates the following alternative to the traditional sequential testing procedures:

Procedure 3.6 ($MJ$ Test). 1. For each $m \in \mathcal{M}$, run regression (3.2) and compute $J_{n,m}$ as in (3.3). Let $\mathcal{J}_n := \{J_{n,m} : m \in \mathcal{M}\}$ and define $MJ_n := \min \mathcal{J}_n$.

2. Reject the hypothesis $H_0: m^* \in \mathcal{M}$ in favor of $H_1: m^* \notin \mathcal{M}$ if $MJ_n > c_{1-\alpha}$, where $c_{1-\alpha}$ is the $1-\alpha$ quantile of the $\chi^2_{M-1}$ distribution.

As a referee points out, the $MJ$ test is an instance of an intersection–union test (Berger, 1982): The null hypothesis is the union of different nulls, and the rejection region is thus the intersection of the rejection regions of the respective nulls. Indeed, the null is that one of the models in $\mathcal{M}$ is correctly specified, but it can be any of them, so that the null is the union of the sets of data-generating processes, each of which constitutes one of the models in $\mathcal{M}$, and, for the null to be false, each of those hypotheses must be false.

The following Theorem shows that the $MJ$ test is indeed a valid test for $m^* \in \mathcal{M}$ and, in particular, the asymptotic distributions of $J_{n,m^*}$ and $MJ_n$ coincide. The reason for this result is that the number of models in $\mathcal{M}$ does not depend on $n$ and the $M$ different $J$ statistics are asymptotically independent since $M - 1$ of them diverge—two features that are typically not available for nested testing problems, but can be exploited in non-nested...
environments to construct simple tests such as Procedure 3.6.

**Theorem 3.7.** Suppose Assumptions 3.1–3.4 are true.

(i) If \( m^* \in \mathcal{M} \), then \( MJ_n \sim \chi^2_{M-1} \) and 
(ii) if \( m^* \not\in \mathcal{M} \), then \( \lim_{n \to \infty} P(MJ_n > B) = 1 \) for all \( B \in \mathbb{R} \).

**Remarks.** 1. The \( J \) test can be generalized to nonlinear regressions (Davidson and MacKinnon, 1981), and to models with dependent errors via standard heteroscedasticity and autocorrelation consistent covariance matrix estimators or fixed-\( b \) asymptotics as in Choi and Kiefer (2008). See also MacKinnon, White, and Davidson (1983) for an extension to models with weakly dependent and endogenous covariates. Since the \( MJ \) test is nothing but a \( J \) test for a specific model, these generalizations are also available for the minimum \( J \) approach as long as an identification condition analogous to (3.4) holds.

2. Related tests for non-nested models such as the \( JA \) test of Fisher and McAleer (1981) can be extended in a similar way.

The \( MJ \) test chooses a model \( \hat{m} = \arg \min J_n \) for the test, and if the test provides evidence that \( m^* \in \mathcal{M} \), then \( \hat{m} \) is the natural candidate for \( m^* \). This can be interpreted as the non-nested analogue of general-to-specific testing: Selecting \( \hat{m} \) amounts to choosing a model for which there is the most evidence that it is not outperformed by the other models under consideration in terms of explanatory power. Indeed, as the following Lemma shows, the \( MJ \) test consistently finds the true model among the specifications, both unconditionally and conditional on having accepted the null hypothesis:

**Lemma 3.8.** If Assumptions 3.1–3.4 hold and \( m^* \in \mathcal{M} \), then \( \lim_{n \to \infty} P(\hat{m} = m^*) = 1 \) and \( \lim_{n \to \infty} P(\hat{m} = m^* \mid MJ_n \leq c_{1-\alpha}) = 1 \).

The Monte Carlo results in Section 3.4 show that \( \hat{m} \) reliably chooses the correct model, but nonetheless one should be careful about how to proceed from there; see Leeb and Pötscher (2009) and the references therein.

The \( MJ \) test might be expected to inherit the well-known size distortion present in the \( J \) test (see Davidson and MacKinnon, 2002, and the references therein). However, I will show in Section 3.4 that these effects can be ameliorated as long as the bootstrap is used. The next section provides the necessary modifications.
3.3 Bootstrapping the MJ Test Statistic

This section presents a bootstrap version of the MJ test (Procedure 3.9). The idea is to replace $J_{n,\hat{m}}$ in $J_n$ by realizations of a bootstrap $J$ statistic such that repeatedly computing the minimum over this new set of $J$ statistics mimics the behavior of $MJ_n$. To account for the possible heterogeneity in the errors, I use the wild bootstrap of Liu (1988) and Mammen (1992); it perturbs the residuals with iid copies $\eta_1, \eta_2, \ldots$ of a random variable $\eta$ with $E(\eta \mid y, F) = 0$, $E(\eta^2 \mid y, F) = 1$ and $E(|\eta|^{2+\delta} \mid y, F) < \infty$ for some $\delta > 0$.

Let also $H := \text{diag}(\eta_1, \ldots, \eta_n)$.

Procedure 3.9 (Bootstrap MJ Test). Let $\hat{m} = \arg\min J_n$. (If $\arg\min J_n$ is not unique, choose $\hat{m}$ to be any element of $\arg\min J_n$.)

