The Weighted Integrated Conditional Moment Test for the Validity of Stationary Time Series Regression Models

Herman J. Bierens
Pennsylvania State University
University Park, PA 16802, USA

August 27, 2015

Abstract
In Bierens (1984) I have proposed a consistent weighted integrated conditional moment (WICM) test for time series regression models. However, at that time I was unable to derive the asymptotic null distribution of the test statistic involves. Therefore, I proposed to use upper bounds of the critical values based in Chebyshev’s inequality for first moments. In this paper I will update the 1984 paper by deriving the asymptotic null distribution of the WICM test, much sharper upper bounds of the critical values than those based on Chebyshev’s inequality, and bootstrap critical values. Also, I will solve the problem how to standardize the conditioning variables before applying a bounded transformation such that all the asymptotic results carry over. This was one of the unsolved problems in Bierens (1984).

*This paper is the basis for the Rolf Mantel Lecture to be delivered at the 2015 annual meeting of the Argentine Economic Association on November 11 to 13 in Salta City, Argentina.
†Professor Emeritus of Economics. Please address any correspondence by e-mail only to hbierens@psu.edu, as I no longer have an office in the Department of Economics of Penn State.
1. Introduction

To the best of my knowledge, my papers Bierens (1982) [B82 hereafter] and its companion paper Bierens (1984) [B84 hereafter] are the first papers ever to propose consistent tests of the null hypothesis that the functional form of a (non)linear (time series) regression model is correctly specified as a conditional expectation, against all deviations from the null hypothesis. I wrote these papers in the fall of 1981 while enjoying the hospitality of the University of Minnesota, Minneapolis, as a postdoc.¹

In B82 I consider cross-section regression models, and in B84 time series regression models. However, at that time I did not know how to derive the limiting null distribution of the test statistics involved, but I did know how to consistently estimate their expectations under the null hypothesis. Therefore, in these papers I proposed to use upper bounds of the critical values based on Chebyshev’s inequality for first moments. Admittedly, these upper bounds are very conservative.

In B82 I proposed two test for the validity of the functional specification of possibly nonlinear cross-section regression models. For reasons to become clear in the next section, test 1 in B82 was named by Bierens and Ploberger (1997) the Integrated Conditional Moment (ICM) test, also known in the statistical literature as a Cramér-von Mises type test. The second test in B82, which is a kind of Kolmogorov-Smirnov type test, grew out of frustration with my inability to derive the null distribution of test 1. However, this test is actually inferior to test 1, and will therefore be ignored.²

But in view of the results in Bierens (1990), Bierens and Ploberger (1997), Bierens and Wang (2012) and Bierens (2014, 2015a,b,c) I now know how the limiting distribution of the ICM test looks like, how to approximate the critical values by a bootstrap method, and how to derive much sharper upper bounds of these critical values than the ones based on Chebyshev’s inequality. In the next section I will briefly review the main results for cross-section regression models, because they are needed to update B84. The details and proofs of these new

¹See Pinkse (2013) for more on how these papers came about.
²In hindsight, test 2 in B82 is a pretty bad idea because it is obvious from the asymptotic normality result (24) in B82, that the rate of convergence of the proposed test statistic under $H_0$ is lower than $\sqrt{n}$, and via the approach in Bierens and Ploberger (1996) it can be shown that this test has no power against $\sqrt{n}$ local alternatives.
results are given in Bierens (2015a).

Paper B84 extends the ICM approach in B82 to time series regression models. In B84 I proposed a weighted integrated conditional moments (WICM) tests, where a sequence of ICM tests for increasing lag lengths are combined in a weighted sum. The main purpose of the current paper is to bring paper B84 up-to-date. This is much more complicated than in the case of B82 because the correctness of the specification of a time series regression models is equivalent to the condition that the model errors form a martingale difference process relative to the σ-algebra generated by the infinite past of the dependent variable and current and past exogenous variables.


Moreover, in Bierens and Wang (2012) and Bierens (2015c) we have generalized the test in B82 to a consistent test for the correctness of parametric specifications of conditional distribution models.

As shown in Bierens and Ploberger (1997), the ICM test has nontrivial power against \( \sqrt{n} \) local alternatives.\(^3\) This desirable property holds for the other ICM type tests in strand (1) as well, although often not mentioned by the authors. On the other hand, all the tests of strand (2) have only trivial power, i.e., the power and size are equal, against these \( \sqrt{n} \) local alternatives, due to the slower rate of convergence of nonparametric estimators than the standard parametric rate \( \sqrt{n} \).

Most specification tests for regression-type time series models proposed in the statistical and econometric literature, including Bierens and Ploberger (1997), only test implications or aspects of the martingale difference hypothesis rather

\(^3\)For example, in the case of a correctly specified nonlinear regression model \( Y = f(X, \theta_0) + U \), where \( E[U|X] = 0 \) a.s., a typical \( \sqrt{n} \) local alternative takes the form \( Y = f(X, \theta_0) + g(X)/\sqrt{n} + U \), where \( n \) is the sample size.
than this hypothesis itself. For example, Hong (1999), Hong and Lee (2005), Escanciano and Velasco (2006) and Su and White (2007) proposed tests for pairwise time series independence. Only if it also assumed that the errors are jointly normal, together with some additional conditions, would pairwise independence imply that the errors are martingale differences. Dominguez and Lobato (2003) and Stute et al. (2006) extend the approach of Stute (1997) to time series models, but with a fixed number of lagged conditioning variables. Therefore, these authors only test an implication of the martingale difference hypothesis rather than the latter hypothesis itself.

To the best of my knowledge the only genuine consistent tests of the martingale difference hypothesis for time series regression errors are the WICM tests in B84 and de Jong’s (1996) generalization to an ICM test for time series regressions of the approach in Bierens (1990). The test in the latter paper is of the Kolmogorov-Smirnov type with a selection trick based on the difference of the rate of convergence to under the null hypothesis and the alternative hypothesis that the null is false, and it has been shown in Bierens (2014) that under some conditions this test has nontrivial power against $\sqrt{n}$ local alternatives. De Jong (1996) qualifies this trick an "abuse of asymptotic theory" (with which I now agree), and therefore uses an infinite dimensional ICM setup, with a bootstrap approach for the critical values involved.

The setup of the current paper is as follows. In section 2 I will briefly discuss the ICM test for cross-section models in B82, with additional results based on Bierens (2015a). These results are the basis for the update of B84. In section 3 I will introduce the WICM test for time series regressions in B84, and in section 4 I will set forth conditions for the consistency and asymptotic normality of nonlinear least squares (NLLS) estimators. In section 5 I will derive the limiting null distribution of the WICM test, and in section 6 I will show that this test is consistent. In sections 7 I will derive upper bounds of the critical values similar to Bierens and Ploberger (1997) and in section 8 I will show how to approximate the critical values by a bootstrap method. In section 9 I will address the initial values problem, because all time series data have a finite starting time, whereas the asymptotic theory of the WICM test uses the entire infinite past of the time series involved. Section 10 briefly reviews the ideas behind de Jong’s (1996) ICM test for time series regressions. In section 11 I will propose a martingale difference structure preserving standardization procedure for the conditioning lagged variables, and show that all the asymptotic results of the WICM test carry over. In section 12 I will discuss briefly how to deal with nonlinear time series models
with ARMA type errors and other linear and nonlinear models involving infinitely many lagged conditioning variables. In section 13 I will show how to implement the test and the bootstrap procedure in practice, and in section 14 the test is applied to a numerical example, where the data generating process is a Gaussian MA(1) process and the null hypothesis is that this is an AR(1) process. In the last section I summarize the results in this paper, and discuss some ideas for improving the finite sample power of the ICM and WICM tests.

2. The ICM test for cross-section regression models.

Cross-section regression models aim to represent the conditional expectation of a dependent variable $Y$ given a vector $X \in \mathbb{R}^k$ of explanatory variables. In particular, consider the possibly nonlinear regression model

$$Y = f(X, \theta_0) + U, \quad \theta_0 \in \Theta \subset \mathbb{R}^m,$$

where $\Theta$ is a given parameter space, $f(x, \theta)$ is an a priori specified real (and usual continuous) function on $\mathbb{R}^k \times \Theta$, and $U$ is the error term. As is well-known, given that $E[Y^2] < \infty$ the conditional mean square error

$$E\left[ (Y - f(X, \theta))^2 | X \right] = E\left[ (Y - E[Y|X])^2 | X \right] + (E[Y|X] - f(X, \theta))^2$$

is minimal if for some $\theta_0 \in \Theta$, $\Pr \left[ E[Y|X] = f(X, \theta_0) \right] = 1$, which is equivalent to the usual assumption that the error term $U$ corresponding to $\theta_0$ satisfies

$$H_0 : \Pr \left[ E[U|X] = 0 \right] = 1.$$

In other words, the regression function $f(x, \theta_0)$ is correctly specified as a conditional expectation function if and only if (2.2) holds.

Given that $E[(f(X, \theta))^2] < \infty$ for all $\theta \in \Theta$, we may without loss of generality interpret $\theta_0$ as

$$\theta_0 = \arg \min_{\theta \in \Theta} E[(Y - f(X, \theta))^2] = \arg \min_{\theta \in \Theta} E \left[ (E[Y|X] - f(X, \theta))^2 \right]$$

regardless whether (2.2) is true or not. The latter case corresponds to

$$H_1 : \Pr \left[ E[U|X] = 0 \right] < 1.$$

The ICM test in B82 tests the null hypothesis (2.2) against the alternative hypothesis (2.4), given a random sample $\{(Y_j, X_j^T)\}_{j=1}^n$ from the distribution of
under the maintained (standard) conditions for the consistency and asymptotic normality of the nonlinear least squares estimator \( \hat{\theta}_n \) of \( \theta_0 \). This test is consistent because it has asymptotic power 1. The test is based on a complex-valued empirical process of the form

\[
\hat{W}_n(\tau) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{U}_{n,j} \exp (i \tau' \Phi(X_j)), \tag{2.5}
\]

where \( i = \sqrt{-1}, \hat{U}_{n,j} = Y_j - f(X_j, \hat{\theta}_n) \) is the nonlinear regression residual of observation \( j \), and \( \Phi \) is a Borel measurable bounded one-to-one mapping with Borel measurable inverse.

Under the standard conditions listed in B82,

\[
\frac{\hat{W}_n(\tau)}{\sqrt{n}} \xrightarrow{a.s.} E[U \exp (i \tau' \Phi(X))] = E[E[U|X] \exp (i \tau' \Phi(X))] = \eta(\tau), \text{ say,} \tag{2.6}
\]

pointwise in \( \tau \in \mathbb{R}^k \) and uniform on any compact subset \( \Upsilon \) of \( \mathbb{R}^k \), regardless whether \( H_0 \) is true or not. Obviously, \( \eta(\tau) \equiv 0 \) under \( H_0 \), whereas under \( H_1 \) we have that \( \eta(\tau) \neq 0 \) for a \( \tau \) in an arbitrary small neighborhood of the origin of \( \mathbb{R}^k \). By the continuity of \( \eta(\tau) \), the latter implies that \( |\eta(\tau)| > 0 \) in an open set arbitrarily close to the origin of \( \mathbb{R}^k \). Note that if we replace \( \Phi(X) \) by \( X \) and if \( ||X|| \) is unbounded then this statement would read: \( |\eta(\tau)| > 0 \) in an open set somewhere in \( \mathbb{R}^k \), but where in \( \mathbb{R}^k \) is undetermined. Also, it has been shown in B82 that under \( H_0 \),

\[
\sup_{\tau \in \Upsilon} |\hat{W}_n(\tau) - W_n(\tau)| = o_p(1) \tag{2.7}
\]

for an arbitrary compact subset \( \Upsilon \) of \( \mathbb{R}^k \), where \( W_n(\tau) \) is a complex valued empirical process of the form

\[
W_n(\tau) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_j \rho(X_j, \tau), \tag{2.8}
\]

with \( U_j = Y_j - f(X_j, \theta_0) \) and \( \rho(X_j, \tau) \) defined similar to the expression (47) in B82. Choosing the compact set \( \Upsilon \) such that the origin of \( \mathbb{R}^k \) is contained in its

\footnote{Note that due to the definition of \( \theta_0 \) as (2.3) and the i.i.d. condition regarding the data the conditions for \( \hat{\theta}_n \xrightarrow{a.s.} \theta_0 \) and \( \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N_m(0, \Sigma) \) are the same under \( H_0 \) and \( H_1 \).}
interior, it follows now that
\[ \int_{\mathcal{Y}} |\hat{W}_n(\tau)|^2 d\tau = \int_{\mathcal{Y}} |W_n(\tau)|^2 d\tau + o_p(1) \quad \text{under } H_0, \]
\[ \frac{1}{n} \int_{\mathcal{Y}} |\hat{W}_n(\tau)|^2 d\tau \xrightarrow{D} \int_{\mathcal{Y}} |\eta(\tau)|^2 d\tau > 0 \quad \text{under } H_1. \]
Moreover, under \( H_0 \),
\[
E \left[ \int_{\mathcal{Y}} |W_n(\tau)|^2 d\tau \right] = \int_{\mathcal{Y}} E \left[ |W_n(\tau)|^2 \right] d\tau = E \left[ U^2 \int_{\mathcal{Y}} \rho(X, \tau) \overline{\rho(X, \tau)} d\tau \right] \]
\[ = E \left[ U^2 \left( \int_{\mathcal{Y}} (\text{Re} [\rho(X, \tau)])^2 d\tau + \int_{\mathcal{Y}} (\text{Im} [\rho(X, \tau)])^2 d\tau \right) \right] \]
\[ = E_0, \text{ say.} \]
This expectation can be estimated consistently. Denoting this estimator by \( \hat{E}_n \),
\[ \hat{T}_n = \int_{\mathcal{Y}} |\hat{W}_n(\tau)|^2 d\tau / \hat{E}_n, \]
it follows that under \( H_0 \),
\[
\limsup_{n \to \infty} \Pr \left[ \hat{T}_n > \alpha^{-1} \right] = \limsup_{n \to \infty} \Pr \left[ \int_{\mathcal{Y}} |W_n(\tau)|^2 d\tau > \alpha^{-1} E_0 \right] \leq \alpha \quad (2.9) \]
for \( \alpha \in (0, 1) \), where the inequality follows from Chebyshev’s inequality for first moments. Thus, \( 1/\alpha \) is an upper bound of the asymptotic \( \alpha \times 100\% \) critical value of \( \hat{T}_n \). This test is consistent because under \( H_1 \),
\[
\lim_{n \to \infty} \Pr \left[ \hat{T}_n > \alpha^{-1} \right] = I \left( \int_{\mathcal{Y}} |\eta(\tau)|^2 d\tau > 0 \right) = 1, \]
where \( I(.) \) is the well-known indicator function.
This is how far I got in 1981!
But I now know (see Bierens 2015a) that under \( H_0 \),
\[
\int_{\mathcal{Y}} |\hat{W}_n(\tau)|^2 d\tau \xrightarrow{d} \int_{\mathcal{Y}} |W(\tau)|^2 d\tau \sim \sum_{i=1}^{\infty} \omega_i \varepsilon_i^2, \quad (2.10) \]
where \( W(\tau) \) is a complex-valued zero-mean continuous Gaussian process on \( \mathcal{Y} \) with covariance function \( \Gamma(\tau_1, \tau_2) = E \left[ W(\tau_1) \overline{W(\tau_2)} \right] \), the \( \varepsilon_i \)’s are i.i.d. standard
normal random variables and the \( \omega_i \)'s are positive constants\(^5\) satisfying
\[
\sum_{i=1}^{\infty} \omega_i = \int_{\mathcal{Y}} E[|W(\tau)|^2] d\tau < \infty.
\]

Hence under \( H_0 \),
\[
\tilde{T}_n \overset{d}{\rightarrow} \frac{\sum_{i=1}^{\infty} \omega_i \varepsilon_i^2}{\sum_{i=1}^{\infty} \omega_i} \leq \sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 = \chi_1^2, \quad \text{say,}
\]
where the inequality follows from Bierens and Ploberger (1997, Theorem 7).

Therefore, for \( \alpha \in (0, 1) \) and \( \overline{\sigma}(\alpha) \) such that \( \Pr[\chi_1^2 > \overline{\sigma}(\alpha)] = \alpha \),
\[
\lim \sup_{n \to \infty} \Pr \left[ \tilde{T}_n > \overline{\sigma}(\alpha) \right] \leq \Pr \left[ \chi_1^2 > \overline{\sigma}(\alpha) \right] = \alpha
\]

The values of the upper bounds \( \overline{\sigma}(\alpha) \) of the \( \alpha \times 100\% \) critical values for \( \alpha = 0.01 \), \( \alpha = 0.05 \) and \( \alpha = 0.10 \) have been calculated in Bierens and Ploberger (1997), i.e.,
\[
\overline{\sigma}(0.01) = 6.81, \quad \overline{\sigma}(0.05) = 4.26, \quad \overline{\sigma}(0.10) = 3.23.
\]

Note that for these \( \alpha \)'s the corresponding upper bounds \( 1/\alpha \) of the critical values derived from Chebyshev’s inequality are much larger than \( \overline{\sigma}(\alpha) \).

As shown in Bierens (2015a), the exact critical values of the ICM test involved can be approximated by a bootstrap method, but this procedure is computational intensive. Therefore, it is recommended to use the new upper bounds (2.12) of the critical values first, and only use bootstrap critical values if in the former case the null hypothesis is not rejected at the 5% significance level or less.

Finally, there is no need to confine the integration range \( \mathcal{Y} \) to compact subsets containing the origin of \( \mathbb{R}^k \) be cause the complex \( \exp(.) \) function satisfies the conditions of Theorem 1 in Bierens and Ploberger (1997). Therefore, under \( H_1 \) the set
\[
S = \{ \tau \in \mathbb{R}^k : E[U \exp(i.\tau'\Phi(X))] = 0 \}
\]
has Lebesgue measure zero and is nowhere dense. Consequently, we may choose for \( \mathcal{Y} \) any compact subset of \( \mathbb{R}^k \) with positive Lebesgue measure. Moreover,\(^5\)

\(^5\)Note that in Bierens and Ploberger (1997) the covariance function \( \Gamma(\tau_1, \tau_2) \) is real valued, and then the \( \omega_i \)'s in (2.10) are the eigenvalues of \( \Gamma(\tau_1, \tau_2) \). However, in the complex case under review this is no longer the case.
as shown in Bierens and Ploberger (1997), Stinchcombe and White (1998) and Bierens (2015b), the following more general result holds.

**Theorem 2.1.** Let \( w(u) \) be a real or complex valued nonpolynomial analytical function on \( \mathbb{R} \), with higher-order derivatives \( w_n(u) = (d/du)^n w(u), n \geq 1 \)
\( w_0(u) = w(u) \), satisfying \( w_n(0) \neq 0 \) for all \( n \geq 0 \). Moreover, let \( U, X \) and \( \Phi \) be the same as before, with \( \Pr (E[U|X] = 0) < 1 \). Then the set
\[
S = \{ \tau \in \mathbb{R}^k : E[U.w(\tau'\Phi(X))] = 0 \}
\]
has Lebesgue measure zero and is nowhere dense.

These conditions apply to \( w(u) = \exp(i.u), w(u) = \exp(u), w(u) = \cos(u) + \sin(u), \) and a wide range of other non-polynomial analytical functions.

Therefore, the results above and below carry over for any weight function \( w(u) \) satisfying the conditions of Theorem 2.1. Moreover, without loss of generality we may replace all Lebesgue integrals involved by integrals with respect to an absolutely continuous probability measure \( \mu \) on \( \mathbb{Y} \), simply by replacing \( d\tau \) with \( d\mu(\tau) \).

**3. The weighted ICM test for time series regression models**

As in B84, I will focus on the nonlinear ARX\((p, q)\) model
\[
Y_t = f(Y_{t-1}, Y_{t-2}, ..., Y_{t-p}, X_{t-1}, X_{t-2}, ..., X_{t-q}, \theta_0) + U_t
\]
(3.1)
\[
= f_{t-1}(\theta_0) + U_t, \text{ \, say,} \]
(3.2)
where (3.2) is merely a short-hand notation for (3.1). The \( X_t \)'s are \( s \)-dimensional vectors of exogenous variables which for notational convenience enter the regression function (3.1) in lagged form. The presence of only lagged exogenous variables is no restriction because we can always define \( X_{t-1} = X_t^* \) where the latter is the actual vector of exogenous variables.

Throughout it will be assumes that the following conditions hold.

**Assumption 3.1.**
(a) \( Z_t = (Y_t, X_t^*)' \in \mathbb{R} \times \mathbb{R}^s \) is a strictly stationary vector time series process, with \( E[Y_t^2] < \infty \).
(b) \( \theta_0 \) is a parameter vector contained in a given compact parameter space \( \Theta \subset \mathbb{R}^m \).
(c) \( f(v, \theta) \) is a given real function on \( \mathbb{R}^{p+q.s} \times \Theta \) which for each \( v \in \mathbb{R}^{p+q.s} \) is continuous in \( \theta \in \Theta \), and for each \( \theta \in \Theta \) Borel measurable in \( v \in \mathbb{R}^{p+q.s} \), hence for each \( \theta \in \Theta \) the sequence \( f_{t-1}(\theta) \) is a well-defined strictly stationary time series process.

(d) \( \sup_{\theta \in \Theta} E[f_{t-1}(\theta)^2] < \infty \).

In general, time series regression models aim to represent the best one-step ahead forecasting scheme, in the sense that given the entire past of the time series involved up to time \( t - 1 \), the mean square forecast error is minimal. For the model (3.1) this is the case if and only if the error process \( U_t \) is a martingale difference sequence w.r.t. the \( \sigma \)-algebra

\[
\mathcal{F}_{-\infty}^t = \sigma \{ Z_{t-j} \}_{j=0}^\infty
\]
generated by the sequence \( \{ Z_{t-j} \}_{j=0}^\infty \), i.e.,

\[
U_t \text{ is measurable } \mathcal{F}_{-\infty}^t \text{ and } E[U_t|\mathcal{F}_{-\infty}^{t-1}] = 0 \text{ a.s.} \quad (3.3)
\]

More formally, using the short-hand notation (3.2), the null hypothesis to be tested is that

\[
H_0 : \text{ There exists a } \theta_0 \in \Theta \text{ such that } \Pr \left( E[Y_t|\mathcal{F}_{-\infty}^{t-1}] = f_{t-1}(\theta_0) \right) = 1, \quad (3.4)
\]

which is equivalent to the hypothesis that \( U_t = Y_t - f_{t-1}(\theta_0) \) is a martingale difference process, against the alternative hypothesis that \( H_0 \) is false, i.e.,

\[
H_1 : \text{ For all } \theta \in \Theta, \Pr \left( E[Y_t|\mathcal{F}_{-\infty}^{t-1}] = f_{t-1}(\theta) \right) < 1. \quad (3.5)
\]

Similar to (2.3), define \( \theta_0 \) as

\[
\theta_0 = \arg \min_{\theta \in \Theta} \mathbb{E} \left[ (Y_t - f_{t-1}(\theta))^2 \right] \quad (3.6)
\]

\[
= \arg \min_{\theta \in \Theta} \left\{ \mathbb{E} \left[ (Y_t - E[Y_t|\mathcal{F}_{-\infty}^{t-1}])^2 \right] + \mathbb{E} \left[ (E[Y_t|\mathcal{F}_{-\infty}^{t-1}] - f_{t-1}(\theta))^2 \right] \right\}
\]

\[
= \arg \min_{\theta \in \Theta} \mathbb{E} \left[ (E[Y_t|\mathcal{F}_{-\infty}^{t-1}] - f_{t-1}(\theta))^2 \right]
\]

regardless whether \( H_0 \) is true or not, and let \( U_t = Y_t - f_{t-1}(\theta_0) \). Then \( H_0 \) and \( H_1 \) become

\[
H_0 : \Pr \left( E[U_t|\mathcal{F}_{-\infty}^{t-1}] = 0 \right) = 1, \quad (3.7)
\]

\[
H_1 : \Pr \left( E[U_t|\mathcal{F}_{-\infty}^{t-1}] = 0 \right) < 1, \quad (3.8)
\]
respectively.

To determine which one of the two hypotheses is true, we need to condition on the one-sided infinite vector time series \( \{Z_{t-j}\}_{j=1}^{\infty} \), which in practice is not possible because \( Z_t \) is usually only observed from a particular time \( t_0 \) onwards. Part of the solution to this problem is the following lemma.

**Lemma 3.1.** Under Assumption 3.1,

\[
E \left[ U_t | \mathcal{F}_{-\infty}^{t-1} \right] = \lim_{k \to \infty} E \left[ U_t | \mathcal{F}_{t-k}^{t-1} \right] \text{ a.s.,} \tag{3.9}
\]

where \( \mathcal{F}_{t-k}^{t-1} = \sigma \left( \{Z_{t-j}\}_{j=1}^{k} \right) \) is the \( \sigma \)-algebra generated by \( Z_{t-1}, Z_{t-2}, ..., Z_{t-k} \). Consequently, under \( H_1 \) there exists an \( k_0 \in \mathbb{N} \) such that for all \( t \),

\[
\sup_{k \geq k_0} \Pr \left( E \left[ U_t | \mathcal{F}_{t-k}^{t-1} \right] = 0 \right) < 1. \tag{3.10}
\]

**Proof.** The result (3.9) is well-known. See for example Theorem 9.4.8 in Chung (1974) or Theorem 3.12 in Bierens (2004). As to part (3.10), let

\[
V_t = \left| E \left[ U_t | \mathcal{F}_{-\infty}^{t-1} \right] \right|, \quad V_{t,k} = \left| E \left[ U_t | \mathcal{F}_{t-k}^{t-1} \right] \right|.
\]

Then by (3.9), \( \lim_{k \to \infty} V_{t,k} = V_t \) a.s., which implies that \( V_{t,k} \overset{d}{\to} V_t \) for \( k \to \infty \) and thus, by definition of convergence in distribution,

\[
\lim_{k \to \infty} \Pr \left[ V_{t,k} > \varepsilon \right] = \Pr \left[ V_t > \varepsilon \right] \tag{3.11}
\]

for all continuity points \( \varepsilon \) of the distribution function of \( V_t \). Moreover, under the alternative hypothesis (3.8), \( \Pr \left[ V_t > 0 \right] > 0 \), which implies that there exists a continuity point \( \varepsilon > 0 \) of the distribution function of \( V_t \) such that \( \Pr \left[ V_t > \varepsilon \right] > 0 \) as well.\(^6\) It follows now from (3.11) that

\[
\lim \inf_{k \to \infty} \Pr \left[ V_{t,k} > 0 \right] \geq \lim_{k \to \infty} \Pr \left[ V_{t,k} > \varepsilon \right] = \Pr \left[ V_t > \varepsilon \right] > 0,
\]

hence

\[
\lim \sup_{k \to \infty} \Pr \left( E \left[ U_t | \mathcal{F}_{t-k}^{t-1} \right] = 0 \right) = \lim \sup_{k \to \infty} \Pr \left[ V_{t,k} = 0 \right] < 1. \tag{3.12}
\]

Since by the strict stationarity condition in Assumption 3.1 the probabilities in (3.12) do not depend on \( t \), the result (3.10) follows. \( \blacksquare \)

\(^6\)Because \( \Pr \left[ V_t > 1/n \right] \uparrow \Pr \left[ V_t > 0 \right] \) monotonically as \( n \to \infty \).
Now let $\Phi : \mathbb{R}^{s+1} \to \mathbb{R}^{s+1}$ be a bounded Borel measurable one-to-one mapping with Borel measurable inverse. Then

$$\mathcal{F}_{t-k}^{-1} = \sigma \left( \{ Z_{t-j} \}_{j=1}^{k} \right) = \sigma \left( \{ \Phi(Z_{t-j}) \}_{j=1}^{k} \right),$$

hence similar to the case (2.13),

$$H_1(k) : \Pr \left( E \left[ U_t | \mathcal{F}_{t-k}^{-1} \right] = 0 \right) < 1 \quad (3.13)$$

implies that the set

$$S_k = \left\{ \tau = (\tau_1', \tau_2', ..., \tau_k')' \in \mathbb{R}^{(s+1)k} : E \left[ U_t \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right] = 0 \right\}$$

has Lebesgue measure zero and is nowhere dense. Consequently, denoting

$$\psi_k(\tau_1, \tau_2, ..., \tau_k) = E \left[ U_t \exp \left( i \sum_{j=1}^{k} \tau_j \Phi(Z_{t-j}) \right) \right] \quad (3.14)$$

the following result holds.

