Long Memory Testing in the Time Domain*

Matei Demetrescu, Vladimir Kuzin, Uwe Hassler
Goethe-University Frankfurt†

May 25, 2007

Abstract

An integration test against fractional alternatives is suggested for univariate time series. The new test is a completely regression based, lag augmented version of the LM test by Robinson (1991, *Journal of Econometrics* 47, 67-84). Our main contributions, however, are the following. First, we let the short memory component follow a general linear process. Second, the innovations driving this process are martingale differences with eventual conditional heteroskedasticity that is accounted for by means of White's standard errors. Third, we assume the number of lags to grow with the sample size, thus approximating the general linear process. Under these assumptions limiting normality of the test statistic is retained. The usefulness of the asymptotic results for finite samples is established in Monte Carlo experiments. In particular, we study several strategies of model selection.

Key words

LM test; general linear process; model selection; White standard errors

JEL classification

C12 (Hypothesis testing), C22 (Time series models)

*An earlier version of this paper was presented at the URCT Conference in Faro, 2005, and at the Econometrics Seminar of the University of Zürich. We are in particular grateful to Peter Robinson and Michael Wolf for helpful comments. Moreover, we wish to thank two anonymous referees for reports that greatly helped to improve the paper.

†Statistics and Econometric Methods, Goethe-University Frankfurt, Gräfstr. 78 / PF 76, D-60054 Frankfurt, Germany, Tel: +49.69.798.23660, Fax: +49.69.798.23662.

1 Introduction

Since the work by Dickey and Fuller (1979, 1981) integration testing has become routine in empirical studies in order to discriminate between time series integrated of order one or zero, I(1) or I(0). The persistence in many economic and financial time series, however, is not well captured by the I(1)/I(0) model. This led to the long memory paradigm introduced to econometrics by Granger and Joyeux (1980), see Henry and Zaffaroni (2003) for a recent review. Most often, long memory is modelled by fractionally integrated processes of order d, I(d). Noninteger orders of integration provoke the concept of fractional cointegration, see Granger (1981). Hence, an increasing number of econometric papers is concerned with (co)integration testing in a fractional surrounding.

The power of Dickey-Fuller [DF] type tests against alternatives of fractional integration is known to be low, see Sowell (1990), Diebold and Rudebusch (1991), Hassler and Wolters (1994) and Krämer (1998). This motivated the development of powerful tests against fractional alternatives. Robinson (1991) pioneered an integration test constructed from the Lagrange Multiplier [LM] principle, which was proven by Robinson (1994) to be locally most powerful under Gaussianity. The test has been further studied and modified by Agiakloglou and Newbold (1994), Tanaka (1999) and Breitung and Hassler (2002). These papers discuss several proposals to capture the short memory, or I(0), component of a time series under investigation. The size and power properties of the LM test crucially hinge on the discrimination of long memory from short memory. In this paper we adopt the approach by Breitung and Hassler (2002). We modify their LM test relying on two regressions by incorporating the second one into the first. The resulting procedure parallels the augmented DF regression [ADF], see also the so-called fractional DF test by Dolado, Gonzalo and Mayoral (2002).

Our main contributions to this literature, however, are the following. The short memory component is allowed to follow a general linear process. The summability conditions required are weaker than those implied by stationary and invertible ARMA processes. Further, we do not require the innovations driving this process to be serially independent. We rather consider a martingale difference sequence allowing for conditional heteroskedasticity. Moreover, the number of lags is assumed to grow with the sample size, thus approximating the general linear process. Similarly to Said and Dickey (1984) for the ADF test, we obtain conditions that leave the limiting distribution unchanged. With the LM type test the asymptotic distribution is standard normal. Finally, we compare the new test with competing procedures in Monte Carlo experiments. In particular, different strategies for model selection in finite samples are investigated. It turns out that simple deterministic rules of thumb outperform data-dependent rules for lag length selection, which is not so surprising given the results by Leeb and Pötscher (2005).

The remainder of this article is structured as follows. In the next section the augmented LM test is described, and our assumptions are discussed in detail. The asymptotic results are presented in Section 3, while Section 4 contains small-sample evidence from a Monte Carlo simulation. The final section includes some guidelines for applied work. Proofs are relegated to the Appendix.

2 Assumptions and procedure

Let an univariate time series y_t follows a fractionally integrated process,

$$(1-L)^{d+\theta} y_t = x_t, \quad t = 1, \ldots, \infty,$$

where L is the lag operator, $(1-L)^{d+\theta}$ is given by the usual series expansion and it is assumed that $y_t = 0$ for $t \leq 0$. We are interested in testing the null hypothesis that y_t is fractionally integrated of order d, or $\mathrm{H}_0 : \theta = 0$. If x_t is independently and identically distributed, $x_t = \varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)$, the process y_t is said to follow fractional white noise. Assuming a normal distribution for ε_t, Robinson (1991) and Tanaka (1999) derive the LM test statistic for a sample of size T as

$$\tau = \frac{1}{\sqrt{T}} \sum_{t=2}^{T} x_t x_{t-1}^* \quad \text{with} \quad x_{t-1}^* = \sum_{j=1}^{t-1} \frac{x_{t-j}}{j}.$$

Under the null hypothesis x_{t-1}^* is asymptotically stationary, and in the i.i.d case, x_t and x_{t-1}^* are independent, thus leading to limiting normality with known variance. In case that x_t follows a stationary and invertible ARMA process, Tanaka (1999) suggests to construct the corresponding test statistic with residuals obtained from fitting the known ARMA model. The variance of the limiting normal distribution, however, depends on the ARMA parameters in a rather complicated way, and closed expressions are provided by Tanaka (1999, p. 564) only for the particular cases of AR(1) and MA(1) models. Robinson (1994) recasts the LM test in the frequency domain, but again the estimation of the limiting variance is not easily accessible for general ARMA processes. In particular, we are not aware of empirical strategies how fit a model to the short-run component. For that reason Breitung and Hassler (2002) propose a regression-based variant of the LM test:

$$x_t = \widehat{\phi} x_{t-1}^* + \widehat{e}_t, \quad t = 2, 3, \ldots .$$

The usual t statistic testing for $\phi = 0$ is asymptotically equivalent to τ under H_0. The regression approach was pioneered by Agiakloglou and Newbold (1994) with $x_{t-1}^{(m)} = \sum_{j=1}^{m} \frac{x_{t-j}}{j}$ for some fixed m as regressor instead of x_{t-1}^*. As the choice of some m is arbitrary we stick to x_{t-1}^* in the following. To account for serial dependence of x_t, the regression-based LM test may be performed in two steps. Assume an autoregressive mechanism of order p for x_t, and denote the AR(p) residuals $\widehat{\varepsilon}_t$. Then the LM principle yields as a second regression

$$\widehat{\varepsilon}_t = \widetilde{\phi}\, \widehat{\varepsilon}_{t-1}^* + \widetilde{\psi}_1 x_{t-1} + \ldots + \widetilde{\psi}_p x_{t-p} + \widetilde{e}_t, \tag{1}$$

with $\widehat{\varepsilon}_{t-1}^* = \sum_{j=1}^{t-1} \frac{\widehat{\varepsilon}_{t-j}}{j}$ and the t statistic t_ϕ is used to test the null hypothesis, where limiting standard normal distribution is recovered. For a rigorous justification, confer Hassler and Breitung (2006, Prop. 1).

The analysis by Breitung and Hassler (2002) of the two-step procedure relying on (1) has three shortcomings. (i) They assume p to be known, which is not the case in practice. However, the choice of p is crucial, especially under the alternative, since in small samples long memory can be easily confused with short memory. (ii) Breitung and Hassler (2002) as well as Tanaka (1999) have to assume that the innovations ε_t driving x_t are i.i.d., while in many applications conditional heteroskedasticity of the ARCH type violates this assumption. (iii) The assumption of an AR(p) model or even an ARMA(p,q) model (implying geometrically declining Wold coefficients) may seem unrealistic for applied purposes. Therefore, we wish to consider a general linear process with some weak summability condition instead.

To overcome the second and third shortcoming, we allow the process x_t to be a stationary and ergodic general linear process with mild summability conditions, driven by a martingale difference sequence [MDS] obeying certain cumulant restrictions.

Assumption 1 *Let x_t follow a stationary invertible general linear process,*

$$x_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \tag{2}$$

where $b_0 = 1$, and ε_t follows a stationary MDS process with absolutely summable 8^{th} cumulants, κ_ε:

$$\sum_{l_1,\dots,l_7=-\infty}^{\infty} |\kappa_\varepsilon(0, l_1, \dots, l_7)| < \infty.$$

Assumption 2 *There exists some $s > 2$ so that the following summability condition holds:*

$$\sum_{j=0}^{\infty} j^s |b_j| < \infty. \tag{3}$$

This summability condition (sometimes called s-summability) implies a quite general linear structure and covers all ARMA processes of finite order. Notice that the representation (2) is equivalent to assuming an infinite order invertible autoregressive process:

$$x_t = \sum_{j=1}^{\infty} a_j x_{t-j} + \varepsilon_t. \tag{4}$$

It is well known (see, e.g., Brillinger, 1975, p. 79) that condition (3) also implies s-summability for the coefficients of the infinite autoregressive representation, namely

$$\sum_{j=0}^{\infty} j^s |a_j| < \infty\,, \tag{5}$$

for the same $s > 2$.

Let us briefly discuss the cumulant condition in Assumption 1, too. In general, it rules out long range dependence in the 2nd moments of the innovations. However, adding a normal distribution this condition holds trivially true because eighth order cumulants are zero. Maintaining the classical i.i.d. assumption one obtains

$$\kappa_\varepsilon(0, l_1, \ldots, l_7) = \begin{cases} \kappa_{\varepsilon,8} & 0 = l_1 = \ldots = l_7 \\ 0 & \text{else} \end{cases} .$$

Hence, in this case the summability reduces to finite 8^{th} order cumulants, $|\kappa_{\varepsilon,8}| < \infty$, and a necessary and sufficient condition for that is the existence of finite 8^{th} moments. Linking other commonly used models to the imposed cumulant conditions turns out as a very difficult task and is beyond the scope of this paper. We notice, however, that the existence of finite eighth moments remains to be a necessary condition. For example from Milhoj (1985) and Bollerslev (1986) it is known that an ARCH(1) model has finite eighth moments, $\mathbb{E}|\varepsilon_t|^8 < \infty$, if $\varphi < (105)^{-1/4} \approx 0.31$, where φ is the corresponding ARCH parameter. In the following we investigate the behaviour of our test under conditional heteroskedasticity by means of Monte Carlo methods.

The assumption of summability of fourth order cumulants is called standard for instance by Andrews (1991), cf. the references and discussion following his Assumption A. In fact, the condition for only fourth order cumulants turns out to be sufficient for consistency in Proposition 1 further down. Limiting normality, however, is established by resorting to eighth order cumulants, although this assumption may be stronger than needed.

To tackle the first shortcoming (i) we propose to test for ϕ and to correct for serial correlation in one regression replacing two steps. This parallels the so-called fractional Dickey-Fuller test by Dolado, Gonzalo and Mayoral (2002). Our augmented LM test thus relies on the regression (under H_0)

$$x_t = \widehat{\phi}\; x^*_{t-1} + \widehat{a}_1 x_{t-1} + \widehat{a}_2 x_{t-2} + \ldots + \widehat{a}_p x_{t-p} + \widehat{\varepsilon}_t, \quad t = p+1, \ldots, T \quad (6)$$

where x^*_{t-1} has been defined above. The null hypothesis is parameterized as $\phi = 0$, and its t statistic is the obvious choice when testing the null hypothesis $\theta = 0$. As will be shown in the following section, large negative values of the test statistic point towards the alternative $\theta < 0$, while large positive values advocate the alternative $\theta > 0$. Note that the test regression (6) has the typical structure of an unbalanced regression, however it is not a Wald-type test, since $\widehat{\phi}$ is not an estimator for θ. Motivated by the classic work of Berk (1974), its multivariate extension due to Lewis and Reinsel (1985) and also by Said and Dickey (1984, for the augmented Dickey-Fuller [ADF] test) and the extension due to Chang and Park (2002) we employ an approximation of the general linear process from Assumption 1 by lagged endogeneous regressors, where the lag order p is increasing with the sample size T as specified by following assumption.

Assumption 3 *Let the order p of the approximating autoregressive process satisfy*

$$p = O(T^{\kappa}),$$

with $p \to \infty$ as $T \to \infty$, where $\kappa \in (1/2s, 1/4)$ and $s > 2$ defined in Assumption 2.