1. (a) Use LS to obtain the residuals $\hat{u} = y - X\hat{\beta}_\hat{m}$ and compute $u^* := H\hat{u}$.
   (b) Generate bootstrap data $y^* := X\hat{\beta}_\hat{m} + u^*$ and calculate
   $\lambda_{n,\hat{m}}^* := n^{-1/2}\left(y^*\top P_l M_{n,\hat{m}} y^*\right)_{l \in M \setminus \{\hat{m}\}}$ and
   $\hat{\Sigma}_{n,\hat{m}}^* := n^{-1}\left(y^*\top P_l M_{n,\hat{m}} \hat{\Omega}_{n,\hat{m}} M_{n,\hat{m}} P_l' y^*\right)_{l,l' \in M \setminus \{\hat{m}\}}$.
   (c) Compute the bootstrap $J$ statistic $J_{n,\hat{m}}^* := \lambda_{n,\hat{m}}^* \hat{\Sigma}_{n,\hat{m}}^{-1} \lambda_{n,\hat{m}}^*$.

2. Let $J_n^* := \{J_{n,\hat{m}}^*\} \cup J_n \setminus \{J_{n,\hat{m}}\}$ and calculate $MJ_n^* := \min J_n^*$.

3. Repeat steps 1–2 $R$ times, each with a new realization of $H$. Reject the hypothesis $m^* \in M$ if $MJ_n$ is larger than $c^*_{n,1-\alpha}$, the $1 - \alpha$ empirical quantile of the $MJ_n^*$.

Remarks. 1. The recommended choice for $\eta$ in practice is a Rademacher variable that takes on the value 1 with probability 1/2 and the value $-1$ with probability 1/2. Distributions other than the Rademacher distribution could be used for $\eta$, in particular if the error distribution is skewed, but there is no evidence that they would lead to better inference; see Davidson et al. (2007) for a discussion.

2. Step 1 is similar to the standard residual bootstrap used in Fan and Li (1995), Godfrey (1998), and Davidson and MacKinnon (2002), who deal with homoscedastic errors; Choi and Kiefer (2008) use the block bootstrap. None of these authors establish the validity of their bootstrap method.

3. There is no need to approximate $\hat{\Omega}_{n,\hat{m}}$ in $\hat{\Sigma}_{n,\hat{m}}$ by a bootstrap version.
because the variance of the bootstrap errors \( u^* \) is \( \text{var}(u^* u^{*\top} | y, \mathcal{F}) = \hat{\Omega}_{n, \hat{m}} \); see also Flachaire (2002).

4. Step 1 only approximates the distribution of \( J_{n, \hat{m}} \) conditional on \( \hat{m} \), and computing the minimum over the \( J \) statistics after replacing \( J_{n, \hat{m}} \) with \( J^*_{n, \hat{m}} \) in step 2 mimics the additional randomness from choosing \( \hat{m} \) out of the \( M \) models. If \( m^* \in \mathcal{M} \), \( \hat{m} \) and the minimizer of \( J^*_{\hat{m}} \) coincide in large samples, and hence this step is not crucial for the validity of the bootstrap. However, the simulation study in Section 3.4 suggests that, if anything, the Bootstrap \( MJ \) test can behave conservatively in small samples. Step 2 can therefore improve both the size and power of the bootstrap test in finite samples because it restricts the magnitude of large realizations of \( J^*_{n, \hat{m}} \).

By choosing the number of bootstrap repetitions \( R \) in Procedure 3.9 large enough, we can approximate the quantiles of the distribution of \( MJ^*_{n} \) with arbitrary precision. I therefore let \( R \to \infty \) in the following and define the quantiles of \( MJ^*_{n} \) directly as

\[
\hat{c}^*_{n, 1-\alpha} := \inf \{ x \in \mathbb{R} : P^*(MJ^*_{n} \leq x) \geq 1 - \alpha \},
\]

where \( P^*(\cdot) \) abbreviates \( P(\cdot | y, \mathcal{F}) \).

To ensure that the bootstrap approximation is well-behaved, a further condition similar to Assumption 3.3 is needed: Since \( \mathbb{E}(X_l^\top u^* | y, \mathcal{F}) = 0 \) and \( \text{var}(X_l^\top u^*/n | y, \mathcal{F}) = X_l^\top \hat{\Omega} X_l/n^2 = o_P(1) \) provided Assumptions 3.1 and 3.2 hold, we can with \( P \)-probability approaching one as \( n \to \infty \) write

\[
\hat{\Sigma}^*_{n,m} = n^{-1} \left( \hat{\beta}_m^\top X_m^\top P_l M_m \hat{\Omega}_{n,m} M_m P_l X_m \hat{\beta}_m \right)_{l,l'} + o_P(1).
\]

Define \( \hat{\Xi}^*_{n,m} \) as the first term on the right-hand side. In the same way as \( \hat{\Sigma}^*_{n,m} = \Sigma_m + \Pi_m + o_P(1) \), it can be seen that \( \hat{\Xi}^*_{n,m} = \Xi_m + \Psi_m + o_P(1) \), where \( \Xi_m \) is \( \hat{\Xi}^*_{n,m} \) with expressions of the form \( n^{-1} X_m^\top X_l \) replaced by \( \mathbb{E} x_{i,m} x_{i,l}^\top \) and expressions of the form \( n^{-1} X_l \hat{\Omega}_{n,m} X_l \) replaced by \( \mathbb{E}(u_{i,m}^2 x_{i,l}^\top x_{i,l}^\top) \); \( \Psi_m \) is a positive semi-definite matrix with \( \Psi_m^* \) equal to zero. Details are provided in Section 3.6. I impose the following condition:

**Assumption 3.10.** For all \( m \in \mathcal{M} \), \( \Xi_m \) is positive definite.

Remarks. In addition to the restrictions imposed by Assumptions 3.3 and 3.4, this condition rules out that the covariates of any two models in \( \mathcal{M} \) are orthogonal. In practice, it should not limit the applicability of the Bootstrap \( MJ \) test because all models under consideration are supposed to explain the same variable \( y \). Therefore, it is not much of a restriction to assume that
they have some common features.

The following result establishes the consistency of the Bootstrap MJ test.

**Theorem 3.11.** Suppose Assumptions 3.1–3.4 and 3.10 are satisfied with \( \mathbb{E}|u_{i,m*}|^{4+\delta} < \infty \) and \( \mathbb{E}|x_{i,m}|^{4+\delta} < \infty \) for all \( m \in \mathcal{M} \cup \{m^*\} \) and some \( \delta > 0 \). Let \( \alpha \in (0,1) \). Procedure 3.9 has the following properties:

(i) If \( m^* \in \mathcal{M} \), then \( \lim_{n \to \infty} \mathbb{P}(MJ_n > c_{n,1-\alpha}^*) = \alpha \), and

(ii) if \( m^* \not\in \mathcal{M} \), then \( \lim_{n \to \infty} \mathbb{P}(MJ_n > c_{n,1-\alpha}^*) = 1 \).

The consistency of \( \hat{m} \) conditional on accepting the null hypothesis of the bootstrap test is then an immediate consequence.

**Corollary 3.12.** Suppose the conditions of Theorem 3.11 are satisfied. If \( m^* \in \mathcal{M} \), then \( \lim_{n \to \infty} \mathbb{P}(\hat{m} = m^* \mid MJ_n \leq c_{n,1-\alpha}^*) = 1 \).

The next section illustrates this consistency property and the finite sample behavior of the MJ and Bootstrap MJ tests in a small simulation study.

### 3.4 Simulation Study

The J test is known to severely over-reject when (1) the sample size is small, (2) the error variance is large, (3) the number of regressors differs among the models, or (4) the correlation between the models under consideration is small; see Davidson and MacKinnon (2002) for a thorough analysis of why this is the case. Focusing on the correlation structure, this section investigates the impact of these properties on the performance of the MJ test and the Bootstrap MJ test.

The true model \( m^* \) for the following experiments is

\[
y_i = x_{i,m^*}^T \beta_{m^*} + u_i, \quad u_i = (v_i - 1)[|x_{i,1,m^*}x_{i,2,m^*}|(|x_{i,3,m^*}| + 2.5)]^{1/2},
\]

where \( \beta_{m^*} \) is a \( d_{m^*} \)-vector of ones, the first element of \( x_{i,m^*} \) is one and the other components, denoted as \( (x_{i,2,m^*}, \ldots, x_{i,d_{m^*},m^*})^T \), are uncorrelated standard normal variables; the \( v_i \) are independently distributed as \( \chi_1^2 \). I experimented with the form of the heteroscedasticity in \( u_i \), but found that it had little effect on the performance of the bootstrap test. I therefore settled for a worst-case scenario and chose an error structure that is known to cause size distortions when heteroscedasticity-robust estimators are employed (see
Long and Ervin, 2000). The covariates $x_{i,m} = (x_{i,1,m}, \ldots, x_{i,d_{m},m})^\top$ of any other model $m \in \mathcal{M} \setminus \{m^*\}$ are given by $x_{i,1,m} = 1,$

$$x_{i,j,m} = \frac{\rho}{\sqrt{1 - \rho^2}} x_{i,j,m^*} + z_{i,j,m}, \quad j = 2, \ldots, \min\{d_{m^*}, d_m\},$$

and $x_{i,j,m} = z_{i,j,m}$ for $d_m > d_{m^*},$ where the $z_{i,j,m}$ are independent standard normals. This ensures that the correlation between the random components of $x_{i,m^*}$ and the corresponding components of $x_{i,m}$ is exactly $\rho.$

I used 10,000 replications for each sample size $n \in \{10, 20, \ldots, 500\}$ to investigate the behavior of the $MJ$ test for $M \in \{2, 3\}$ and $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ under both the null and the alternative hypothesis. For each of these cases, I recorded the test decision of the $MJ$ test at a significance level of 5% and the model $\hat{m}$ with the smallest $J$ statistic. I used the warp-speed method of Giacomini et al. (2007) to estimate size and power of the Bootstrap $MJ$ test; this method considerably sped up the simulations because only one bootstrap replication per Monte Carlo replication was needed. For notational simplicity, let also $m^* = 1$ when $m^* \in \mathcal{M}.$