**Lemma 3.2.** Assume that the conditions of Lemma 3.1 hold. Let $\Upsilon$ be a compact subset of $\mathbb{R}^{s+1}$ with positive Lebesgue measure and let $\mu$ be an absolutely continuous probability measure on $\Upsilon$. Then

$$\int_\Upsilon |\psi_k(\tau_1, \tau_2, ..., \tau_k)|^2 \, d\mu(\tau_1) \, d\mu(\tau_2) ... \, d\mu(\tau_k) \begin{cases} > 0 \text{ under } H_1(k), \\ = 0 \text{ under } H_0. \end{cases}$$

Given that $Z_t$ is observed for $1 - t_0 \leq t \leq n$, where $t_0 = \max(p, q)$, the empirical counter-part of $\psi_k(\tau_1, \tau_2, ..., \tau_k)$ defined in (3.14) is

$$\tilde{\psi}_{k,n}(\tau_1, \tau_2, ..., \tau_k) = \frac{1}{n} \sum_{t=1}^{n} (Y_t - f_{t-1}(\hat{\theta}_n)) \exp \left( i \sum_{j=1}^{\min(k, t-t_0)} \tau_j \Phi(Z_{t-j}) \right) \quad (3.15)$$

for $n > k + t_0$, where $\hat{\theta}_n$ is the nonlinear least estimator of $\theta_0$. 


However, for the time being I will assume that all the lagged \( Z_t \)'s are observed, so that

\[
\hat{\psi}_{k,n}(\tau_1, \tau_2, ..., \tau_k) = \frac{1}{n} \sum_{t=1}^{n} (Y_t - f_{t-1}(\hat{\theta}_n)) \exp \left( i \sum_{j=1}^{k} \tau_j \Phi(Z_{t-j}) \right)
\]  

(3.16)

will be treated as the empirical counter-part of \( \psi_k(\tau_1, \tau_2, ..., \tau_k) \). In section 9 I will set forth conditions under which the asymptotic results below on the basis of (3.16) are the same as for (3.15).

In order to keep this paper in tune with Bierens and Ploberger (1997), Bierens and Wang (2012) and Bierens (2015a,b,c), I will from this point onwards denote \( \sqrt{n} \hat{\psi}_{k,n}(\tau_1, \tau_2, ..., \tau_k) \) by \( \sqrt{n} \hat{W}_{k,n}(\tau_1, \tau_2, ..., \tau_k) \). Thus,

\[
\hat{W}_{k,n}(\tau_1, \tau_2, ..., \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t - f_{t-1}(\hat{\theta}_n)) \exp \left( i \sum_{j=1}^{k} \tau_j \Phi(Z_{t-j}) \right).
\]  

(3.17)

I will next set forth further conditions such that under \( H_0 \) the following results hold.

- For each \( k \in \mathbb{N} \),

\[
\hat{B}_{n,k} \overset{\text{def.}}{=} \int_{\Upsilon_k} \left| \hat{W}_{k,n}(\tau_1, \tau_2, ..., \tau_k) \right|^2 \text{d}\mu(\tau_1)\text{d}\mu(\tau_2)\cdots\text{d}\mu(\tau_k)
\]  

(3.18)

\[
\hat{B}_{k} \overset{\text{def.}}{=} B_k \overset{\text{d}}{\rightarrow} \int_{\Upsilon_k} \left| W_k(\tau_1, \tau_2, ..., \tau_k) \right|^2 \text{d}\mu(\tau_1)\text{d}\mu(\tau_2)\cdots\text{d}\mu(\tau_k),
\]

where \( W_k(\tau_1, \tau_2, ..., \tau_k) \) is a complex-valued zero-mean Gaussian process on \( \Upsilon_k \);

- For any positive sequence \( \gamma_k \) satisfying \( \sum_{k=1}^{\infty} \gamma_k < \infty \) and any subsequence \( L_n \) of \( n \) satisfying \( \lim_{n \to \infty} L_n = \infty \),

\[
\sum_{k=1}^{L_n} \gamma_k \hat{B}_{n,k} \overset{\text{d}}{\rightarrow} \sum_{k=1}^{\infty} \gamma_k B_k.
\]

- The critical values of \( \sum_{k=1}^{\infty} \gamma_k B_k \) can be approximated by a bootstrap method.
The upper bounds (2.12) of the critical values apply to
\[ \frac{\sum_{k=1}^{\infty} \gamma_k B_k}{\sum_{k=1}^{\infty} \gamma_k E[B_k]} \]
as well.

Moreover, under \( H_1 \) the following results hold.

- For each \( k \in \mathbb{N} \),
  \[ \hat{B}_{n,k}/n \xrightarrow{p} \eta_k \overset{\text{def.}}{=} \int_{\mathcal{T}_k} |\psi_k(\tau_1, \tau_2, ..., \tau_k)|^2 \, d\mu(\tau_1) \, d\mu(\tau_2) \cdots d\mu(\tau_k) \]
  (3.19)
  where \( \eta_k > 0 \) for all but a finite number of \( k \)'s,
- \( (1/n) \sum_{k=1}^{L_n} \gamma_k \hat{B}_{n,k} \xrightarrow{p} \sum_{k=1}^{\infty} \gamma_k \eta_k > 0. \)

The statistic
\[ \hat{T}_n = \sum_{k=1}^{L_n} \gamma_k \hat{B}_{n,k} \]
will be referred to as the test statistic of the Weighted Integrated Conditional Moment (WICM) test.

4. Nonlinear least squares estimation

4.1. Consistency

In B84 the consistency of the NLLS estimator was \( \hat{\theta}_n \) derived under data heterogeneity. However, under the current strict stationarity condition in Assumption 3.1 the following conditions suffice.

**Assumption 4.1.**

(a) The vector time series process \( Z_t \) in Assumption 3.1(a) has a vanishing memory, in the sense that the sets in its remote \( \sigma \)-algebra \( \mathcal{F}_{-\infty} = \cap_t \mathcal{F}_t \) have either probability zero or one;
(b) \( E[\sup_{\theta \in \Theta} f_{t-1}(\theta)^2] < \infty \);
(c) The solution \( \theta_0 \) of (3.6) is unique.
See Bierens (2004, Ch. 7) for the motivation of the vanishing memory concept. In particular this concept plays a key-role in the following uniform weak law of large numbers (UWLLN).

**Lemma 4.1.** Let \( V_t \in \mathbb{R}^k \) be a strictly stationary time series process with vanishing memory, and let \( \Theta \) be a compact subset of a Euclidean space. Let \( g(v, \theta) \) be a real or complex-valued function on \( \mathbb{R}^k \times \Theta \) satisfying the following conditions.

1. For each \( \theta \in \Theta \), \( g(v, \theta) \) is Borel measurable in \( v \in \mathbb{R}^k \);
2. For each \( v \in \mathbb{R}^k \), \( g(v, \theta) \) is continuous in \( \theta \in \Theta \);
3. \( E[\sup_{\theta \in \Theta} |g(V_1, \theta)|] < \infty \).

Then \( p \lim_{n \to \infty} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^{n} g(V_t, \theta) - E[g(V_1, \theta)] \right| = 0 \).

**Proof.** Similar to Theorem 7.8(a) in Bierens (2004, p. 187).

This lemma mimics the uniform strong law of large number of Jennrich (1969), with Kolmogorov’s strong law of large numbers replaced by the following weak version for time series.

**Lemma 4.2.** Let \( V_t \in \mathbb{R}^k \) be a strictly stationary time series process with vanishing memory, with \( E[||V_t||] < \infty \). Then

\[
p \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} V_t = E[V_1].
\]

**Proof.** See Bierens (2004, Theorem 7.4, p. 184).

Denoting \( V_t = (Z'_t, Z'_{t-1}, ..., Z'_{t-\max(p,q)})' \), it follows from Assumptions 3.1 and 4.1(a) that \( V_t \) is strictly stationary with vanishing memory because

\[
\mathcal{F}_{-\infty}^t = \sigma \left( \{Z_{t-j}\}_{j=0}^{\infty} \right) = \sigma \left( \{V_{t-j}\}_{j=0}^{\infty} \right).
\]

Next, let \( g(V_t, \theta) = (Y_t - f_{t-1}(\theta))^2 \) and note that by Assumptions 3.1 and 4.1(b), \( E[\sup_{\theta \in \Theta} |g(V_1, \theta)|] < \infty \). Therefore, it follows from Lemma 4.1 that under Assumption 3.1 and parts (a) and (b) of Assumption 4.1,

\[
p \lim_{n \to \infty} \sup_{\theta \in \Theta} \left| \hat{Q}_n(\theta) - Q(\theta) \right| = 0,
\]

(4.1)
where
\[
\hat{Q}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} (Y_t - f_{t-1}(\theta))^2, \quad Q(\theta) = E \left[ (Y_t - f_{t-1}(\theta))^2 \right].
\] (4.2)

It is now a standard nonlinear regression exercise\(^7\) to verify that

**Theorem 4.1.** Under Assumptions 3.1 and 4.1 the NLLS estimator \(\hat{\theta}_n\) satisfies
\[
\lim_{n \to \infty} \hat{\theta}_n = \theta_0 \quad \text{regardless whether } H_0 \text{ is true or not.}
\]

Of course, if \(H_0\) is not true then \(U_t = Y_t - f_{t-1}(\theta_0)\) is no longer a martingale difference process.

### 4.2. Asymptotic normality

In addition to Assumptions 3.1 and 4.1 the following conditions are sufficient for the asymptotic normality of \(\sqrt{n}(\hat{\theta}_n - \theta_0)\) under \(H_0\), but some of them are also necessary under \(H_1\).

**Assumption 4.2.**
(a) \(\Theta\) is convex and \(\theta_0\) is an interior point of \(\Theta\).
(b) \(f_{t-1}(\theta)\) is a.s. twice continuously differentiable in the components \(\theta_1, \theta_2, \ldots, \theta_m\) of \(\theta\).
(c) For \(i_1, i_2 = 1, 2, \ldots, m\),
\[
E \left[ (Y_t - f_{t-1}(\theta))^2 \cdot |(\partial/\partial \theta_{i_1}) f_{t-1}(\theta) \cdot |(\partial/\partial \theta_{i_2}) f_{t-1}(\theta)\right]_{\theta=\theta_0} < \infty, \quad (4.3)
\]
\[
E \left[ \sup_{\theta \in \Theta_0} |(\partial/\partial \theta_{i_1}) f_{t-1}(\theta) \cdot |(\partial/\partial \theta_{i_2}) f_{t-1}(\theta)\right] < \infty, \quad (4.4)
\]
\[
E \left[ \sup_{\theta \in \Theta_0} |Y_t - f_{t-1}(\theta)\cdot |(\partial/\partial \theta_{i_1})(\partial/\partial \theta_{i_2}) f_{t-1}(\theta)\right] < \infty, \quad (4.5)
\]

where \(\Theta_0 = \{\theta \in \Theta : ||\theta - \theta_0|| \leq \varepsilon\}\) with \(\varepsilon > 0\) arbitrarily small.
(d) The matrix
\[
A_2 = E \left[ ((\partial/\partial \theta')(f_{t-1}(\theta))(\partial/\partial \theta) f_{t-1}(\theta))_{\theta=\theta_0}\right]
\]
is nonsingular.

\(^7\)See for example Bierens (2004, Ch. 6).
Denoting
\[ \nabla f_{t-1}(\theta) = (\partial/\partial \theta') f_{t-1}(\theta) \]
it follows then from the first-order conditions for a minimum of \( \hat{Q}_n(\theta) \) and the mean value theorem that under \( H_0 \),
\[ \sqrt{n}(\hat{\theta}_n - \theta_0) = A_2^{-1} \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_t. \nabla f_{t-1}(\theta_0) \right) + o_p(1) \]

Finally, note that under \( H_0 \), \( U_t. \nabla f_{t-1}(\theta_0) \) is a vector-valued martingale difference process, so that by the martingale difference central limit theorem of McLeish (1974) and condition (4.3),
\[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t. \nabla f_{t-1}(\theta_0) \xrightarrow{d} N_m[0, A_1], \text{ where } \\
A_1 = E \left[ U_1^2. (\nabla f_{t-1}(\theta_0)) (\nabla f_{t-1}(\theta_0))' \right]. \]

Consequently,

**Theorem 4.2.** Under \( H_0 \) and Assumptions 3.1, 4.1 and 4.2,
\[ \sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_t. A_2^{-1} \nabla f_{t-1}(\theta_0) + o_p(1) \]
\[ \xrightarrow{d} N_m \left[ 0, A_2^{-1} A_1 A_2^{-1} \right]. \]

**5. The limiting null distribution of the WICM test**

I will first show that

**Lemma 5.1.** Under \( H_0 \) and Assumptions 3.1, 4.1 and 4.2, \( \hat{B}_{n,k} \leq 2.\hat{B}_{1,n,k} + 2.\hat{B}_{2,n,k} \), where \( E[\hat{B}_{1,n,k}] = E[U_1^2] \) and \( \sup_{k \in \mathbb{N}} \hat{B}_{2,n,k} = O_p(1) \), hence for any positive sequence \( \gamma_k \) satisfying \( \sum_{k=1}^{\infty} \gamma_k < \infty \) and any subsequence \( L_n \) of \( n \) such that \( \lim_{n \to \infty} L_n = \infty \),
\[ \sum_{k=1}^{\infty} \gamma_k \hat{B}_{n,k} - \sum_{k=1}^{L_n} \gamma_k \hat{B}_{n,k} = o_p(1). \quad (5.1) \]
Proof. It follows from (3.17) that
\[
\hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)
= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right)
- \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0) \right) \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right)
\]
so that
\[
\hat{B}_{n,k} = \int_{\mathcal{T}^k} \left| \hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)
= \int_{\mathcal{T}^k} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \cos \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right)^2
\times d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)
+ \int_{\mathcal{T}^k} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \sin \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right)^2
\times d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)
\leq 2 \int_{\mathcal{T}^k} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \cos \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right)^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)
+ 2 \int_{\mathcal{T}^k} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0) \right) \cos \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right)^2
\times d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)
+ 2 \int_{\mathcal{T}^k} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \sin \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right)^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)
where

\[
\begin{align*}
\hat{B}_{1,n,k} &= \int_{\mathcal{T}^k} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right|^2 \, d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k) \\
&= \int_{\mathcal{T}^k} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \cos \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right)^2 \, d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k) \\
&+ \int_{\mathcal{T}^k} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \sin \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right)^2 \, d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)
\end{align*}
\]

\[
\hat{B}_{2,n,k} = \int_{\mathcal{T}^k} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0) \right) \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right|^2 \\
\times d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)
\]

\[
\begin{align*}
&= \int_{\mathcal{T}^k} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0) \right) \cos \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right)^2 \\
&\times d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k) \\
&+ \int_{\mathcal{T}^k} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0) \right) \sin \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right)^2 \\
&\times d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)
\end{align*}
\]

Obviously,

\[
E[\hat{B}_{1,n,k}] = E[U_t^2] = E[U_1^2].
\] (5.5)

To prove that \(\sup_{k \in \mathbb{N}} \hat{B}_{2,n,k} = O_p(1)\), observe that by the mean value theorem,

\[
\begin{align*}
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0) \right) \cos \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \\
= \sqrt{n}(\hat{\theta}_n - \theta_0) \frac{1}{n} \sum_{t=1}^{n} \nabla f_{t-1}(\hat{\theta}_n(\tau)) \cos \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right)
\end{align*}
\] (5.6)
where $\tilde{\theta}_n(\tau)$ is a mean value such that $||\tilde{\theta}_n(\tau) - \theta_0|| \leq ||\tilde{\theta}_n - \theta_0||$, hence

$$\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\tilde{\theta}_n) - f_{t-1}(\theta_0) \right) \cos \left( \sum_{j=1}^{k} \tau'_j \Phi(Z_{t-j}) \right) \right|$$

$$\leq ||\sqrt{n}(\tilde{\theta}_n - \theta_0)|| \cdot \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \Theta} ||\nabla f_{t-1}(\theta)||$$

and thus,

$$\int_{\mathcal{X}_k} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\tilde{\theta}_n) - f_{t-1}(\theta_0) \right) \cos \left( \sum_{j=1}^{k} \tau'_j \Phi(Z_{t-j}) \right) \right)^2$$

$$\times d\mu(\tau_1)d\mu(\tau_2)...d\mu(\tau_k)$$

$$\leq \left( ||\sqrt{n}(\tilde{\theta}_n - \theta_0)|| \cdot \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \Theta} ||\nabla f_{t-1}(\theta)|| \right)^2$$

$$= O_p(1), \text{ uniformly in } k.$$  

The latter follows from the fact that by Theorem 4.2, $||\sqrt{n}(\tilde{\theta}_n - \theta_0)|| = O_p(1)$ and by condition (4.4), $E[\sup_{\theta \in \Theta} ||\nabla f_{t-1}(\theta)||^2] < \infty$, hence $\frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \Theta} ||\nabla f_{t-1}(\theta)|| = O_p(1)$.

Similarly,

$$\int_{\mathcal{X}_k} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\tilde{\theta}_n) - f_{t-1}(\theta_0) \right) \sin \left( \sum_{j=1}^{k} \tau'_j \Phi(Z_{t-j}) \right) \right)^2$$

$$\times d\mu(\tau_1)d\mu(\tau_2)...d\mu(\tau_k)$$

$$\leq ||\sqrt{n}(\tilde{\theta}_n - \theta_0)|| \cdot \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \Theta} ||\nabla f_{t-1}(\theta)||$$

$$= O_p(1), \text{ uniformly in } k.$$  

hence

$$\sup_{k \in \mathbb{N}} \hat{B}_{2,n,k} \leq 2||\sqrt{n}(\tilde{\theta}_n - \theta_0)|| \cdot \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \Theta} ||\nabla f_{t-1}(\theta)|| = O_p(1). \quad (5.7)$$

The result (5.1) follows now trivially from the fact that $\sum_{k=1}^{\infty} \gamma_k < \infty$ implies $\lim_{k \to \infty} \sum_{k=\ell+1}^{\infty} \gamma_k = 0.$
Next, denote

\[ b_k(\tau_1, \tau_2, ..., \tau_k) = E \left[ \nabla f_{t-1}(\theta_0) \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right], \quad (5.8) \]

\[ \phi_{k,t-1}(\tau_1, \tau_2, ..., \tau_k) = \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \]

\[ - b_k(\tau_1, \tau_2, ..., \tau_k)' A^{-1}_2 \nabla f_{t-1}(\theta_0), \quad (5.9) \]

\[ W_{k,n}(\tau_1, \tau_2, ..., \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \phi_{k,t-1}(\tau_1, \tau_2, ..., \tau_k), \quad (5.10) \]

\[ B_{k,n} = \int_{\mathbf{T}^k} |W_{k,n}(\tau_1, \tau_2, ..., \tau_k)|^2 \mu(\tau_1) \mu(\tau_2) ... \mu(\tau_k) \, d\mu(\tau_1) d\mu(\tau_2) ... d\mu(\tau_k). \quad (5.11) \]

Then

**Lemma 5.2.** Under \( H_0 \) and Assumptions 3.1, 4.1 and 5.2, and for fixed \( k \in \mathbb{N} \),

\[ \sup_{(\tau_1, \tau_2, ..., \tau_k) \in \mathbf{T}^k} \left| \hat{W}_{k,n}(\tau_1, \tau_2, ..., \tau_k) - W_{k,n}(\tau_1, \tau_2, ..., \tau_k) \right| = o_p(1), \quad (5.12) \]

*\textit{hence}*

\[ \hat{B}_{n,k} = B_{k,n} + o_p(1). \quad (5.13) \]

**Proof.** Observe from (5.6) that

\[ \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0) \right) \cos \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right| \]

\[ - \sqrt{n}(\hat{\theta}_n - \theta_0)' \frac{1}{n} \sum_{t=1}^{n} \nabla f_{t-1}(\theta_0) \cos \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right| \]

\[ \leq \left| \sqrt{n}(\hat{\theta}_n - \theta_0) \right| \frac{1}{n} \sum_{t=1}^{n} \sup_{|\theta - \theta_0| \leq |\hat{\theta}_n - \theta_0|} \left| \nabla f_{t-1}(\theta) - \nabla f_{t-1}(\theta_0) \right| \]

\[ = o_p(1) \] \quad (5.14)
because \( |\sqrt{n}(\hat{\theta}_n - \theta_0)| = O_p(1) \) and
\[
\frac{1}{n} \sum_{i=1}^{n} \sup_{||\theta - \theta_0|| \leq ||\hat{\theta}_n - \theta_0||} ||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| = o_p(1). \tag{5.15}
\]
To prove the latter, observe that for an arbitrary small \( \varepsilon > 0 \),
\[
E \left[ \frac{1}{n} \sum_{i=1}^{n} \sup_{||\theta - \theta_0|| \leq ||\hat{\theta}_n - \theta_0||} ||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| \right] \\
\leq E \left[ \sup_{||\theta - \theta_0|| \leq \varepsilon} ||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| \times I \left( ||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| \leq \varepsilon \right) \right] \\
+ 2E \left[ \sup_{\theta \in \Theta} ||\nabla f_{i-1}(\theta)|| \times I \left( ||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| > \varepsilon \right) \right] \\
\leq E \left[ \sup_{||\theta - \theta_0|| \leq \varepsilon} ||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| \right] \\
+ 2\sqrt{E \left[ \sup_{\theta \in \Theta} ||\nabla f_{i-1}(\theta)||^2 \right]} \sqrt{\Pr \left[ ||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| > \varepsilon \right]} \\
= E \left[ \sup_{||\theta - \theta_0|| \leq \varepsilon} ||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| \right] + o(1), \tag{5.16}
\]
where the \( o(1) \) term is due to the facts that by Theorem 4.1, \( \Pr[||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| > \varepsilon] \to 0 \) and that by condition (4.4), \( E[\sup_{\theta \in \Theta} ||\nabla f_{i-1}(\theta)||^2] < \infty \). Moreover, since
\[
E \left[ \sup_{||\theta - \theta_0|| \leq \varepsilon} ||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| \right] \\
\leq 2E \left[ \sup_{\theta \in \Theta} ||\nabla f_{i-1}(\theta)|| \right] \\
\leq \sqrt{E \left[ \sup_{\theta \in \Theta} ||\nabla f_{i-1}(\theta)||^2 \right]} < \infty
\]
and by the a.s. continuity of \( \nabla f_{i-1}(\theta) \),
\[
\lim_{\varepsilon \downarrow 0} \sup_{||\theta - \theta_0|| \leq \varepsilon} ||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| = 0 \text{ a.s.,}
\]
it follows from the dominated convergence theorem that
\[
\lim_{\varepsilon \downarrow 0} E \left[ \sup_{||\theta - \theta_0|| \leq \varepsilon} ||\nabla f_{i-1}(\theta) - \nabla f_{i-1}(\theta_0)|| \right] = 0. \tag{5.17}
\]
It follows now straightforwardly from (5.16) and (5.17) that (5.15) holds.

Next, observe from Theorem 4.1 and Assumption 4.1(b) that for fixed $k$,

$$
\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left\| \frac{1}{n} \sum_{t=1}^{n} \nabla f_{t-1}(\theta_0) \cos \left( \sum_{j=1}^{k} \tau_j^t \Phi(Z_{t-j}) \right) - \text{Re}[b_k(\tau_1, \tau_2, \ldots, \tau_k)] \right\| = o_p(1)
$$

and thus,

$$
\sqrt{n} (\hat{\theta}_n - \theta_0) \left( \frac{1}{n} \sum_{t=1}^{n} \nabla f_{t-1}(\theta_0) \cos \left( \sum_{j=1}^{k} \tau_j^t \Phi(Z_{t-j}) \right) \right)
= \sqrt{n} (\hat{\theta}_n - \theta_0) \text{Re}[b_k(\tau_1, \tau_2, \ldots, \tau_k)] + o_p(1)
= \text{Re}[b_k(\tau_1, \tau_2, \ldots, \tau_k)] A_2^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_t \nabla f_{t-1}(\theta_0) + o_p(1)
$$

uniformly on $\Upsilon^k$, where the latter equality follows from Theorem 2. Hence by (5.2) and (5.14),

$$
\text{Re} \left[ \hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] = \text{Re} \left[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \exp \left( i \sum_{j=1}^{k} \tau_j^t \Phi(Z_{t-j}) \right) \right]
- \text{Re}[b_k(\tau_1, \tau_2, \ldots, \tau_k)] A_2^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_t \nabla f_{t-1}(\theta_0) + o_p(1)
= \text{Re} \left[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \phi_{k,t-1}(\tau_1, \tau_2, \ldots, \tau_k) \right] + o_p(1)
= \text{Re} \left[ W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] + o_p(1)
$$

uniformly on $\Upsilon^k$. The same applies to the Im[.] case. This proves (5.12).

The proof that (5.12) implies (5.13) is not hard and is therefore left to the reader. 