Including lags of increasing order we hope to obtain innovations, ε_{tp}, that are asymptotically MDS:

$$\varepsilon_{tp} = \varepsilon_t + \sum_{j=p+1}^{\infty} a_j x_{t-j}. \tag{7}$$

With p going to infinity, the bias from using a finite autoregressive approximation will vanish asymptotically. The lag length p must be prevented from increasing too fast, in order not to induce excess variability in the estimators. Hence, we adopt in Assumption 3 the condition of Gonçalves and Kilian (2003) and assume that $p = o(T^{1/4})$ as $p \to \infty$ and $T \to \infty$. This is more restrictive than the assumption of Berk (1974) working with i.i.d. innovations. We also need a lower bound for the growth of the lag order p, in order to prevent an inadequate description of the model. The lower bound is taken to be $p = O(T^{1/2s})$, where s is from Assumption 2. Clearly, we require $s > 2$ in condition (5) to guarantee that the lower bound of κ in Assumption 3 does not exceed the upper bound. Note that the lower bound condition $p = O(T^{1/2s})$ prohibits the lag order to increase at logarithmic rate, $\ln(T)$, even if x_t is a finite-order ARMA process. There is a trade-off between the s summability and the freedom one has in the choice of p. The conditions stated in Assumption 3 are comparable to those in Said and Dickey (1984), where their lower bound does not depend on the summability properties of the process x_t. Chang and Park (2002) allow for a wider range of values for p, in particular they allow for logarithmic rates, but both papers work in a simpler framework, given the presence of the integrated regressor in the ADF regression. In contrast, our test exhibits stationary regressors under the null hypothesis, thus constraining our possibilities.

As is well known, the choice of p is of crucial importance for the performance of tests. Ng and Perron (1995) proposed information criteria and strategies of significance testing for the ADF test. Given the findings by Leeb and Pötscher (2005), these rules only work because lagged differences and lagged levels are asymptotically uncorrelated in the ADF case. In contrast, the estimators $\widehat{\phi}$ and $\widehat{a}_p$ from (6) are correlated due to the stationarity of all regressors. Hence, a data-driven model selection will affect the finite sample distribution of $\widehat{\phi}$ and its t statistic, see Leeb and Pötscher (2005) and the references there. Therefore, we will discuss automated, deterministic lag length selection. Schwert (1989) proposed a rule of thumb, often used with integration tests. For a given constant K, the truncation lag, p, is chosen as

$$p_K = \left[K(T/100)^{1/4}\right] \tag{8}$$

where $[\cdot]$ denotes the largest integer part a real number. Values of $K = 4$ and 12 were used in Schwert's Monte Carlo analysis. Notice that p_K increases with the boundary rate from Assumption 3.

3 Asymptotic results

Under the null hypothesis it holds because of (4) $x_t = \sum_{j=1}^{p} a_j x_{t-j} + \varepsilon_{tp}$, with ε_{tp} from Equation (7). Adding x_{t-1}^* to the set of regressors denoted as $V_{tp} = \left(x_{t-1}^*, x_{t-1}, x_{t-2}, \ldots, x_{t-p}\right)'$ the true parameter vector becomes $\underline{\beta}_p = (0, a_1, \ldots, a_p)'$. Further, we denote $\widehat{\underline{\beta}}_p = \left(\widehat{\phi}, \widehat{a}_1, \ldots, \widehat{a}_p\right)'$ as the $(p+1)$-vector of the estimated coefficients from the regression (6):

$$x_t = V_{tp}' \widehat{\underline{\beta}}_p + \widehat{\varepsilon}_t \,.$$

The first result is consistency of the estimators, a proof of which is given in the Appendix.

Proposition 1 *Assume $\theta = 0$ and let Assumption 1, weakened to summability of only 4^{th} cumulants, as well as Assumptions 2 and 3 be fulfilled. Then, the following relationships hold true as $T \to \infty$:*

$$\left\| \sqrt{T}(\widehat{\underline{\beta}}_p - \underline{\beta}_p) \right\| = O_p(p^{1/2}) \text{ and } \sqrt{T}(\widehat{\beta}_{pi} - \beta_{pi}) = O_p(1),$$

where $i = 1, \ldots, p+1$, and $\|\cdot\|$ is the Euclidean vector norm.

Next, we establish limiting normality of $\widehat{\phi} = \widehat{\beta}_{p1}$. The asymptotically normal test statistic, however, will not rely on the conventional studentization. As ε_t is allowed to be serially dependent, the squared errors may be correlated with $V_{tp}V_{tp}'$. Hence, the covariance matrix of $\widehat{\underline{\beta}}_p$ requires some heteroskedasticity robust estimation. Specifically, we consider White standard errors:

$$\hat{s}(\widehat{\phi}) = \left[\left(\sum_{t=p+1}^{T} V_{tp}V_{tp}' \right)^{-1} \left(\sum_{t=p+1}^{T} V_{tp}V_{tp}'\widehat{\varepsilon}_t^{\,2} \right) \left(\sum_{t=p+1}^{T} V_{tp}V_{tp}' \right)^{-1} \right]_{11}^{1/2}, \qquad (9)$$

where $[\cdot]_{11}$ denotes the first diagonal element. White standard errors were considered by Haldrup (1994) for the Dickey-Fuller test, see also Demetrescu (2006) for corresponding asymptotic results. The test statistic is

$$\tilde{t}_\phi = \frac{\widehat{\phi}}{\hat{s}(\widehat{\phi})}$$

Asymptotic standard normal distribution is proven in the Appendix.

Proposition 2 *Under Assumptions 1, 2 and 3, it holds as $T \to \infty$:*

$$\tilde{t}_\phi \xrightarrow{d} \mathbf{N}(0,1).$$

Remark 1 *If a researcher is willing to assume serial independence of ε_t then he or she may employ the standard t statistic t_ϕ without White's correction. The limiting distribution under i.i.d. innovations will remain unchanged.*

Now, we consider our test problem under the alternative, where y_t is integrated of order $d + \theta$, and it thus holds

$$(1 - L)^d y_t = \widetilde{x}_t,$$

with $\widetilde{x}_t$ being fractionally integrated of order $\theta \neq 0$. Consequently

$$\widetilde{x}_t = (1-L)^{-\theta} x_t = \left(1 + \theta L + \frac{\theta(\theta+1)L^2}{2!} + \frac{\theta(\theta+1)(\theta+2)L^3}{3!} + \ldots \right) x_t.$$

Let us consider a sequence of local alternatives, $\theta = \delta/\sqrt{T}$, so that, see Tanaka (1999, p. 579), it holds

$$\widetilde{x}_t = x_t + \frac{\delta}{\sqrt{T}} x^*_{t-1} + O_p(T^{-1}), \tag{10}$$

for $t = 1, 2, \ldots, T$, with $x^*_0 = 0$. The test regression becomes instead of (6)

$$\widetilde{x}_t = \widehat{\phi}\ \widetilde{x}^*_{t-1} + \widehat{a}_1 \widetilde{x}_{t-1} + \widehat{a}_2 \widetilde{x}_{t-2} + \ldots + \widehat{a}_p \widetilde{x}_{t-p} + \widehat{\varepsilon}_t,$$

and it follows

$$\widetilde{x}^*_{t-1} = x^*_{t-1} + \frac{\delta}{\sqrt{T}} x^{\delta}_{t-2} + o_p(1), \text{ where } x^{\delta}_{t-2} = \sum_{j=1}^{t-1} \frac{x^*_{t-1-j}}{j}. \tag{11}$$

Heuristically, $\widetilde{x}_t$ will correlate with $\widetilde{x}^*_{t-1}$, due to the common component x^*_{t-1}, leading to a value of ϕ different from zero and thus rejection of the null hypothesis. This intuition is formalized in the following proposition, whose proof is given in the Appendix.

Proposition 3 *Under $\theta = \delta/\sqrt{T}$ and Assumptions 1 through 3, it holds as $T \to \infty$*

$$\tilde{t}_{\phi} \xrightarrow{d} \mathbf{N}(\delta/s_{\phi}, 1),$$

with $s_{\phi} = \text{plim}\left(\sqrt{T}\widehat{s}(\widehat{\phi})\right)$ defined in (9) being positive.

In practical applications, time series typically exhibit non-zero mean, or seasonally varying means. Of course, their presence distorts the outcome of the testing procedure. Therefore, one needs to account for deterministic components in the observed series. In the following, we allow for a general trend function f, where

$$f(s) = s^{\alpha} + o(s^{\alpha}) \text{ as } s \to \infty.$$

The observed process is modelled then by

$$y_{t,obs} = y_t + \eta\, f(t).$$

This formulation encompasses most of the usual trend functions. For instance, a non-zero mean corresponds to $\alpha = 0$, as does a break in the mean with known break date. A linear trend corresponds to $\alpha = 1$.

Instead of removing the trend from $y_{t,obs}$, we follow Robinson (1994) suggesting to regress fractional differences of $y_{t,obs}$ on fractional differences of the deterministic component. Concretely, one simply builds differences, $x_{t,obs} := (1-L)^d y_{t,obs}$, and regresses them on the differenced trend function,

$$(1-L)_t^d\, f(t) = \sum_{i=0}^{t-1} d_i\, f(t-i), \text{ where } d_i = \frac{i-1-d}{i} d_{i-1},\ d_0 = 1.$$

The residuals from this regression, $\widehat{x}_t = x_{t,obs} - \widehat{\eta}\,(1-L)_t^d f(t)$, are then used for regression (6). This may be of advantage; for instance, under the null of $d = 1$, a time trend reduces to a non-zero mean of the differenced process, while a non-zero mean in levels disappears.

The following proposition ensures that the parameter vector $\underline{\beta}_p$ is consistently estimated in the case of detrended series where the deterministic component has been removed. In particular there is no difference between the estimator computed from demeaned series and the one computed from the unobserved series without deterministics asymptotically. Also, the corresponding test statistic $\widetilde{t}_{\phi r}$ computed with differences after properly removing deterministics, $\widehat{x}_t$, has again standard normal asymptotic distribution. The proof is provided in the Appendix.

Proposition 4 *Let $\widehat{\underline{\beta}}_{pr}$ be the estimator obtained by estimating the test regression (6) with $\widehat{x}_t = x_{t,obs} - \widehat{\eta}\,(1-L)_t^d f(t)$ instead of x_t. Under the assumptions of Proposition 2, it holds notwithstanding removal of deterministics*

$$\left\| \widehat{\underline{\beta}}_{pr} - \widehat{\underline{\beta}}_p \right\| = o_p\left(T^{-1/2}\right) \ \ \text{and} \ \ \widetilde{t}_{\phi r} \xrightarrow{d} \mathbf{N}(0,1)$$

as long as $\alpha > -1$ and $d > 0$.

Remark 2 *Proposition 4 can be directly generalized to the case of several deterministic components, for instance a constant and a trend, or seasonally varying means.*

Remark 3 *With a slightly modified proof, Proposition 4 can also be extended to cover the case where y_t are the error components in a regression of stochastic quantities independent of y_t.*

4 Finite sample properties

Finite sample properties of our test, denoted by A-LM (augmented LM), and the two-step test by Breitung and Hassler (2002) as its natural competitor [2S-LM] are studied. Moveover, we compare the performance with the so-called fractional Dickey-Fuller test proposed by Dolado, Gonzalo and Mayoral (2002) [F-DF], which we briefly review now. Assuming $d = 1$, the regression similar to (6) becomes

$$\Delta y_t = \widehat{\phi}\ \Delta^{d^*} y_{t-1} + \widehat{a}_1 \Delta y_{t-1} + \widehat{a}_2 \Delta y_{t-2} + \ldots + \widehat{a}_p \Delta y_{t-p} + \widehat{\varepsilon}_t.$$

For $0.5 \leq d^* < 1$, the limiting distribution of the t-statistic testing for $\phi = 0$ is again normal. Lobato and Velasco (2005) show that the optimal value of d^* is given by $-0.03 + 0.717(d + \theta)$. Throughout, we work with this optimal choice, although in practice $d + \theta$ from this formula is not known and has to be estimated. In fact, we employ a slight modification: In order to maintain a standard normal limiting distribution of the F-DF test statistic, we use a trimming rule based on the one suggested in Dolado, Gonzalo and Mayoral (2002) with $c = 0.02$: $d^* = 0.52$, if $d^* < 0.52$ and $d^* = 0.98$, if $d^* > 0.98$. In the latter case the A-LM and the F-DF tests are very similar in the following sense:

$$\Delta^{0.98} y_{t-1} = \Delta^{-0.02} x_{t-1} = \sum_{j=0} d_j x_{t-1-j} \quad \text{with} \quad d_j = O(j^{-0.98}),$$

and hence $\Delta^{0.98} y_{t-1} \approx x^*_{t-1}$, where x^*_{t-1} is the sum weighted with j^{-1} from the A-LM test relying on $x_t = \Delta y_t$.