Panels (a)–(e) of Figure 3.1 show the rejection frequencies of the $MJ$ test and the bootstrap version for the case $M = 2$ when $H_0: m^* \in \mathcal{M}$ is true as a function of $n$ and $\rho.$ Here and in the following, I used the same sequence of random numbers for each correlation structure in order to make the results comparable. The number of regressors in the true model was $d_{m^*} = 5$ and the other specification had $d_2 = 7$ covariates. The $MJ$ test over-rejected severely in all cases at all sample sizes, although the size of the test improved considerably in larger samples for $\rho \in \{0.3, 0.5, 0.7, 0.9\}.$ The case $\rho = 0.9$ was the slowest to adjust because the two models under consideration were so similar. The $MJ$ test broke down completely for $\rho = 0.1$ because the distribution of the $J$ test statistic is not well approximated by the $\chi_1^2$ distribution when the correlation among the models is small; see Davidson and MacKinnon (2002). In contrast, the Bootstrap $MJ$ test was almost exact for $\rho \in \{0.5, 0.7, 0.9\}$ at all sample sizes, and behaved mildly conservatively for the low correlation structures. Experiments with larger values of $d_2$ and larger variation in $v_i$ increased the size distortion of the $MJ$ test even further, whereas the Bootstrap $MJ$ test remained unaffected.

The last panel in Figure 3.1 shows the relative number of times $\hat{m}$ was indeed the true model in the experiments presented in panels (a)–(e). Se-
Figure 3.1: Panels (a)–(e) show rejection frequencies of the $MJ$ test (dashed line) and Bootstrap $MJ$ test (solid) under $H_0$ with $M = 2$ at the 5% level (dotted) for different values of $\rho$ as a function of sample size $n$. Panel (f) plots the relative frequency of the event $\hat{m} = m^*$ for the experiments in panels (a)–(e) as a function of $n$. 
lecting the model with the smallest $J$ statistic worked well in small samples even when the models were highly correlated. In samples larger than 100 (not reported), the selection frequency of the true model was essentially one.

Table 3.1: Conditional and Unconditional Selection Frequencies

<table>
<thead>
<tr>
<th></th>
<th>Relative Frequency of $\hat{m} = m^*$</th>
<th>Unconditional</th>
<th>$MJ_n \leq c_{1-\alpha}$</th>
<th>$MJ_n \leq c^*_{n,1-\alpha}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 30$</td>
<td>$\rho = 0.9$</td>
<td>0.855</td>
<td>0.921</td>
<td>0.858</td>
</tr>
<tr>
<td></td>
<td>$\rho = 0.5$</td>
<td>0.940</td>
<td>0.996</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>$\rho = 0.1$</td>
<td>0.902</td>
<td>0.998</td>
<td>0.908</td>
</tr>
<tr>
<td>$n = 60$</td>
<td>$\rho = 0.9$</td>
<td>0.933</td>
<td>0.979</td>
<td>0.936</td>
</tr>
<tr>
<td></td>
<td>$\rho = 0.5$</td>
<td>0.986</td>
<td>1.000</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>$\rho = 0.1$</td>
<td>0.950</td>
<td>1.000</td>
<td>0.958</td>
</tr>
<tr>
<td>$n = 90$</td>
<td>$\rho = 0.9$</td>
<td>0.967</td>
<td>0.997</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>$\rho = 0.5$</td>
<td>0.996</td>
<td>1.000</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>$\rho = 0.1$</td>
<td>0.967</td>
<td>1.000</td>
<td>0.976</td>
</tr>
</tbody>
</table>

Figure 3.1 does not show the selection frequencies conditional on accepting the null hypothesis since the warp-speed method cannot estimate this quantity for the Bootstrap $MJ$ test. I therefore calculated the actual rejection and selection frequencies for $n \in \{30, 60, 90\}$; the bootstrap test was based on $R = 399$ bootstrap replications. As Table 3.1 shows, the conditional selection frequencies for the Bootstrap $MJ$ test and the relative frequencies of the event $\hat{m} = m^*$ were almost identical. The discrepancy was larger for the conditional selection frequencies based on the $MJ$ test, which was mostly driven by its inability to control the nominal size.

Figure 3.2 plots the rejection frequency of the Bootstrap $MJ$ test for $M = 2$ under $H_1: m^* \notin \mathcal{M}$ as a function of $n$ and $\rho$. I do not report power estimates for the $MJ$ test because its size distortion invalidates the power estimates. As before, the true model had $d_{m^*} = 5$ covariates, and the other specifications had $d_1 = 5$ and $d_2 = 7$ regressors. The bootstrap test had good power in moderately large samples as long as the correlation between the models in $\mathcal{M}$ and the true models was high. However, the power dropped considerably for $\rho = 0.5$ and was essentially zero for smaller correlations. Hence, if the researcher uses models that are very different from the correct model, the Bootstrap $MJ$ test will most likely not be able to detect this problem. Experiments with larger error variances gave qualitatively similar
results, but larger sample sizes were required to reach the same level of power. I also experimented with the number of covariates of the models in $\mathcal{M}$, but found that the impact was relatively small.

Figure 3.3 repeats the experiments shown in Figure 3.1 for $M = 3$. The models under consideration now had $d_{m^*} = 5$, $d_2 = 5$, and $d_3 = 7$ regressors. The size distortion of the $MJ$ test was even more extreme than before, whereas the Bootstrap $MJ$ test was less conservative and even more precise than in the case $M = 2$. The selection frequency of the true model was slightly worse and larger samples were needed to reliably detect the true model.