---

Denote the left-hand side of (5.12) by $d_{k,n}$, and verify that

$$
\left| \hat{B}_{k,n} - \hat{B}_{k,n} \right| \leq d_{k,n}^2 + 4d_{k,n} \sqrt{B_{k,n}}.
$$

23
Next, I will set forth conditions such that \( \sup_{k \in \mathbb{N}} E[B_{k,n}] < \infty \), as follows. Observe from (5.8), (5.9) and (5.10) that

\[
E \left[ |W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|^2 \right] = E \left[ U_t^2 \phi_{k,t-1}(\tau_1, \tau_2, \ldots, \tau_k) \phi_{k,t-1}^*(\tau_1, \tau_2, \ldots, \tau_k) \right] = E \left[ U_t^2 \left( \exp \left( i \sum_{j=1}^k \tau_j^* \Phi(Z_{t-j}) \right) \right) \right. \\
E \left[ U_t^2 \left( \exp \left( -i \sum_{j=1}^k \tau_j \Phi(Z_{t-j}) \right) \right) \right] \left( \frac{\partial f_{t-1}(\theta_0)}{\partial \theta_0} + A_2^{-1} b_k(\tau_1, \tau_2, \ldots, \tau_k) \right) \\
= E[U_t^2] - b_k(\tau_1, \tau_2, \ldots, \tau_k)' A_2^{-1} \left[ U_t^2 \nabla f_{t-1}(\theta_0) \exp \left( -i \sum_{j=1}^k \tau_j^* \Phi(Z_{t-j}) \right) \right] \\
- E \left[ U_t^2 \left( \nabla f_{t-1}(\theta_0) \right) \right] \left( \frac{\partial f_{t-1}(\theta_0)}{\partial \theta_0} + A_2^{-1} b_k(\tau_1, \tau_2, \ldots, \tau_k) \right) \\
E[U_t^2] - b_k(\tau_1, \tau_2, \ldots, \tau_k)' A_2^{-1} c_k(\tau_1, \tau_2, \ldots, \tau_k) \\
- c_k(\tau_1, \tau_2, \ldots, \tau_k)' A_2^{-1} b_k(\tau_1, \tau_2, \ldots, \tau_k) \\
+ b_k(\tau_1, \tau_2, \ldots, \tau_k)' A_2^{-1} A_1 A_2^{-1} b_k(\tau_1, \tau_2, \ldots, \tau_k),
\]

where the bar denotes the complex-conjugate, and

\[
c_k(\tau_1, \tau_2, \ldots, \tau_k) = E \left[ U_t^2 \nabla f_{t-1}(\theta_0) \exp \left( i \sum_{j=1}^k \tau_j \Phi(Z_{t-j}) \right) \right].
\]

It is an easy complex calculus exercise to verify that

\[
b_k(\tau_1, \tau_2, \ldots, \tau_k)' A_2^{-1} c_k(\tau_1, \tau_2, \ldots, \tau_k) \\
+ c_k(\tau_1, \tau_2, \ldots, \tau_k)' A_2^{-1} b_k(\tau_1, \tau_2, \ldots, \tau_k) \\
= 2 \Re \left[ b_k(\tau_1, \tau_2, \ldots, \tau_k)' A_2^{-1} \Re \left[ c_k(\tau_1, \tau_2, \ldots, \tau_k) \right] \right] \\
+ 2 \Im \left[ b_k(\tau_1, \tau_2, \ldots, \tau_k)' A_2^{-1} \Im \left[ c_k(\tau_1, \tau_2, \ldots, \tau_k) \right] \right]
\]

24
and

\[ b_k(\tau_1, \tau_2, \ldots, \tau_k)'A_2^{-1}A_1A_2^{-1}b_k(\tau_1, \tau_2, \ldots, \tau_k) \]

\[ = \text{Re} [b_k(\tau_1, \tau_2, \ldots, \tau_k)'A_2^{-1}A_1A_2^{-1}b_k(\tau_1, \tau_2, \ldots, \tau_k)] \]

\[ + \text{Im} [b_k(\tau_1, \tau_2, \ldots, \tau_k)'A_2^{-1}A_1A_2^{-1}b_k(\tau_1, \tau_2, \ldots, \tau_k)], \]

hence

\[ E \left[ |W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|^2 \right] \leq E[U_1^2] \]

\[ + 2 \text{Re} [b_k(\tau_1, \tau_2, \ldots, \tau_k)'A_2^{-1}b_k(\tau_1, \tau_2, \ldots, \tau_k)] \]

\[ + 2 \text{Im} [b_k(\tau_1, \tau_2, \ldots, \tau_k)'A_2^{-1}b_k(\tau_1, \tau_2, \ldots, \tau_k)] \]

\[ + \text{Re} [b_k(\tau_1, \tau_2, \ldots, \tau_k)'A_2^{-1}A_1A_2^{-1}b_k(\tau_1, \tau_2, \ldots, \tau_k)] \]

\[ + \text{Im} [b_k(\tau_1, \tau_2, \ldots, \tau_k)'A_2^{-1}A_1A_2^{-1}b_k(\tau_1, \tau_2, \ldots, \tau_k)] \]

(5.18)

Let \( \lambda_1, \ldots, \lambda_m \) be the eigenvalues of \( A_2^{-1} \), with corresponding orthonormal eigenvectors \( q_1, \ldots, q_m \). Then

\[ \text{Re} [b_k(\tau_1, \tau_2, \ldots, \tau_k)'A_2^{-1}b_k(\tau_1, \tau_2, \ldots, \tau_k)] \]

\[ \leq \sum_{i=1}^{m} \lambda_i \left| q_i' \text{Re} [b_k(\tau_1, \tau_2, \ldots, \tau_k)] \right| \left| \text{Re} [q_i'b_k(\tau_1, \tau_2, \ldots, \tau_k)] \right| \]

\[ = \sum_{i=1}^{m} \lambda_i \left| E \left[ (\nabla f_{t-1}(\theta_0))'q_i \cos \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right] \right| \]

\[ \times \left| E \left[ U_t^2(\nabla f_{t-1}(\theta_0))'q_i \cos \left( \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) \right] \right| \]

\[ \leq \sum_{i=1}^{m} \lambda_i E \left[ |(\nabla f_{t-1}(\theta_0))'q_i| \right] E \left[ U_t^2 \right] \left| (\nabla f_{t-1}(\theta_0))'q_i \right| \]

\[ \leq \sum_{i=1}^{m} \lambda_i E \left[ |\nabla f_{t-1}(\theta_0)| \right] \left| E \left[ U_t^2 |\nabla f_{t-1}(\theta_0)| \right] \right| \]

\[ \leq \sum_{i=1}^{m} \lambda_i \sqrt{E \left[ |\nabla f_{t-1}(\theta_0)| \right]^{2}} E \left[ U_t^2 |\nabla f_{t-1}(\theta_0)| \right] \]

\[ = \sum_{i=1}^{m} \lambda_i \sqrt{\text{trace}(A_2)} E \left[ U_t^2 |\nabla f_{t-1}(\theta_0)| \right] \]

\[ = \text{trace}(A_2)\sqrt{\text{trace}(A_2)} E \left[ U_t^2 |\nabla f_{t-1}(\theta_0)| \right] \]
and similarly

\[
\mathrm{Im}\left[b_k(\tau_1, \tau_2, \ldots, \tau_k)\right] A_{2}^{-1} \mathrm{Im}\left[c_k(\tau_1, \tau_2, \ldots, \tau_k)\right] \\
\leq \text{trace}(A_2) \sqrt{\text{trace}(A_2)} E\left[U_1^2||\nabla f_{t-1}(\theta_0)||\right]
\]

Moreover, with \(\lambda_1^*, \ldots, \lambda_m^*\) the eigenvalues of \(A_2^{-1}A_1A_2^{-1}\) we have similarly,

\[
\mathrm{Re}\left[b_k(\tau_1, \tau_2, \ldots, \tau_k)\right] A_{2}^{-1}A_1A_{2}^{-1} \mathrm{Re}\left[b_k(\tau_1, \tau_2, \ldots, \tau_k)\right] \\
\leq \sum_{i=1}^{m} \lambda_i^* E\left[||\nabla f_{t-1}(\theta_0)||\right]^2 \\
\leq \sum_{i=1}^{m} \lambda_i^* E\left[||\nabla f_{t-1}(\theta_0)||^2\right] \\
= \sum_{i=1}^{m} \lambda_i^* \text{trace}(A_2) \\
= \text{trace}(A_2^{-1}A_1A_2^{-1}).\text{trace}(A_2)
\]

and

\[
\mathrm{Im}\left[b_k(\tau_1, \tau_2, \ldots, \tau_k)\right] A_{2}^{-1}A_1A_{2}^{-1} \mathrm{Im}\left[b_k(\tau_1, \tau_2, \ldots, \tau_k)\right] \\
\leq \text{trace}(A_2^{-1}A_1A_2^{-1}).\text{trace}(A_2).
\]

Hence,

\[
E\left[B_{k,n}\right] = \int_{\mathbf{T}^k} E\left[|W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|^2\right] d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k) \\
\leq E[U_1^2] + 4\text{trace}(A_2) \sqrt{\text{trace}(A_2)} E\left[U_1^2||\nabla f_{t-1}(\theta_0)||\right] \\
+ 2\text{trace}(A_2^{-1}A_1A_2^{-1}).\text{trace}(A_2).
\]

Thus, if

**Assumption 5.1.** \(E[U_1^2||\nabla f_{t-1}(\theta_0)||] < \infty\)

then

**Lemma 5.3.** Under \(H_0\) and Assumptions 3.1, 4.1, 4.2 and 5.1, \(\sup_{k \in \mathbb{N}} E[B_{k,n}] < \infty\).
Lemmas 5.1-5.3 now yield the following corollary.

**Lemma 5.4.** Under $H_0$ and Assumptions 3.1, 4.1, 4.2 and 5.1,

$$\sum_{k=1}^{\infty} \gamma_k \hat{B}_{k,n} = \sum_{k=1}^{\infty} \gamma_k B_{k,n} + o_p(1)$$

for any positive sequence $\gamma_k$ satisfying $\sum_{k=1}^{\infty} \gamma_k < \infty$.

**Proof.** For any $K \in \mathbb{N}$,

$$\left| \sum_{k=1}^{\infty} \gamma_k \hat{B}_{k,n} - \sum_{k=1}^{\infty} \gamma_k B_{k,n} \right| \leq R_{0,n,K} + R_{1,n,K} + R_{2,n,K}$$

where

$$R_{0,n,K} = \left| \sum_{k=1}^{K} \gamma_k \hat{B}_{k,n} - \sum_{k=1}^{K} \gamma_k B_{k,n} \right|$$

$$R_{1,n,K} = 2 \sum_{k=K+1}^{\infty} \gamma_k \hat{B}_{1,k,n} + \sum_{k=1}^{\infty} \gamma_k B_{k,n}$$

$$R_{2,n,K} = 2 \left( \sum_{k=K+1}^{\infty} \gamma_k \right) \sup_{k \in \mathbb{N}} \hat{B}_{2,k,n}.$$

Let $\varepsilon > 0$ and $\delta > 0$ be arbitrary. By Chebyshev’s inequality for first moments and Lemmas 5.1 and 5.3 there exists a $K_1(\delta, \varepsilon) \in \mathbb{N}$ such that for all $K \geq K_1(\delta, \varepsilon)$,

$$\Pr[R_{1,n,K} > \delta/2] \leq 2E[R_{1,n,K}]/\delta \leq 2 \frac{\sum_{k=K+1}^{\infty} \gamma_k E[\hat{B}_{1,k,n}] + \sum_{k=K+1}^{\infty} \gamma_k E[B_{k,n}]}{\delta} \leq \delta^{-1} \left( \sum_{k=K+1}^{\infty} \gamma_k \right) \left( 4E[U_1] + 2 \sup_{k \in \mathbb{N}} E[B_{k,n}] \right) \leq \varepsilon$$

Moreover, by Lemma 5.1 there exists a $K_2(\delta, \varepsilon) \in \mathbb{N}$ such that for all $K \geq K_2(\delta, \varepsilon)$,

$$\Pr[R_{2,n,K} > \delta/2] = \Pr \left[ \sup_{k \in \mathbb{N}} \hat{B}_{2,k,n} > \frac{\delta}{4 \sum_{k=K+1}^{\infty} \gamma_k} \right] < \varepsilon$$
Using the easy inequality
\[ \Pr[R_1 + R_2 > \delta] \leq \Pr[R_1 > \delta / 2] + \Pr[R_2 > \delta / 2] \] (5.20)
for nonnegative random variables \( R_1 \) and \( R_2 \), it follows now that for a fixed \( K_0 \geq \max\{K_1(\delta, \varepsilon), K_2(\delta, \varepsilon)\}, \)
\[ \Pr[R_{1,n,K_0} + R_{2,n,K_0} > \delta] < 2\varepsilon \]

Finally, it follows from Lemma 5.2 that \( \lim_{n \to \infty} R_{0,n,K_0} = 0 \), so that there exists an \( n_0 \in \mathbb{N} \) such that for all \( n \geq n_0, \)
\[ \Pr[R_{2,n,K_0} > \delta] < \varepsilon. \]
Hence by (5.20),
\[
\Pr\left[ \left| \sum_{k=1}^{\infty} \gamma_k \tilde{B}_{k,n} \right| > 2\delta \right] \\
\leq \Pr\left[ R_{2,n,K_0} + R_{1,n,K_0} + R_{2,n,K_0} > 2\delta \right] \\
\leq \Pr\left[ R_{2,n,K_0} + R_{1,n,K_0} + R_{2,n,K_0} > 2\delta \right] \\
\leq \Pr[R_{2,n,K_0} > \delta] + \Pr[R_{1,n,K_0} + R_{2,n,K_0} > \delta] \\
< 3\varepsilon,
\]
which implies that \( \lim_{n \to \infty} \left| \sum_{k=1}^{\infty} \gamma_k \tilde{B}_{k,n} \right| = 0. \]

The next step is to prove that for each \( k \in \mathbb{N}, W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \) converges weakly to a zero-mean complex-valued Gaussian process \( W_k(\tau_1, \tau_2, \ldots, \tau_k) \) on \( \Upsilon^k \), denoted by \( W_{k,n} \Rightarrow W_k \), by showing that \( W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \) is tight on \( \Upsilon^k \) and its finite distributions converge to the corresponding finite distribution of \( W_k(\tau_1, \tau_2, \ldots, \tau_k) \). See Bierens (2014a) and Billingsley (1968) for these concepts. However, it follows from Bierens and Ploberger (1997, Lemma A.1) that \( \text{Re}[W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)] \) and \( \text{Im}[W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)] \) are tight on \( \Upsilon^k \), hence so is \( W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \). Now it suffices to verify that the finite distributions of \( W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \) converge to the corresponding finite distribution of \( W_k(\tau_1, \tau_2, \ldots, \tau_k) \), which follows straightforwardly from the martingale difference central limit theorem of McLeish (1974). Thus,
Lemma 5.5. Under $H_0$ and Assumptions 3.1, 4.1 and 4.2, and for each $k \in \mathbb{N}$, $W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \Rightarrow W_k(\tau_1, \tau_2, \ldots, \tau_k)$ on $\Upsilon^k$, where the latter is a zero-mean complex-valued Gaussian process on $\Upsilon^k$ with covariance function

$$
\Gamma_k (((\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,k})), ((\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,k}))) \\
= E \left[ W_k(\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,k}) W_k(\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,k}) \right] \\
= E \left[ U_t^2 \phi_{k,t-1}(\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,k}) \phi_{k,t-1}(\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,k}) \right].
$$

(5.21)

Consequently, by the continuous mapping theorem,

$$
B_{k,n} \xrightarrow{d} B_k = \int_{\Upsilon^k} |W_k(\tau_1, \tau_2, \ldots, \tau_k)|^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)
$$

Note that

$$
W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = W_{k+1,n}(\tau_1, \tau_2, \ldots, \tau_k, 0)
$$

and similarly,

$$
W_k(\tau_1, \tau_2, \ldots, \tau_k) = W_{k+1}(\tau_1, \tau_2, \ldots, \tau_k, 0).
$$

Hence

$$
\begin{pmatrix}
W_{1,n}(\tau_1) \\
W_{2,n}(\tau_1, \tau_2) \\
\vdots \\
W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)
\end{pmatrix}
\Rightarrow
\begin{pmatrix}
W_k(\tau_1, 0, 0, \ldots, 0) \\
W_k(\tau_1, \tau_2, 0, \ldots, 0) \\
\vdots \\
W_k(\tau_1, \tau_2, \ldots, \tau_k)
\end{pmatrix}
= 
\begin{pmatrix}
W_{1,n}(\tau_1, 0, 0, \ldots, 0) \\
W_{k,n}(\tau_1, 0, 0, \ldots, 0) \\
\vdots \\
W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)
\end{pmatrix}
= 
\begin{pmatrix}
W_1(\tau_1) \\
W_k(\tau_1, \tau_2) \\
\vdots \\
W_k(\tau_1, \tau_2, \ldots, \tau_k)
\end{pmatrix}.
$$

(5.22)

where the weak convergence result involved is not hard to verify. It follows now from the continuous mapping theorem that

Lemma 5.6. Under the conditions of Lemma 5.5, $(B_{1,n}, B_{2,n}, \ldots, B_{k,n})' \xrightarrow{d} (B_1, B_2, \ldots, B_k)'$. 

It remains to show that
Lemma 5.7. Under $H_0$ and Assumptions 3.1, 4.1, 4.2 and 5.1, and for any positive sequence $\gamma_k$ satisfying $\sum_{k=1}^{\infty} \gamma_k < \infty$,

$$\sum_{k=1}^{\infty} \gamma_k B_{k,n} \xrightarrow{d} \sum_{k=1}^{\infty} \gamma_k B_k.$$

Proof. It follows from Lemma 5.5 that

$$E[B_k] = \int_{\Upsilon^k} E \left[ |W_k(\tau_1, \tau_2, \ldots, \tau_k)|^2 \right] \, d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k)$$

$$= \int_{\Upsilon^k} E \left[ W_k(\tau_1, \tau_2, \ldots, \tau_k) \overline{W_k(\tau_1, \tau_2, \ldots, \tau_k)} \right]$$

$$\times d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k)$$

$$= \int_{\Upsilon^k} E \left[ U_t^2 \phi_{k,t-1}(\tau_1, \tau_2, \ldots, \tau_k) \overline{\phi_{k,t-1}(\tau_1, \tau_2, \ldots, \tau_k)} \right]$$

$$\times d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k)$$

hence, similar to Lemma 5.3 we have

$$\sup_{k \in \mathbb{N}} E[B_k] < \infty. \quad (5.23)$$

Moreover, Lemma 5.6 implies that for each $\ell \in \mathbb{N}$,

$$\sum_{k=1}^{\ell} \gamma_k B_{k,n} \xrightarrow{d} \sum_{k=1}^{\ell} \gamma_k B_k. \quad (5.24)$$

Let $x$, $x - \delta$ and $x + \delta$, with $\delta > 0$, be continuity points of the distribution of $\sum_{k=1}^{\infty} \gamma_k B_k$, Then

$$\limsup_{n \to \infty} \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_{k,n} \leq x \right]$$

$$\leq \lim_{n \to \infty} \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x \right] = \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_k \leq x \right]$$

$$= \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_k \leq x \text{ and } \sum_{k=\ell+1}^{\infty} \gamma_k B_k \leq \delta \right]$$
\[ + \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_k \leq x \text{ and } \sum_{k=\ell+1}^{\infty} \gamma_k B_k > \delta \right] \]
\[ \leq \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_k \leq x + \delta \right] + \Pr \left[ \sum_{k=\ell+1}^{\infty} \gamma_k B_k > \delta \right] \]
\[ \leq \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_k \leq x + \delta \right] + \delta^{-1} \left( \sum_{k=\ell+1}^{\infty} \gamma_k \right) \sup_{k \in \mathbb{N}} E[B_k], \]

where first equality follows from (5.24) and the last inequality follows from (5.23) and Chebyshev’s inequality for first moments. Hence, letting \( \ell \to \infty \) first and then letting \( \delta \downarrow 0 \) it follows that

\[ \limsup_{n \to \infty} \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_{k,n} \leq x \right] \leq \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_k \leq x \right]. \quad (5.25) \]

Moreover,

\[ \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_{k,n} \leq x \right] \]
\[ = \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x - \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} \text{ and } \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} \leq \delta \right] \]
\[ + \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x - \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} \text{ and } \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} > \delta \right] \]
\[ \geq \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x - \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} \text{ and } \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} \leq \delta \right] \]
\[ \geq \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x - \delta \text{ and } \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} \leq \delta \right] \]
\[ = \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x - \delta \right] \]
\[ - \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x - \delta \text{ and } \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} > \delta \right] \]
\[
\geq \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x - \delta \right] - \Pr \left[ \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} > \delta \right]
\]

Hence
\[
\lim \inf_{n \to \infty} \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_{k,n} \leq x \right] \geq \lim \inf_{n \to \infty} \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x - \delta \right] - \lim \sup_{n \to \infty} \Pr \left[ \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} > \delta \right]
\]
\[
= \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x - \delta \right] - \lim \sup_{n \to \infty} \Pr \left[ \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} > \delta \right]
\]
\[
\geq \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x - \delta \right] - \lim \sup_{n \to \infty} \Pr \left[ \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} > \delta \right]
\]

By Lemma 5.1, (5.20) and Chebyshev’s inequality for first moments,
\[
\Pr \left[ \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} > \delta \right] \leq \Pr \left[ 2 \sum_{k=\ell+1}^{\infty} \gamma_k \tilde{B}_{1,k,n} + 2 \sum_{k=\ell+1}^{\infty} \gamma_k \tilde{B}_{2,k,n} > \delta \right]
\]
\[
\leq \Pr \left[ \sum_{k=\ell+1}^{\infty} \gamma_k \tilde{B}_{1,k,n} > \delta/4 \right] + \Pr \left[ \sum_{k=\ell+1}^{\infty} \gamma_k \tilde{B}_{2,k,n} > \delta/4 \right]
\]
\[
\leq 4 \delta^{-1} \left( \sum_{k=\ell+1}^{\infty} \gamma_k \right) E[U_1^2] + \Pr \left[ \sup_{k \in \mathbb{N}} \tilde{B}_{2,k,n} > \frac{\delta}{4 \left( \sum_{k=\ell+1}^{\infty} \gamma_k \right)} \right],
\]

hence \( \lim_{\ell \to \infty} \lim \sup_{n \to \infty} \Pr \left[ \sum_{k=\ell+1}^{\infty} \gamma_k B_{k,n} > \delta \right] = 0 \) and thus
\[
\lim \inf_{n \to \infty} \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_{k,n} \leq x \right] \geq \Pr \left[ \sum_{k=1}^{\ell} \gamma_k B_{k,n} \leq x - \delta \right].
\]

Letting \( \delta \downarrow 0 \) it follows now that
\[
\lim \inf_{n \to \infty} \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_{k,n} \leq x \right] \geq \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_{k} \leq x \right] \quad (5.26)
\]

32
Combining (5.25) and (5.26) yields

$$\lim_{n \to \infty} \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_{k,n} \leq x \right] = \Pr \left[ \sum_{k=1}^{\infty} \gamma_k B_k \leq x \right],$$

which proves the lemma. ■

Summarizing, the following theorem has been proved.

**Theorem 5.1.** Let $\gamma_k$ be an a priori chosen positive sequence satisfying $\sum_{k=1}^{\infty} \gamma_k < \infty$, and let $L_n < n$ be a subsequence of $n$ satisfying $\lim_{n \to \infty} L_n = \infty$. The test statistic of the WICM test takes the form

$$\hat{T}_n = \sum_{k=1}^{L_n} \gamma_k \hat{B}_{n,k},$$

(5.27)

where the ICM test statistics $\hat{B}_{n,k}$ are defined by (3.18). Under $H_0$ and Assumptions 3.1, 4.1, 4.2 and 5.1,

$$\hat{T}_n \overset{d}{\to} T = \sum_{k=1}^{\infty} \gamma_k B_k,$$

(5.28)

where each $B_k$ takes the form

$$B_k = \int_{\Upsilon^k} |W_k(\tau_1, \tau_2, \ldots, \tau_k)|^2 \, d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k),$$

with $W_k(\tau_1, \tau_2, \ldots, \tau_k)$ a complex-valued zero-mean Gaussian process on $\Upsilon^k$ with covariance function given in (5.21).

**6. Consistency of the WICM test**

In this section I will show that under $H_1$ the WICM test statistic $\hat{T}_n$ in Theorem 5.1 converges in probability to infinity, so that this test is consistent. This will be done in the following three steps.
Lemma 6.1. Denote

\[
\psi_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{n} \sum_{t=1}^{n} U_t \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right),
\]

(6.1)

\[
\eta_{k,n} = \int_{\Upsilon_k} |\psi_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k).
\]

(6.2)

Under $H_1$ and Assumptions 3.1 and 4.1,

\[
\sup_{k \in \mathbb{N}} \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left| \frac{\hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)}{\sqrt{n}} - \psi_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right| = o_p(1),
\]

(6.3)

hence

\[
\sup_{k \in \mathbb{N}} \left| \frac{\hat{B}_{k,n}}{n} - \eta_{k,n} \right| = o_p(1).
\]

(6.4)

Moreover, for each $k \in \mathbb{N},$

\[
\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left| \psi_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) - \psi_k(\tau_1, \tau_2, \ldots, \tau_k) \right| = o_p(1),
\]

(6.5)

where $\psi_k(\tau_1, \tau_2, \ldots, \tau_k)$ is defined by (3.14), hence

\[
\eta_{k,n} = \eta_k + o_p(1),
\]

(6.6)

where $\eta_k$ is defined in (3.19).

Proof. Observe from (5.2) that under $H_1,$

\[
\frac{\hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)}{\sqrt{n}} \approx \frac{1}{n} \sum_{t=1}^{n} \left( Y_t - f_{t-1}(\hat{\theta}_n) \right) \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right)
\]

\[
= \frac{1}{n} \sum_{t=1}^{n} U_t \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right)
\]

\[
- \frac{1}{n} \sum_{t=1}^{n} \left( f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0) \right) \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right),
\]

34
where now \( U_t = Y_t - f_{t-1}(\theta_0) \) is no longer a martingale difference process. Hence,

\[
\sup_{k \in \mathbb{N}} \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon_k} |\hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)/\sqrt{n} - \psi_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)| \leq \frac{1}{n} \sum_{t=1}^{n} |f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0)|.
\]

Since by Lemma 5.1,

\[
\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^{n} |f_{t-1}(\theta) - f_{t-1}(\theta_0)| - E[|f_{t-1}(\theta) - f_{t-1}(\theta_0)|] \right| = o_p(1)
\]

and by Theorem 5.1, \( \hat{\theta}_n \overset{p}{\to} \theta_0 \), it follows from the continuity of \( E[|f_{t-1}(\theta) - f_{t-1}(\theta_0)|] \) that

\[
\frac{1}{n} \sum_{t=1}^{n} |f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0)| = E[|f_{t-1}(\theta) - f_{t-1}(\theta_0)|]_{\theta = \hat{\theta}_n} + o_p(1)
\]

which proves (6.3). The latter implies trivially the result (6.4).