To begin with we work with linear processes driven by ε_t that are i.i.d. and normal, which meets Assumption 1. More precisely, we employ the following ARMA(1,1) process:

$$x_t = 0.5 x_{t-1} + \varepsilon_t + 0.5 \varepsilon_{t-1}, \qquad t = 1, \ldots, T. \tag{12}$$

Clearly, this is an AR(∞) process with geometrically declining coefficients satisfying Assumption 2.

We present results for two different sample sizes: $T = 100$ and $T = 500$; further results for $T = 200$ and $T = 1000$ are available upon request. No constant is included in the test regressions, and $d = 1$. Only two sided tests were computed at the nominal level of 5%. We simulate with $\theta = 0$ to explore size properties. Different lag length selection algorithms are employed to approximate the $AR(\infty)$ processes by autoregressions of order p: Schwert's deterministic rules with $K = 12$ and $K = 4$, see (8), sequential testing, see for example Ng and Perron (1995), and also information criteria, in particular AIC and SIC. Sequential testing was carried out by individual t-tests on the last lag at 10% significance level, where Schwert's rule with $K = 12$ was also applied to set the maximal lag for sequential testing and information criteria. Table 1 contains rejection frequencies as well as means of the selected lag length over 10000 replications. All tests rely on standard t statistics without heteroskedasticity consistent standard errors.

		$T = 100$					$T = 500$				
		$p4$	$p12$	10%	AIC	SIC	$p4$	$p12$	10%	AIC	SIC
A-LM	Pr	4.93	6.26	13.10	12.24	16.25	5.46	4.75	8.00	8.15	11.99
	$\overline{p}$	4.00	12.00	6.58	4.52	2.22	5.00	17.00	10.36	5.46	3.02
2S-LM	Pr	5.01	5.94	2.54	2.55	3.79	5.19	4.70	3.01	2.95	4.40
	$\overline{p}$	4.00	12.00	6.54	4.50	2.34	5.00	17.00	10.34	5.51	3.10
F-DF	Pr	5.14	6.53	12.09	11.24	14.53	5.36	4.87	7.28	7.24	9.35
	$\overline{p}$	4.00	12.00	6.62	4.55	2.24	5.00	17.00	10.35	5.50	3.06

Table 1: 5% size ($\theta = 0$) for ARMA(1,1) as in (12), $p_K = \left[K(T/100)^{1/4}\right]$; no ARCH-effects, $N = 10000$ replications, Pr denotes the resulting size and $\bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i$ the mean of all selected lags; 10%, AIC and SIC denote the stochastic lag selection criteria described in the text .

All three tests perform similarly well in terms of size if the lag length is selected deterministically following Schwert's rule with $K = 4$ and $K = 12$ (the choice of only p_2 lags yields experimental size distortions not reported here). For small samples ($T = 100$), p_{12} is dominated by p_4 in Table 1. But then we observe a result that is rather surprising at first glance: none of the three data-driven lag length selection methods does really work! The A-LM and F-DF tests are always oversized, and the 2S-LM test is undersized in all cases where lag length is estimated from the data, even in large samples (it is worth noting that there are no differences in the means of selected numbers of lags between the tests). At first sight this seems to be counter-intuitive, especially if we think about the successful use of information criteria and significance testing applying the ADF test. But in contrast to the ADF test, only (asymptotically) stationary variables are included in the A-LM, F-DF or 2S-LM test regressions. In particular, $x^*_{t-1} = \sum_{j=1}^{t-1} j^{-1} x_{t-j}$ is square summable and therefore stationary as $t \to \infty$. Hence, the test statistics t_ϕ and t_{a_i}, $i = 1, \ldots, p$, from regression (6) exhibit a certain amount of stochastic dependence even asymptotically, so that the selected lag length $\widehat{p}$ and the test statistic t_ϕ are dependent random variables. Our simulation evidence is related to the results of Leeb and Pötscher (2005), where the potentially disastrous impact of data-driven model selection on subsequent inference is explored.

This problem is not restricted to the regression-based tests investigated here. The same problem affects the tests due to Tanaka (1999) and Robinson (1994). To illustrate this, we simulate $x_t = \varepsilon_t$ as normal white noise, and test at the 10% level whether a fitted AR(1) parameter is significant or not. Tanaka's test is then performed with an AR(0) specification, an AR(1) specification and with the specification indicated by the outcome of the significance test. While both AR(0) and AR(1) specifications produce sizes close to the 5% nominal level, the variant choosing between AR(0) and AR(1) does not work. In the latter case, Tanaka's test is just as conservative (even for T=1000) as the 2S-LM test in Table 1. Actually, this result is not surprising, since both tests are asymptotically equivalent. Similar results are available for Robinson's test.

In Table 2 we turn our attention to power for the special case of an AR(1) process, and in particular to the question how much power is lost by implementing some deterministic rule p_K given the true lag length is $p = 1$. The following features are observed: Firstly, all tests are more powerful when $\theta < 0$ compared to $\theta > 0$. Secondly, and not surprisingly, the augmentation with the true lag length $p = 1$ dominates the rule p_4, which outperforms p_{12}. In practice, however, the true lag length is not known. Given the superiority of the rule p_4 over p_{12}, we, thirdly, rank the tests for p_4: A-LM and (optimal) F-DF have very similar empirical (size and) power functions. F-DF is slightly superior for $\theta < 0$, which is no longer the case for $\theta > 0$.

θ		$T=100$			$T=200$			$T=500$		
		1	p_4	p_{12}	1	p_4	p_{12}	1	p_4	p_{12}
	A-LM	99.53	93.36	59.70	100.00	99.98	89.22	100.00	100.00	99.93
-1.00	2S-LM	98.47	79.11	32.10	100.00	99.22	64.99	100.00	100.00	97.12
	F-DF	99.95	98.24	69.98	100.00	100.00	96.13	100.00	100.00	100.00
	A-LM	36.49	26.32	13.79	66.09	53.18	22.41	97.66	91.37	49.15
−0.40	2S-LM	37.39	22.03	8.74	68.49	46.60	15.85	98.22	86.43	38.57
	F-DF	44.38	29.72	15.20	80.10	60.85	24.99	99.78	96.21	55.30
	A-LM	21.18	17.08	10.63	39.62	33.20	14.95	81.48	67.34	29.93
−0.30	2S-LM	22.43	14.97	7.26	43.27	29.80	10.93	85.80	62.81	24.91
	F-DF	25.36	18.37	11.62	49.58	36.36	15.57	92.76	73.44	32.32
	A-LM	11.18	10.19	8.06	18.82	16.69	9.47	44.14	34.74	15.86
−0.20	2S-LM	12.19	9.17	6.64	21.13	15.67	7.39	50.63	33.04	12.94
	F-DF	12.58	11.01	8.81	22.24	18.03	10.33	53.99	36.39	16.35
	A-LM	6.89	5.97	6.38	8.14	7.92	6.22	13.27	12.14	7.92
−0.10	2S-LM	6.90	6.27	5.81	8.54	7.54	5.48	15.53	12.04	7.40
	F-DF	7.01	6.66	6.68	8.44	8.32	6.72	14.85	12.70	8.28
	A-LM	5.09	5.30	6.45	5.04	4.63	4.88	4.68	4.60	4.82
0.00	2S-LM	4.80	5.25	5.60	4.42	4.62	4.99	4.82	4.78	4.81
	F-DF	5.12	5.48	6.58	4.82	4.63	5.20	4.52	4.69	5.18
	A-LM	6.40	5.77	6.06	7.33	6.02	5.72	13.10	10.20	6.09
0.10	2S-LM	5.71	5.58	6.33	6.76	5.86	5.15	13.09	9.53	5.81
	F-DF	6.37	5.92	6.22	7.08	5.77	6.00	12.85	9.15	6.03
	A-LM	8.73	7.12	6.73	13.07	11.39	6.08	31.17	23.91	9.53
0.20	2S-LM	6.77	6.21	6.80	10.79	9.06	5.63	27.32	19.34	7.92
	F-DF	8.53	7.19	6.99	12.82	10.92	6.22	29.86	22.08	9.49
	A-LM	8.86	8.84	7.64	17.90	17.33	7.71	42.65	37.15	12.41
0.30	2S-LM	6.95	6.93	7.62	11.03	10.55	6.80	27.19	21.26	7.29
	F-DF	8.55	8.82	7.69	17.02	17.19	7.70	40.07	35.73	12.23
	A-LM	8.97	10.44	7.76	17.58	20.35	8.15	47.84	46.22	16.05
0.40	2S-LM	8.83	7.73	8.41	11.79	9.23	7.00	20.75	15.88	8.41
	F-DF	8.28	10.38	7.79	15.44	20.44	8.15	42.90	46.37	16.03
	A-LM	63.15	7.84	10.52	76.80	6.89	8.18	91.71	6.72	6.97
1.00	2S-LM	98.22	6.77	9.95	99.99	6.13	7.68	100.00	5.29	6.22
	F-DF	71.97	7.76	10.47	85.61	6.88	8.20	97.05	6.66	6.95

Table 2: 5% size and power for AR(1) process, $x_t = 0.5x_{t-1} + \varepsilon_t$, with deterministic lag length selection according to $p = 1$ (the true model), $p_4 = \left[4(T/100)^{1/4}\right]$, $p_{12} = \left[12(T/100)^{1/4}\right]$; no ARCH effects, $N = 10000$ replications.

Generally, both tests dominate 2S-LM whenever p_4 is applied. The power functions are not only asymmetric, but also nonmonotonic for $\theta > 0$. In particular, power collapses for $\theta = 1$, unless $p = 1$ is employed. The reason for that is that this alternative is not fractional but purely autoregressive and hence not detected by the test. The dynamics is rather fully captured by the autoregression so that power is reduced to size. Further Monte Carlo results were collected but are not reported here. In particular, we observed more power for AR(1) processes with negative serial correlation, and we found that Schwert's p_4 rule is recommendable also with the more persistent linear process $x_t = \sum_{j=1}^{t-1} j^{-4} \varepsilon_{t-j}$.

Given the superior performance of p_4 over p_{12} with respect to size and power, we apply p_4 to determine the lag length in what follows. Two further tests are now included. The first one builds on a semi-parametric estimator of d which is asymptotically normally distributed. In particular, the local Whittle [LW] estimator is investigated because Robinson (1995) shows that it is more efficient than competing semi-parametric estimators. Moreover, it has a couple of nice properties: It is robust to conditional heteroskedasticity, see Robinson and Henry (1999), it remains consistent up to $d < 1$ and asymptotically normal up to $d < 0.75$ (Velasco, 1999), and a procedure to compute the required bandwidth m optimally from the data may follow Henry (2001). The second test is the A-LM variant with White standard errors from (9), [A-LM-W]. One further aspect of Table 3 is conditional heteroskedasticity. In addition to i.i.d. sequences we allow for ARCH(1) innovations that are close to nonstationarity with conditional variance given as $1 + 0.95\, \varepsilon_{t-1}^2$.