Finally, Figure 3.4 shows the power experiment from Figure 3.2 for $M = 3$. The true model had $d_{m^*} = 5$ regressors, and the models in $\mathcal{M}$ now had $d_1 = 5$, $d_2 = 6$, and $d_3 = 7$ covariates, but the results remained almost unchanged.

3.5 Conclusion

In this chapter, I introduced a simple test for the presence of the data-generating process among several non-nested regression models. In contrast to classical approaches to non-nested testing, the $MJ$ test does not require the correct model to be among the considered specifications and avoids ambiguous test outcomes.
Figure 3.3: Panels (a)–(e) show rejection frequencies of the $MJ$ test (dashed line) and Bootstrap $MJ$ test (solid) under $H_0$ with $M = 3$ at the 5% level (dotted) for different values of $\rho$ as a function $n$. Panel (f) plots the relative frequency of the event $\hat{m} = m^*$ for the experiments in panels (a)–(e) as a function of $n$.

Figure 3.4: Rejection frequencies of the Bootstrap $MJ$ test under $H_1$ with $M = 3$ at the 5% level for different values of $\rho$ as a function of sample size $n$. 

77
3.6 Auxiliary Results and Definitions

This section states two results that are needed for the proofs below and defines \( \Sigma_{n,m} = \Sigma_m + \Pi_m + \sigma(1) \) and \( \hat{\Sigma}_{n,m} = \Xi_m + \Psi_m + \sigma(1) \) explicitly. To simplify notation, for a random variable \( Z \) and random sequences \( Z_n, Z'_n \), the expression \( Z_n \rightsquigarrow Z \) in probability abbreviates \( \mathbb{P}(\sup_z |Z_n - z| > \varepsilon) = o(1) \) for every \( \varepsilon > 0 \), and \( Z_n = Z'_n + o(1) \) in probability abbreviates \( \mathbb{P}(|Z_n - Z'_n| > \varepsilon) = o(1) \) for every \( \varepsilon > 0 \).

The Frobenius norm \( \sqrt{\text{trace}(A^\top A)} \) of a matrix \( A \) is denoted by \( \|A\| \).

Given any \( m, l, l' \in \mathcal{M} \), repeated application of the Law of Large Numbers yields \( n^{-1}X_i^\top \hat{\Omega}_{m,n}X_{l'} = \mathbb{E}(u_{i,m,x_i,l'}^2) + \mathbb{E}[(x_{i,m}^\top \beta - x_{i,m}^\top \beta_m)^2] + o_0(1) \), provided that Assumptions 3.1 and 3.2 are satisfied, where \( \beta_m := (\mathbb{E} x_{i,m} x_{i,m}^\top)^{-1} \mathbb{E} x_{i,m} x_{i,m}^\top \beta_m \). Define \( \Gamma_{m,l} := (\mathbb{E} x_{i,m} x_{i,m}^\top)^{-1} \mathbb{E} x_{i,m} x_{i,m}^\top \) and \( v_{i,m}^2 := (x_{i,m}^\top \beta - x_{i,m}^\top \beta_m)^2 \), then

\[
n^{-1}y^\top P M_m \hat{\Omega}_{m,n} M_n P^\top y
= \left[ \beta_m^\top \Gamma_{l,m}^\top \left( \mathbb{E}(u_{i,m,x_i,l'}^2) - \Gamma_{m,l} \mathbb{E}(u_{i,m,x_i,l'}^2) \right) \right. \\
+ \left. \Gamma_{m,l} \mathbb{E}(u_{i,m,x_i,l'}^2 \mathbb{E}(u_{i,m,x_i,l'}^2) \Gamma_{l,m} \mathbb{E}(u_{i,m,x_i,l'}^2) \right) \right] + o_0(1).
\]

Denote the first term in square brackets by \( \sigma_{m,l,l'} \) and the second term in square brackets by \( \tau_{m,l,l'} \). Define \( \Sigma_m := (\sigma_{m,l,l'})_{l,l' \in \mathcal{M} \setminus \{m\}} \) and \( \Pi_m := (\tau_{m,l,l'})_{l,l' \in \mathcal{M} \setminus \{m\}} \). Similarly, replace \( y \) by \( P_m y \), \( \Gamma_{l,m} \beta_m \) by \( \Gamma_{l,m} \Gamma_{m,m} \beta_m \), and \( \Gamma_{l,m} \beta_m \) by \( \Gamma_{l,m} \Gamma_{m,m} \beta_m \) in the preceding display to define \( \xi_{m,l,l'} \) as the resulting first term in square brackets and \( \psi_{m,l,l'} \) as the new second term in square brackets. Let \( \Xi_m := (\xi_{m,l,l'})_{l,l' \in \mathcal{M} \setminus \{m\}} \) and \( \Psi_m := (\psi_{m,l,l'})_{l,l' \in \mathcal{M} \setminus \{m\}} \). By construction, for all \( m \in \mathcal{M} \) we then have \( \hat{\Sigma}_{n,m} = \Sigma_m + \Pi_m + o_0(1) \) and \( \hat{\Sigma}_{n,m} = \Xi_m + \Psi_m + o_0(1) \), and \( \Sigma_m, \Pi_m, \Xi_m, \) and \( \Psi_m \) are positive semi-definite with \( \Pi_m = \Psi_m = 0 \).