The result (6.5) follows straightforwardly from Lemma 4.1, and (6.6) follows straightforwardly from (6.5). \( \blacksquare \)

**Lemma 6.2.** Under \( H_1 \) and Assumptions 3.1 and 4.1,

\[
\sup_{k \in \mathbb{N}} \hat{B}_{k,n}/n = o_p(1), \quad \sup_{k \in \mathbb{N}} \eta_{k,n} = o_p(1), \quad \sup_{k \in \mathbb{N}} \eta_{k} = O(1).
\]

**Proof.** It follows from (6.1) that

\[
|\psi_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|^2 \leq \left( \frac{1}{n} \sum_{t=1}^{n} |U_t| \right)^2 \leq \frac{1}{n} \sum_{t=1}^{n} U_t^2
\]

hence by (6.2),

\[
\sup_{k \in \mathbb{N}} \eta_{k,n} \leq \frac{1}{n} \sum_{t=1}^{n} U_t^2 = \hat{Q}_n(\theta_0) = Q(\theta_0) + o_p(1) = E[U_t^2] + o_p(1) \quad (6.7)
\]
where the equalities follow from (4.1) and (4.2). Obviously, (6.4) and (6.7) imply that
\[ \sup_{k \in \mathbb{N}} \frac{B_{k,n}}{n} = O_p(1). \]

Similarly, it follows from (3.14) that
\[ |\psi_k(\tau_1, \tau_2, \ldots, \tau_k)|^2 \leq (E[|U_1|])^2 \leq E[U_1^2] \]
hence \[ \sup_{k \in \mathbb{N}} \eta_k \leq E[U_1^2] = O(1). \quad \blacksquare \]

Using Lemmas 6.1 and 6.2, I will now prove that the WICM test is consistent.

**Theorem 6.1.** Under \( H_1 \) and Assumptions 3.1 and 4.1, and with \( \gamma_k \) and \( L_n \) as in Theorem 5.1,
\[ \tilde{T}_n/n = \sum_{k=1}^{L_n} \gamma_k \frac{\hat{B}_{k,n}}{n} \xrightarrow{p} \sum_{k=1}^{\infty} \gamma_k \eta_k > 0. \] (6.8)

**Proof.** For fixed \( \ell \leq L_n \),
\[ \left| \sum_{k=1}^{L_n} \gamma_k \frac{\hat{B}_{k,n}}{n} - \sum_{k=1}^{\infty} \gamma_k \frac{\hat{B}_{k,n}}{n} - \eta \right| \leq \sum_{k=1}^{\ell} \gamma_k \left| \frac{\hat{B}_{k,n}}{n} - \eta \right| \\
\quad + \left( \sum_{k=\ell+1}^{\infty} \gamma_k \right) \left( \sup_{k \in \mathbb{N}} \frac{\hat{B}_{k,n}}{n} + \sup_{k \in \mathbb{N}} \eta_{k,n} + \sup_{k \in \mathbb{N}} \eta_k \right). \] (6.9)

Let \( \varepsilon > 0 \) and \( \delta > 0 \) be arbitrary. By Lemma 6.2 we can choose \( \ell \) so large that
\[ \limsup_{n \to \infty} \Pr \left( \left( \sum_{k=\ell+1}^{\infty} \gamma_k \right) \left( \sup_{k \in \mathbb{N}} \frac{\hat{B}_{k,n}}{n} + \sup_{k \in \mathbb{N}} \eta_{k,n} + \sup_{k \in \mathbb{N}} \eta_k \right) > \delta/2 \right) < \varepsilon \] (6.10)
and for this \( \ell \) it follows from parts (6.5) and (6.6) of Lemma 6.1 that
\[ p \lim_{n \to \infty} \sum_{k=1}^{\ell} \gamma_k \left| \frac{\hat{B}_{k,n}}{n} - \eta_k \right| = 0, \]

hence
\[ \lim_{n \to \infty} \Pr \left( \sum_{k=1}^{\ell} \gamma_k \left| \frac{\hat{B}_{k,n}}{n} - \eta_k \right| > \delta/2 \right) = 0. \] (6.11)
It follows now from (5.20), (6.9), (6.10) and (6.11) that
\[
\limsup_{n \to \infty} \Pr \left[ \left| \sum_{k=1}^{L_n} \gamma_k \hat{B}_{k,n}/n - \sum_{k=1}^{\infty} \gamma_k \eta_k \right| > \delta \right] < \varepsilon,
\]
which by the arbitrariness of \(\varepsilon\) implies that
\[
\lim_{n \to \infty} \Pr \left[ \left| \sum_{k=1}^{L_n} \gamma_k \hat{B}_{k,n}/n - \sum_{k=1}^{\infty} \gamma_k \eta_k \right| > \delta \right] = 0.
\]
This result and the result of Lemma 3.2 imply (6.8).

7. Upper bounds of the critical values

I will now show that (2.10) applies to \(T = \sum_{k=1}^{\infty} \gamma_k B_k\) in Theorem 3.1 as well.

**Theorem 7.1.** Under the conditions of Theorem 5.1, and for fixed \(K \in \mathbb{N}\),
\[
T_K = \sum_{k=1}^{K} \gamma_k B_k = \sum_{j=1}^{\infty} \omega_{K,j} \varepsilon_{K,j}^2
\]
(7.1)
where the \(\omega_{K,j}\)'s are non-negative and satisfy \(\sum_{j=1}^{\infty} \omega_{K,j} = E[T_K]\), and the \(\varepsilon_{K,j}\)'s are independent standard normally distributed random variables. Hence,
\[
\Pr \left[ T_K / E[T_K] > y \right] \leq \Pr \left[ \chi_1^2 > y \right]
\]
(7.2)
for all \(y > 0\), with \(\chi_1^2\) the same as in (2.11). Moreover, letting \(K \to \infty\), it follows that for all \(y > 0\),
\[
\Pr \left[ T / E[T] > y \right] \leq \Pr \left[ \chi_1^2 > y \right].
\]
(7.3)

**Proof.** Recall that \(B_k = \int_{\Upsilon_k} \left| W_k(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)\), where \(W_k(.)\) is a continuous complex-valued zero-mean Gaussian process on \(\Upsilon^k\).

Denote
\[
W_k^+(\tau_1, \tau_2, \ldots, \tau_k, \tau_{k+1}) = \sqrt{\gamma_k} W_k(\tau_1, \tau_2, \ldots, \tau_k) \rho(\tau_{k+1}),
\]
where \(\rho(\tau)\) is a continuous real function on \(\Upsilon\) such that
\[
\int_{\Upsilon} \rho(\tau)^2 d\mu(\tau) = 1, \quad \int_{\Upsilon} \rho(\tau) d\mu(\tau) = 0.
\]
Then for \(1 \leq m < k \leq K\),

\[
\int_{\mathcal{T}_{K+1}} W_m^+(\tau_1, \tau_2, \ldots, \tau_{m+1}) W_k^+(\tau_1, \tau_2, \ldots, \tau_{k+1}) \, d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_{K+1})
\]

\[
= \sqrt{\gamma_m} \sqrt{\gamma_k} \int_{\mathcal{T}_k} W_m(\tau_1, \tau_2, \ldots, \tau_m) \rho(\tau_{m+1}) W_k(\tau_1, \tau_2, \ldots, \tau_k) \, d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \times \int_{\mathcal{T}} \rho(\tau) \, d\mu(\tau)
\]

\[
= 0
\]

whereas for \(m = k\),

\[
\int_{\mathcal{T}_{K+1}} W_k^+(\tau_1, \tau_2, \ldots, \tau_{k+1}) W_k^+(\tau_1, \tau_2, \ldots, \tau_{k+1}) \, d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_{K+1})
\]

\[
= \gamma_k \int_{\mathcal{T}_k} W_k(\tau_1, \tau_2, \ldots, \tau_m) W_k(\tau_1, \tau_2, \ldots, \tau_k) \, d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \times \int_{\mathcal{T}} \rho(\tau)^2 \, d\mu(\tau)
\]

\[
= \gamma_k \int_{\mathcal{T}_k} |W_k(\tau_1, \tau_2, \ldots, \tau_m)|^2 \, d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k)
\]

\[
= \gamma_k B_k
\]

Consequently, denoting

\[
\mathcal{W}_K(\tau_1, \tau_2, \ldots, \tau_{K+1}) = \sum_{k=1}^{K} W_k^+(\tau_1, \tau_2, \ldots, \tau_{k+1}),
\]

we have

\[
\int_{\mathcal{T}_{K+1}} |\mathcal{W}_K(\tau_1, \tau_2, \ldots, \tau_{K+1})|^2 \, d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_{K+1}) = \sum_{k=1}^{K} \gamma_k B_k. \quad (7.4)
\]

Note that for \(k < K\), \(W_k(\tau_1, \tau_2, \ldots, \tau_k) = W_K(\tau_1, \tau_2, \ldots, \tau_k, 0, \ldots, 0)\), hence

\[
\mathcal{W}_K(\tau_1, \tau_2, \ldots, \tau_{K+1}) = \sum_{k=1}^{K-1} \sqrt{\gamma_k} W_K(\tau_1, \tau_2, \ldots, \tau_k, 0, \ldots, 0) \rho(\tau_{k+1})
\]

\[
+ \sqrt{\gamma_K} W_K(\tau_1, \tau_2, \ldots, \tau_K) \rho(\tau_{K+1})
\]

38
is a continuous complex-valued zero mean Gaussian process on $\mathcal{Y}^{K+1}$ because $W_K(\tau_1, \tau_2, \ldots, \tau_K)$ is a continuous complex-valued zero mean Gaussian process on $\mathcal{Y}^K$ and the function $\rho(\tau)$ is continuous and nonrandom. Therefore, similar to (2.10), $T_K = \sum_{k=1}^{K} \gamma_k B_k$ has the representation (7.1), so that similar to (2.11),

$$\Pr \left[ \frac{T_K}{E[T_K]} > y \right] \leq \Pr \left[ \sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{j=1}^{n} \varepsilon_{K,j}^2 > y \right] = \Pr \left[ X_1^2 > y \right]. \tag{7.5}$$

Finally, recall from (5.23) that $\sup_{k \in \mathbb{N}} E[B_k] < \infty$, which implies that

$$\lim_{K \to \infty} T_K = T \text{ and } \lim_{K \to \infty} E[T_K] = E[T]$$

and thus $T_K/E[T_K] \overset{d}{\to} T/E[T]$ as $K \to \infty$. The latter is equivalent to

$$\lim_{K \to \infty} \Pr \left[ T_K/E[T_K] \leq y \right] = \Pr \left[ T/E[T] \leq y \right] \tag{7.6}$$

in the continuity points $y$ of the distribution function of $T/E[T]$, which proves inequality (7.6) for all continuity points $y$ of the distribution of $T/E[T]$.

However, in Theorem 7.3 at the end of this subsection it will be shown that the distribution function of $T$ is continuous on $(0, \infty)$, and therefore the same applies to $T/E[T]$. Consequently, inequality (7.6) holds for all $y > 0$. ■

To make these result operational for the WICM test statistic $\hat{T}_n = \sum_{k=1}^{L_n} \gamma_k \hat{B}_{n,k}$ we need estimates $\hat{\Gamma}_{n,k}$ of the covariance functions $\Gamma_k$ of the $W_k(\tau_1, \tau_2, \ldots, \tau_K)$'s such that

$$\hat{E}(\hat{T}_n) \overset{\text{def}}{=} \sum_{k=1}^{L_n} \gamma_k \int_{\mathcal{Y}^k} \hat{\Gamma}_{n,k} ( (\tau_1, \tau_2, \ldots, \tau_k), (\tau_1, \tau_2, \ldots, \tau_k) ) \, d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)$$

$$\overset{p}{\to} \sum_{k=1}^{\infty} \gamma_k \int_{\mathcal{Y}^k} \Gamma_k ( (\tau_1, \tau_2, \ldots, \tau_k), (\tau_1, \tau_2, \ldots, \tau_k) ) \, d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k) = E[T].$$

In Bierens (2014a, Section 5.1) I have shown how to derive closed form expressions of the $\hat{B}_{n,k}$'s and the consistent estimates $\hat{\Gamma}_{n,k}$ of $\Gamma_k$ for the case that $\mathcal{Y}$ is a hypercube centered around the zero vector and $\mu$ is the uniform probability measure on $\mathcal{Y}$. The formulas involved can easily adapted to the present case, which is left to the reader.
Now the following results hold for the standardized WICM test $\hat{T}_n / \hat{E}(\hat{T}_n)$.

**Theorem 7.2.** For $\alpha \in (0, 1)$, let $\tau(\alpha)$ be such that $\Pr[\chi^2_1 > \tau(\alpha)] = \alpha$, and let $\hat{E}(\hat{T}_n)$ be a consistent estimator of $E[T]$ under $H_0$, and $\hat{E}(\hat{T}_n) = O_p(1)$ under $H_1$. Under the conditions of Theorems 5.1 and 6.1, respectively,

\[
\limsup_{n \to \infty} \Pr \left[ \frac{\hat{T}_n}{\hat{E}(\hat{T}_n)} > \tau(\alpha) \right] \leq \alpha \text{ under } H_0,
\]

\[
\lim_{n \to \infty} \Pr \left[ \frac{\hat{T}_n}{\hat{E}(\hat{T}_n)} > \tau(\alpha) \right] = 1 \text{ under } H_1.
\]

**Remarks.**

- The condition that $\hat{E}(\hat{T}_n) = O_p(1)$ under $H_1$ is not restrictive, because without loss of generality we may base $\hat{E}(\hat{T}_n)$ on a bootstrap version of $\hat{T}_n$ for which $\hat{E}(\hat{T}_n) = O_p(1)$ under $H_1$. See (8.3) in the next section.

- The values of $\tau(\alpha)$ for $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$ are given in (2.12) and are the same as in Bierens and Ploberger (1997).

Finally, I will show that the limiting distribution of the WICM test is continuous.

**Theorem 7.3.** Let $G(x) = \Pr[T \leq x]$, where $T$ is the asymptotic null distribution of the WICM test in Theorem 5.1. Under the conditions of Theorem 5.1, $G(x)$ is continuous on $(0, \infty)$.

**Proof.** It follows from (7.4) and Theorem 6.4 in Bierens (2014a) that the distribution function

\[
G_K(x) = \Pr \left[ \sum_{k=1}^{K} \gamma_k B_k \leq x \right]
\]

is continuous and strictly monotonic on $(0, \infty)$. Moreover, note that

\[
\sum_{k=1}^{K} \gamma_k B_k \xrightarrow{d} \sum_{k=1}^{\infty} \gamma_k B_k
\]
as $K \to \infty$, hence
\[
\lim_{K \to \infty} G_K(x) = G(x)
\]
in the continuity points of $G$.

To prove that $G(x)$ is continuous on $(0, \infty)$, suppose that $x_0 > 0$ is a discontinuity point of $G$, i.e.,
\[
G(x_0) - \lim_{\delta \downarrow 0} G(x_0 - \delta) = \varepsilon > 0.
\]
Hence, for all $\delta > 0$ such that $x_0 - \delta$ is a continuity point of $G$,
\[
\varepsilon \leq G(x_0) - G(x_0 - \delta) = G(x_0) - \lim_{K \to \infty} G_K(x_0 - \delta) \tag{7.7}
\]
Since $G_K$ is continuous, for each $K$ there exists a $\delta_K > 0$ such that $G_K(x_0) - G_K(x_0 - \delta_K) < \varepsilon/2$, and since for $\delta \in (0, \delta_K]$, $G_K(x_0) - G_K(x_0 - \delta) < \varepsilon/2$ as well, we may without loss of generality assume that $\lim_{K \to \infty} \delta_K = 0$. Then for arbitrary $\delta > 0$ and $K$ so large that $\delta > \delta_K$, $G_K(x_0 - \delta_K) \geq G_K(x_0 - \delta)$, hence
\[
G(x_0) - G_K(x_0 - \delta) & \leq G(x_0) - G_K(x_0 - \delta_K) \\
& = G(x_0) - G_K(x_0) + G_K(x_0) - G_K(x_0 - \delta_K) \\
& \leq G(x_0) - G_K(x_0) + \varepsilon/2 \\
& \leq \varepsilon/2
\]
where that last inequality follows from $G(x_0) \leq G_K(x_0)$. Again, assuming that $x_0 - \delta$ is a continuity point of $G$, we have
\[
G(x_0) - G(x_0 - \delta) = G(x_0) - \lim_{K \to \infty} G_K(x_0 - \delta) \leq \varepsilon/2. \tag{7.8}
\]
However, (7.7) and (7.8) contradict, which proves that $G(x)$ does not have any discontinuity points.

8. Bootstrap critical values

In this section I will extend the bootstrap procedure in Bierens (2014a) to the present case. For that purpose it is necessary to strengthen some of the conditions in Assumptions 4.2 and the condition in Assumption 5.1 a little bit, as follows.
Assumption 8.1. Let $\theta_0$ be defined by (3.6). The following conditions hold, regardless whether $H_0$ is true or not.

(a) $\Theta$ is convex and $\theta_0$ is an interior point of $\Theta$.

(b) $f_{t-1}(\theta)$ is a.s. twice continuously differentiable in the components $\theta_1, \theta_2, ..., \theta_m$ of $\theta \in \Theta$.

(c) $E[\sup_{\theta \in \Theta} ||\nabla f_{t-1}(\theta)||^2] < \infty$.

(d) $E[\sup_{\theta \in \Theta} (Y_t - f_{t-1}(\theta))^2 ||\nabla f_{t-1}(\theta)||^2] < \infty$.

(e) $E[\sup_{\theta \in \Theta} (Y_t - f_{t-1}(\theta))^2 ||\nabla f_{t-1}(\theta)||] < \infty$.

(f) For $i_1, i_2 = 1, 2, ..., m$, $E[\sup_{\theta \in \Theta} |Y_t - f_{t-1}(\theta)|(\partial/\partial \theta_{i_1})(\partial/\partial \theta_{i_2})f_{t-1}(\theta)] < \infty$.

(g) There exists a sequence $M_t$ of random variables such that for all $\theta_1, \theta_2 \in \Theta$,

$$|| (Y_t - f_{t-1}(\theta_1)) \nabla f_{t-1}(\theta_1) - (Y_t - f_{t-1}(\theta_2)) \nabla f_{t-1}(\theta_2) || \leq M_t || \theta_1 - \theta_2 ||$$

and

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E[M_t^2] < \infty.$$ 

(h) For all $\theta \in \Theta$ the matrix $A_2(\theta) = E[|(\nabla f_{t-1}(\theta))(\nabla f_{t-1}(\theta))^\prime|]$ is nonsingular.

Except for part (g), this assumption encompasses the conditions in Assumptions 3.2 and 4.1, for ease of reference. Using the mean value theorem, Assumption 8.1(g) can be broken down in more primitive conditions. In particular, denoting $\nabla^2 f_{t-1}(\theta) = (\partial/\partial \theta_0)(\partial/\partial \theta')f_{t-1}(\theta)$, and defining the matrix norm as $||A|| = \sqrt{\text{trace}(AA')}$, it follows from the mean value theorem that for $\theta_1, \theta_2 \in \Theta$,

$$|| (Y_t - f_{t-1}(\theta_1)) \nabla f_{t-1}(\theta_1) - (Y_t - f_{t-1}(\theta_2)) \nabla f_{t-1}(\theta_2) ||$$

$$\leq |Y_t - f_{t-1}(\theta_1)| ||\nabla f_{t-1}(\theta_1) - \nabla f_{t-1}(\theta_2)|| + |f_{t-1}(\theta_1) - f_{t-1}(\theta_2)| ||\nabla f_{t-1}(\theta_2)||$$

$$\leq \left(\sup_{\theta \in \Theta} |Y_t - f_{t-1}(\theta)|, \sup_{\theta \in \Theta} ||\nabla^2 f_{t-1}(\theta)|| + \sup_{\theta \in \Theta} ||\nabla f_{t-1}(\theta)||^2 \right) || \theta_1 - \theta_2 ||.$$

Theorem 8.1. Denote

$$W_{i,k,n}(\tau_1, \tau_2, ..., \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_{t,i}(Y_t - f_{t-1}(\theta_0))\phi_{k,t-1}(\tau_1, \tau_2, ..., \tau_k),$$

$$B_{i,k,n}^* = \int_{\Theta} |W_{i,k,n}^*(\tau_1, \tau_2, ..., \tau_k)|^2 d\mu(\tau_1)d\mu(\tau_2)...d\mu(\tau_k),$$

$$T_{i,n}^* = \sum_{k=1}^{L_\mu} \gamma_k B_{i,k,n}^*, \ i = 1, 2, ..., M,$$
where the $\varepsilon_{i,t}$'s are independent random drawings from the standard normal distribution, $M$ is the number of bootstraps, $\phi_{k,t-1}(\tau_1, \tau_2, \ldots, \tau_k)$ is defined by (5.9), and $\gamma_k$ and $L_n$ are the same as in Theorem 5.1. Under Assumptions 3.1, 4.1 and 8.1 the following results hold.

(a) For each $k \in \mathbb{N}$ and $i = 1, 2, \ldots, M$,

$$W_{i,k,n}^*(\tau_1, \tau_2, \ldots, \tau_k) \Rightarrow W_{i,k}^*(\tau_1, \tau_2, \ldots, \tau_k)$$

on $\mathcal{T}^k$, where $W_{i,k}^*(\tau_1, \tau_2, \ldots, \tau_k)$ is a zero-mean complex-valued Gaussian process on $\mathcal{T}^k$ with covariance function

$$
\Gamma_k^*((\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,k}), (\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,k}))
= E \left[ W_{i,k}^*(\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,k}) W_{i,k}^*(\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,k}) \right]
= E \left[ (Y_t - f_{t-1}(\theta_0))^2 \phi_{k,t-1}(\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,k}) \phi_{k,t-1}(\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,k}) \right]
= E \left[ (Y_t - E[Y_t|\mathcal{F}_{t-1}^{\infty}])^2 \phi_{k,t-1}(\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,k}) \phi_{k,t-1}(\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,k}) \right]
+ E \left[ (f_{t-1}(\theta_0) - E[Y_t|\mathcal{F}_{t-1}^{\infty}])^2 \right.
\times \phi_{k,t-1}(\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,k}) \phi_{k,t-1}(\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,k}) \bigg]
= \Gamma_k^*((\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,k}), (\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,k}))
+ E \left[ (f_{t-1}(\theta_0) - E[Y_t|\mathcal{F}_{t-1}^{\infty}])^2 \right.
\times \phi_{k,t-1}(\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,k}) \phi_{k,t-1}(\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,k}) \bigg] \quad (8.1)
$$

with $\Gamma_k$ the covariance function (5.21). Hence, $B_{i,k,n}^* \overset{d}{\rightarrow} B_{i,k}^*$, and

$$T_{i,n} \overset{d}{\rightarrow} T_i^* = \sum_{k=1}^{\infty} \gamma_k B_{i,k}^*$$

where

$$B_{i,k}^* = \int_{\mathcal{T}^k} |W_{i,k}^*(\tau_1, \tau_2, \ldots, \tau_k)|^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k). \quad (8.2)$$

(b) Moreover,

$$(T_{i,n}^*, T_{2,n}^*, \ldots, T_{M,n}^*)' \overset{d}{\rightarrow} (T_{1}^*, T_{2}^*, \ldots, T_{M}^*)'$$

43
where \( T_1^*, T_2^*, \ldots, T_M^* \) are i.i.d. Furthermore, under \( H_0 \),
\[
(T_{1,n}^*, T_{2,n}^*, \ldots, T_{M,n}^*)' \overset{d}{\to} (T_1, T_2, \ldots, T_M)',
\]
where \( T_1, T_2, \ldots, T_M \) are i.i.d. as \( T \) in (5.28) and are independent of \( T \).

**Proof.** Part (a) follows similar to Lemma 5.5 and Theorem 5.1. As to part (b), it follows from the independence of \( \varepsilon_{i,t} \)'s that, conditional on the data, \( T_1^*, T_2^*, \ldots, T_M^* \) are i.i.d. and are therefore i.i.d. unconditionally as well. The same applies to \( T_1, T_2, \ldots, T_M \). Moreover, under \( H_0 \) the covariance function (8.1) is exactly the same as (5.21), hence \( T_i \sim T \). Furthermore, since the \( \varepsilon_{i,t} \)'s are drawn independently of the data, \((T_1, T_2, \ldots, T_M)\)' is independent of \( T \). ■

If it were possible to compute the \( T_{i,n}^* \)'s then the asymptotic \( \alpha \times 100\% \) critical values \( c(\alpha) \) of the WICM test can be approximated by sorting the \( T_{i,n}^* \)'s in decreasing order and then use \( c_{n,M}(\alpha) = T^*_{[\alpha M],n} \) as the approximation of \( c(\alpha) \), where \([\alpha M]\) denotes the largest natural number less or equal to \( \alpha M \).

However, obviously the computation of the \( T_{i,n}^* \)'s is not possible. Therefore, to make this bootstrap procedure feasible, we need to construct feasible bootstrap WICM statistics \( \widetilde{T}_{i,n} \) such that \( \widetilde{T}_{i,n} = T_{i,n}^* + o_p(1) \), as follows.

Denote for \( i = 1, 2, \ldots, M \),
\[
\widetilde{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k|\theta) = \frac{1}{n} \sum_{t=1}^{n} \nabla f_{t-1}(\theta) \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right),
\]
\[
\widetilde{A}_{2,n}(\theta) = \frac{1}{n} \sum_{t=1}^{n} \nabla f_{t-1}(\theta)(\nabla f_{t-1}(\theta))',
\]
\[
\widetilde{\phi}_{k,n,t-1}(\tau_1, \tau_2, \ldots, \tau_k|\theta) = \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right)
\]
\[
-\tilde{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k|\theta)' \tilde{A}_{2,n}^{-1}(\theta) \nabla f_{t-1}(\theta),
\]
\[
\widetilde{W}_{i,k,n}(\tau_1, \tau_2, \ldots, \tau_k|\theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_{i,t}(Y_t - f_{t-1}(\theta)) \widetilde{\phi}_{k,n,t-1}(\tau_1, \tau_2, \ldots, \tau_k|\theta),
\]
\[
\widetilde{B}_{i,k,n} = \int_{T^n} \left| \widetilde{W}_{i,k,n}(\tau_1, \tau_2, \ldots, \tau_k|\theta) \right|^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k),
\]
\[
\widetilde{T}_{i,n} = \sum_{k=1}^{L_n} \gamma_k \widetilde{B}_{i,k,n}, \quad (8.3)
\]
where the $\varepsilon_{i,j}$'s are the same as in Theorem 8.1.