The local Whittle estimator (or test) is included only under the null hypothesis, and it is computed from the differences of the series, which are I(0). Gross size distortions are observed still for $T = 500$, and hence the test is not investigated under the alternative. Given i.i.d. innovations we observe for A-LM (and A-LM-W), 2S-LM and F-DF similar results as in Table 2: Tests against $\theta < 0$ are more powerful than against $\theta > 0$, A-LM, A-LM-W and F-DF display very similar power functions and dominate 2S-LM. Under the assumption of the rather persistent ARCH(1) volatility we find that the White corrected version A-LM-W is closest to the nominal level, although the other test (except for LW) work remarkably well under the null. At the same time A-LM-W is most powerful as long as $|\theta| \leq 0.2$. We conclude that

θ		I.I.D.			ARCH		
		$T = 100$	$T = 200$	$T = 500$	$T = 100$	$T = 200$	$T = 500$
	A-LM	20.86	41.89	90.92	20.68	42.80	91.30
-0.40	A-LM-W	22.62	42.95	90.98	23.26	40.17	80.98
	2S-LM	15.72	33.59	87.30	15.42	33.78	87.18
	F-DF	23.94	50.34	95.94	24.06	52.00	95.97
	A-LM	12.58	23.40	67.82	12.60	24.13	70.41
-0.30	A-LM-W	13.69	24.51	67.98	14.74	23.83	64.10
	2S-LM	10.09	20.00	64.69	10.17	19.68	66.63
	F-DF	14.41	28.21	72.24	14.20	27.87	74.29
	A-LM	7.97	11.20	36.86	7.69	11.98	38.82
-0.20	A-LM-W	9.03	11.92	37.18	9.22	12.21	38.83
	2S-LM	7.14	9.62	35.18	6.79	10.75	35.94
	F-DF	9.03	13.31	37.53	8.42	13.60	38.20
	A-LM	4.97	5.51	13.71	4.75	5.32	15.06
-0.10	A-LM-W	5.77	6.01	14.06	5.33	5.15	15.83
	2S-LM	4.91	5.30	12.83	4.65	5.32	13.70
	F-DF	5.48	6.24	13.46	5.12	5.64	14.51
	A-LM	4.59	4.05	5.35	4.16	3.69	5.96
	A-LM-W	5.25	4.37	5.62	5.08	3.78	5.58
0.00	2S-LM	4.75	4.37	5.24	4.76	4.43	5.61
	F-DF	4.57	4.08	5.26	4.21	3.55	5.46
	LW	29.37	27.80	20.44	28.58	25.37	18.37
	A-LM	5.88	7.16	8.49	5.38	6.90	8.24
0.10	A-LM-W	6.49	7.50	8.72	6.72	8.39	8.67
	2S-LM	5.40	6.85	8.27	6.03	7.67	8.90
	F-DF	5.87	6.56	8.16	5.37	6.32	7.47
	A-LM	7.87	12.58	19.40	7.29	13.29	19.31
0.20	A-LM-W	8.74	13.24	19.76	9.64	14.99	19.41
	2S-LM	6.66	10.49	16.43	7.31	12.77	18.68
	F-DF	7.88	11.96	18.19	7.19	12.79	18.28
	A-LM	9.46	18.27	31.31	10.41	19.49	32.65
0.30	A-LM-W	10.84	18.86	31.69	12.47	20.81	30.09
	2S-LM	8.05	11.96	18.44	9.01	15.17	23.11
	F-DF	9.56	17.83	30.51	10.44	19.13	31.88
	A-LM	11.77	21.94	39.31	11.41	24.77	41.79
0.40	A-LM-W	12.80	22.63	39.61	13.91	25.08	36.91
	2S-LM	8.87	12.56	12.90	10.06	16.43	16.99
	F-DF	11.79	22.25	39.28	11.61	25.13	41.83

Table 3: 5% size and power for ARMA(1,1)-process as in (12) with deterministic lag length selection, $p_4 = \left[4(T/100)^{1/4}\right]$. A-LM-W denotes the A-LM test with White's standard errors, see (9), while LW stands for the local Whittle estimator, with automatic bandwidth selection and starting bandwidth $[T^{0.8}]$. We consider i.i.d. errors and ARCH(1) errors with conditional variance $1 + \varphi\,\varepsilon_{t-1}^2$, $\varphi = 0.95$, $N = 10000$ replications.

use of White standard errors seems to be a good strategy given the fact that they have no negative affect with i.i.d. innovations while they improve the test under ARCH errors.

5 Summary

Several variants of the LM test for integration against fractional alternatives have been studied since Robinson (1991). An open challenge still is how to account for serial short-run correlation as opposed to long memory, and how to correct for heteroskedasticity. Those aspects are tackled in the present paper where we present a lag augmented LM test, which is completely regression based and recommended with White's standard errors.

Our findings are the following. Lag augmentation leaves the limiting normal distribution unchanged. The short memory component does not have to be an ARMA process, it may follow a general linear process with weaker summability conditions. The limiting distribution is retained as long as the number of lags is growing with the sample size at appropriate rate. Further, the innovations driving this process are not required to be serially independent. We rather consider a martingale difference sequence allowing for conditional heteroskedasticity. Finally, the new test is compared with competing procedures in Monte Carlo experiments. Special focus is on different strategies for model selection accounting for serial short-run correlation. It turns out that data-dependent rules for lag length selection do not provide valid inference for any of the tests, which is in line with findings by Leeb and Pötscher (2005). Therefore, we propose a simple deterministic rules of thumb. Moreover, the investigated correction for heteroskedasticity seems to improve the behaviour in the presence of ARCH.

We only considered regression-based tests since we are not aware of empirical strategies how fit a model to the short-run component otherwise. For empirical work our findings may be cast into the following guidelines. (i) The number of lagged differences should not be determined following information criteria or significance of their coefficients. Such a stochastic, data-driven model selection (which is usually employed with the augmented Dickey-Fuller test) implies invalid inference. (ii) Therefore, we explored deterministic choices of lag length selection, in particular $p_K = \left[K(T/100)^{1/4}\right]$

for a sample of length T. For a variety of samples sizes and simulated processes we observed that $K = 4$ is large enough to yield experimental sizes of the tests close to the nominal one under the null hypothesis. At the same time the power is reduced under the alternative as p_K grows for given T. Therefore, we recommend to work with $p_4 = \left[4(T/100)^{1/4}\right]$ in practice and eventually to vary the lag length around this value to check whether test results are robust and hence reliable with respect to the number of lags. (iii) Our proposed lag augmented LM [ALM] test performed in one step is in finite samples superior to the two-step variant proposed by Breitung and Hassler (2002). (iv) We investigated the fractional Dickey-Fuller [FDF] test by Dolado, Gonzalo and Mayoral (2002), too. It depends on a parameter d^* one has to choose. Throughout, we simulated with the optimal choice following Lobato and Velasco (2005). This variant behaves very similarly to our new test and is mildly more powerful under some alternatives. In practice, however, the optimal choice is not known but has to be estimated. The effect of uncertainty due to estimating d^* on the FDF test has not been addressed. (v) The power function is asymmetric. The tests are clearly more powerful when testing against $d < d_0$ than under alternatives $d > d_0$. In particular, under the alternative $d_0 + 1$ power reduces to size because the alternative is not fractional but purely autoregressive and hence not detected when lags are included. (vi) Monte Carlo experiments suggest that the considered tests are fairly robust with respect to some conditional heteroskedasticity. Nevertheless, we observe that our proposal to compute the ALM test with White standard errors may improve the performance. At the same time the correction for heteroskedasticity does no harm when no heteroskedasticity is present.

Appendix

Throughout the Appendix, we use for matrices the matrix norm induced by the Euclidean vector norm, defined for any $m \times n$ matrix A as $\|A\| = \sup\left\{ \|Ax\| \,/\, \|x\| \,|\, x \in \mathbb{R}^n \right\}$. Further, $\{\gamma_h^u\}_{h\in\mathbb{Z}}$ denotes the sequence of autocovariances of any stationary process u_t, and $\{\gamma_h^{uv}\}_{h\in\mathbb{Z}}$ the sequence of cross-covariances of u_t with v_t (also stationary). Explicitly, γ_j^{*x} denotes cross-covariances of x_t^{**} (defined in (13) below) and x_t. Finally, all sums are over $\{p+1,\ldots,T\}$, unless stated otherwise.

A Auxiliary Lemmas

Lemma 1 *Let Assumptions 1 and 2 hold true and define the process* x_{t-1}^{**}

$$x_{t-1}^{**}=\sum_{j=1}^{\infty}\frac{x_{t-j}}{j}=\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-1-j}. \tag{13}$$

Then it follows that the process x_{t-1}^{**} *is square summable,* $\sum_{j=0}^{\infty}\psi_j^2<\infty$.

Proof of Lemma 1

The sequence $\{\psi_j\}_{j\geq 0}$, the convolution of $\left\{(j+1)^{-1}\right\}_{j\geq 0}$ and $\{b_j\}_{j\geq 0}$, is given by

$$\psi_j=\sum_{k=0}^{j}\frac{b_k}{j-k+1}.$$

Consider now the sequence $\{\xi_j\}_{j\geq 0}$, $\xi_j=j|\psi_j|$. Then, if $2\leq k\leq (j+1)/2$, it holds $j/(j-k+1)\leq 2\leq k$. If, on the other hand, $(j+1)/2\leq k\leq j-1$, it follows $j/(j-k+1)\leq j/2<k$. Hence,

$$\frac{j}{j-k+1}\leq k,\ \forall k\in\{1,2,3,\ldots,j\},$$

since the inequality obviously holds for $k=1$ and $k=j$. Then, we can write

$$\xi_j\leq\sum_{k=0}^{j}\frac{j}{j-k+1}|b_k|\leq|b_0|\frac{j}{j+1}+\sum_{k=1}^{j}k|b_k|.$$

But $\{b_j\}$ is assumed to be (at least) one-summable (see Assumption 2), so the sequence $\{\xi_j\}$ converges to a constant as $j\to\infty$, and we thus have

$$\psi_j=O(j^{-1}),\text{ or }\psi_j^2=O(j^{-2}), \tag{14}$$

as needed for the result. ■

Lemma 2 *Consider the regression equation* (6) *and its OLS estimator* $\widehat{\underline{\beta}}_p$, *but use the stationary process* x_{t-1}^{**} *instead of* x_{t-1}^{*} *and denote the corresponding OLS estimator as* $\widetilde{\underline{\beta}}_p$. *Then it follows:*

$$(a)\left\|\widehat{\underline{\beta}}_p-\widetilde{\underline{\beta}}_p\right\|=o_p(1),\ \text{and}\ (b)\sqrt{T}(\widehat{\phi}-\widetilde{\phi})=o_p(1).$$

Proof of Lemma 2

Define $\Delta_{t-1}^{*}=x_{t-1}^{**}-x_{t-1}^{*}$. It holds

$$\Delta_{t-1}^{*}=\sum_{j=t}^{\infty}\frac{x_{t-j}}{j}=\sum_{j=t}^{\infty}\frac{1}{j}\sum_{k\geq 0}b_k\varepsilon_{t-j-k}=\sum_{j=t-1}^{\infty}\widetilde{\psi}_j\varepsilon_{t-1-j}$$

with $\widetilde{\psi}_j=\sum_{k=0}^{j-t+1}\frac{1}{j-k+1}b_k$. Holding t fixed and letting $j\to\infty$, it follows immediately that

$$\widetilde{\psi}_j=O\left(\psi_j\right)=O\left(\frac{1}{j}\right).$$

Thanks to the MDS property of ε_t, we can write

$$\mathbb{E}\left(\Delta_{t-1}^{*}\right)^2=\mathbb{E}\left(\sum_{j=t-1}^{\infty}\widetilde{\psi}_j\varepsilon_{t-1-j}\right)^2=O\left(\sum_{j=t-1}^{\infty}\frac{1}{j^2}\mathbb{E}\left(\varepsilon_{t-j}^2\right)\right)=O\left(\frac{1}{t}\right),$$

or, equivalently, $\sqrt{t}\Delta_{t-1}^{*}=O_p\left(1\right)$.