Finally, I state two results that are needed below. In particular, Lemma 3.13 establishes the asymptotic distribution of \( J^*_{n,m} \) defined in step 1 of Procedure 3.9 with the random index \( \hat{m} \) equal to a fixed index \( m \in \mathcal{M} \); this statistic is denoted by \( J^*_{n,m} \).

**Lemma 3.13.** Suppose we are in the situation of Theorem 3.11. For all \( m \in \mathcal{M} \),
$\mathcal{M}$, we have $J_{n,m}^* \rightsquigarrow \mathcal{X}$ in probability, where $\mathcal{X} \sim \chi_{M-1}^2$.

**Lemma 3.14.** Suppose the conditions of Theorem 3.11 hold. If $m^* \in \mathcal{M}$, then

$$\limsup_{n \to \infty} \mathbb{P}^*(MJ_n^* \leq x) - \mathbb{P}(MJ_n \leq x) = 0. \quad (3.6)$$

### 3.7 Proofs

**Proof of Lemma 3.5.** Both parts can be established by routine arguments for Wald statistics. The details are therefore omitted.

**Proof of Theorem 3.7.** (i) In view of Lemma 3.5, this can be shown by arguments similar to those used to prove Theorem 1 of Berger (1982). (ii) Use Lemma 3.5(ii) and the continuity of the minimum function.

**Proof of Lemma 3.8.** By Lemma 3.5, for every $\epsilon > 0$ we can find $B > 0$ such that

$$\limsup_{n \to \infty} \mathbb{P}(\hat{m} \neq m^*) = \limsup_{n \to \infty} \mathbb{P}(\exists m \in \mathcal{M} \setminus \{m^*\}: J_{n,m} < J_{n,m^*})$$

$$\leq \sum_{m \in \mathcal{M} \setminus \{m^*\}} \lim_{n \to \infty} \mathbb{P}(J_{n,m} \leq B) + \sup_{n \in \mathbb{N}} \mathbb{P}(J_{n,m^*} > B) < \epsilon,$$

which can be made arbitrarily small by choosing $B$ large enough. To see the second part, note that

$$\mathbb{P}(\hat{m} = m^* \mid MJ_n \leq c_{1-\alpha}) = \frac{\mathbb{P}(J_{n,m^*} \leq c_{1-\alpha})}{\mathbb{P}(MJ_n \leq c_{1-\alpha})},$$

which converges to one by Lemma 3.5(i) and Theorem 3.7(i).

**Proof of Lemma 3.13.** I first show $\hat{\Sigma}_{n,m}^{*-1/2} \hat{\lambda}_{n,m}^* \rightsquigarrow N_{M-1}(0, I_{M-1})$ in probability. Since

$$\hat{\Sigma}_{n,m}^{*-1/2} \hat{\lambda}_{n,m}^* = \hat{\Sigma}_{n,m}^{-1/2} \left(n^{-1/2} \hat{\beta}_{m}^T X_m^T P_m H \hat{u} \right)_{l \in \mathcal{M} \setminus \{m\}} + o_{p^*}(1) \text{ in probability},$$

it suffices to argue that the first term on the right-hand side satisfies a Liapounov condition in probability.
Let $E^r(\cdot) := E(\cdot \mid y, \mathcal{F})$, $\hat{\Gamma}_{m,l} := \left( X_m^T X_m \right)^{-1} X_m^T X_l$, and note that

\[
\begin{align*}
n^{-1 - \delta} \sum_{i=1}^n E^r \left[ \left( \hat{\beta}_m^T \hat{\Gamma}_{l,m}^T (x_{i,l} - \hat{\Gamma}_{l,m} x_{i,m}) \eta_i \hat{u}_{i,m} \right)_{l \in M \setminus \{m\}} \right]^{2 + \delta} \\
&\leq \left( \frac{2(M - 1)}{n} \right)^{1 + \delta} \sum_{i=1}^n \sum_{l \in M \setminus \{m\}} E^r \left[ \left| \hat{\beta}_m^T \hat{\Gamma}_{l,m}^T x_{i,l} \eta_i \hat{u}_{i,m} \right|^{2 + \delta} \right] \\
&\quad + \left| \hat{\beta}_m^T \hat{\Gamma}_{l,m}^T \hat{\Gamma}_{l,m} x_{i,m} \eta_i \hat{u}_{i,m} \right|^{2 + \delta}) \\
&\leq \left( \frac{2(M - 1)}{n} \right)^{1 + \delta} \sum_{l \in M \setminus \{m\}} \left( (|\hat{\beta}_m| ||\hat{\Gamma}_{l,m}||)^{2 + \delta} E^r |\eta|^{2 + \delta} \sum_{i=1}^n |x_{i,l} \hat{u}_{i,m}|^{2 + \delta} \right) \\
&\quad + \left( (|\hat{\beta}_m| ||\hat{\Gamma}_{l,m}||^{2 + \delta} E^r |\eta|^{2 + \delta} \sum_{i=1}^n |x_{i,m} \hat{u}_{i,m}|^{2 + \delta} \right) = O_P(n^{-\delta}),
\end{align*}
\]

where the first inequality uses the $c_r$ inequality and the second inequality applies the fact that $\eta$ is iid and submultiplicativity of the Frobenius norm. The equality follows from another application of the $c_r$ inequality to the sums involving $n$ such that