We can write

$$
\widehat{W}_{i,k,n}(\tau_1, \tau_2, \ldots, \tau_k | \hat{\theta}_n)
= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_{i,t} (Y_t - f_{t-1}(\theta_0)) \exp \left( i \sum_{j=1}^{k} \tau_j \Phi(Z_{t-j}) \right) \\
- \hat{a}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k)
- \hat{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k \hat{\theta}_n) A_{2,n}^{-1}(\hat{\theta}_n) \hat{d}_{n,k,i}(\hat{\theta}_n, \tau_1, \tau_2, \ldots, \tau_k)
$$

where

$$
\hat{a}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_{i,t} (f_{t-1}(\theta) - f_{t-1}(\theta_0)) \\
\times \exp \left( i \sum_{j=1}^{k} \tau_j \Phi(Z_{t-j}) \right),
$$

$$
\hat{d}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_{i,t} \Psi_{k,t}(\theta, \tau_1, \tau_2, \ldots, \tau_k), \text{ with}
$$

$$
\Psi_{k,t}(\theta, \tau_1, \tau_2, \ldots, \tau_k) = (Y_t - f_{t-1}(\theta)) \nabla f_{t-1}(\theta) \exp \left( i \sum_{j=1}^{k} \tau_j \Phi(Z_{t-j}) \right).
$$

Similarly,

$$
W_{i,k,n}^{*}(\tau_1, \tau_2, \ldots, \tau_k)
= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_{i,t} (Y_t - f_{t-1}(\theta_0)) \exp \left( i \sum_{j=1}^{k} \tau_j \Phi(Z_{t-j}) \right) \\
- b_k(\tau_1, \tau_2, \ldots, \tau_k) A_{2,n}^{-1} \hat{a}_{n,k,i}(\theta_0, \tau_1, \tau_2, \ldots, \tau_k).
$$

I will show first that

$$
\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon_k} \left| \widehat{W}_{i,k,n}(\tau_1, \tau_2, \ldots, \tau_k | \hat{\theta}_n) - W_{i,k,n}^{*}(\tau_1, \tau_2, \ldots, \tau_k) \right| = o_p(1), \quad (8.4)
$$

using the following three lemmas, and then that

$$
\widehat{T}_{i,n} = T_{i,n}^{*} + o_p(1). \quad (8.5)
$$
Lemma 8.1. Under the conditions of Theorem 8.1 and for each \( k \) and \( i \), \( \hat{d}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots \tau_k) \) is tight on \( \Theta \times \Upsilon^k \). Consequently,

\[
\sup_{(\tau_1, \tau_2, \ldots \tau_k) \in \Upsilon^k} \left\| \hat{d}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots \tau_k) - \hat{d}_{n,k,i}(\theta_0, \tau_1, \tau_2, \ldots \tau_k) \right\| = o_p(1), \tag{8.6}
\]

where here and in the sequel the norm \( ||.|| \) on \( \mathbb{C}^m \) is defined as \( ||a + i.b|| = \sqrt{a^2 + b^2} \).

Proof. According to Lemma A.1 in Bierens and Ploberger (1997), in multivariate form, \( \text{Re}[\hat{d}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots \tau_k)] \) is tight on \( \Theta \times \Upsilon^k \) if for any pair

\[
(\theta_1, \tau_{1,1}, \tau_{1,2}, \ldots \tau_{1,k}), \ (\theta_2, \tau_{2,1}, \tau_{2,2}, \ldots \tau_{2,k}) \in \Theta \times \Upsilon^k,
\]

the following Lipschitz condition holds:

\[
\left| \left| \text{Re}[\Psi_{k,t}(\theta_1, \tau_{1,1}, \tau_{1,2}, \ldots \tau_{1,k})] - \text{Re}[\Psi_{k,t}(\theta_2, \tau_{2,1}, \tau_{2,2}, \ldots \tau_{2,k})] \right| \right| \leq K_{k,t} \sqrt{||\theta_1 - \theta_2^2 + \sum_{j=1}^k ||\tau_{2,j} - \tau_{1,j}||^2}, \tag{8.7}
\]

where \( K_{k,t} \) is a sequence of random variables such that

\[
\limsup_{n \to \infty} \frac{1}{n} \sum_{t=1}^n E \left[ \varepsilon_{i,t}^2 K_{k,t}^2 \right] < \infty, \tag{8.8}
\]

and for an arbitrary point \( (\theta, \tau_1, \tau_2, \ldots \tau_k) \in \Theta \times \Upsilon^k \),

\[
\limsup_{n \to \infty} \frac{1}{n} \sum_{t=1}^n E \left[ \varepsilon_{i,t}^2 \left| \text{Re}(\Psi_{k,t}(\theta, \tau_1, \tau_2, \ldots \tau_k)) \right|^2 \right] < \infty. \tag{8.9}
\]

Condition (8.7) follows from

\[
\left| \left| \text{Re}[\Psi_{k,t}(\theta_1, \tau_{1,1}, \tau_{1,2}, \ldots \tau_{1,k})] - \text{Re}[\Psi_{k,t}(\theta_2, \tau_{2,1}, \tau_{2,2}, \ldots \tau_{2,k})] \right| \right|
\]

\[
= \left| \left| (Y_t - f_{t-1}(\theta_1)) \nabla f_{t-1}(\theta_1) \cos \left( \sum_{j=1}^k \tau'_{1,j} \Phi(Z_{t-j}) \right) \right| \right|
\]

46
\(- (Y_t - f_{t-1}(\theta_2)) \nabla f_{t-1}(\theta_2) \cos \left( \sum_{j=1}^{k} \tau_{2,j} \Phi(Z_{t-j}) \right) \leq ||(Y_t - f_{t-1}(\theta_1)) \nabla f_{t-1}(\theta_1) - (Y_t - f_{t-1}(\theta_2)) \nabla f_{t-1}(\theta_2)||
\]
\[+ \left| \left| (Y_t - f_{t-1}(\theta_2)) \nabla f_{t-1}(\theta_2) \sum_{j=1}^{k} (\tau_{2,j} - \tau_{1,j})' \Phi(Z_{t-j}) \right| \right|
\leq M_{t-1} ||\theta_1 - \theta_2||
\[+ \left( \sup_z ||\Phi(z)|| \right) \left( \sup_{\theta \in \Theta} (|Y_t - f_{t-1}(\theta)| \cdot ||\nabla f_{t-1}(\theta)||) \left( \sum_{j=1}^{k} ||\tau_{2,j} - \tau_{1,j}|| \right) \right)
\leq K_{k,t} \sqrt{||\theta_1 - \theta_2||^2 + \sum_{j=1}^{k} ||\tau_{2,j} - \tau_{1,j}||^2},
\]
where the second inequality follows from Assumption 8.1(g), and the last inequality from
\[K_{k,t} = \sqrt{k + 1} \left( M_t + \left( \sup_z ||\Phi(z)|| \right) \left( \sup_{\theta \in \Theta} (|Y_t - f_{t-1}(\theta)| \cdot ||\nabla f_{t-1}(\theta)||) \right) \right)
\]
and
\[||\theta_1 - \theta_2|| + \sum_{j=1}^{k} ||\tau_{2,j} - \tau_{1,j}|| \leq \sqrt{k + 1} \sqrt{||\theta_1 - \theta_2||^2 + \sum_{j=1}^{k} ||\tau_{2,j} - \tau_{1,j}||^2}.
\]
Moreover, condition (8.8) follows from
\[\lim_{n \to \infty} \sup \frac{1}{n} \sum_{t=1}^{n} E \left[ \varepsilon_{i,t}^2 K_{k,t}^2 \right] = \lim_{n \to \infty} \sup \frac{1}{n} \sum_{t=1}^{n} E \left[ K_{k,t}^2 \right]
\leq 2(k + 1) \lim_{n \to \infty} \sup \frac{1}{n} \sum_{t=1}^{n} E[M_t^2]
\[+ 2(k + 1) \sup_z ||\Phi(z)||^2 E \left[ \sup_{\theta \in \Theta} (Y_t - f_{t-1}(\theta))^2 \cdot ||\nabla f_{t-1}(\theta)||^2 \right]
\leq 2(k + 1) \lim_{n \to \infty} \sup \frac{1}{n} \sum_{t=1}^{n} E[M_t^2]
\[+ 2(k + 1) \sup_z ||\Phi(z)||^2 E \left[ \sup_{\theta \in \Theta} (Y_t - f_{t-1}(\theta))^2 \cdot ||\nabla f_{t-1}(\theta)||^2 \right]
\leq 2(k + 1) \lim_{n \to \infty} \sup \frac{1}{n} \sum_{t=1}^{n} E[M_t^2]
\leq 2(k + 1) \lim_{n \to \infty} \sup \frac{1}{n} \sum_{t=1}^{n} E[M_t^2]
\[+ 2(k + 1) \sup_z ||\Phi(z)||^2 E \left[ \sup_{\theta \in \Theta} (Y_t - f_{t-1}(\theta))^2 \cdot ||\nabla f_{t-1}(\theta)||^2 \right]
\leq \infty,
\]
where the last inequality is due to parts (d) and (g) of Assumption 8.1.
Finally, condition (8.9) follows from part (d) of Assumption 8.1 and the easy inequality

\[ E \left[ \left| \text{Re}(\Psi_{k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k)) \right|^2 \right] \leq E \left[ \sup_{\theta \in \Theta} (Y_i - f_{t-1}(\theta))^2 \left| \nabla f_{t-1}(\theta) \right|^2 \right]. \]

Thus, \( \text{Re}[\hat{d}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k)] \) is tight on \( \Theta \times \Upsilon^k \). Obviously, the same holds for \( \text{Im}[\hat{d}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k)] \) and thus for \( \hat{d}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) \) as well.

Using the martingale difference central limit theorem of McLeish (1974) it follows straightforwardly that the finite distributions of \( \hat{d}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) \) converge to multivariate zero-mean complex-valued normal distributions, hence

\[ \hat{d}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) \Rightarrow d_{k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) \]
on \( \Theta \times \Upsilon^k \), where \( d_{k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) \) is a \( m \)-dimensional zero-mean complex-valued Gaussian process on \( \Theta \times \Upsilon^k \). Similarly,

\[ \hat{d}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) - \hat{d}_{n,k,i}(\theta_0, \tau_1, \tau_2, \ldots, \tau_k) \Rightarrow d_{k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) - d_{k,i}(\theta_0, \tau_1, \tau_2, \ldots, \tau_k) \]

(8.10) on \( \Theta \times \Upsilon^k \).

Now let \( \Theta_0(\varepsilon) = \{ \theta \in \mathbb{R}^m : \| \theta - \theta_0 \| \leq \varepsilon \} \), where \( \varepsilon > 0 \) is so small that \( \Theta_0(\varepsilon) \subset \Theta \). and denote

\[ \hat{\Delta}_{n,k,i}(\theta) = \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left\| \hat{d}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) - \hat{d}_{n,k,i}(\theta_0, \tau_1, \tau_2, \ldots, \tau_k) \right\|, \]

\[ \Delta_{k,i}(\theta) = \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left\| d_{k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) - d_{k,i}(\theta_0, \tau_1, \tau_2, \ldots, \tau_k) \right\|. \]

Then by the continuous mapping theorem, (8.10) implies

\[ \sup_{\theta \in \Theta_0(\varepsilon)} \hat{\Delta}_{n,k,i}(\theta) \overset{d}{=} \sup_{\theta \in \Theta_0(\varepsilon)} \Delta_{k,i}(\theta) \]

and thus for any \( \delta > 0 \),

\[ \Pr \left[ \hat{\Delta}_{n,k,i}(\hat{\theta}_n) > \delta \right] = \Pr \left[ \hat{\Delta}_{n,k,i}(\hat{\theta}_n) > \delta \text{ and } \hat{\theta}_n \in \Theta_0(\varepsilon) \right] \]

\[ + \Pr \left[ \hat{\Delta}_{n,k,i}(\hat{\theta}_n) > \delta \text{ and } \hat{\theta}_n \notin \Theta_0(\varepsilon) \right] \]

\[ \leq \Pr \left[ \sup_{\theta \in \Theta_0(\varepsilon)} \hat{\Delta}_{n,k,i}(\theta) > \delta \right] + \Pr \left[ \left\| \hat{\theta}_n - \theta_0 \right\| > \varepsilon \right] \]

\[ \to \Pr \left[ \sup_{\theta \in \Theta_0(\varepsilon)} \Delta_{k,i}(\theta) > \delta \right] \]

48
Since \( \lim_{\varepsilon \downarrow 0} \sup_{\theta \in \Theta(\varepsilon)} \Delta_{k,i}(\theta) = \Delta_{k,i}(\theta_0) = 0 \) a.s. and thus
\[
\lim_{\varepsilon \downarrow 0} \Pr \left( \sup_{\theta \in \Theta(\varepsilon)} \Delta_{k,i}(\theta) > \delta \right) = 0,
\]
it follows now that for any \( \delta > 0 \), \( \lim_{n \to \infty} \Pr \left[ \Delta_{n,k,i}(\hat{\theta}_n) > \delta \right] = 0 \). The latter proves (8.6). ■

Similarly,

**Lemma 8.2.** Under the conditions of Theorem 8.1 and for each \( k \) and \( i \), \( \hat{a}_{n,k,i}(\theta, \tau_1, \tau_2, \ldots, \tau_k) \) is tight on \( \Theta \times \Upsilon^k \). Consequently,
\[
\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left| \hat{a}_{n,k,i}(\hat{\theta}_n, \tau_1, \tau_2, \ldots, \tau_k) \right| = o_p(1).
\]

Moreover, it follows from Assumption 8.1(c) and Theorem 2.1 that
\[
\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \sup_{\theta \in \Theta} \left\| \Re \left[ \hat{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k | \theta) \right] - E \left( \Re \left[ \hat{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k | \theta) \right] \right) \right\| = o_p(1),
\]
\[
\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \sup_{\theta \in \Theta} \left\| \Im \left[ \hat{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k | \theta) \right] - E \left( \Im \left[ \hat{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k | \theta) \right] \right) \right\| = o_p(1),
\]
and
\[
\sup_{\theta \in \Theta} \left\| \hat{A}_{2,n}(\theta) - E \left[ \hat{A}_{2,n}(\theta) \right] \right\| = o_p(1),
\]
where in the latter case the matrix norm involved is defined by \( ||A|| = \sqrt{\text{trace}(AA^T)} \), hence by Lemma 2.1 and Assumption 8.1(h) it follows that

**Lemma 8.3.** Under the conditions of Theorem 8.1 and for each \( k \),
\[
\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left\| \hat{A}_{2,n}(\hat{\theta}_n)^{-1} \hat{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k | \hat{\theta}_n) - A_2^{-1}(\theta_0) b_k(\tau_1, \tau_2, \ldots, \tau_k) \right\| = o_p(1).
\]

Combining the results of Lemmas 8.1-8.3 it follows that (8.4) holds, which in its turn implies that,
Lemma 8.4. Under the conditions of Theorem 8.1, $|\bar{B}_{i,k,n} - B_{i,k,n}^*| = o_p(1)$ for each $k \in \mathbb{N}$ and each $i = 1, 2, \ldots, M$.

Proof. Denote
\[
\delta_{i,k,n} = \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \mathcal{T}^k} \left| \tilde{W}_{i,k,n}(\tau_1, \tau_2, \ldots, \tau_k| \hat{\theta}_n) - W_{i,k,n}^*(\tau_1, \tau_2, \ldots, \tau_k) \right|.
\]

It is easy to verify that
\[
|\bar{B}_{i,k,n} - B_{i,k,n}^*| \leq \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \mathcal{T}^k} \left| \tilde{W}_{i,k,n}(\tau_1, \tau_2, \ldots, \tau_k| \hat{\theta}_n) - W_{i,k,n}^*(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 \leq \delta_{i,k,n}^2 + 2\delta_{i,k,n} \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \mathcal{T}^k} \left| W_{i,k,n}^*(\tau_1, \tau_2, \ldots, \tau_k) \right|.
\]

Since by part (a) of Theorem 8.1,
\[
W_{i,k,n}^*(\tau_1, \tau_2, \ldots, \tau_k) \Rightarrow W_{i,k}^*(\tau_1, \tau_2, \ldots, \tau_k)
\]
on $\mathcal{T}^k$, it follows from the continuous mapping theorem that
\[
\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \mathcal{T}^k} \left| W_{i,k,n}^*(\tau_1, \tau_2, \ldots, \tau_k) \right| \overset{d}{\to} \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \mathcal{T}^k} \left| W_{i,k}^*(\tau_1, \tau_2, \ldots, \tau_k) \right|,
\]
which implies that
\[
\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \mathcal{T}^k} \left| W_{i,k,n}^*(\tau_1, \tau_2, \ldots, \tau_k) \right| = O_p(1).
\]
The lemma now follows straightforwardly from the latter and $\delta_{i,k,n} = o_p(1)$. \hfill \blacksquare

Lemma 8.5. Under the conditions of Theorem 8.1,
\[
\sum_{k=1}^{L_n} \gamma_k \tilde{B}_{i,k,n} = \sum_{k=1}^{\infty} \gamma_k B_{i,k,n}^* = o_p(1),
\]
hence
\[
(\tilde{T}_{1,n}, \tilde{T}_{2,n}, \ldots, \tilde{T}_{M,n})' = (T_{1,n}^*, T_{2,n}^*, \ldots, T_{M,n}^*)' + o_p(1),
\]
50
where $T_{i,n}^*$ is defined in Theorem 8.1 and $\tilde{T}_{i,n}$ in (8.3).

**Proof.** Note that similar to (5.18),
\[
E \left[ |\tilde{W}_{i,k,n}(\tau_1, \tau_2, \ldots, \tau_k|\theta)|^2 \middle| \mathcal{F}_{-\infty}^\infty \right]
\]
\[
= \frac{1}{n} \sum_{t=1}^{n} (Y_t - f_{t-1}(\theta))^2
\]
\[
- 2 \text{Re} [\tilde{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k|\theta)'] \tilde{A}_{2,1,n}(\theta) \text{Re} [\tilde{c}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k|\theta)]
\]
\[
- 2 \text{Im} [\tilde{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k|\theta)'] \tilde{A}_{2,1,n}(\theta) \text{Im} [\tilde{c}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k|\theta)]
\]
\[
+ \text{Re} [\tilde{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k|\theta)'] \tilde{A}_{2,1,n}(\theta) \tilde{A}_{2,1,n}(\theta) \text{Re} [\tilde{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k|\theta)]
\]
\[
+ \text{Im} [\tilde{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k|\theta)'] \tilde{A}_{2,1,n}(\theta) \tilde{A}_{2,1,n}(\theta) \text{Im} [\tilde{b}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k|\theta)]
\]

where
\[
\tilde{c}_{n,k}(\tau_1, \tau_2, \ldots, \tau_k|\theta) = \frac{1}{n} \sum_{t=1}^{n} (Y_t - f_{t-1}(\theta))^2 \nabla f_{t-1}(\theta) \exp \left( i \sum_{j=1}^{k} \tau^j \Phi(Z_{t-j}) \right),
\]
\[
\tilde{A}_{2,1,n}(\theta) = \frac{1}{n} \sum_{t=1}^{n} (Y_t - f_{t-1}(\theta))^2 (\nabla f_{t-1}(\theta))(\nabla f_{t-1}(\theta))'.
\]

Then similar to (5.19),
\[
\int_{\mathcal{T}^k} E \left[ |\tilde{W}_{i,k,n}(\tau_1, \tau_2, \ldots, \tau_k|\theta)|^2 \middle| \mathcal{F}_{-\infty}^\infty \right] d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k)
\]
\[
\leq \frac{1}{n} \sum_{t=1}^{n} (Y_t - f_{t-1}(\theta))^2 + 4 \text{trace}(\tilde{A}_{2,n}(\theta)) \sqrt{\text{trace}(\tilde{A}_{2,n}(\theta))}
\]
\[
+ 4 \text{trace}(\tilde{A}_{2,n}(\theta)) \sqrt{\text{trace}(\tilde{A}_{2,n}(\theta))} \frac{1}{n} \sum_{t=1}^{n} (Y_t - f_{t-1}(\theta))^2 ||\nabla f_{t-1}(\theta)||
\]
\[
+ 2 \text{trace} \left( \tilde{A}_{2,n}(\theta) \tilde{A}_{1,n}(\theta) \tilde{A}_{2,n}(\theta) \right) \text{trace} (\tilde{A}_{2,n}(\theta))
\]
\[
= \hat{R}_n(\theta),
\]

say, hence $E \left[ |\tilde{B}_{i,k,n}| \mathcal{F}_{-\infty}^\infty \right] \leq \hat{R}_n(\hat{\theta}_n)$ and thus for $K \in \mathbb{N},$
\[
E \left[ \sum_{k=K}^{\infty} \gamma_k |\tilde{B}_{i,k,n}| \mathcal{F}_{-\infty}^\infty \right] \leq \hat{R}_n(\hat{\theta}_n) \sum_{k=K}^{\infty} \gamma_k.
\]

51
It is not hard (but somewhat tedious) to verify from Assumption 8.1 and
Theorem 2.1 that $\sup_{\theta \in \Theta} \left| \hat{R}_n(\theta) - R(\theta) \right| = o_p(1)$ and thus by Lemma 2.1,
$$\hat{R}_n(\hat{\theta}_n) \xrightarrow{p} R(\theta_0),$$
where
$$R(\theta) = E[(Y_t - f_{t-1}(\theta))^2] + 4.\text{trace}(A_2(\theta))\sqrt{\text{trace}(A_2(\theta))}$$
$$+ 4.\text{trace}(A_2(\theta))\sqrt{\text{trace}(A_2(\theta))}E[(Y_t - f_{t-1}(\theta))^2||\nabla f_{t-1}(\theta)||]$$
$$+ 2.\text{trace}(A_2^1(\theta)A_1(\theta)A_2^{-1}(\theta))\text{trace}(A_2(\theta))$$
with $A_2(\theta)$ defined in Assumption 8.1(h) and
$$A_1(\theta) = E[(Y_t - f_{t-1}(\theta))^2(\nabla f_{t-1}(\theta))(\nabla f_{t-1}(\theta))^\prime].$$

Now by Chebyshev’s inequality for conditional probabilities and first moments, it follows that for arbitrary $\beta > 0$,
$$\Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta \hat{R}_n(\hat{\theta}_n) \right| \mathcal{F}_{-\infty}^\infty \right] \leq \frac{E \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} \mathcal{F}_{-\infty}^\infty \right]}{\beta \hat{R}_n(\hat{\theta}_n)} \leq \sum_{k=K}^{\infty} \frac{\gamma_k}{\beta},$$
hence
$$\sum_{k=K}^{\infty} \frac{\gamma_k}{\beta} \geq E \left( \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta \hat{R}_n(\hat{\theta}_n) \right| \mathcal{F}_{-\infty}^\infty \right)$$
$$= \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta \hat{R}_n(\hat{\theta}_n) \right]$$
$$= \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta \hat{R}_n(\hat{\theta}_n) \text{ and } |\hat{R}_n(\hat{\theta}_n) - R(\theta_0)| > 1 \right]$$
$$+ \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta \hat{R}_n(\hat{\theta}_n) \text{ and } |\hat{R}_n(\hat{\theta}_n) - R(\theta_0)| \leq 1 \right]$$
$$\geq \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta \hat{R}_n(\hat{\theta}_n) \text{ and } |\hat{R}_n(\hat{\theta}_n) - R(\theta_0)| > 1 \right]$$
$$+ \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta \hat{R}_n(\hat{\theta}_n) \text{ and } |\hat{R}_n(\hat{\theta}_n) - R(\theta_0)| \leq 1 \right]$$

52
\[
\begin{align*}
&= \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta \hat{R}_n(\hat{\theta}_n) \text{ and } |\hat{R}_n(\hat{\theta}_n) - R(\theta_0)| > 1 \right] \\
&+ \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} = \beta (R(\theta_0) + 1) \right] \\
&- \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta (R(\theta_0) + 1) \text{ and } |\hat{R}_n(\hat{\theta}_n) - R(\theta_0)| > 1 \right]
\end{align*}
\]

(8.11)

Since
\[
\limsup_{n \to \infty} \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta \hat{R}_n(\hat{\theta}_n) \text{ and } |\hat{R}_n(\hat{\theta}_n) - R(\theta_0)| > 1 \right] \\
\leq \limsup_{n \to \infty} \Pr \left[ |\hat{R}_n(\hat{\theta}_n) - R(\theta_0)| > 1 \right] = 0
\]

and
\[
\limsup_{n \to \infty} \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta (R(\theta_0) + 1) \text{ and } |\hat{R}_n(\hat{\theta}_n) - R(\theta_0)| > 1 \right] \\
\leq \limsup_{n \to \infty} \Pr \left[ |\hat{R}_n(\hat{\theta}_n) - R(\theta_0)| > 1 \right] = 0
\]

it follows now from (8.11) that for arbitrary \( \beta > 0 \),
\[
\limsup_{n \to \infty} \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \beta (R(\theta_0) + 1) \right] \leq \frac{\sum_{k=K}^{\infty} \gamma_k}{\beta}
\]

For arbitrary \( \varepsilon > 0 \) and \( \delta \in (0, 1) \), choose \( \beta = \varepsilon / (R(\theta_0) + 1) \) and let \( K_0(\varepsilon, \delta) \in \mathbb{N} \) be such that \( \sum_{k=K_0(\varepsilon, \delta)}^{\infty} \gamma_k < \delta \beta \). Then for \( K \geq K_0(\varepsilon, \delta) \),
\[
\limsup_{n \to \infty} \Pr \left[ \sum_{k=K}^{\infty} \gamma_k \tilde{B}_{i,k,n} > \varepsilon \right] < \delta. \quad (8.12)
\]

Note that similar to (5.19), \( E[B_{i,k,n}^*] \leq R(\theta_0) \), so that \( K_0(\varepsilon, \delta) \) can be chosen such that for \( K \geq K_0(\varepsilon, \delta) \),
\[
\limsup_{n \to \infty} \Pr \left[ \sum_{k=K}^{\infty} \gamma_k B_{i,k,n}^* > \varepsilon \right] < \delta. \quad (8.13)
\]
It follows now from (8.12), (8.13) and Lemma 8.4, similar to the proof of Lemma 5.4, that 
\[ \sum_{k=1}^{\infty} \gamma_k B_{i,k,n} - \sum_{k=1}^{L_n} \gamma_k B^\ast_{i,k,n} = o_p(1) \] and 
\[ \sum_{k=L_n+1}^{\infty} \gamma_k B_{i,k,n} = o_p(1). \]

Summarizing, the following results have been proved.

**Theorem 8.2.** Under the conditions and notations in Theorem 8.1 it follows that 
\[ (\widetilde{T}_{i,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n})' \overset{d}{\rightarrow} (T^\ast_1, T^\ast_2, ..., T^\ast_M)', \] 
where \( T^\ast_1, T^\ast_2, ..., T^\ast_M \) are i.i.d.. Moreover, with \( \hat{T}_n \) the WICM test statistic, it follows that under \( H_0 \), 
\[ (\hat{T}_n, \widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n})' \overset{d}{\rightarrow} (T, T_1, T_2, ..., T_M)', \] 
where \( T, T_1, T_2, ..., T_M \) are i.i.d.