Denote $V_{tp}=\left(x_{t-1}^{*},x_{t-1},x_{t-2},\ldots,x_{t-p}\right)'$ and $W_{tp}=\left(x_{t-1}^{**},x_{t-1},x_{t-2},\ldots,x_{t-p}\right)'$. The OLS estimator $\widehat{\underline{\beta}}_p$ of $\underline{\beta}_p$ is given by

$$\widehat{\underline{\beta}}_p=\left[\frac{1}{T}\sum V_{tp}V_{tp}'\right]^{-1}\left(\frac{1}{T}\sum V_{tp}x_t\right).$$

Obviously, it holds that

$$\widetilde{\underline{\beta}}_p=\left[\frac{1}{T}\sum W_{tp}W_{tp}'\right]^{-1}\left(\frac{1}{T}\sum W_{tp}x_t\right).$$

For proving (a) note that, for any fixed p, the matrix $\frac{1}{T}\sum W_{tp}W_{tp}'$ has a nonsingular limit (see Lemma 5). Then it suffices to show that

$$\left\|\frac{1}{T}\sum V_{tp}V_{tp}'-\frac{1}{T}\sum W_{tp}W_{tp}'\right\|=o_p\left(p^{-1/2}\right)$$

and

$$\left\| \frac{1}{T} \sum V_{tp} x_t - \frac{1}{T} \sum W_{tp} x_t \right\| = o_p(1),$$

since inverting a matrix and building a matrix norm are continuous operations. One obtains

$$\frac{1}{T} \sum V_{tp} V'_{tp} - \frac{1}{T} \sum W_{tp} W'_{tp} =$$

$$\begin{bmatrix} \frac{1}{T} \sum \left(\left(x^*_{t-1}\right)^2 - \left(x^{**}_{t-1}\right)^2 \right) & \frac{1}{T} \sum \Delta^*_{t-1} x_{t-1} & \cdots & \frac{1}{T} \sum \Delta^*_{t-1} x_{t-p} \\ \frac{1}{T} \sum x_{t-1} \Delta^*_{t-1} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots \vdots & \vdots \\ \frac{1}{T} \sum x_{t-p} \Delta^*_{t-1} & 0 & \cdots & 0 \end{bmatrix}.$$

Further, by applying the Cauchy-Schwarz inequality, one obtains for $h = 1, \ldots, p$

$$\begin{aligned} \left(\frac{1}{T} \sum x_{t-h} \Delta^*_{t-1} \right)^2 &\leq \frac{1}{T^2} \sum x^2_{t-h} \sum (\Delta^*_{t-1})^2 \\ &= \frac{1}{T} \left(\frac{1}{T} \sum x^2_{t-h} \right) \cdot O_p\left(\sum \frac{1}{t} \right) = O_p\left(\frac{\ln T}{T} \right), \end{aligned}$$

due to $\sum_{t=1}^{T} t^{-1} = C + \ln T + O\left(T^{-1}\right)$, where C is Euler 's constant. Similarly, following relationship also holds

$$\frac{1}{T} \sum \left(\left(x^*_{t-1}\right)^2 - \left(x^{**}_{t-1}\right)^2 \right) = o_p\left(p^{-1/2}\right).$$

Since the used matrix norm is bounded by the Euclidean norm, it follows with $p^4/T \to 0$

$$\left\| \frac{1}{T} \sum V_{tp} V'_{tp} - \frac{1}{T} \sum W_{tp} W'_{tp} \right\|^2 \leq o_p\left(p^{-1/2}\right) + 2p\left(O_p\left(\frac{\ln T}{T} \right) \right) = o_p\left(p^{-1/2}\right).$$

To complete the result, note that

$$\frac{1}{T} \sum V_{tp} x_t - \frac{1}{T} \sum W_{tp} x_t = \left(\frac{1}{T} \sum \Delta^*_{t-1} x_t, 0, \ldots, 0 \right)',$$

onto which the Cauchy-Schwarz inequality is again applied as before.
For (b), one needs to show

$$\left\| \widehat{\beta}_p - \widetilde{\beta}_p \right\| = o_p\left(T^{-1/2}\right),$$

or

$$\left\| \sqrt{T}\left(\widehat{\beta}_p - \beta_p \right) - \sqrt{T}\left(\widetilde{\beta}_p - \beta_p \right) \right\| = o_p(1).$$

For this to hold, it suffices

$$\left\| \frac{1}{T} \sum V_{tp} V_{tp}' - \frac{1}{T} \sum W_{tp} W_{tp}' \right\| = o_p \left(p^{-1/2} \right),$$

which holds true from (a), and

$$\left\| \frac{1}{\sqrt{T}} \sum V_{tp} \varepsilon_{tp} - \frac{1}{\sqrt{T}} \sum W_{tp} \varepsilon_{tp} \right\| = o_p \left(1 \right),$$

with ε_{tp} from (7). For the latter to hold, one needs to show that

$$\frac{1}{\sqrt{T}} \sum \Delta_{t-1}^* \varepsilon_{tp} \xrightarrow{p} 0.$$

We know from Chang and Park (2002, p. 434) that $\varepsilon_{tp} - \varepsilon_t = o_p \left(p^{-s} \right)$, so it can be checked that

$$\frac{1}{\sqrt{T}} \sum \Delta_{t-1}^* \varepsilon_t - \frac{1}{\sqrt{T}} \sum \Delta_{t-1}^* \varepsilon_{tp} = o_p \left(1 \right).$$

Finally,

$$\mathbb{E} \left(\frac{1}{\sqrt{T}} \sum \Delta_{t-1}^* \varepsilon_t \right)^2 = \frac{1}{T} \sum_i \sum_j \mathbb{E} \left(\Delta_{i-1}^* \varepsilon_i \Delta_{j-1}^* \varepsilon_j \right).$$

By applying the law of iterated expectations [LIE] and recalling that ε_t possesses MDS property, only terms with $i = j$ have non-zero expected value. Therefore,

$$\mathbb{E} \left(\frac{1}{\sqrt{T}} \sum \Delta_{t-1}^* \varepsilon_t \right)^2 = \frac{1}{T} \sum \mathbb{E} \left(\left(\Delta_{t-1}^* \right)^2 \varepsilon_t^2 \right).$$

Note that, since the unconditional variance of ε_t is finite, its conditional variance is bounded in probability. Then, by applying LIE again, one obtains, as needed for the result,

$$\mathbb{E} \left(\frac{1}{\sqrt{T}} \sum \Delta_{t-1}^* \varepsilon_t \right)^2 = O_p \left(\frac{1}{T} \sum \frac{1}{t} \right) = O_p \left(\frac{\ln T}{T} \right).$$

■

Lemma 3 *For absolutely summable sequences* $\{d_j\}_{j \geq 0}$*, it holds* $d_j = o(j^{-1})$*. For one summable sequences, it holds* $d_j = o(j^{-2})$*. Finally, for absolutely summable multidimensional sequences,* $\sum_{i_1=-\infty}^{\infty} \cdots \sum_{i_k=-\infty}^{\infty} |d_{i_1 \ldots i_k}| < \infty$*, it holds* $|d_{i_1 \ldots i_k}| = o(|i_1|^{-1} \cdot \ldots \cdot |i_k|^{-1})$*.*

Proof of Lemma 3

Assume $d_j = O\left(j^{-1}\right)$ or has an even lower convergence rate. Then, $\sum_{j\geq 0} d_j$ diverges, so $\{d_j\}_{j\geq 0}$ must be $o\left(j^{-1}\right)$. The proof proceeds similarly for one summable and multidimensional sequences. ■

Lemma 4 *The cross-covariances $\gamma_h^{*x} = \mathbb{E}(x_{t-1}^{**} x_{t-1-h})$ and the autocovariances $\gamma_h^{*} = \mathbb{E}(x_{t-1}^{**} x_{t-1-h}^{**})$ are square summable, $\sum_{h=0}^{\infty} (\gamma_h^{*x})^2 < \infty$, as well as $\sum_{h=0}^{\infty} (\gamma_h^{*})^2 < \infty$.*

Proof of Lemma 4

It holds

$$\begin{aligned}\gamma_h^{*x} &= \mathbb{E}\left(x_{t-1}^{**} x_{t-1-h}\right) = \mathbb{E}\left(\left(\sum_{i\geq 0} b_i \varepsilon_{t-1-h-i}\right)\left(\sum_{j\geq 0} \psi_j \varepsilon_{t-1-j}\right)\right) \\ &= \sum_{i\geq 0}\sum_{j\geq 0} b_i \psi_j \mathbb{E}\left(\varepsilon_{t-1-h-i}\varepsilon_{t-1-j}\right)\end{aligned}$$

Because of MDS property of ε_t, only terms for which $t-h-i = t-1-j$ have non-zero expected value, so

$$\gamma_h^{*x} = \gamma_0^{\varepsilon} \sum_{i\geq 0} b_i \psi_{i+h}.$$

Further, $b_i = o\left(i^{-1}\right)$ due to Lemma 3 and $\psi_{i+h} = O\left((i+h)^{-1}\right)$, see (14). Then,

$$\gamma_h^{*x} = o\left(\sum_{i\geq 1} \frac{1}{i\,(i+h)}\right).$$

With

$$\sum_{i\geq 1} \frac{1}{i\,(i+h)} = \frac{1}{h}\sum_{i\geq 1}\left(\frac{1}{i} - \frac{1}{i+h}\right)$$

it follows

$$\gamma_h^{*x} = o\left(\frac{\ln h}{h}\right),$$

which leads to the desired result. The proof runs similarly for the autocovariance sequence γ_h^{*}. ■

Lemma 5 *Consider the vector* $W_{tp} = (x_{t-1}^{**}, x_{t-1}, \ldots, x_{t-p})'$ *and define the variance-covariance matrix* $\Sigma_p = \mathbb{E}\left(W_{tp}W_{tp}'\right)$. *Then,* Σ_p *is non-singular and* $\|\Sigma_p\|$, *as well as* $\left\|\Sigma_p^{-1}\right\|$ *are bounded as* $p \to \infty$.

Proof of Lemma 5

Note that Σ_p is singular if and only if the components of W_{tp} are linearly dependent, which is not the case. Then, we may obviously write

$$\Sigma_p = \begin{bmatrix} \gamma_0^* & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \Sigma_{*p}' \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \Sigma_{*p} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{xp} \end{bmatrix},$$

where $\Sigma_{*p} = \left(\gamma_0^{*x}, \ldots, \gamma_{p-1}^{*x}\right)'$ and Σ_{xp} is the $p \times p$ autocovariance matrix of x_t. Then, by triangle inequality, we obtain

$$\|\Sigma_p\| \leq \left\| \begin{matrix} \gamma_0^* & 0 \\ 0 & 0 \end{matrix} \right\| + \left\| \begin{matrix} 0 & \Sigma_{*p}' \\ 0 & 0 \end{matrix} \right\| + \left\| \begin{matrix} 0 & 0 \\ \Sigma_{*p} & 0 \end{matrix} \right\| + \left\| \begin{matrix} 0 & 0 \\ 0 & \Sigma_{xp} \end{matrix} \right\|.$$

From the definition of the norm if follows that the norm of a matrix B obtained by bordering a matrix A with zeros equals the norm of A, so

$$\|\Sigma_p\| \leq |\gamma_0^*| + \left\|\Sigma_{*p}'\right\| + \|\Sigma_{*p}\| + \|\Sigma_{xp}\|.$$

The Euclidean norm of Σ_{*p} converges to a constant as $p \to \infty$ due to square summability of γ_h^{*x}, whereas the norm $\|\Sigma_{xp}\|$ is bounded (Berk, 1974, p. 493). The inverse can be computed (Lütkepohl, 1996, p. 148) from

$$\begin{bmatrix} \gamma_0^* & \Sigma_{*p}' \\ \Sigma_{*p} & \Sigma_{xp} \end{bmatrix}^{-1} = \begin{bmatrix} E & G' \\ G & H \end{bmatrix},$$

with

$$\begin{aligned} E &= \left(\gamma_0^* - \Sigma_{*p}'\Sigma_{xp}^{-1}\Sigma_{*p}\right)^{-1} \\ G &= -\Sigma_{xp}^{-1}\Sigma_{*p}\left(\gamma_0^* - \Sigma_{*p}'\Sigma_{xp}^{-1}\Sigma_{*p}\right)^{-1} \\ H &= \Sigma_{xp}^{-1} + \Sigma_{xp}^{-1}\Sigma_{*p}\left(\gamma_0^* - \Sigma_{*p}'\Sigma_{xp}^{-1}\Sigma_{*p}\right)^{-1}\Sigma_{*p}'\Sigma_{xp}^{-1}. \end{aligned}$$

Then,

$$\left\|\Sigma_p^{-1}\right\| \leq \|E\| + \|G\| + \left\|G'\right\| + \|H\|$$

and the desired result follows from

$$\begin{aligned} \|E\| &= \frac{1}{\sigma_{\omega p}^2} < \infty \\ \|G\| &\leq \left\|\Sigma_{xp}^{-1}\right\| \|\Sigma_{*p}\| \|E\| < \infty \\ \|H\| &\leq \left\|\Sigma_{xp}^{-1}\right\| + \left\|\Sigma_{xp}^{-1}\right\| \|\Sigma_{*p}\| \|E\| \left\|\Sigma_{*p}'\right\| \left\|\Sigma_{xp}^{-1}\right\| < \infty, \end{aligned}$$

since $\left\|\Sigma_{xp}^{-1}\right\| < \infty$ (Berk, 1974, p. 493), and $\sigma_{\omega p}^2$ (defined in Lemma 6) converges to a positive constant. ■

Lemma 6 *The sequence* $\sigma_{\omega p}^2 = \gamma_0^* - \Sigma_{*p}' \Sigma_{xp}^{-1} \Sigma_{*p}$ *converges to a positive constant as* $p \to \infty$*.*