\[
\sum_{i=1}^n |x_{i,l} \hat{u}_{i,m}|^{2 + \delta} \leq 3^{1 + \delta} \sum_{i=1}^n \left( |x_{i,l} \hat{u}_{i,m}|^{2 + \delta} + ||x_{i,l} x_{i,m}^T||^{2 + \delta} |\beta_m|^{2 + \delta} \right)
\]

which is $O_P(n)$, and $|\hat{\beta}_m|$, $||\hat{\Gamma}_{l,m}||$, and $||\hat{\Gamma}_{l,m}||$ are $O_P(1)$. Pólya’s Theorem (Theorem 11.2.9 of Lehmann and Romano, 2005, p. 429) then implies $\hat{\Sigma}_{n,m}^{1/2} \lambda_{n,m}^* \rightsquigarrow N_{M-1}(0, I_{M-1})$ in probability and therefore $J_{n,m}^* \rightsquigarrow \chi_{M-1}^2$ in probability by continuity. \hfill \square

**Proof of Lemma 3.14.** Let $\hat{m}^* := \arg \min J_n^*$. I first show that if $m^* \in \mathcal{M}$, then $\hat{m}^*$ approximates $m^*$. Notice that for all $m \in \mathcal{M} \setminus \{\hat{m}\}$, we have $\mathbb{P}^*(J_{n,m} \leq B) = 1(J_{n,m} \leq B)$ since $J_{n,m}$ is constant with respect to $\mathbb{P}^*$ and therefore

\[
\mathbb{P}^*(\hat{m}^* \neq \hat{m}) \leq 1(\exists m \in \mathcal{M} \setminus \{\hat{m}\} : J_{n,m} \leq B) + \mathbb{P}^*(J_{n,\hat{m}}^* > B). \tag{3.7}
\]

Without loss of generality, fix any $0 < \varepsilon < 1$; then $\mathbb{P}[1(\exists m \in \mathcal{M} \setminus \{\hat{m}\} : J_{n,m} \leq B) > \varepsilon] \leq \mathbb{P}(\exists m \in \mathcal{M} \setminus \{m^*\} : J_{n,m} \leq B) + \mathbb{P}(\hat{m} \neq m^*)$, which converges to zero as $n \to \infty$ by Lemmas 3.5(ii) and 3.8. Further, $\mathbb{P}[\mathbb{P}^*(J_{n,\hat{m}}^* > B)$
$B > \varepsilon \leq \mathbb{P}(\hat{m} \neq m^*) + \sup_{n \in \mathbb{N}} \mathbb{P}[\mathbb{P}^*(J_{n,m}^* > B) > \varepsilon]$. Since the second term on the right-hand side can be shown to be small for large $B$ by arguments similar to those used to prove Prohorov's Theorem (Theorem 2.4 of van der Vaart, 1998, p. 8), conclude that (3.7) and therefore also similar to those used to prove Prohorov’s Theorem (Theorem 2.4 of van der Vaart, 1998, p. 8), conclude that (3.7) and therefore also $\mathbb{P}^*(\hat{m}^* \neq \hat{m})$ is $o_p(1)$; but then $\mathbb{P}[\mathbb{P}^*(\hat{m}^* \neq \hat{m}) > \varepsilon] + \mathbb{P}(\hat{m} \neq m^*) = o(1)$ for all $\varepsilon > 0$. Since $\mathbb{P}^*(J_{n,m}^* \leq x) = \mathbb{P}^*(J_{n,\hat{m}}^* \leq x) + \mathbb{P}^*(\{J_{n,m}^* \leq x\} \cap \{\hat{m}^* \neq m^*\})$ for all $x \in \mathbb{R}$, and $\mathbb{P}^*(\{J_{n,m}^* \leq x\} \cap \{\hat{m}^* \neq m^*\}) \leq \mathbb{P}^*(\hat{m}^* \neq m^*) = o_\varepsilon(1)$, we have $\lim_{n \to \infty} \mathbb{P}^*(J_{n,\hat{m}}^* \leq x) = K(x)$, where $K$ is the distribution function of a $\chi^2_{M-1}$ variable. Pólya’s Theorem then implies $MJ_n = J_{n,\hat{m}}^* \rightsquigarrow \chi^2_{M-1}$ in probability. Finally, because the supremum in (3.6) is bounded above by $\sup_x |\mathbb{P}^*(MJ_n \leq x) - K(x)| + \sup_x |\mathbb{P}(MJ_n \leq x) - K(x)|$, the result follows from Theorem 3.7(i). \hfill \Box

**Proof of Theorem 3.11.** (i) Lemma 3.14 combined with Lehmann and Romano’s (2005, p. 430) Lemma 11.2.1 implies $\lim_{n \to \infty} c^*_{n,1-\alpha} = K^{-1}(1-\alpha)$, and therefore $MJ_n - c^*_{n,1-\alpha} \rightsquigarrow \mathcal{X} - K^{-1}(1-\alpha)$ by the Slutsky Lemma, where $\mathcal{X} \sim \chi^2_{M-1}$; but $\mathbb{P}[\mathcal{X} - K^{-1}(1-\alpha) \leq x] = K[K^{-1}(1-\alpha) + x]$ is continuous in $x$ and thus

$$|\mathbb{P}(MJ_n > c^*_{n,1-\alpha}) - \alpha| \leq \sup_{x \in \mathbb{R}} |\mathbb{P}(MJ_n - c^*_{n,1-\alpha} \leq x) - K[K^{-1}(1-\alpha) + x]|$$

converges to zero by Pólya’s Theorem.