Again, the asymptotic \( \alpha \times 100\% \) critical values \( c(\alpha) \) of the WICM test, i.e., 
\[ \Pr[T > c(\alpha)] = \alpha, \] can be approximated by sorting the \( \widetilde{T}_{i,n} \)'s in decreasing order and then use \( \tilde{c}_{n,M}(\alpha) = \widetilde{T}_{[\alpha M],n} \) as the approximation of \( c(\alpha) \). More precisely,

**Theorem 8.3.** Let \( \tilde{c}_{n,M}(\alpha) \) be the \( 1 - \alpha \) quantile of the empirical distribution function 
\[ \tilde{G}_{n,M}(x) = \frac{1}{M} \sum_{i=1}^{M} I(\widetilde{T}_{i,n} \leq x), \] 
i.e., \( \tilde{c}_{n,M}(\alpha) = \arg\min_{G_{n,M}(x) \geq 1 - \alpha} x \), and let \( c(\alpha) \) be the \( 1 - \alpha \) quantile of the distribution function \( G(x) = \Pr[T \leq x] \). Then under \( H_0 \), the conditions of Theorem 8.2 and for arbitrary \( \delta > 0 \),
\[ \lim_{n \to \infty} \Pr[|\tilde{G}_{n,M}(x) - G(x)| > \delta] \leq \frac{1}{4\delta M} \] 
hence \[ \lim_{M \to \infty} \limsup_{n \to \infty} \Pr[|\tilde{c}_{n,M}(\alpha) - c(\alpha)| > \delta] = 0. \] 
Moreover, under \( H_1 \), \( \Pr[\hat{T}_n > \tilde{c}_{n,M}(\alpha)] \to 1 \) for any \( M \in \mathbb{N} \).

**Proof.** Similar to Bierens (2015a, Theorem 6.3), using Theorem 7.3. ■
9. The initial value problem

The results in this paper were based on the assumption that all lagged $Z_t$’s are observed, which is obviously not realistic. In particular, given that $Z_t$ is observed for $1 - t_0 \leq t \leq n$, where $t_0 \geq \max(p, q)$, the empirical counterpart of $\psi_k(\tau_1, \tau_2, \ldots, \tau_k)$ defined in (3.14) is

$$\tilde{\psi}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{n} \sum_{t=1}^{n} (Y_t - f_{t-1}(\hat{\theta}_n)) \exp \left( i \sum_{j=1}^{\min(k, t-t_0)} \tau_j' \Phi(Z_{t-j}) \right)$$

whereas all the asymptotic results were based on

$$\psi_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{n} \sum_{t=1}^{n} (Y_t - f_{t-1}(\theta_n)) \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right)$$

As alluded to in section 3, we need to set forth conditions under which the asymptotic results above on the basis of $\tilde{W}_{k,n}$ are the same as for $W_{k,n}$. The condition involved is condition (9.1) in the following theorem.

**Theorem 9.1.** Denote

$$\tilde{B}_{n,k} = \int_{\Theta_k} \left| \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k),$$

$$\hat{B}_{n,k} = \int_{\Theta_k} \left| \hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k).$$

If the sequence of weights $\gamma_k$ is chosen such that

$$\sum_{k=1}^{\infty} k^2 \sqrt{\gamma_k} < \infty \quad (9.1)$$

then under $H_0$ and the conditions and notations in Theorem 5.1,

$$\sum_{k=1}^{L_n} \gamma_k \tilde{B}_{n,k} = \sum_{k=1}^{L_m} \gamma_k \hat{B}_{n,k} + O_p(1/\sqrt{n}),$$
whereas under $H_1$,
\[
\sum_{k=1}^{L_n} \gamma_k \tilde{B}_{n,k}/n = \sum_{k=1}^{L_n} \gamma_k \hat{B}_{n,k}/n + O_p(1/n),
\]

**Proof.** Observe that
\[
\left| \exp \left( i \sum_{j=1}^{k} \tau_j' \Phi(Z_{t-j}) \right) - \exp \left( i \sum_{j=1}^{\min(k,t-t_0)} \tau_j' \Phi(Z_{t-j}) \right) \right|
\leq \left| \sum_{j=\min(k,t-t_0)+1}^{k} \tau_j' \Phi(Z_{t-j}) \right|
\leq 0 \text{ for } t \geq k + t_0
\leq C(k + t_0 - 1) \text{ for } 1 \leq t < k + t_0
\]
where $C = \sup_{\tau \in \mathcal{T}} \|\tau\| \cdot \sup_{z \in \mathbb{R}^{+1}} \|\Phi(z)\|$. Hence,
\[
\left| \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) - \hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right| \leq C \frac{1}{\sqrt{n}} \sum_{t=1}^{k+t_0-1} |Y_t - f_{t-1}(\hat{\theta}_n)| (k + t_0 - 1)
\]
and thus
\[
\left| \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) - \hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2
\leq \left| \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) - \hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2
+ 2\sqrt{2} \left| \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) - \hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right| \times |\tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|
\leq C^2 \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{k+t_0-1} |Y_t - f_{t-1}(\hat{\theta}_n)| \right)^2 (k + t_0 - 1)^2
+ 2\sqrt{2} C \frac{1}{\sqrt{n}} \sum_{t=1}^{k+t_0-1} |Y_t - f_{t-1}(\hat{\theta}_n)| (k + t_0 - 1) \times |\tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|
\]
where the first inequality follow from the easy inequality $||z_1|^2 - |z_2|^2| \leq |z_1 - z_2|^2 + 2\sqrt{2}|z_1 - z_2| |z_2|$ for complex numbers $z_1$ and $z_2$. Consequently,
\[
\left| \hat{B}_{n,k} - \tilde{B}_{n,k} \right| \leq \frac{1}{n} C^2 (k + t_0 - 1)^4 \left( \frac{1}{k + t_0 - 1} \sum_{t=1}^{k+t_0-1} |Y_t - f_{t-1}(\hat{\theta}_n)| \right)^2
\]

56
which implies that

\[
\left| \sum_{k=1}^{L_n} \gamma_k \bar{B}_{n,k} - \sum_{k=1}^{L_n} \gamma_k \hat{B}_{n,k} \right| 
\leq \frac{1}{n} C^2 \sum_{k=1}^{L_n} \gamma_k (k + t_0 - 1)^4 \left( \frac{1}{k + t_0 - 1} \sum_{t=1}^{k+t_0-1} \sup_{\theta \in \Theta} (Y_t - f_{t-1}(\theta))^2 \right)
\]

\[
+ \frac{1}{\sqrt{n}} 2 \sqrt{2} C \sum_{k=1}^{L_n} \sqrt{\gamma_k} (k + t_0 - 1) \left( \frac{1}{k + t_0 - 1} \sum_{t=1}^{k+t_0-1} \sup_{\theta \in \Theta} |Y_t - f_{t-1}(\theta)| \right)
\times \sqrt{\sum_{k=1}^{L_n} \gamma_k \hat{B}_{n,k}}
\]

Next, note that (9.1) implies

\[
\sum_{k=1}^{\infty} k^4 \gamma_k < \infty
\]

(9.2)

Since by Assumption 3.1 and (9.1) and (9.2),

\[
E \left[ \sum_{k=1}^{L_n} \gamma_k (k + t_0 - 1)^4 \left( \frac{1}{k + t_0 - 1} \sum_{t=1}^{k+t_0-1} \sup_{\theta \in \Theta} (Y_t - f_{t-1}(\theta))^2 \right) \right]
\]

\[
= E \left[ \sup_{\theta \in \Theta} (Y_t - f_{t-1}(\theta))^2 \right] \sum_{k=1}^{\infty} \gamma_k (k + t_0 - 1)^4 < \infty
\]

and

\[
E \left[ \sum_{k=1}^{L_n} \sqrt{\gamma_k} (k + t_0 - 1)^2 \left( \frac{1}{k + t_0 - 1} \sum_{t=1}^{k+t_0-1} \sup_{\theta \in \Theta} |Y_t - f_{t-1}(\theta)| \right) \right]
\]

57
\[
E \left[ \sup_{\theta \in \Theta} |Y_t - f_{t-1}(\theta)| \right] \sum_{k=1}^{\infty} \sqrt{\gamma_k}(k + t_0 - 1)^2 < \infty,
\]

whereas by Theorem 5.1, \( \sum_{k=1}^{L_n} \gamma_k \tilde{B}_{n,k} \xrightarrow{d} \sum_{k=1}^{\infty} \gamma_k B_k \) under \( H_0 \), which implies that \( \sum_{k=1}^{L_n} \gamma_k \tilde{B}_{n,k} = O_p(1) \), it follows now that under \( H_0 \),

\[
\left| \sum_{k=1}^{L_n} \gamma_k \tilde{B}_{n,k} - \sum_{k=1}^{L_n} \gamma_k \hat{B}_{n,k} \right| = O_p(1/\sqrt{n}).
\]

Finally, the result under \( H_1 \) follows easily from

\[
\left| \sum_{k=1}^{L_n} \gamma_k \tilde{B}_{n,k}/n - \sum_{k=1}^{L_n} \gamma_k \hat{B}_{n,k}/n \right|
\leq \frac{1}{n^2} C^2 \sum_{k=1}^{L_n} \gamma_k(k + t_0 - 1)^4 \left( \frac{1}{k + t_0 - 1} \sum_{t=1}^{k+t_0-1} \sup_{\theta \in \Theta} (Y_t - f_{t-1}(\theta))^2 \right)
\]

\[
+ \frac{1}{n} 2\sqrt{2} C \sum_{k=1}^{L_n} \sqrt{\gamma_k}(k + t_0 - 1)^2 \left( \frac{1}{k + t_0 - 1} \sum_{t=1}^{k+t_0-1} \sup_{\theta \in \Theta} |Y_t - f_{t-1}(\theta)| \right)
\]

\[
\times \left( \sum_{k=1}^{L_n} \gamma_k \tilde{B}_{n,k}/n \right)
\]

and Theorem 6.1.


De Jong (1996) identifies the null hypothesis \( \Pr \left( E \left[ U_t | \mathcal{F}_{t-\infty}^t \right] = 0 \right) = 1 \) versus the alternative hypothesis \( \Pr \left( E \left[ U_t | \mathcal{F}_{t-\infty}^t \right] = 0 \right) < 1 \) for the errors \( U_t \) of a time series regression via the contents of a set \( S \subset \mathbb{R}^\infty \) of the type

\[
S = \left\{ \xi = (\xi_1, \xi_2, \xi_3, \ldots)' \in \Xi : E \left[ U_t \exp \left( \sum_{j=1}^{\infty} \xi_j' \Phi(Z_{t-j}) \right) \right] = 0 \right\},
\]

where \( \Xi \) is a compact metric space in \( \mathbb{R}^\infty \), and \( \Phi \) is a bounded one-to-one mapping.

In particular, de Jong specifies

\[
\Xi = \mathbb{X}_{j=1}^{\infty} \left[ -c.j^{-2}, c.j^{-2} \right]^k
\]
for some constant $c > 0$, where $k$ is the dimension of $\Phi(Z_{t-j})$, and shows that under an appropriate norm and metric the space $\Xi$ is compact. Under the null hypothesis, $S = \Xi$, whereas under the alternative, similar to Theorem 2.1, $S$ is ”almost empty”. Therefore, a consistent ICM test of the null hypothesis can be based on the ICM statistic

$$\int_{\Xi} \left( \tilde{W}_n(\xi) \right)^2 d\xi,$$

where

$$\tilde{W}_n(\xi) = \frac{1}{\sqrt{n}} \sum_{t=2}^{n} \hat{U}_t \exp \left( \sum_{j=1}^{t-1} \xi_j' \Phi(Z_{t-j}) \right)$$

with $\hat{U}_t$’s are the regression residuals and $n$ is the sample size.\(^9\) De Jong shows that under the null hypothesis, $\tilde{W}_n \Rightarrow W$ on $\Xi$, where $W$ is a zero-mean Gaussian process, so that

$$\int_{\Xi} \left( \tilde{W}_n(\xi) \right)^2 d\xi \xrightarrow{d} \int_{\Xi} (W(\xi))^2 d\xi.$$

Moreover, under the alternative hypothesis,

$$\int_{\Xi} \left( \frac{\tilde{W}_n(\xi)}{\sqrt{n}} \right)^2 d\xi \xrightarrow{p} \int_{\Xi} \left( \mathbb{E} \left[ U_t \exp \left( \sum_{j=1}^{\infty} \xi_j' \Phi(Z_{t-j}) \right) \right] \right)^2 d\xi > 0.$$

Similar to Bierens and Ploberger (1997) it can be shown that this test has non-trivial power against $\sqrt{n}$ local alternatives, the same upper bounds of the critical values as (2.12), and that bootstrap critical values can be derived similar to Bierens (2015a). Note however that the bootstrap procedure in de Jong (1996) is not correct.

Recall that the integration range $\Xi$ is, and needs to be, compact because otherwise the weak convergence result $\tilde{W}_n \Rightarrow W$ on $\Xi$ may not hold. However, infinite dimensional compact metric spaces are extremely small, which may affect the small sample power of the test.

\(^9\)Actually, de Jong (1996) uses $\int_{\Xi} \left( \tilde{W}_n(\xi) \right)^2 f(\xi_1) d\xi$ as his ICM test statistic, where $f(.)$ is a density function on $[-c,c]^k$. The role of this density is to reduce the computational burden a little bit.
11. Standardization of the conditioning variables

Another unresolved issue in B84 is how to standardize the conditioning lagged variables in \( Z_t \) before taking the bounded transformation \( \Phi \) in order to preserve enough variation in \( \Phi(Z_t) \). See section 6 in B84.

To discuss this issue, let us for the time being assume that \( Z_t = Y_t \), so that the nonlinear ARX model (3.1) becomes a nonlinear AR model:

\[
Y_t = f(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}, \theta_0) + U_t = f_{t-1}(\theta_0) + U_t, \text{ say.}
\]

Moreover, as in B82, let us assume that

\[
\Phi(x) = \arctan(x), \quad (11.1)
\]

and note that

\[
\Phi'(x) = (1 + x^2)^{-1}. \quad (11.2)
\]

11.1. A wrong way to standardize

Following B82, one may be tempted to standardize each lagged \( Y_t \) as

\[
\tilde{Y}_{n,t} = \hat{\sigma}_n^{-1}(Y_t - \hat{\mu}_n) = \hat{\alpha}_n Y_t - \hat{\beta}_n,
\]

where

\[
\hat{\mu}_n = \frac{1}{(n + t_0)} \sum_{t=1-t_0}^n Y_t, \quad \hat{\sigma}_n = \sqrt{(1/(n + t_0 - 1)) \sum_{t=1-t_0}^n (Y_t - \hat{\mu}_n)^2},
\]

\[
\hat{\alpha}_n = 1/\hat{\sigma}_n, \quad \hat{\beta}_n = \hat{\mu}_n/\hat{\sigma}_n,
\]

and \( 1 - t_0 \) is the time index of the first observed \( Y_t \). Then (3.17) becomes

\[
\tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^n (Y_t - f_{t-1}(\hat{\theta}_n)) \exp \left( i \sum_{j=1}^k \tau_j \Phi(\hat{\alpha}_n Y_{t-j} - \hat{\beta}_n) \right). \quad (11.3)
\]

Similarly, denoting

\[
\mu_0 = E[Y_t], \quad \sigma_0 = \sqrt{E[(Y_t - \mu)^2]}, \quad \alpha_0 = 1/\sigma_0, \quad \beta_0 = \mu_0/\sigma_0, \quad (11.4)
\]

replace \( \Phi(Y_{t-j}) \) in (3.17) by \( \Phi(\alpha_0 Y_{t-j} - \beta_0) \), i.e.,

\[
\tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^n (Y_t - f_{t-1}(\hat{\theta}_n)) \exp \left( i \sum_{j=1}^k \tau_j \Phi(\alpha_0 Y_{t-j} - \beta_0) \right). \quad (11.5)
\]
Now the hope is that under $H_0$, $\tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)$ and $\tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)$ are asymptotically equivalent, in the sense the WICM test based $\tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)$ has the same limiting null distribution as the (now infeasible) WICM test based on $\tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)$.

It can be shown that it does if $b_{\alpha_n} = \alpha_0 + O_p(n^{-1/2})$ and $b_{\beta_n} = \beta_0 + O_p(n^{-1/2})$, but it is unlikely that these conditions hold in the time series case. Moreover, even if it does (which is doubtful), this particular standardization is not recommended because it destroys the martingale difference structure of the first term in (11.3) under the null hypothesis and therefore may lead to rejection of the true null hypothesis in the case of small and medium size macro-economics data.

11.2. Martingale difference structure preserving standardization

Therefore, a better standardization procedure is the following. Denote

\[
\begin{align*}
\hat{\mu}_t &= (t + t_0)^{-1} \sum_{i=1-t_0}^{t} Y_i, \quad \hat{\sigma}_t = \sqrt{(t + t_0)^{-1} \sum_{i=1-t_0}^{t} (Y_i - \hat{\mu}_t)^2}, \quad \text{if } t > 1 - t_0, \\
\hat{\mu}_t &= 0, \quad \hat{\sigma}_t = 1 \quad \text{if } t \leq 1 - t_0, \\
\hat{\alpha}_t &= 1/\hat{\sigma}_t, \quad \hat{\beta}_t = \hat{\mu}_t/\hat{\sigma}_t, \\
\tilde{\sum}_t &= \hat{\alpha}_t Y_t - \hat{\beta}_t, \quad \tilde{\sum}_t = \alpha_0 Y_t - \beta_0,
\end{align*}
\]

(11.6)

where $\alpha_0$ and $\beta_0$ are the same as in (11.4). Then

**Lemma 11.1** Under Assumptions 3.1(a) and 4.1(a) and with $\Phi$ as in (11.1), it follows that

\[
p \lim_{t \to \infty} \left| \Phi \left( \tilde{\sum}_t \right) - \Phi \left( Y_t \right) \right| = 0.
\]

(11.7)

Consequently, for an arbitrary strictly stationary time series process $V_t \in \mathbb{R}$ satisfying $E[|V_t|] < \infty$ and each $j \in \mathbb{N}$,

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E \left[ |V_t| \cdot \Phi \left( \tilde{\sum}_{t-j} \right) - \Phi \left( Y_{t-j} \right) \right] = 0,
\]

(11.8)

which by Chebyshev’s inequality for first moments implies that

\[
p \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} |V_t| \cdot \left| \Phi \left( \tilde{\sum}_{t-j} \right) - \Phi \left( Y_{t-j} \right) \right| = 0.
\]
Proof. Recall from Assumptions 3.1(a) and 4.1(a) that $Y_t$ and $Y_t^2$ are strictly stationary with vanishing memory, and $E[Y_t^2] < \infty$, so that by Lemma 4.2,
\[ p \lim_{t \to \infty} \mu_t = E[Y_1] = \mu_0, \quad p \lim_{t \to \infty} \sigma_t^2 = \text{var}(Y_1) = \sigma_0^2, \]
\[ \text{hence} \quad p \lim_{t \to \infty} \alpha_t = \alpha_0, \quad p \lim_{t \to \infty} \beta_t = \beta_0. \quad (11.9) \]
Moreover, with $\Phi$ as in (11.1), it follows from the mean value theorem and (11.2) that
\[ \left| \Phi \left( \tilde{Y}_t \right) - \Phi \left( Y_t \right) \right| \leq \left| \tilde{Y}_t - Y_t \right| \leq |\alpha_t - \alpha_0| |Y_t| + |\beta_t - \beta_0| \]
so that (11.7) holds by (11.9) and the fact that by strict stationarity, $|Y_t| = O_p(1)$.

As to (11.8), observe that for an arbitrary $M > 0$, and with $I(.)$ the indicator function,
\[ \frac{1}{n} \sum_{t=1}^{n} E \left[ |V_t|, \Phi \left( \tilde{Y}_t \right) - \Phi \left( Y_t \right) \right] \]
\[ \leq \frac{1}{n} \sum_{t=1}^{n} E \left[ |V_t| I \left( |V_t| \leq M \right) \left| \Phi \left( \tilde{Y}_t \right) - \Phi \left( Y_t \right) \right| \right] \]
\[ + \frac{1}{n} \sum_{t=1}^{n} E \left[ |V_t| I \left( |V_t| > M \right) \left| \Phi \left( \tilde{Y}_t \right) - \Phi \left( Y_t \right) \right| \right] \]
\[ \leq M \frac{1}{n} \sum_{t=1}^{n} E \left[ \Phi \left( \tilde{Y}_t \right) - \Phi \left( Y_t \right) \right] + \pi E \left[ |V_1| I \left( |V_1| > M \right) \right], \]
where the $\pi$ follows from the fact that by (11.1), $\sup_{x \in \mathbb{R}} |\Phi(x)| = \pi/2$. For an arbitrary $\varepsilon > 0$ we can choose $M$ so large that $\pi E \left[ |V_1| I \left( |V_1| > M \right) \right] < \varepsilon$. Moreover, it follows from the bounded convergence theorem and (11.7) that
\[ \lim_{t \to \infty} E \left[ \Phi \left( \tilde{Y}_t \right) - \Phi \left( Y_t \right) \right] = 0, \]
which trivially implies that $\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E \left[ \Phi \left( \tilde{Y}_t \right) - \Phi \left( Y_t \right) \right]$. Hence,
\[ \limsup_{n \to \infty} E \left[ \frac{1}{n} \sum_{t=1}^{n} |V_t|, \Phi \left( \tilde{Y}_t \right) - \Phi \left( Y_t \right) \right] < \varepsilon, \]
which, by the arbitrariness of $\varepsilon > 0$ implies that (11.8) holds. \[\blacksquare\]
Now (11.5) reads

\[
\widehat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t - f_{t-1}(\hat{\theta}_n)) \exp \left( i \sum_{j=1}^{k} \tau_j \Phi (Y_{t-j}) \right)
\]

\[
= \widehat{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) - \widehat{W}^{(2)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \tag{11.10}
\]

where

\[
\widehat{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \exp \left( i \sum_{j=1}^{k} \tau_j \Phi (Y_{t-j}) \right), \tag{11.11}
\]

\[
\widehat{W}^{(2)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0) \right) \exp \left( i \sum_{j=1}^{k} \tau_j \Phi (Y_{t-j}) \right). \tag{11.12}
\]

Similarly, (11.3) now becomes

\[
\widehat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t - f_{t-1}(\hat{\theta}_n)) \exp \left( i \sum_{j=1}^{k} \tau_j \Phi (Y_{t-j}) \right)
\]

\[
= \widehat{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) - \widehat{W}^{(2)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \tag{11.13}
\]

where

\[
\widehat{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \exp \left( i \sum_{j=1}^{k} \tau_j \Phi (Y_{t-j}) \right), \tag{11.14}
\]

\[
\widehat{W}^{(2)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0) \right) \exp \left( i \sum_{j=1}^{k} \tau_j \Phi (Y_{t-j}) \right). \tag{11.15}
\]

11.2.1. The case \(H_0\)

Using Lemma 11.1, I will show now that
Lemma 11.2. Under the conditions of Theorem 4.2, and with $\Phi$ the arctan function,

$$\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left| \widetilde{W}_{k,n}^{(2)}(\tau_1, \tau_2, \ldots, \tau_k) - \widetilde{W}_{k,n}^{(2)}(\tau_1, \tau_2, \ldots, \tau_k) \right| = o_p(1).$$

Proof. It follows from the mean value theorem that $f_{t-1}(\tilde{\theta}_n) - f_{t-1}(\theta_0) = (\tilde{\theta}_n - \theta_0)' \nabla f_{t-1}(\tilde{\theta}_t)$, where $\tilde{\theta}_t$ is a mean value satisfying $||\tilde{\theta}_t - \theta_0|| \leq ||\tilde{\theta}_n - \theta_0||$, hence

$$\left| \text{Re} \left[ \widetilde{W}_{k,n}^{(2)}(\tau_1, \tau_2, \ldots, \tau_k) \right] - \text{Re} \left[ \widetilde{W}_{k,n}^{(2)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right|$$

$$\leq \sqrt{n} \left( \tilde{\theta}_n - \theta_0 \right)' \frac{1}{n} \sum_{t=1}^{n} \left( \nabla f_{t-1}(\tilde{\theta}_t) - \nabla f_{t-1}(\theta_0) \right)$$

$$\times \left( \cos \left( \sum_{j=1}^{k} \tau_j \Phi \left( \tilde{Y}_{t-j} \right) \right) - \cos \left( \sum_{j=1}^{k} \tau_j \Phi \left( Y_{t-j} \right) \right) \right)$$

$$\leq 2 \left\| \sqrt{n} \left( \tilde{\theta}_n - \theta_0 \right) \right\| \frac{1}{n} \sum_{t=1}^{n} \sup_{||\theta - \theta_0|| \leq ||\tilde{\theta}_n - \theta_0||} \left\| \nabla f_{t-1}(\theta) - \nabla f_{t-1}(\theta_0) \right\|$$

$$+ \sup_{\tau \in \Upsilon} |\tau| \times \left\| \sqrt{n} \left( \tilde{\theta}_n - \theta_0 \right) \right\|$$

$$\times \sum_{j=1}^{k} \frac{1}{n} \sum_{t=1}^{n} \left\| \nabla f_{t-1}(\theta_0) \right\| \left| \Phi \left( \tilde{Y}_{t-j} \right) - \Phi \left( Y_{t-j} \right) \right|$$

Theorem 4.2 implies that $\left\| \sqrt{n} \left( \tilde{\theta}_n - \theta_0 \right) \right\| = O_p(1)$, and it follows from Lemma 64.
11.1 that for fixed $k$, \( \sum_{j=1}^{k} \frac{1}{n} \sum_{t=1}^{n} \| \nabla f_{t-1}(\theta_0) \| \cdot \left| \Phi(\tilde{\Sigma}_{t-j}) - \Phi(\Sigma_{t-j}) \right| = o_p(1) \).

Moreover, part (4.4) of Assumption 4.2 implies that for an arbitrary small \( \varepsilon > 0 \), \( \kappa(\varepsilon) = E \left[ \sup_{\| \theta - \theta_0 \| \leq \varepsilon} \| \nabla f_{t-1}(\theta) - \nabla f_{t-1}(\theta_0) \| \right] < \infty \), so that by Lemma 3.2, 

\[
 p \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \sup_{\| \theta - \theta_0 \| \leq \| \hat{\theta}_n - \theta_0 \|} \| \nabla f_{t-1}(\theta) - \nabla f_{t-1}(\theta_0) \| = p \lim_{n \to \infty} \kappa \left( \| \hat{\theta}_n - \theta_0 \| \right) = \kappa(0) = 0. \tag{11.16}
\]

Thus, \( \sup_{(\tau_1, \tau_2, ..., \tau_k) \in \Upsilon^k} \left| \text{Re} \left[ \hat{W}_{k,n}(\tau_1, \tau_2, ..., \tau_k) \right] - \text{Re} \left[ \hat{W}_{k,n}(\tau_1, \tau_2, ..., \tau_k) \right] \right| = o_p(1) \), and the same applies to the \( \text{Im}[\cdot] \) case.