Proof of Lemma 6

Begin by expressing $\sigma_{\omega p}^2$ as a series. To this end, one can use the decomposition

$$\Sigma_{xp}^{-1} = L_p' D_p^{-1} L_p,$$

given in Berk (1974, p. 491) with

$$L_p = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\widetilde{a}_{11} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\widetilde{a}_{p-1,p-1} & -\widetilde{a}_{p-2,p-1} & \cdots & 1 \end{bmatrix},$$

where, for given $j \in \{1, \ldots, p\}$, $\widetilde{a}_j = (\widetilde{a}_{1j}, \ldots, \widetilde{a}_{jj})'$ represent the solutions to the minimum problem

$$\min_{(c_1,\ldots,c_j)' \in \mathbb{R}^j} \left\{ \mathbb{E} \left(x_t - c_1 x_{t-1} - c_2 x_{t-2} - \ldots - c_j x_{t-j} \right)^2 \right\},$$

or, equivalently, the values the OLS estimators $\widehat{c}_j = (\widehat{c}_{1j}, \ldots, \widehat{c}_{jj})'$

$$\widehat{c}_j = \underset{(c_1,\ldots,c_j)' \in \mathbb{R}^j}{\arg\min} \left\{ \frac{1}{T-j} \sum_{t=j+1}^{T} \left(x_t - c_1 x_{t-1} - c_2 x_{t-2} - \ldots - c_j x_{t-j} \right)^2 \right\}$$

are consistent for as $T \to \infty$ and j fixed. Since x_t is ergodic with finite variance, the necessary conditions for Theorem 4.1.1 from Amemyia (1985), theorem which ensures the claimed consistency, can be shown to be met. The matrix D_p has diagonal structure, $D_p = \operatorname{diag}(\widetilde{\gamma}_0, \ldots, \widetilde{\gamma}_{p-1})$ where $\widetilde{\gamma}_j$ is the respective minimum,

$$\widetilde{\gamma}_j = \min_{(c_1,\ldots,c_j)' \in \mathbb{R}^j} \left\{ \mathbb{E} \left(x_t - c_1 x_{t-1} - c_2 x_{t-2} - \ldots - c_j x_{t-j} \right)^2 \right\}.$$

Obviously, $\widetilde{\gamma}_p \to \sigma^2$ and $\widetilde{a}_{i,p-1} \to a_i$ as $p \to \infty$. Then,

$$\Sigma_{*p}' \Sigma_{xp}^{-1} \Sigma_{*p} = \left(L_p \Sigma_{*p} \right)' D_p^{-1} L_p \Sigma_{*p}$$

and

$$(L_p\Sigma_{*p})' = \left(\widetilde{l}_{1p}, \ldots, \widetilde{l}_{pp}\right) \text{ with } \widetilde{l}_{jp} = \gamma_j^* + \sum_{i=1}^{j-1} \gamma_{j-i}^* \widetilde{a}_{ij}.$$

It follows,

$$\sigma_{\omega p}^2 = \gamma_0^* - \sum_{j=1}^{p} \widetilde{l}_{jp}^2 \widetilde{\gamma}_{j-1}^{-1}. \tag{15}$$

Note that $\sigma_{\omega p}^2$ is bounded, $\left|\sigma_{\omega p}^2\right| \leq |\gamma_0^*| + \left\|\Sigma'_{*p}\right\| \left\|\Sigma_{xp}^{-1}\right\| \|\Sigma_{*p}\| < \infty$. Hence, a sufficient condition for convergence in (15) is $\widetilde{l}_{jp}\widetilde{\gamma}_{j-1}^{-1} \to 0$ as $p \to \infty$. Since $\widetilde{\gamma}_j$ converges to a nonzero constant, the sequence $\widetilde{\gamma}_j^{-1}$ is bounded and this condition reduces to $\widetilde{l}_{jp} \to 0$. Standard OLS algebra leads to

$$(\widehat{c}_{ij} - a_i)_{1\leq i\leq j} = \left[\frac{1}{T-j}\sum X_{tj}X'_{tj}\right]^{-1} \left(\frac{1}{T-j}\sum X_{tj}\varepsilon_{tj}\right),$$

with $X_{tp} = (x_{t-1}, x_{t-2}, \ldots, x_{t-p})'$. Since x_t is ergodic, it holds that

$$\frac{1}{T-j}\sum X_{tj}X'_{tj} \xrightarrow{p} \Sigma_{xj}$$

and

$$\frac{1}{T-j}\sum X_{tj}\varepsilon_{tj} \xrightarrow{p} \left(\sum_{k\geq j+1} a_k \gamma_{k-i}\right)_{1\leq i\leq j}.$$

For any i,

$$\sum_{k\geq j+1} a_k\gamma_{k-i} = o\left(\sum_{k\geq j+1} \frac{1}{k^2}\frac{1}{k-i}\right) = o\left(\frac{1}{ij}\sum_{k=1}^{i}\frac{1}{j-k}\right),$$

or

$$\sum_{k\geq j+1} a_k\gamma_{k-i} = o\left(\frac{\ln j}{j}\right).$$

The norm of Σ_{xj}^{-1} is bounded, so $\widetilde{a}_{ij} - a_i = o\left(\frac{\ln j}{j}\right)$. Since

$$\widetilde{l}_{jp} = \gamma_j^* + \sum_{i=1}^{j-1}\gamma_{j-i}^*\widetilde{a}_{ij} = \gamma_j^* + \sum_{i=1}^{j-1}\gamma_{j-i}^* a_i + (\widetilde{a}_{ij} - a_i)\sum_{i=1}^{j-1}\gamma_{j-i}^*,$$

and $a_i = o\left(i^{-2}\right) = o\left(i^{-1}\right)$ we have

$$\begin{aligned}\sum_{i=1}^{j-1} \gamma_{j-i}^{*} a_i &= o\left(\sum_{i=1}^{j-1} \frac{\ln(j-i)}{j-i}\frac{1}{i}\right) = o\left(\frac{\ln j}{j}\sum_{i=1}^{j-1}\left(\frac{1}{i}+\frac{1}{j-i}\right)\right)\\ &= o\left(\frac{2\ln j}{j}\sum_{i=1}^{j-1}\frac{1}{i}\right) = o\left(\frac{\ln^2 j}{j}\right)\end{aligned}$$

as well as

$$\sum_{i=1}^{j-1} \gamma_{j-i}^{*} = o\left(\sum_{i=1}^{j-1}\frac{\ln i}{i}\right) = o\left(\ln^2 j\right),$$

so it holds $\widetilde{l}_{jp} = o(1)$, as needed for the convergence result. Being the diagonal element of a positive definite matrix, Σ_p^{-1}, the limit of $\sigma_{\omega p}^2$ must be positive. ■

Lemma 7 *The process $u_{tp} = \Sigma_{*p}' \Sigma_{xp}^{-1} X_{tp}$, with $X_{tp} = (x_{t-1}, x_{t-2}, \ldots, x_{t-p})'$, converges in probability as $p \to \infty$ to a square summable process, $u_t = \sum_{j=0}^{\infty} \zeta_j \varepsilon_{t-1-j}$.*

Proof of Lemma 7

Represent X_{tp} as a countable linear combination of the MDS innovations ε_{t-1-j}, $j \geq 0$:

$$X_{tp} = B_p \epsilon_{t-1},$$

where B_p and ϵ_{t-1} have infinitely many columns, and lines, respectively:

$$B_p = \begin{bmatrix} b_0 & b_1 & \cdots & b_{p-1} & b_p & \cdots \\ 0 & b_0 & \cdots & b_{p-2} & b_{p-1} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ 0 & 0 & \cdots & b_0 & b_1 & \cdots \end{bmatrix}$$

$$\epsilon_{t-1} = (\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)'.$$

Note that, since the sequence $\{b_j\}_{j\geq 0}$ is s-summable, the norm of B_p is bounded as $p \to \infty$. Then, the process u_{tp} is nothing more than a linear process with innovations ε_{t-1-j}, $j \geq 0$:

$$u_{tp} = \sum_{j\geq 0} \widetilde{\zeta}_{jp} \varepsilon_{t-1-j}, \tag{16}$$

with coefficients that can be obtained by calculating the multipication $\Sigma_{*p}' \Sigma_{xp}^{-1} B_p$. This is just another way of expressing the convolution of the filters $\Sigma_{*p}' \Sigma_{xp}^{-1}$, which is a $MA(p)$, and $\{b_j\}$, which is $MA(\infty)$. Denote u_t the asymptotic counterpart of u_{tp},

$$u_t = \sum_{j \geq 0} \zeta_j \varepsilon_{t-1-j}, \tag{17}$$

where the coefficients ζ_j are obtained as a limit of the convolution $\Sigma'_{*p}\Sigma_{xp}^{-1}B_p$ for $p \to \infty$. The coefficients ζ_j are square summable, since

$$\sqrt{\sum_{j \geq 0} \zeta_j^2} \leq \sup_{p \in \mathbb{N}} \left\{ \left\| \Sigma'_{*p} \right\| \left\| \Sigma_{xp}^{-1} \right\| \left\| B_p \right\| \right\} < \infty,$$

the matrices having bounded norms as $p \to \infty$ (see Lemma 5 and the properties of B_p). Consider now the difference of the processes given by (16) and (17):

$$u_t - u_{tp} = \sum_{j \geq 0} \zeta_j \varepsilon_{t-1-j} - \sum_{j \geq 0} \widetilde{\zeta}_{jp} \varepsilon_{t-1-j} = \sum_{j \geq 0} \left(\zeta_j - \widetilde{\zeta}_{jp} \right) \varepsilon_{t-1-j},$$

with second moment

$$\mathbb{E}\left(u_t - u_{tp}\right)^2 = \sigma^2 \sum_{j \geq 0} \left(\zeta_j - \widetilde{\zeta}_{jp} \right)^2.$$

From the definition of u_t, it follows

$$\lim_{p \to \infty} \sum_{j \geq 0} \widetilde{\zeta}_{jp} = \sum_{j \geq 0} \zeta_j, \text{ or } \lim_{p \to \infty} \sum_{j \geq 0} \left(\widetilde{\zeta}_{jp} - \zeta_j \right) = 0.$$

This leads to

$$\lim_{p \to \infty} \sum_{j \geq 0} \left(\widetilde{\zeta}_{jp} - \zeta_j \right)^2 = 0, \text{ or } \sum_{j \geq 0} \left(\widetilde{\zeta}_{jp} - \zeta_j \right)^2 = o(1)$$

as $p \to \infty$, from which the result follows. ■

Lemma 8 *Assume the 4^{th} cumulants of the innovations ε_t are absolutely summable in the sense that*

$$\sum_{t_1, t_2, t_3 = -\infty}^{\infty} \left| \kappa_\varepsilon(0, t_1, t_2, t_3) \right| < \infty.$$

Then, the 4^{th} cumulants of a linear process y_t, $y_t = \sum_{j\geq 0} \psi_j \varepsilon_{t-j}$, with $\psi_0 = 1$ and $|\psi_j| = O\left(\frac{1}{j+1}\right)$ (a) are finite, and (b) satisfy following the summability condition:

$$\sum_{t=-\infty}^{\infty} |\kappa_y(0, s, t, t+s)| < \infty \quad \forall s \in \mathbb{Z}.$$

Proof of Lemma 8

(a) From Lemma 3, we know that

$$|\kappa_\varepsilon(0, t_1, t_2, t_3)| = o\left(\frac{1}{(|t_1|+1)(|t_2|+1)(|t_3|+1)}\right).$$

It is straightforward to show that

$$|\kappa_y(0, t_1, t_2, t_3)| \leq \sum_{i,j,k,l=0} |\psi_i \psi_j \psi_k \psi_l| \, |\kappa_\varepsilon(0, t_1 + j - i, t_2 + k - i, t_3 + l - i)|$$

Then, $|\kappa_y(0, t_1, t_2, t_3)|$ is bounded by

$$C \sum_{i,j,k,l=0}^{\infty} \frac{1}{(i+1)(j+1)(k+1)(l+1)} \cdot \frac{1}{(|t_1 + j - i| + 1)(|t_2 + k - i| + 1)(|t_3 + l - i| + 1)},$$

where C is a positive constant. To see how this sum behaves, define $p = j - i$, $q = k - i$, $r = l - i$. Since t_i are fixed, it follows that an upper bound for $|\kappa_y(0, t_1, t_2, t_3)|$ is

$$C \sum_{i=0}^{\infty} \sum_{p=-i}^{\infty} \sum_{q=-i}^{\infty} \sum_{r=-i}^{\infty} \frac{1}{(i+1)(p+i+1)(q+i+1)(r+i+1)} \cdot \frac{1}{(|p|+1)(|q|+1)(|r|+1)}.$$