(ii) Suppose for now that the random index $\hat{m}$ that selects the model that is bootstrapped is equal to some fixed $l \in \mathcal{M}$. Define $\mathcal{J}_n^*(l) := \{J_{n,l}^*\} \cup \mathcal{J}_n \setminus \{J_{n,l}\}$ and notice that $\mathcal{J}_n^*(\hat{m}) = \mathcal{J}_n^*$. Lemma 3.13 implies that $J_{n,l}^* \rightsquigarrow \chi^2_{M-1}$ in probability. To deal with the other elements of $\mathcal{J}_n^*(l)$, pick any $m \in \mathcal{M} \setminus \{l\}$ and note that for any $0 < \varepsilon < 1$, we can find a $B > 0$ such that $K(B) > 1 - \varepsilon$. Thus, $\mathbb{P}^*(J_{n,m} \leq B) \geq \mathbb{P}^*|1 - K(J_{n,m})| > \varepsilon$. Because $\mathbb{P}(J_{n,m} \leq B) = o_p(1)$ in probability by Lemma 3.5(ii), the asymptotic distribution of $K(J_{n,m})$, $m \in \mathcal{M} \setminus \{l\}$, therefore converges to $\mathbb{P}^*$-point mass at 1 in probability.

By an “in probability” version of Theorem 2.7(v) of van der Vaart (1998, p. 10), we can now strengthen the marginal convergence of each element of $\mathcal{J}_n^*(l)$ to the joint convergence of the vector

$$\mathcal{K}_l := (K(J_{n,1}), \ldots, K(J_{n,l}), \ldots, K(J_{n,M}))$$

81
such that
\[ \text{plim}_{n \to \infty} \mathbb{P}(K_l \leq x) = \mathbb{P}[(1, \ldots, K(X_l), \ldots, 1) \leq x] \quad (3.8) \]
for each \( x \in \mathbb{R}^M \) at which the right-hand side is continuous, where \( X_l \sim \chi^2_{M-1} \).

The Continuous Mapping Theorem and Pólya’s Theorem in probability then yield
\[ \min_{l \in \mathcal{M}} K_l \rightsquigarrow \min(K(X_l), 1) \sim U_l \sim \text{Uniform}(0, 1) \text{ in probability}. \]

There are \( M \) different ways of choosing the index \( l \) and therefore there are \( M \) different bootstrap procedures that can be carried out. The bootstrapped variables of each of the procedures, i.e., \( (J^*_{n,1}, \ldots, J^*_{n,M}) \), are independent conditional on \( y \) and \( \mathcal{F} \), and therefore the components of the vector \( (\min K_1, \ldots, \min K_M) \) are also conditionally independent. By Lévy’s Continuity Theorem (Theorem 11.2.2 of Lehmann and Romano, 2005, p. 426) and Pólya’s Theorem in probability, the marginal convergence then implies the joint convergence \( (\min K_1, \ldots, \min K_M) \rightsquigarrow (U_1, \ldots, U_M) \) in probability, where the \( U_l \), \( l \in \mathcal{M} \), are independent. Since \( K^{-1} \) is continuous and \( K \) is increasing, conclude that
\[ MJ_n^* := \max_{l \in \mathcal{M}} \min_l J^*_n(l) = K^{-1} \left( \max_{l \in \mathcal{M}} \min_l K_l \right) \rightsquigarrow K^{-1} \left( \max_{l \in \mathcal{M}} U_l \right) \text{ in probability}. \]

Notice that \( MJ_n^* = \min J^*_n(\hat{m}) \) and thus \( \mathbb{P}^*(MJ_n^* > B) \leq \mathbb{P}^*(\overline{MJ}_n^* > B) \). By an “in probability” version of Prohorov’s Theorem, for any \( 0 < \varepsilon < 1 \) and \( \delta > 0 \), we can then find a \( B \) such that \( \sup_{n \in \mathbb{N}} \mathbb{P}[\mathbb{P}^*(MJ_n^* > B) > \varepsilon] \leq \sup_{n \in \mathbb{N}} \mathbb{P}[\mathbb{P}^*(\overline{MJ}_n^* > B) > \varepsilon] < \delta \). Conclude from Lemma 21.1(i) of van der Vaart (1998, p. 304) that \( \mathbb{P}(c_{n,1-\alpha}^* > B) = \mathbb{P}[\mathbb{P}^*(MJ_n^* > B) > \alpha] < \delta \) uniformly in \( n \) for any fixed \( \alpha \in (0, 1) \). This result and Lemma 3.5(ii) imply that for large enough \( n \) the right-hand side of the inequality \( \mathbb{P}(MJ_n \leq c_{n,1-\alpha}^*) \leq \mathbb{P}(MJ_n \leq B) + \mathbb{P}(c_{n,1-\alpha}^* > B) \) can be made as small as desired, which completes the proof.

**Proof of Corollary 3.12.** Identical to the proof of Lemma 3.8, mutatis mutandis.
REFERENCES