Next, denote

\[
\hat{B}_{n,k} = \int_{\Upsilon^k} \left| \hat{W}_{k,n}(\tau_1, \tau_2, ..., \tau_k) \right|^2 \, d\mu(\tau_1)d\mu(\tau_2)...d\mu(\tau_k),
\]

\[
\tilde{B}_{n,k} = \int_{\Upsilon^k} \left| \hat{W}_{k,n}(\tau_1, \tau_2, ..., \tau_k) \right|^2 \, d\mu(\tau_1)d\mu(\tau_2)...d\mu(\tau_k).
\]

Then

**Lemma 11.3.** Under \( H_0 \) and Assumptions 3.1, 4.1 and 4.2, \( \hat{B}_{n,k} = \tilde{B}_{n,k} + o_p(1) \) for each \( k \in \mathbb{N} \).

**Proof.** Lemma 11.2 implies that

\[
\hat{W}_{k,n}(\tau_1, \tau_2, ..., \tau_k) - \hat{W}_{k,n}(\tau_1, \tau_2, ..., \tau_k) = \Delta W_{k,n}(\tau_1, \tau_2, ..., \tau_k) + o_p(1),
\]

uniformly in \( (\tau_1, \tau_2, ..., \tau_k) \in \Upsilon^k \), where

\[
\Delta W_{k,n}(\tau_1, \tau_2, ..., \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \left( \exp \left( \sum_{j=1}^{k} \tau_j \Phi(\tilde{\Sigma}_{t-j}) \right) - \exp \left( \sum_{j=1}^{k} \tau_j \Phi(\Sigma_{t-j}) \right) \right).
\]

Hence

\[
\tilde{B}_{n,k} = \int_{\Upsilon^k} \left| \hat{W}_{k,n}(\tau_1, \tau_2, ..., \tau_k) + \Delta W_{k,n}(\tau_1, \tau_2, ..., \tau_k) + o_p(1) \right|^2
\]

65
and thus,

\[
\left| \tilde{B}_{n,k} - \hat{B}_{n,k} \right| \leq \int_{\mathcal{T}_k} (\text{Re} [\Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)])^2 \, d\mu(\tau_1) d\mu(\tau_2) \cdots d\mu(\tau_k) \\
+ \int_{\mathcal{T}_k} (\text{Im} [\Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)])^2 \, d\mu(\tau_1) d\mu(\tau_2) \cdots d\mu(\tau_k) \\
+ 2 \sqrt{\int_{\mathcal{T}_k} (\text{Re} [\hat{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)])^2 \, d\mu(\tau_1) d\mu(\tau_2) \cdots d\mu(\tau_k)}
\]
that $Z$ as follows. Note that for $E \leq \sum_{i=1}^{k} \mathbb{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)$

\[ \times \int_{\Upsilon^k} \left( \text{Re} \left[ \Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \]

\[ + 2 \sqrt{\int_{\Upsilon^k} \left( \text{Im} \left[ \Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \]

\[ \times \int_{\Upsilon^k} \left( \text{Im} \left[ \Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \]

\[ + 2 \sqrt{\int_{\Upsilon^k} \left( \text{Re} \left[ \Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \times o_p(1) \]

\[ + 2 \sqrt{\int_{\Upsilon^k} \left( \text{Im} \left[ \Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \times o_p(1) \]

\[ + o_p(1) \]

\[ \leq \int_{\Upsilon^k} |\Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \]

\[ + 4 \sqrt{\mathcal{B}_{n,k} \times \int_{\Upsilon^k} |\Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \]

\[ + 4 \sqrt{\mathcal{B}_{n,k} \times o_p(1) \]

\[ + 4 \sqrt{\int_{\Upsilon^k} |\Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \times o_p(1) \]

\[ + o_p(1) \]

Since by Lemma 5.5, $\mathcal{B}_{n,k} \overset{d}{=} B_k$, which implies $B_{n,k} = O_p(1)$, it suffices to prove that

\[ \int_{\Upsilon^k} |\Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) = o_p(1), \quad (11.17) \]

as follows. Note that for $(\tau_1, \tau_2, \ldots, \tau_k)' \in \Upsilon^k$,

\[ E \left[ |\Delta W_{k,n}(\tau_1, \tau_2, \ldots, \tau_k)|^2 \right] \]
\[
= \frac{1}{n} \sum_{t=1}^{n} E \left[ U_t^2 \left( \cos \left( \sum_{j=1}^{k} \tau_j \Phi \left( \tilde{Y}_{t-j} \right) \right) - \cos \left( \sum_{j=1}^{k} \tau_j \Phi \left( Y_{t-j} \right) \right) \right)^2 \right] \\
+ \int_{\Upsilon} \frac{1}{n} \sum_{t=1}^{n} E \left[ U_t^2 \left( \sin \left( \sum_{j=1}^{k} \tau_j \Phi \left( \tilde{Y}_{t-j} \right) \right) - \sin \left( \sum_{j=1}^{k} \tau_j \Phi \left( Y_{t-j} \right) \right) \right)^2 \right] \\
= 2E[U_t^2] - \frac{1}{n} \sum_{t=1}^{n} E \left[ U_t^2 \cos \left( \sum_{j=1}^{k} \tau_j \Phi \left( \tilde{Y}_{t-j} \right) \right) \right] \times \cos \left( \sum_{j=1}^{k} \tau_j \Phi \left( Y_{t-j} \right) \right) \\
- \frac{2}{n} \sum_{t=1}^{n} E \left[ U_t^2 \sin \left( \sum_{j=1}^{k} \tau_j \Phi \left( \tilde{Y}_{t-j} \right) \right) \right] \times \sin \left( \sum_{j=1}^{k} \tau_j \Phi \left( Y_{t-j} \right) \right) \\
= 2E[U_t^2] - 2 \frac{1}{n} \sum_{t=1}^{n} E \left[ U_t^2 \left( 1 - \cos \left( \sum_{j=1}^{k} \tau_j \left( \Phi \left( \tilde{Y}_{t-j} \right) - \Phi \left( Y_{t-j} \right) \right) \right) \right) \right] \\
\leq 2 \sum_{j=1}^{k} |\tau_j| \frac{1}{n} \sum_{t=1}^{n} E \left[ U_t^2 \Phi \left( \tilde{Y}_{t-j} \right) - \Phi \left( Y_{t-j} \right) \right] \\
\leq 2 \sup_{\tau \in \Upsilon} |\tau| \sum_{j=1}^{k} \frac{1}{n} \sum_{t=1}^{n} E \left[ U_t^2 \Phi \left( \tilde{Y}_{t-j} \right) - \Phi \left( Y_{t-j} \right) \right] \\
= o(1)
\]

where the first inequality follows from the mean value theorem and the last equality is due to Lemma 11.1. Thus,

\[
\int_{\Upsilon} E \left[ |\Delta W_{k,n}(\tau_1, \tau_2, ..., \tau_k)|^2 \right] d\mu(\tau_1)d\mu(\tau_2)...d\mu(\tau_k) = o(1)
\]

which implies (11.17). ■

It follows now from Theorem 5.1, Lemma 5.1 and Lemma 11.3 that the following general result holds.
Theorem 11.1. Denote
\[ \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t - f_{t-1}(\hat{\theta}_n)) \exp \left( i \sum_{j=1}^{k} \tau_j \Phi(\tilde{Z}_{t-j}) \right), \]
\[ \tilde{B}_{n,k} = \int_{\mathcal{T}_k} \left| \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k), \]
\[ \tilde{T}_n = \sum_{k=1}^{L_n} \gamma_k \tilde{B}_{n,k}, \]
where each component \( \tilde{Z}_{i,t} \) of \( \tilde{Z}_t \) is a standardized version of component \( Z_{i,t} \) of \( Z_t \), similar to \( \tilde{Y}_t \) in (11.6) with \( Y_t \) replaced by \( Z_{i,t} \). Similarly, denote
\[ \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t - f_{t-1}(\hat{\theta}_n)) \exp \left( i \sum_{j=1}^{k} \tau_j \Phi(Z_{t-j}) \right), \]
\[ \tilde{B}_{n,k} = \int_{\mathcal{T}_k} \left| \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k), \]
\[ \tilde{T}_n = \sum_{k=1}^{L_n} \gamma_k \tilde{B}_{n,k}. \]
where each component \( Z_{i,t} \) of \( Z_t \) is a standardized version of component \( Z_{i,t} \) of \( Z_t \), similar to \( \tilde{Y}_t \) in (11.6) with \( Y_t \) replaced by \( Z_{i,t} \). Moreover, let all the components of the one-to one mapping \( \Phi \) be \( \arctan(\cdot) \) functions. Then under the conditions of Theorem 5.1,
\[ \tilde{T}_n = \tilde{T}_n + o_p(1) \xrightarrow{d} \sum_{k=1}^{\infty} \gamma_k B_k = T, \]
where for each \( k \) the random variable \( B_k \) represents the limiting distribution of \( \tilde{B}_{n,k} \). Therefore, all the previous result for the WICM test statistic \( \tilde{T}_n \) under \( H_0 \) carry over to \( \tilde{T}_n \).

Proof. Recall from Theorem 5.1 that
\[ \tilde{T}_n \overset{d}{=} \sum_{k=1}^{L_n} \gamma_k \tilde{B}_{n,k} \overset{d}{=} \sum_{k=1}^{\infty} \gamma_k B_k = T \quad (11.18) \]
and note that Lemma 5.1 carries over to \( \tilde{B}_{n,k} \), i.e., \( \tilde{B}_{n,k} \leq 2 \tilde{B}_{1,n,k} + 2 \tilde{B}_{2,n,k} \), where \( E[\tilde{B}_{1,n,k}] = E[U_1^2] \) and \( \sup_{k \in \mathbb{N}} \tilde{B}_{2,n,k} = o_p(1) \). Of course, Lemma 5.1 is directly applicable to the current \( \hat{B}_{n,k} \)'s as well. Then with \( \hat{B}_{1,n,k} \) and \( \hat{B}_{2,n,k} \) as in Lemma 5.1, it follows that for each \( K \in \mathbb{N} \),

\[
\sum_{k=K+1}^{\infty} \gamma_k \left| \tilde{B}_{n,k} - \hat{B}_{n,k} \right| \leq 2 \sum_{k=K+1}^{\infty} \gamma_k \left( \tilde{B}_{1,n,k} + \hat{B}_{1,n,k} \right) + 2 \sum_{k=K+1}^{\infty} \gamma_k \sup_{k \in \mathbb{N}} \left( \tilde{B}_{2,n,k} + \hat{B}_{2,n,k} \right) = \sum_{k=K+1}^{\infty} \gamma_k \times O_p(1) \tag{11.19}
\]

whereas by Lemma 11.3,

\[
\sum_{k=1}^{K} \gamma_k \left| \tilde{B}_{n,k} - \hat{B}_{n,k} \right| = o_p(1). \tag{11.20}
\]

The result (11.19) implies that for arbitrary \( \varepsilon \in (0, 1) \) and \( \delta > 0 \) we can choose \( K \) so large that

\[
\limsup_{n \to \infty} \Pr \left[ \sum_{k=K+1}^{\infty} \gamma_k \left| \tilde{B}_{n,k} - \hat{B}_{n,k} \right| > \delta/2 \right] < \varepsilon/2,
\]

whereas (11.20) implies that for this \( K \),

\[
\limsup_{n \to \infty} \Pr \left[ \sum_{k=1}^{K} \gamma_k \left| \tilde{B}_{n,k} - \hat{B}_{n,k} \right| > \delta/2 \right] < \varepsilon/2.
\]

It follows now from the inequality (5.20) that \( \limsup_{n \to \infty} \Pr \left[ \sum_{k=1}^{\infty} \gamma_k \left| \tilde{B}_{n,k} - \hat{B}_{n,k} \right| > \delta \right] < \varepsilon \), hence

\[
\sum_{k=1}^{L_n} \gamma_k \left| \tilde{B}_{n,k} - \hat{B}_{n,k} \right| \leq \sum_{k=1}^{\infty} \gamma_k \left| \tilde{B}_{n,k} - \hat{B}_{n,k} \right| = o_p(1). \tag{11.21}
\]

The theorem under review now follows from (11.18) and (11.21). □
11.2.2. The case $H_1$

Also the consistency result in Theorem 6.1 carries over to the WICM test statistic $\tilde{T}_n$.

**Theorem 11.2.** Under $H_1$ and Assumptions 3.1 and 4.1, and under the notations in Theorems 6.1 and 11.1,

$$\frac{\tilde{T}_n}{n} = \sum_{k=1}^{L_n} \gamma_k \tilde{B}_{k,n}/n = \tilde{T}_n/n + o_p(1) \xrightarrow{p} \sum_{k=1}^{\infty} \gamma_k \eta_k > 0.$$  

**Proof.** Denote

$$\tilde{B}_{(1),n,k} = \int_{\mathcal{Y}_k} \left| \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k),$$

$$\tilde{B}_{(2),n,k} = \int_{\mathcal{Y}_k} \left| \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k),$$

and let $\tilde{B}_{n,k}$ and $\tilde{B}_{n,k}$ be the same as in Theorem 11.1. Moreover, denote

$$\psi_k(\tau_1, \tau_2, \ldots, \tau_k) = E \left[ U_t \exp \left( i \sum_{j=1}^{k} \tau_j \Phi(Y_{t-j}) \right) \right]$$

and

$$\eta_k = \int_{\mathcal{Y}_k} |\psi_k(\tau_1, \tau_2, \ldots, \tau_k)|^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k).$$

I will prove the theorem under review by showing that

$$\sup_{k \in \mathbb{N}} \tilde{B}_{(1),n,k} = O_p(1), \quad \sup_{k \in \mathbb{N}} \tilde{B}_{(2),n,k} = O_p(1),$$

$$\sup_{k \in \mathbb{N}} \tilde{B}_{n,k} = o_p(1), \quad \sup_{k \in \mathbb{N}} \tilde{B}_{n,k} = o_p(1),$$

$$\sup_{k \in \mathbb{N}} \left| \tilde{B}_{n,k}/n - \tilde{B}_{(1),n,k}/n \right| = o_p(1), \quad \sup_{k \in \mathbb{N}} \left| \tilde{B}_{n,k}/n - \tilde{B}_{(1),n,k}/n \right| = o_p(1),$$

$$\tilde{B}_{n,k}/n \xrightarrow{p} \eta_k, \quad \text{for each } k \in \mathbb{N},$$

$$\tilde{B}_{n,k}/n \xrightarrow{p} \eta_k, \quad \text{for each } k \in \mathbb{N}.$$  

71
Then it not hard to verify that Theorem 11.2 holds.

**Proof of (11.22).** It is easy to verify from (11.11) and (11.14) that

\[
\sup_{k \in \mathbb{N}} \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left| n^{-1/2} \tilde{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right| \leq C_{1,n},
\]

\[
\sup_{k \in \mathbb{N}} \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left| n^{-1/2} \tilde{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right| \leq C_{1,n},
\]

where

\[ C_{1,n} = \sqrt{2} \frac{1}{n} \sum_{t=1}^{n} |U_t| = O_p(1). \]

Hence,

\[
\sup_{k \in \mathbb{N}} \hat{B}^{(1)}_{n,k} \leq C_{1,n} = O_p(1), \quad \sup_{k \in \mathbb{N}} \hat{B}^{(1)}_{n,k} \leq C_{1,n} = O_p(1).
\] (11.27)

**Proof of (11.23).** Again, it is easy to verify from (11.12) and (11.15) that

\[
\sup_{k \in \mathbb{N}} \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left| n^{-1/2} \tilde{W}^{(2)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right| \leq C_{2,n},
\]

\[
\sup_{k \in \mathbb{N}} \sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \Upsilon^k} \left| n^{-1/2} \tilde{W}^{(2)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right| \leq C_{2,n},
\]

where

\[ C_{2,n} = \sqrt{2} \frac{1}{n} \sum_{t=1}^{n} \left| f_{t-1}(\hat{\theta}_n) - f_{t-1}(\theta_0) \right| \]

\[ \leq \sqrt{2} \frac{1}{n} \sum_{t=1}^{n} \sup_{||\theta - \theta_0|| \leq ||\hat{\theta}_n - \theta_0||} \left| f_{t-1}(\theta) - f_{t-1}(\theta_0) \right| \]

\[ = o_p(1). \]

The latter follows from $||\hat{\theta}_n - \theta_0|| \xrightarrow{p} 0$ and (11.16). Hence

\[
\sup_{k \in \mathbb{N}} \hat{B}^{(2)}_{n,k} \leq C_{2,n} = o_p(1), \quad \sup_{k \in \mathbb{N}} \hat{B}^{(2)}_{n,k} \leq C_{2,n} = o_p(1).
\] (11.28)

**Proof of (11.24).** Observe from (11.3) that

\[
\tilde{B}_{n,k} / n = \int_{\Upsilon^k} \left| n^{-1/2} \tilde{W}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 d\mu(\tau_1)d\mu(\tau_2)d\mu(\tau_k)
\]

\[ = \int_{\Upsilon^k} \left( \Re \left[ n^{-1/2} \tilde{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] - \Re \left[ n^{-1/2} \tilde{W}^{(2)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2
\]

72
\[ \times \mu(\tau_1) \mu(\tau_2) \ldots \mu(\tau_k) \]
\[ + \int_{\mathcal{Y}^h} \left( \Im \left[ n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] - \Im \left[ n^{-1/2} \tilde{W}_{k,n}^{(2)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 \times \mu(\tau_1) \mu(\tau_2) \ldots \mu(\tau_k) \]
\[ = \tilde{B}_{n,k}^{(1)}/n + \int_{\mathcal{Y}^h} \left| n^{-1/2} \tilde{W}_{k,n}^{(2)}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 \mu(\tau_1) \mu(\tau_2) \ldots \mu(\tau_k) \]
\[ - 2 \int_{\mathcal{Y}^h} \Re \left[ n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \times \Re \left[ n^{-1/2} \tilde{W}_{k,n}^{(2)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \times \mu(\tau_1) \mu(\tau_2) \ldots \mu(\tau_k) \]
\[ - 2 \int_{\mathcal{Y}^h} \Im \left[ n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \times \Im \left[ n^{-1/2} \tilde{W}_{k,n}^{(2)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \times \mu(\tau_1) \mu(\tau_2) \ldots \mu(\tau_k), \]

hence
\[
\left| \tilde{B}_{n,k}/n - \tilde{B}_{n,k}^{(1)}/n \right| \leq \tilde{B}_{n,k}^{(2)}/n
\]
\[ + 2 \int_{\mathcal{Y}^h} \left| \Re \left[ n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right| \times \left| \Re \left[ n^{-1/2} \tilde{W}_{k,n}^{(2)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right| \times \mu(\tau_1) \mu(\tau_2) \ldots \mu(\tau_k) \]
\[ + 2 \int_{\mathcal{Y}^h} \left| \Im \left[ n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right| \times \left| \Im \left[ n^{-1/2} \tilde{W}_{k,n}^{(2)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right| \times \mu(\tau_1) \mu(\tau_2) \ldots \mu(\tau_k) \]
\[ \leq \tilde{B}_{n,k}^{(2)}/n + 4 \sqrt{\tilde{B}_{n,k}^{(1)}/n} \sqrt{\tilde{B}_{n,k}^{(2)}/n} \]
\[ \leq C_{2,n}^2 + 4C_{1,n}C_{2,n} = o_p(1), \]

where the latter follows from (5.19) and (11.28). Obviously, the last inequality also holds for \( \tilde{B}_{n,k}/n - \tilde{B}_{n,k}^{(1)}/n \). This proves (11.24).

**Proof of (11.25).** Denote
\[
\Delta W_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) = \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) - \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k).
\]

and
\[
\Delta W_{k,n}^{\Delta} = \int_{\mathcal{Y}^h} \left| \Delta W_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 \mu(\tau_1) \mu(\tau_2) \ldots \mu(\tau_k).
\]

It is not hard to verify from (11.11) and (11.14) and the mean value theorem that
\[
\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in \mathcal{Y}^h} \left| n^{-1/2} \Delta W_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right| \leq D_{k,n},
\]

73
where

\[ D_{k,n} = \sqrt{2} \sup_{\tau \in \mathcal{T}} \left| \sum_{j=1}^{k} \frac{1}{n} \sum_{t=1}^{n} |U_t| \right| \Phi (\tilde{Y}_{t-j}) - \Phi (Y_{t-j}) \right| = o_p(1), \]

where the latter follows from Lemma 11.1. Hence

\[ B_{k,n}^2 / n \leq D_{k,n}^2 = o_p(1). \quad (11.29) \]

Then

\[
\begin{align*}
\tilde{B}_{n,k}^{(1)} / n \\
= & \int_{\mathcal{T}^k} \left| n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) + n^{-1/2} \Delta W_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right|^2 \times d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \\
= & \left( \text{Re} \left[ n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] + \text{Re} \left[ n^{-1/2} \Delta W_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 \times d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \\
+ & \left( \text{Im} \left[ n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] + \text{Im} \left[ n^{-1/2} \Delta W_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 \times d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \\
= & \int_{\mathcal{T}^k} \left( \text{Re} \left[ n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \\
+ & \int_{\mathcal{T}^k} \left( \text{Im} \left[ n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \\
+ & \int_{\mathcal{T}^k} \left( \text{Re} \left[ n^{-1/2} \Delta W_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \\
+ & \int_{\mathcal{T}^k} \left( \text{Im} \left[ n^{-1/2} \Delta W_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \\
+ & 2 \int_{\mathcal{T}^k} \left( \text{Re} \left[ n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right) \left( \text{Re} \left[ n^{-1/2} \Delta W_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right) \times d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \\
+ & 2 \int_{\mathcal{T}^k} \left( \text{Im} \left[ n^{-1/2} \tilde{W}_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right) \left( \text{Im} \left[ n^{-1/2} \Delta W_{k,n}^{(1)}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right) \times d\mu(\tau_1) d\mu(\tau_2) \ldots d\mu(\tau_k) \\
= & \tilde{B}_{n,k}^{(1)} / n + B_{k,n}^\Delta / n.
\end{align*}
\]
+2 \int_{Y^k} \left( \text{Re} \left[ n^{-1/2} \hat{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right) \left( \text{Re} \left[ n^{-1/2} \Delta W^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right) \\
\times d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k) \\
+2 \int_{Y^k} \left( \text{Im} \left[ n^{-1/2} \hat{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right) \left( \text{Im} \left[ n^{-1/2} \Delta W^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right) \\
\times d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k),

$$
\begin{align*}
&= B^\Delta_{k,n}/n \\
&+ 2 \sqrt{\int_{Y^k} \left( \text{Re} \left[ n^{-1/2} \hat{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)} \\
&\times \sqrt{\int_{Y^k} \left( \text{Re} \left[ n^{-1/2} \Delta W^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)} \\
&+ 2 \sqrt{\int_{Y^k} \left( \text{Im} \left[ n^{-1/2} \hat{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)} \\
&\times \sqrt{\int_{Y^k} \left( \text{Im} \left[ n^{-1/2} \Delta W^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) \right] \right)^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k)} \\
&\leq B^\Delta_{k,n}/n + 4 \sqrt{B^1_{n,k}/n} \sqrt{B^\Delta_{k,n}/n} \leq D^2_{k,n} + 4C_{1,n}D_{k,n} = o_p(1),
\end{align*}

where the latter follows from (5.19) and (11.29).

**Proof of (11.28).** Next, denote

$$
\psi_k(\tau_1, \tau_2, \ldots, \tau_k) = E \left[ U_t \exp \left( i\sum_{j=1}^k \tau_j \Phi (Y_{r-j}) \right) \right] \right],
$$

$$
\eta_k = \int_{Y^k} |\psi_k(\tau_1, \tau_2, \ldots, \tau_k)|^2 d\mu(\tau_1)d\mu(\tau_2)\ldots d\mu(\tau_k),
$$

and observe from Lemma 4.1 that under Assumptions 3.1 and 4.1,

$$
\sup_{(\tau_1, \tau_2, \ldots, \tau_k) \in Y^k} \left| n^{-1/2} \hat{W}^{(1)}_{k,n}(\tau_1, \tau_2, \ldots, \tau_k) - \psi_k(\tau_1, \tau_2, \ldots, \tau_k) \right| = o_p(1)
$$

for each \( k \in \mathbb{N} \), which implies that \( \hat{B}^{(1)}_{n,k}/n \overset{p}{\to} \eta_k \) for each \( k \in \mathbb{N} \).
12. Models with infinitely many lagged conditioning variables

So far the above results have been derived for nonlinear ARX models for which the conditional expectation function \( f_{t-1}(\theta_0) \) depend on a finite number of lagged dependent variables \( Y_t \) and a finite number of lagged exogenous variables \( X_t \). However, since \( f_{t-1}(\theta_0) \) aims to represent the conditional expectation \( E[Y_t|\mathcal{F}^{t-1}_{-\infty}] \), which in general may depend on all lagged \( Y_t \)'s and lagged \( X_t \)'s, it may be more realistic to allow infinitely many lags in the specification of \( f_{t-1}(\theta_0) \). A convenient way to do that is to specify the errors \( U_t \) as an ARMA process, for example an ARMA(1,1) process \( U_t = \kappa_0 U_{t-1} + e_t - \delta_0 e_{t-1} \), where |\( \kappa_0 \)| < 1, |\( \delta_0 \)| < 1 and \( \delta_0 \neq \kappa_0 \), and now under \( H_0 \), \( E[e_t|\mathcal{F}^{t-1}_{-\infty}] = 0 \) a.s. Then model (3.2) becomes a nonlinear ARMAX model,

\[
Y_t = f_{t-1}(\theta_0) + \kappa_0 (Y_{t-1} - f_{t-2}(\theta_0)) + e_t - \delta_0 e_{t-1},
\]

which by inverting the lag polynomial \( 1 - \delta_0 L \) can be written as an infinite-order ARX model,

\[
Y_t = f_{t-1}(\theta_0) + (\kappa_0 - \delta_0) \sum_{j=0}^{\infty} \delta^j_0 (Y_{t-1-j} - f_{t-2-j}(\theta_0)) + e_t
\]

say.

Previously it was assumed that \( Z_t = (Y_t, X'_t)' \) is observed for \( 1 - \max(p, q) \leq t \leq n \), where \( p \) is the maximum lag of \( Y_t \) in \( f_{t-1}(\theta_0) \), and \( q \) is the maximum lag of \( X_t \) in \( f_{t-1}(\theta_0) \). Adopting this assumption in the present case as well we need to truncate \( g_{t-1}(\theta_0, \delta_0, \kappa_0) \), for example by

\[
\overline{g}_0(\theta_0, \delta_0, \kappa_0) = f_0(\theta_0),
\]

\[
\overline{g}_{t-1}(\theta_0, \delta_0, \kappa_0) = f_{t-1}(\theta_0) + (\kappa_0 - \delta_0) \sum_{j=0}^{t-2} \delta^j_0 (Y_{t-1-j} - f_{t-2-j}(\theta_0))
\]

for \( t \geq 2 \).