Therefore, the cumulants are finite if this sum is. In turn, this sum equals

$$\lim_{n_1, n_2, n_3, n_4 \to \infty} \sum_{i=0}^{n_1} \sum_{p=-i}^{n_2} \sum_{q=-i}^{n_3} \sum_{r=-i}^{n_4} \frac{1}{i+1} \frac{1}{(p+i+1)(|p|+1)} \cdot \frac{1}{(q+i+1)(|q|+1)} \frac{1}{(r+i+1)(|r|+1)},$$

or, since all its elements are positive,

$$\lim_{n_1,n_2,n_3,n_4\to\infty}\sum_{i=0}^{n_1}\left(\frac{1}{i+1}\left(\sum_{p=-i}^{n_2}\frac{1}{(p+i+1)}\frac{1}{(|p|+1)}\right)\cdot\right.$$

$$\left.\cdot\left(\sum_{q=-i}^{n_3}\frac{1}{(q+i+1)}\frac{1}{(|q|+1)}\right)\left(\sum_{r=-i}^{n_4}\frac{1}{(r+i+1)}\frac{1}{(|r|+1)}\right)\right).$$

Further, the sum indexed by p, $\sum_{p=-i}^{n_2}\frac{1}{(p+i+1)}\frac{1}{(|p|+1)}$, equals

$$\sum_{p=-i}^{0}\frac{1}{(p+i+1)}\frac{1}{(-p+1)}+\sum_{p=1}^{n_2}\frac{1}{(p+i+1)}\frac{1}{(p+1)},$$

or

$$\frac{1}{i+2}\sum_{p=-i}^{0}\left(\frac{1}{(p+i+1)}+\frac{1}{(-p+1)}\right)+\frac{1}{i}\sum_{p=1}^{n_2}\left(\frac{1}{(p+1)}-\frac{1}{(p+i+1)}\right),$$

where it obviously holds

$$\sum_{p=1}^{n_2}\left(\frac{1}{(p+1)}-\frac{1}{(p+i+1)}\right)<\sum_{p=1}^{i}\frac{1}{p+1}.$$

Thus, following upper bound results:

$$\sum_{p=-i}^{n_2}\frac{1}{(p+i+1)}\frac{1}{(|p|+1)}<\frac{3}{i}\sum_{p=0}^{i}\frac{1}{p+1}.$$

Similar relationships hold for the sums indexed by q and r. Hence

$$\sum_{i=0}^{n_1}\left(\frac{1}{i+1}\left(\sum_{p=-i}^{n_2}\frac{1}{(p+i+1)}\frac{1}{(|p|+1)}\right)\left(\sum_{q=-i}^{n_3}\frac{1}{(q+i+1)}\frac{1}{(|q|+1)}\right)\cdot\right.$$

$$\left.\cdot\left(\sum_{r=-i}^{n_4}\frac{1}{(r+i+1)}\frac{1}{(|r|+1)}\right)\right)=O\left(\sum_{i=0}^{n_1}\frac{\ln^3(i+1)}{(i+1)^4}\right)<\infty.$$

For (b), observe that $|\kappa_y(0,s,t,t+s)|$ is bounded by

$$C\sum_{i,j,k,l=0}^{\infty}\frac{1}{(i+1)(j+1)(k+1)(l+1)}\cdot$$

$$\cdot\frac{1}{(|s+j-i|+1)(|t+k-i|+1)(|t+s+l-i|+1)}.$$

Bearing in mind that s is finite, we can conclude that this expression is finite, if the sum

$$\sum_{i,j,k,l=0}^{\infty} \frac{1}{(i+1)(j+1)(k+1)(l+1)} \cdot \frac{1}{(|j-i|+1)(|t+k-i|+1)(|t+l-i|+1)}$$

is. Again, obtain by reindexing

$$\sum_{i=0}^{\infty} \sum_{p=-i}^{\infty} \sum_{q=-i}^{\infty} \sum_{r=-i}^{\infty} \qquad \frac{1}{(i+1)(p+i+1)(q+i+1)(r+i+1)} \cdot$$
$$\cdot \frac{1}{(|p|+1)(|t+q|+1)(|t+r|+1)},$$

which, written as a limit, is

$$\lim_{n_1,n_2,n_3,n_4 \to \infty} \sum_{i=0}^{n_1} \sum_{p=-i}^{n_2} \sum_{q=-i}^{n_3} \sum_{r=-i}^{n_4} \qquad \frac{1}{(i+1)} \frac{1}{(p+i+1)(|p|+1)} \cdot$$
$$\cdot \frac{1}{(q+i+1)(|t+q|+1)} \frac{1}{(r+i+1)(|t+r|+1)},$$

or

$$\lim_{n_1,n_2,n_3,n_4 \to \infty} \sum_{i=0}^{n_1} \left(\frac{1}{(i+1)} \left(\sum_{p=-i}^{n_2} \frac{1}{(p+i+1)} \frac{1}{(|p|+1)} \right) \cdot \right.$$
$$\left. \cdot \left(\sum_{q=-i}^{n_3} \frac{1}{(q+i+1)} \frac{1}{(|t+q|+1)} \right) \left(\sum_{r=-i}^{n_4} \frac{1}{(r+i+1)} \frac{1}{(|t+r|+1)} \right) \right).$$

Recall from (a) that

$$\sum_{p=-i}^{n_2} \frac{1}{(p+i+1)} \frac{1}{(|p|+1)} < \frac{3}{i} \sum_{p=0}^{i} \frac{1}{p+1}.$$

Now, we need to find an upper bound for the sums indexed by q and r: We need to distingush three cases, $t < i$, $t > i$ and $t = i$. In the first case, the sum indexed by q, $\sum_{q=-i}^{n_3} \frac{1}{(q+i+1)} \frac{1}{(|t+q|+1)}$, equals

$$\sum_{q=-i}^{-t} \frac{1}{(q+i+1)} \frac{1}{(-t-q+1)} + \sum_{q=-t+1}^{n_3} \frac{1}{(q+i+1)} \frac{1}{(t+q+1)},$$

or

$$\frac{1}{i-t+2}\sum_{q=-i}^{-t}\left(\frac{1}{(q+i+1)}+\frac{1}{(-t-q+1)}\right)$$
$$+\frac{1}{i-t}\sum_{q=-t+1}^{n_3}\left(\frac{1}{(t+q+1)}-\frac{1}{(q+i+1)}\right),$$

which is bounded by

$$\frac{3}{i-t}\sum_{q=0}^{i-t}\frac{1}{q+1}.$$

In the second case,

$$\begin{aligned}\sum_{q=-i}^{n_3}\frac{1}{(q+i+1)}\frac{1}{(|t+q|+1)} &= \frac{1}{t-i}\sum_{q=-i}^{n_3}\left(\frac{1}{(q+i+1)}-\frac{1}{(t+q+1)}\right)\\ &< \frac{1}{t-i}\sum_{q=0}^{t-i}\frac{1}{q+1}\end{aligned}$$

With $t=i$, we are talking about a single cumulant, which is obviously finite due to (a) and only influences the value of the sum, but not its finiteness. Therefore, it suffices to show that

$$\sum_{t=-\infty}^{\infty}\sum_{i=0}^{\infty}\frac{\ln(i+1)}{(i+1)^2}\frac{\ln^2(|t-i|+1)}{(|t-i|+1)^2}<\infty.$$

To this purpose, let $p=t-i$ and obtain

$$\sum_{p=-\infty}^{\infty}\sum_{i=0}^{\infty}\frac{\ln(i+1)}{(i+1)^2}\frac{\ln^2(|p|+1)}{(|p|+1)^2}=\left(\sum_{i=0}^{\infty}\frac{\ln(i+1)}{(i+1)^2}\right)\left(\sum_{p=-\infty}^{\infty}\frac{\ln^2(|p|+1)}{(|p|+1)^2}\right),$$

from which the desired result follows. ■

Lemma 9 *The expression* $\left\|\mathbb{E}(X_{tp}x^{**}_{t-1}\varepsilon_t^2)\right\|$ *is bounded for all* p.

Proof of Lemma 9

Gonçalves and Kilian (2003) show that $\mathbb{E}\left(X_{tp}X'_{tp}\varepsilon_t^2\right)$ is bounded for all p and we use the same arguments here. Consider $\mathbb{E}\left(X_{tp}x_{t-1}^{**}\varepsilon_t^2\right)$ and denote $\underline{b}_{j,p} = (b_{j-1},\ldots,b_{j-p})'$ so that

$$X_{tp} = \sum_{j=1}^{\infty} \underline{b}_{j,p}\varepsilon_{t-j}$$

and

$$\mathbb{E}\left(X_{tp}x_{t-1}^{**}\varepsilon_t^2\right) = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \underline{b}_{j,p}\psi_i \mathbb{E}(\varepsilon_{t-j}\varepsilon_{t-i}\varepsilon_t^2),$$

where $\mathbb{E}(\varepsilon_{t-j}\varepsilon_{t-i}\varepsilon_t^2) = \kappa_\varepsilon(0,-i,-j,0)$ for $i \neq j$ as well as $\mathbb{E}(\varepsilon_{t-j}\varepsilon_{t-i}\varepsilon_t^2) = \sigma^4 + \kappa_\varepsilon(0,-i,-j,0)$ for $i = j$. We can also write $\Sigma_{*p} = \sigma^2 \sum_{j=1}^{\infty} \underline{b}_{j,p}\psi_j$ and we obtain

$$\left\|\mathbb{E}\left(X_{tp}x_{t-1}^{**}\varepsilon_t^2\right)\right\| \leq \sigma^2 \left\|\Sigma_{*p}\right\| + \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \left\|\underline{b}_{j,p}\psi_i\right\| |\kappa_\varepsilon(0,-i,-j,0)|.$$

$\|\Sigma_{*p}\|$ is bounded by Lemma 4 and $\left\|\underline{b}_{j,p}\psi_i\right\|$ is bounded due to the absolute summability of x_t and finally the assumption of absolutely summable 8th cumulants leads to the desired result. ■

Lemma 10 *The sum*

$$\sum_{\tau}\sum_{i,j,k,l} \psi_i\psi_j\psi_k\psi_l\kappa_\varepsilon(0, i-j, 1+i, 1+i, \tau-k+i, \tau-l+i, \tau+1+i, \tau+1+i)$$

is bounded.

Proof of Lemma 10

Observe that, due to the stationarity of ε_t, the sum can be expressed as

$$\sum_{\tau}\sum_{i,j,k,l} \psi_i\psi_j\psi_k\psi_l\kappa_\varepsilon\left(-i,-j,1,1,\tau-k,\tau-l,\tau+1,\tau+1\right),$$

where

$$|\kappa_\varepsilon(-i,-j,1,1,\tau-k,\tau-l,\tau+1,\tau+1)| = \\ = o\left(\frac{1}{(i+1)(j+1)(|\tau-k|+1)(|\tau-l|+1)(|\tau|+1)^2}\right)$$

because of Lemmas 1 and 3. And the sum can be bounded as follows

$$\left|\sum_{i,j,k,l} \psi_i \psi_j \psi_k \psi_l \kappa_\varepsilon(-i,-j,1,1,\tau-k,\tau-l,\tau+1,\tau+1)\right| \leq$$

$$\leq C \left(\sum_i \frac{1}{(i+1)^2}\right) \left(\sum_j \frac{1}{(j+1)^2}\right) \left(\sum_k \frac{1}{k+1} \cdot \frac{1}{|\tau-k|+1}\right) \cdot$$

$$\cdot \left(\sum_l \frac{1}{l+1} \cdot \frac{1}{|\tau-l|+1}\right) \frac{1}{(|\tau|+1)^2}.$$

Now, consider

$$\sum_{k=0}^{\infty} \frac{1}{k+1} \cdot \frac{1}{|\tau-k|+1}, \tag{18}$$

and, for convenience, distinguish three cases:

(a) $\tau < 0$, then we have

$$\sum_{k=0}^{\infty} \frac{1}{k+1} \frac{1}{|\tau|+k+1} = \frac{1}{|\tau|} \sum_{k=0}^{\infty} \left(\frac{1}{k+1} - \frac{1}{|\tau|+k+1}\right) = \frac{1}{|\tau|} \sum_{k=1}^{|\tau|} \frac{1}{k} = \frac{\ln(|\tau|)}{|\tau|};$$

(b) $\tau > 0$, then (18) can be written as

$$\sum_{k=0}^{|\tau|} \frac{1}{k+1} \cdot \frac{1}{|\tau|-k+1} + \sum_{k=|\tau|+1}^{\infty} \frac{1}{k+1} \cdot \frac{1}{k-|\tau|+1}$$

$$= \frac{1}{|\tau|+2} \sum_{k=0}^{|\tau|} \left(\frac{1}{k+1} + \frac{1}{|\tau|-k+1}\right) + \frac{1}{|\tau|} \sum_{k=|\tau|+1}^{\infty} \left(\frac{1}{k-|\tau|+1} - \frac{1}{k+1}\right)$$

$$\leq \frac{2}{|\tau|} \sum_{k=1}^{|\tau|+1} \frac{1}{k} + \frac{1}{|\tau|} \sum_{k=1}^{|\tau|+1} \frac{1}{k} = \frac{3}{|\tau|} \sum_{k=1}^{|\tau|+1} \frac{1}{k} = \frac{3\ln(|\tau|+1)}{|\tau|};$$

(c) $\tau = 0$, then $\sum_{k=0}^{\infty}(k+1)^{-2} < \infty$.