Since \( g_{t-1}(\theta, \delta, \kappa) - \overline{g}_{t-1}(\theta, \delta, \kappa) \to 0 \) exponentially for \( t \to \infty \), using \( \overline{g}_{t-1}(\theta, \delta, \kappa) \) as a proxy for \( g_{t-1}(\theta, \delta, \kappa) \) yields, loosely speaking, asymptotically the same results as if the \( g_{t-1}(\theta, \delta, \kappa) \)'s were used.

76
More generally, suppose that \( f_{t-1}(\theta) \) depends on the entire sequence \( \{Z_{t-j}\}_{j=1}^{\infty} \), and that we can specify a truncated version \( f_{T-1}^*(\theta) \) of \( f_{t-1}(\theta) \) that only depend on the observed lagged \( Y_t \) and \( X_t \). One can formulate conditions regarding the speed of convergence to zero for \( t \to \infty \) of \( \sup_{\theta \in \Theta} |f_{t-1}(\theta) - f_{T-1}^*(\theta)| \), \( \sup_{\theta \in \Theta} |(\partial/\partial \theta') f_{t-1}(\theta) - (\partial/\partial \theta') f_{T-1}^*(\theta) - (\partial/\partial \theta)(\partial/\partial \theta') f_{t-1}^*(\theta)| \) such that, after some modifications of the assumptions to take into account that \( f_{T-1}^*(\theta) \) is no longer strictly stationary, the above results carry over, with \( f_{t-1}(\theta) \) replaced by \( f_{T-1}^*(\theta) \). This problem is related to the initial value problem discussed in section 9. The details are left to the reader as an exercise.

### 13. Implementation

Choosing

\[
\Upsilon = X_{s+1}^{s+1}[-c, c]
\]

for some \( c > 0 \), the WICM test statistic can be written in slightly different but asymptotically equivalent form as

\[
\hat{T}_{n}^{(1)} = \sum_{k=1}^{L_n} \frac{1}{\sqrt{n - L_n}} \sum_{t=L_n+1}^{n} \bar{U}_t \exp \left( i \sum_{j=1}^{k} \tau_j' \bar{Z}_{t-j} \right) \times d\mu(\tau_1)d\mu(\tau_2)...d\mu(\tau_k)
\]

where as before, the \( \hat{U}_t \)'s are the NLLS residuals, the \( \bar{Z}_t \)'s are vectors in \( \mathbb{R}^{s+1} \) of bounded transformations of the standardized dependent and exogeneous variables, and \( L_n \) is a subsequence of the effective sample size \( n \). Throughout it will be assumed that \( \mu \) is the uniform probability measure on \( \Upsilon \). The time shift accommodates the standardization procedure as well as the initial value issue.

Recall that under \( H_0 \), \( \hat{T}_{n}^{(1)} \overset{d}{\to} T \), and that the actual WICM test is

\[
\hat{T}_n = \hat{T}_{n}^{(1)}/\hat{T}_{n}^{(2)},
\]

where under \( H_0 \), \( \hat{T}_{n}^{(2)} \) is a consistent estimator of \( E[T] \), and under \( H_1 \), \( p \lim_{n \to \infty} \hat{T}_{n}^{(2)} \in (0, \infty) \).

**Lemma 13.1.** The statistic \( \hat{T}_{n}^{(1)} \) has the closed form expression

\[
\hat{T}_{n}^{(1)} = \left( \sum_{k=1}^{L_n} \gamma_k \right) \left( \frac{1}{n - L_n} \sum_{t=L_n+1}^{n} \bar{U}_t^2 \right)
\]

77
\[ +2 \frac{1}{n-L_n} \sum_{t_1=L_n+2}^{n} \sum_{t_2=L_n+1}^{t_1-1} \hat{U}_{t_1} \hat{U}_{t_2} P_{t_1,t_2} \]

where

\[ P_{t_1,t_2} = \sum_{k=1}^{L_n} \gamma_k \prod_{j=1}^{k} \prod_{i=1}^{s+1} \frac{\sin(c(\tilde{Z}_{i,t_1-j} - \tilde{Z}_{i,t_2-j}))}{c(\tilde{Z}_{i,t_1-j} - \tilde{Z}_{i,t_2-j})} \]

with \( \tilde{Z}_{i,t} \) component \( i \) of \( \tilde{Z}_t \). Moreover, the estimator \( \hat{T}_n^{(2)} \) has the closed form expression

\[ \hat{T}_n^{(2)} = \left( \sum_{k=1}^{L_n} \gamma_k \right) \hat{\sigma}_n^2 - 2.\text{trace} \left[ \hat{A}_{1,n}^{-1} \hat{C}_{2,n} \right] + \text{trace} \left[ A_{1,n}^{-1} A_{2,n} A_{1,n}^{-1} \hat{C}_{1,n} \right], \]

where

\[ \hat{\sigma}_n^2 = \frac{1}{n-L_n} \sum_{t=L_n+1}^{n} \hat{U}_t^2, \]

\[ \hat{A}_{1,n} = \frac{1}{n-L_n} \sum_{t=L_n+1}^{n} \nabla f_t(\hat{\theta}_n) \nabla f_t(\hat{\theta}_n)', \]

\[ \hat{A}_{2,n} = \frac{1}{n-L_n} \sum_{t=L_n+1}^{n} \hat{U}_t^2 \nabla f_t(\hat{\theta}_n) \nabla f_t(\hat{\theta}_n)', \]

\[ \hat{C}_{1,n} = \frac{1}{(n-L_n)^2} \sum_{t_1=L_n+1}^{n} \sum_{t_2=L_n+1}^{n} \nabla f_{t_1}(\hat{\theta}_n) \nabla f_{t_2}(\hat{\theta}_n)' P_{t_1,t_2}, \]

\[ \hat{C}_{2,n} = \frac{1}{(n-L_n)^2} \sum_{t_1=L_n+1}^{n} \sum_{t_2=L_n+1}^{n} \hat{U}_{t_1}^2 \nabla f_{t_1}(\hat{\theta}_n) \nabla f_{t_2}(\hat{\theta}_n)' P_{t_1,t_2}. \]

**Proof.** Similar to Lemma 5.2 in Bierens (2015a). □

Finally, it is not too hard, but rather tedious, to show that

**Lemma 13.2.** A typical bootstrap version \( \tilde{T}_n^{(1)} \) of \( \hat{T}_n^{(1)} \) takes the form

\[ \tilde{T}_n^{(1)} = \left( \sum_{k=1}^{L_n} \gamma_k \right) \frac{1}{n-L_n} \sum_{t=L_n+1}^{n} \varepsilon_t^2 \hat{U}_t^2 \]

78
\[ +2 \frac{1}{n - L_n} \sum_{t_1 = L_n + 2}^n \sum_{t_2 = L_n + 1}^{t_1 - 1} \varepsilon_{t_1} \varepsilon_{t_2} \hat{U}_{t_1} \hat{U}_{t_2} P_{t_1,t_2} \]

\[ -2d_n \hat{A}_{1,n}^{-1} \tilde{e}_n + \varepsilon_t \hat{A}_{1,n}^{-1} \hat{C}_{1,n} \hat{A}_{1,n}^{-1} \tilde{e}_n \]

where the \( \varepsilon_t \)'s are i.i.d. \( N(0, 1) \), \( \hat{A}_{1,n} \) and \( \hat{C}_{1,n} \) are the same as before, and

\[
\tilde{e}_n = \frac{1}{\sqrt{n - L_n}} \sum_{t = L_n + 1}^n \varepsilon_t \nabla f_t (\hat{\theta}_n),
\]

\[
\tilde{d}_n = \frac{1}{(n - L_n) \sqrt{n - L_n}} \sum_{t_1 = L_n + 1}^n \sum_{t_2 = L_n + 1}^n \varepsilon_{t_1} \hat{U}_{t_1} \nabla f_{t_2} (\hat{\theta}_n). P_{t_1,t_2}.
\]

**Remark.** Since \( \hat{T}_n \) \( \xrightarrow{d} T^* \), where under \( H_0 \), \( T^* \sim T \), and in general \( p \lim_{n \to \infty} \hat{T}_n(2) = E[T^*] \), we may use \( \hat{T}_n = \hat{T}_n(1)/\hat{T}_n(2) \) as a bootstrap version of \( \hat{T}_n = \hat{T}_n(1) / \hat{T}_n(2) \).

### 14. A numerical example

#### 14.1. Fitting an AR(1) model to an MA(1) process

In order to verify the finite sample performance of the WICM test, consider the zero-mean Gaussian MA(1) process

\[ Y_t = U_t - \rho U_{t-1}, \text{ where } U_t \sim \text{i.i.d. } N(0, 1), \ |\rho| < 1, \quad (14.1) \]

and suppose that it is incorrectly assumed that this is an AR(1) process i.e.,

\[ H_0 : E [ Y_t | F_{t-1} ] = \theta_1 Y_{t-1} + \theta_2 \text{ a.s.} \quad (14.2) \]

for some \( \theta = (\theta_1, \theta_2)' \), where \( F_{t-1} = \sigma \left( \{ Y_{t-j} \}_{j=1}^{\infty} \right) = \sigma \left( \{ U_{t-j} \}_{j=1}^{\infty} \right) \). Note that

\[
\theta_0 = (\theta_{0,1}, \theta_{0,2})' = \arg \min_{(\theta_1, \theta_2) \in \mathbb{R}^2} E \left[ (Y_t - \theta_1 Y_{t-1} - \theta_2)^2 \right] = \arg \min_{(\theta_1, \theta_2) \in \mathbb{R}^2} E \left[ (U_t - (\rho + \theta_1) U_{t-1} + \rho \theta_1 U_{t-2} - \theta_2)^2 \right] = \arg \min_{(\theta_1, \theta_2) \in \mathbb{R}^2} (1 + \rho^2 + 2\rho \theta_1 + (1 + \rho^2) \theta_1^2 + \theta_2^2) = (-\rho/(1 + \rho^2), 0)',
\]

79
hence, the error term of the AR(1) model involved is

\[ U_t^* = Y_t + \frac{\rho}{1 + \rho^2} Y_{t-1} \]

\[ = U_t - \frac{\rho^3}{1 + \rho^2} U_{t-1} - \frac{\rho^2}{1 + \rho^2} U_{t-2}. \]

It is easy to verify that

\[ E[U_t^* Y_{t-m}] = \begin{cases} 0 & \text{for } m \neq 2, \\ -\frac{\rho^2}{1 + \rho^2} & \text{for } m = 2, \end{cases} \]

so that \( U_t^* \) is independent of all but one lagged \( Y_t \), with \( Y_{t-2} \) the exception. Moreover

\[ E[U_t^* U_{t-m}] = \begin{cases} -\rho^3/(1 + \rho^2) & \text{for } m = 1, \\ -\rho^2/(1 + \rho^2) & \text{for } m = 2, \\ 0 & \text{for } m \geq 3, \end{cases} \]

hence by joint normality,

\[ E[U_t^* | U_{t-1}, U_{t-2}, \ldots, U_{t-m}] = -\frac{\rho^3}{1 + \rho^2} U_{t-1} - \frac{\rho^2}{1 + \rho^2} U_{t-2} \]

for \( m \geq 2 \) and thus

\[ E[U_t^* | \mathcal{F}_{-\infty}^{-1}] = \lim_{m \to \infty} E[U_t^* | U_{t-1}, U_{t-2}, \ldots, U_{t-m}] \]

\[ = -\frac{\rho^2}{1 + \rho^2} (\rho U_{t-1} + U_{t-2}) \]

\[ = -\frac{\rho^3}{1 + \rho^2} Y_{t-1} - \rho^2 \sum_{j=0}^{\infty} \rho^j Y_{t-2-j} \]

The latter equality follows from inverting the MA(1) process (14.1), yielding \( U_t = \sum_{j=0}^{\infty} \rho^j Y_{t-j} \).

Furthermore, note that

\[ E \left[ (E[U_t^* | \mathcal{F}_{-\infty}])^2 \right] = \frac{\rho^4}{1 + \rho^2} \quad (14.3) \]
The values of this expectation are listed in the following table, for $\rho = 0.2, 0.3, ..., 0.9$.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\rho^4/(1 + \rho^2)$</th>
<th>$\rho$</th>
<th>$\rho^4/(1 + \rho^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.0015</td>
<td>0.6</td>
<td>0.0953</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0074</td>
<td>0.7</td>
<td>0.1611</td>
</tr>
<tr>
<td>0.4</td>
<td>0.0221</td>
<td>0.8</td>
<td>0.2498</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0500</td>
<td>0.9</td>
<td>0.3625</td>
</tr>
</tbody>
</table>

Clearly, for low values of $\rho$ it will be difficult to detect misspecification of the AR(1) model in finite samples.

14.2. WICM test results

I have generated independently four time series $Y_t$, $t = 1, 2, ..., 500$, according to (14.1) for $\rho = 0.8$, $\rho = 0.6$, $\rho = 0.4$ and $\rho = 0.2$, respectively, and conducted for each of these four time series the WICM test for the (false) null hypothesis (14.2).

The WICM test has been implemented as in the previous section, with $c = 5$, $\gamma_k = (0.9)^k$ and $L_n = [\sqrt{n}]$, where $[x]$ denotes the largest integer $\leq x$. Thus, $L_n = 22$. Moreover, the bootstrap sample size is $M = 500$.

In the following table, WICM is the test statistic $\hat{T}_n = \hat{T}_n^{(1)}/\hat{T}_n^{(2)}$ as defined in Lemma 10.1, $b(0.01)$, $b(0.05)$, $b(0.10)$ are the bootstrap critical values for significance levels 1%, 5% and 10%, respectively, based on the bootstrap versions $\hat{T}_n = \hat{T}_n^{(1)}/\hat{T}_n^{(2)}$ of $\hat{T}_n$, with $\hat{T}_n^{(1)}$ defined in Lemma 10.2, and BPV is the bootstrap p-value.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>WICM</th>
<th>$b(0.01)$</th>
<th>$b(0.05)$</th>
<th>$b(0.10)$</th>
<th>BPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>2.11</td>
<td>2.05</td>
<td>1.53</td>
<td>1.35</td>
<td>0.006</td>
</tr>
<tr>
<td>0.6</td>
<td>2.16</td>
<td>1.78</td>
<td>1.47</td>
<td>1.32</td>
<td>0.002</td>
</tr>
<tr>
<td>0.4</td>
<td>1.55</td>
<td>1.77</td>
<td>1.52</td>
<td>1.38</td>
<td>0.042</td>
</tr>
<tr>
<td>0.2</td>
<td>0.56</td>
<td>1.82</td>
<td>1.46</td>
<td>1.30</td>
<td>1.000</td>
</tr>
</tbody>
</table>

First note that the WICM test statistics are lower than each of the three the upper bounds in (2.12). Thus, the tentative conclusion is that the false null hypothesis is not rejected. However, the results demonstrate that the upper bounds (2.12) are way too conservative. They are only informative if they indicate rejection at the 1% or 5% significance levels, because then we may conclude without doing the bootstrap that the null hypothesis will be rejected on the basis of the bootstrap results as well.

For $\rho = 0.8$ and $\rho = 0.6$ the false AR(1) null hypothesis is strongly rejected and even for $\rho = 0.4$ the null hypothesis is rejected at the 5% significance level.
In view of (14.3) and (14.4) the result for \( \rho = 0.2 \) is not surprising. In this case we need a much longer time serie to detect the mispecification.

It would have been much more informative if I had done a complete Monte Carlo analysis of the finite sample power of the WICM test for the above MA(1)-AR(1) set up. However, each bootstrap round took about two hours computing time, so such a Monte Carlo analysis with only 100 replications, for example, will take about one month non-stop!

15. Concluding remarks

In this paper I have brought my old paper B84 up-to-date, by

- deriving the actual limiting null distribution of the proposed WICM test,

- deriving much sharper upper bounds of the critical values than those in B84 based on Chebyshev’s inequality for first moments,

- showing how to approximate the actual asymptotic critical values via a bootstrap method, and

- proposing a martingale difference preserving standardization of the lagged conditioning variables such that all the asymptotic properties of the WICM test carry over.

However, the WICM test requires to make a number of choices, namely regarding

1. the absolutely continuous (with respect to Lebesgue measure) probability measure \( \mu \) on \( \Upsilon \),

2. the compact set \( \Upsilon \) itself, and

3. the positive sequence \( \{\gamma_k\}_{k=1}^\infty \).

Under the null hypothesis that the time series regression model is correctly specified these choices do not matter too much. However, they do affect the finite sample power of the test. But under the alternative hypothesis that the model is incorrectly specified we do not know how the correct model looks like, so that we cannot choose \( \Upsilon \), \( \mu \) and \( \{\gamma_k\}_{k=1}^\infty \) such that the finite sample power of the WICM is "optimal". However, there are some practical considerations as well as sub-optimal adaptive procedures for these choices, which I will now discuss.
15.1. The probability measure $\mu$

Boning and Sowell (1999) have shown that with $\mu$ the uniform probability measure on $\Upsilon$ the ICM test in Bierens and Ploberger (1997) is optimal in the sense of having the greatest weighted average local power as defined in Andrews and Ploberger (1994). Also, with $\mu$ the uniform probability measure and $\Upsilon$ a hypercube the WICM test statistic has a closed form expression. Therefore, it is recommended to choose for $\mu$ the uniform probability measure, i.e., $d\mu(\tau) = (\int_\Upsilon 1du)^{-1}d\tau$.

15.2. The compact set $\Upsilon$

In B82 and B84 it was recommended to choose $\Upsilon$ around the origin of the Euclidean space involved. However, for linear and nonlinear regression models with a constant term (as is usual the case) it is well-know that the least squares residuals sum up to zero, regardless whether the model is correctly specified or not. Consequently, in this case the empirical process $\hat{W}_n(\tau)$ in (2.5) is identically zero in $\tau = 0$, and the same applies to the empirical process $\hat{W}_{k,n}$ in (3.17): $\hat{W}_{k,n}(0,0,...,0) \equiv 0$. Therefore, it seems better to choose $\Upsilon$ away from the origin of its Euclidean space.

According to Theorem 2.1, we may choose $\Upsilon$ anywhere in its Euclidean space, but the best place where depends obviously on the correct model under the alternative hypothesis. A possible solution is to make $\Upsilon$ dependent on parameters representing the location of $\Upsilon$, and maximizing the resulting (W)ICM test statistic to these location parameters. For example, consider the ICM test statistic $\int_\Upsilon |\hat{W}_n(\tau)|^2d\tau$ for the i.i.d. case in section 2, and replace $\Upsilon \subset \mathbb{R}^k$ by $\Upsilon(\xi) = \bigcup_{\xi=1}^k [\xi_i - c, \xi_i + c]$, where $c > 0$ is a given constant and the location parameter vector $\xi = (\xi_1, \xi_2, ..., \xi_k)'$ is confined to a compact subset $\Xi$ of $\mathbb{R}^k$. Now denote

$$\hat{B}_n(\xi) = \int_{\Upsilon(\xi)} |\hat{W}_n(\tau)|^2d\tau$$

and similarly, $B_n(\xi) = \int_{\Upsilon(\xi)} |W_n(\tau)|^2d\tau$ and $B(\xi) = \int_{\Upsilon(\xi)} |W(\tau)|^2d\tau$, where $W_n(\tau)$ is defined by (2.8), and $W$ is defined by $W_n \Rightarrow W$. It is easy to verify from (2.7) that under $H_0$, $\sup_{\xi \in \Xi} |\hat{B}_n(\xi) - B_n(\xi)| = o_p(1)$. Moreover, it can be shown that $B_n(\xi)$ is tight on $\Xi$ and so is $\hat{B}_n(\xi)$. It is now easy to verify (see for example Billingsley 1968) that $\hat{B}_n(\xi) \Rightarrow B(\xi)$ on $\Xi$, so that by the continuous mapping

\[\text{Let } \Upsilon = \bar{\bigcup_{\xi \in \Xi} \Upsilon(\xi)}, \text{ where the bar in the latter case denotes the closure. Note that } \Upsilon \text{ is} \]
Theorem,
\[ \sup_{\xi \in \Xi} \hat{B}_n(\xi) \xrightarrow{d} \sup_{\xi \in \Xi} B(\xi) \text{ under } H_0, \]

whereas
\[ \sup_{\xi \in \Xi} \hat{B}_n(\xi)/n \xrightarrow{p} \sup_{\xi \in \Xi} \int_{\tau(\xi)} |\eta(\tau)|^2 d\tau > 0 \text{ under } H_1, \]

where \( \eta(\tau) \) is defined in (2.6).

In principle we can approximate the critical values of \( \sup_{\xi \in \Xi} B(\xi) \) by the same bootstrap approach as in Bierens (2015a), because it is based on bootstrap versions of \( \hat{W}_n(\tau) \) only. Also, upper bounds of these critical values can be derived on the basis of the inequality \( \sup_{\xi \in \Xi} B(\xi) \leq \int_{\Xi} |W(\tau)|^2 d\tau \), where \( \Xi \) is the closure of \( \bigcup_{\xi \in \Xi} \gamma(\xi) \).

In practice the exact computation of \( \sup_{\xi \in \Xi} \hat{B}_n(\xi) \) is too tedious a numerical exercise. However, if we choose for \( \Xi \) a finite set of distinct points in \( \mathbb{R}^k \) then (15.1) and (15.2) carry over, and in this case the computation of \( \max_{\xi \in \Xi} \hat{B}_n(\xi) \) is feasible.

A similar approach is applicable to the WICM test as well.

### 15.3. The sequence of weights

The ideal weight sequence \( \{\gamma_k\}_{k=1}^{\infty} \) for the WICM test is such that under \( H_1 \), \( \gamma_k \) is maximal when \( \eta_k \) is maximal, where \( \eta_k \) is defined in (3.19). But we don’t know the \( \eta_k \)’s. However, what we can do is to make the \( \gamma_k \)’s dependent on a parameter compact. Then for \( \xi_1, \xi_2 \in \Xi \) and \( \varepsilon > 0 \),

\[ \sup_{||\xi_1 - \xi_2|| \leq \varepsilon} |B_n(\xi_1) - B_n(\xi_2)| \]

\[ = \sup_{||\xi_1 - \xi_2|| \leq \varepsilon} \left| \int \limits_{\tau(\xi_1) \setminus \gamma(\xi_2)} |W_n(\tau)|^2 d\tau - \int \limits_{\gamma(\xi_2) \setminus \tau(\xi_1)} |W_n(\tau)|^2 d\tau \right| \]

\[ \leq \sup_{||\xi_1 - \xi_2|| \leq \varepsilon} \left| \lambda((\gamma(\xi_1) \setminus \gamma(\xi_2)) \cup (\gamma(\xi_2) \setminus \gamma(\xi_1))) \sup_{\tau \in \gamma} |W_n(\tau)|^2 \right| \]

\[ = \sup_{||\xi_1 - \xi_2|| \leq \varepsilon} \lambda((\gamma(\xi_1) \setminus \gamma(\xi_2)) \cup (\gamma(\xi_2) \setminus \gamma(\xi_1))) \times O_p(1) \to 0 \text{ as } \varepsilon \downarrow 0, \]

where \( \lambda \) is the Lebesgue measure on \( \Xi \) and the \( O_p(1) \) factor is due to the fact that by the continuous mapping theorem, \( W_n \Rightarrow W \) implies \( \sup_{\tau \in \gamma} |W_n(\tau)|^2 \xrightarrow{d} \sup_{\tau \in \gamma} |W(\tau)|^2 \). This shows that \( B_n(\xi) \) is equicontinuous on \( \Xi \), which in its turn implies that \( B_n(\xi) \) is tight on \( \Xi \). C.f. Billingsley (1968).
with sufficient wide range to control the $k$ for which $\gamma_k$ is maximal. For example, choose for $\gamma_k$ the probability of the Poisson($\omega$) distribution for $k-1$, i.e.,

$$\gamma_k(\omega) = \exp(-\omega) \frac{\omega^{k-1}}{(k-1)!}, \quad k \in \mathbb{N},$$

with $\omega$ confined to a compact set $\Omega$ in $(0, \infty)$. Note that in this case the condition $\sum_{k=1}^{\infty} k^2 \sqrt{\gamma_k(\omega)} < \infty$ in Theorem 9.1 holds. Then the WICM test statistic takes the form

$$\hat{T}_n(\omega) = \sum_{k=1}^{L_n} \gamma_k(\omega) \hat{B}_{n,k}.$$

It is not hard to verify, similar to (15.1) and (15.2), that

$$\sup_{\omega \in \Omega} \hat{T}_n(\omega) \xrightarrow{d} \sup_{\omega \in \Omega} \sum_{k=1}^{\infty} \gamma_k(\omega) B_k \text{ under } H_0,$$

$$\sup_{\omega \in \Omega} \frac{\hat{T}_n(\omega)}{n} \xrightarrow{p} \sup_{\omega \in \Omega} \sum_{k=1}^{\infty} \gamma_k(\omega) \eta_k > 0 \text{ under } H_1.$$

However, it seems unlikely that in this case the upper bounds (2.12) of the critical values are applicable to $\sup_{\omega \in \Omega} \hat{T}_n(\omega)$, but the bootstrap procedure still works. Note that then for each bootstrap version $\hat{T}_n(\omega) = \sum_{k=1}^{L_n} \gamma_k(\omega) \hat{B}_{n,k}$ of $\hat{T}_n(\omega)$ we need to use $\sup_{\omega \in \Omega} \hat{T}_n(\omega)$ as the bootstrap version of $\sup_{\omega \in \Omega} \hat{T}_n(\omega)$, which may take a lot of computing time. Therefore this approach is only feasible in practice if $\Omega$ is chosen finite.

References


\footnote{We still have

$$\frac{\sum_{k=1}^{\infty} \gamma_k(\omega) B_k}{\sum_{k=1}^{\infty} \gamma_k(\omega) E[B_k]} \leq \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^{n} (\varepsilon_i(\omega))^2,$$

where for each $\omega$ the $\varepsilon_i(\omega)$’s are i.i.d. standard normal, but they may depend on $\omega$.}
s", *Journal of Econometrics* 26, 323-353.

Form", *Econometrica* 58, 1443-1458.

Bierens, H. J. (2004): *Introduction to the Mathematical and Statistical Founda-
tions of Econometrics*, Cambridge University Press.

PDF.

http://grizzly.la.psu.edu/~hbierens/ADDENDUM_B1982.PDF.

Conditional Moment Tests", http://grizzly.la.psu.edu/~hbierens/ADDENDUM_-
BP1997.PDF.

for Parametric Conditional Distributions", http://grizzly.la.psu.edu/~hbierens/
ADDENDUM_BW2012.PDF.


Moment Test", *Econometric Theory* 15, 710-719.


Econometrics* 72, 1-32.


using Projections", *Econometric Theory* 22, 1030-1051.

the Martingale Difference Hypothesis test", *Journal of Econometrics* 134, 151-
185.

Härdle, W., and E. Mammen (1993): "Comparing Nonparametric versus Para-

Hong, Y. (1999): "Hypothesis Testing in Time Series via the Empirical Char-