And it follows

$$\sum_{\tau=-\infty}^{\infty} \left(\sum_{k=0}^{\infty} \frac{1}{k+1} \cdot \frac{1}{|\tau-k|+1}\right) \left(\sum_{l=0}^{\infty} \frac{1}{l+1} \cdot \frac{1}{|\tau-l|+1}\right) \frac{1}{(|\tau|+1)^2}$$

$$= \sum_{\tau=-\infty}^{-1} \left(\frac{\ln(|\tau|)}{|\tau|}\right)^2 \frac{1}{(|\tau|+1)^2} + \sum_{\tau=1}^{\infty} \left(\frac{3\ln(|\tau|+1)}{|\tau|}\right)^2 \frac{1}{(|\tau|+1)^2} + K < \infty.$$

■

B Propositions

Proof of Proposition 1

From Lemma 2 we know that there is no difference between the variable x^*_{t-1} and its stationary counterpart x^{**}_{t-1} asymptotically, so we may assume x^{**}_{t-1} to be observable and investigate the consistency of the corresponding OLS estimator. Let $W_{tp} = (x^{**}_{t-1}, x_{t-1}, \ldots, x_{t-p})$ and we get the OLS estimator as follows

$$\widetilde{\underline{\beta}}_p = \left[\frac{1}{T} \sum_{t=p+1}^{T} W_{tp} W'_{tp} \right]^{-1} \left(\frac{1}{T} \sum_{t=p+1}^{T} W_{tp} x_t \right)$$

and accordingly

$$\widetilde{\underline{\beta}}_p - \underline{\beta}_p = \left[\frac{1}{T} \sum_{t=p+1}^{T} W_{tp} W'_{tp} \right]^{-1} \left(\frac{1}{T} \sum_{t=p+1}^{T} W_{tp} \varepsilon_{tp} \right)$$

where $\varepsilon_{tp} = \varepsilon_t - \sum_{k=p+1}^{\infty} a_k x_{t-k}$. Denote $\widehat{\Sigma}_p = T^{-1} \sum_{t=p+1}^{T} W_{tp} W'_{tp}$ and Σ_p the corresponding covariance matrix, where $\gamma^x_k = \mathbb{E}(x_t x_{t-k})$, $\gamma^*_0 = \mathbb{E}(x^{**}_{t-1})^2$ and $\gamma^{*x}_k = \mathbb{E}(x^{**}_{t-1} x_{t-1-k})$ so that

$$\Sigma_p = \begin{bmatrix} \gamma^*_0 & \gamma^{*x}_0 & \cdots & \gamma^{*x}_{p-1} \\ \gamma^{*x}_0 & \gamma_0 & \cdots & \gamma_{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma^{*x}_{p-1} & \gamma_{p-1} & \cdots & \gamma_0 \end{bmatrix} = \begin{bmatrix} \gamma^*_0 & \Sigma'_{*p} \\ \Sigma_{*p} & \Sigma_{xp} \end{bmatrix}.$$

Recall, $\|\Sigma_p\|$, $\|\Sigma_p^{-1}\|$ are bounded due to Lemma 5. We assumed that $p = o(T^{1/4})$, or $p^4/T \to 0$, and adopt Lemma 3 from Berk (1974, p. 493) and auxiliary Lemma A.2 from Gonçalves and Kilian (2003, p. 14) in our case to show that

$$\left\| \widehat{\Sigma}_p^{-1} - \Sigma_p^{-1} \right\| = o_p(p^{-1}). \tag{19}$$

By the use of spectral matrix norm we know that $\|A\|^2 \leq \sum_{i,j} a^2_{i,j}$ for any real matrix $A = a_{i,j}$. The convergence of all elements included in Σ_{xp}, given the absolute summability of the process x_t, is proven Berk (1974), for the i.i.d. case, and in Gonçalves and Kilian (2003, Lemma A.2), under MDS assumption for ε_t. Therefore, we only are concerned about the components γ^*_0 and Σ_{*p}, which are generated by the square (but not absolutely) summable process x^{**}_t. We consider

$$T\mathbb{E}\left(\frac{1}{T} \sum_{t=p+1}^{T} \left[(x^{**}_{t-1})^2 - \gamma^*_0 \right] \right)^2 \tag{20}$$

which according to Gonçalves and Kilian (2003, Lemma A.2) is bounded by

$$\sum_{l=-\infty}^{\infty} |\kappa_*(0,0,l,l)| + 2\sum_{l=-\infty}^{\infty} (\gamma_l^*)^2 \tag{21}$$

where $\kappa_*(l_1,l_2,l_3,l_4)$ and γ_0^* denote the fourth order cumulants of x_t^{**}, and its autocovariances, respectively. The first sum is bounded, see Lemma 8, just like the second one, due to square summability of the autocovariances of x_t^{**}, see Lemma 4. The analogues results in the case of Σ_{*p} can be shown in the similar way. Due to (21), we now have

$$\mathbb{E}\left\|\widehat{\Sigma}_p - \Sigma_p\right\|^2 \leq C(p^2+2p-1)/T$$

where C is some positive constant. Then, $\left\|\widehat{\Sigma}_p - \Sigma_p\right\| = o_p(p^{-1})$ because $p^4/T \to 0$ by assumption. The desired result follows now by

$$\left\|\widehat{\Sigma}_p^{-1} - \Sigma_p^{-1}\right\| \leq \frac{C^2\left\|\widehat{\Sigma}_p - \Sigma_p\right\|}{1 - C\left\|\widehat{\Sigma}_p - \Sigma_p\right\|} = o_p(p^{-1}).$$

This result is used to prove consistency of the parameter estimates as follows

$$\left\|\underline{\widetilde{\beta}}_p - \underline{\beta}_p\right\| \leq \left\|\widehat{\Sigma}_p^{-1}\right\| \left\|\frac{1}{T}\sum_{t=p+1}^{T} W_{tp}(\varepsilon_t - \varepsilon_{tp})\right\| \tag{22}$$

$$+ \left\|\widehat{\Sigma}_p^{-1}\right\| \left\|\frac{1}{T}\sum_{t=p+1}^{T} W_{tp}\varepsilon_t\right\|. \tag{23}$$

Obviously $\left\|\widehat{\Sigma}_p^{-1}\right\| = O_p(1)$. To guarantee the $\sqrt{T}$-consistency we assume that

$$\sum_{k=p+1}^{\infty} |a_{p+k}| = o(T^{-1/2}). \tag{24}$$

Equation (24) determines also our summability condition from Assumption 2, because as we know from Brillinger (1975, p. 52) $\sum_{k=1}^{\infty} k^s|a_k| < \infty$ implies $\sum_{k=p+1}^{\infty} |a_k| = o(p^{-s})$. Thus, p must be of order higher than $O(T^{1/2s})$ to allow for (24) and we need $s > 2$ in Assumption 2 in order to holds the upper bound $p = o(T^{1/4})$. The results in Gonçalves and Kilian (2003, Theorem 2.1) and (24) leads without any considerable modification to

$$\left\|\frac{1}{T}\sum_{t=p+1}^{T} W_{tp}(\varepsilon_t - \varepsilon_{tp})\right\| = O_p(p^{1/2})o_p(T^{-1/2}). \tag{25}$$

Now consider (23). We have

$$\mathbb{E}\left\|\frac{1}{T}\sum_{t=p+1}^{T}W_{tp}\varepsilon_t\right\|^2 = T^{-2}\mathbb{E}\left(\sum_{t=p+1}^{T}x_{t-1}^{**}\varepsilon_t\right)^2 \tag{26}$$

$$+ T^{-2}\sum_{l=1}^{p}\mathbb{E}\left(\sum_{t=p+1}^{T}x_{t-l}\varepsilon_t\right)^2 \tag{27}$$

where (27) is $O(pT^{-1})$ by Gonçalves and Kilian (2003, Theorem 2.1) and we only need to examine (26):

$$\mathbb{E}\left(\sum_{t=p+1}^{T}x_{t-1}^{**}\varepsilon_t\right)^2 = \sum_{t=p+1}^{T}\sum_{s=p+1}^{T}\mathbb{E}(x_{t-1}^{**}x_{s-1}^{**}\varepsilon_t\varepsilon_s) = \sum_{t=p+1}^{T}\mathbb{E}\left((x_{t-1}^{**})^2\varepsilon_t^2\right)$$

where

$$\mathbb{E}\left((x_{t-1}^{**})^2\varepsilon_t^2\right) = \mathbb{E}\left(\left(\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}\right)^2\varepsilon_t^2\right) = \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\psi_j\psi_i\mathbb{E}(\varepsilon_{t-1-j}\varepsilon_{t-1-i}\varepsilon_t^2)$$

which is bounded by the absolute summability of eight-order cumulants of ε_t and Lemma 10. The following result is now obvious:

$$\left\|\frac{1}{T}\sum_{t=p+1}^{T}W_{tp}\varepsilon_t\right\| = O_p(p^{1/2}T^{-1/2}). \tag{28}$$

By combining (25) and (28), we have

$$\left\|\underline{\widetilde{\beta}}_p - \underline{\beta}_p\right\| = O_p(1)O_p(p^{1/2})o_p(T^{-1/2}) + O_p(1)O_p(p^{1/2}T^{-1/2}) = O_p(p^{1/2}T^{-1/2}).$$

Now it is easy to see that $\left\|\sqrt{T}(\underline{\widetilde{\beta}}_p - \underline{\beta}_p)\right\| = O_p(p^{1/2})$ and therefore also $\left\|\sqrt{T}(\underline{\widehat{\beta}}_p - \underline{\beta}_p)\right\| = O_p(p^{1/2})$.

Proof of Proposition 2

Define $X_{tp} = (x_{t-1}, \ldots, x_{t-p})'$ so that $W_{tp} = (x_{t-1}^*, X_{tp}')'$. Now we have

$$\underline{\widehat{\beta}}_p - \underline{\beta}_p = \begin{bmatrix}\sum(x_{t-1}^*)^2 & \sum x_{t-1}^*X_{tp}' \\ \sum X_{tp}x_{t-1}^* & \sum X_{tp}X_{tp}'\end{bmatrix}^{-1}\begin{bmatrix}\sum x_{t-1}^*\varepsilon_{tp} \\ \sum X_{tp}\varepsilon_{tp}\end{bmatrix}.$$

where $\underline{\beta}_p = (0, a_1, \ldots, a_p)'$. We investigate the asymptotic distribution of $\sqrt{T}\widehat{\phi}$, but from Lemma 2 we know that the difference to $\sqrt{T}\widetilde{\phi}$ vanishes asymptotically, so we only have to study the asymptotic distribution of the latter term. Now we find an expression for $\sqrt{T}\widetilde{\phi}$ as follows:

$$\begin{aligned}\sqrt{T}\widetilde{\phi} &= \widehat{A}\left(\frac{1}{\sqrt{T}}\sum x_{t-1}^{**}\varepsilon_{tp}\right) \\ &- \widehat{A}\left(\frac{1}{T}\sum x_{t-1}^{**}X_{tp}'\right)\left(\frac{1}{T}\sum X_{tp}X_{tp}'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum X_{tp}\varepsilon_{tp}\right),\end{aligned}$$

where

$$\widehat{A} = \left(\frac{1}{T}\sum (x_{t-1}^{**})^2 - \frac{1}{T}\sum x_{t-1}^{**}X_{tp}'\left(\frac{1}{T}\sum X_{tp}X_{tp}'\right)^{-1}\frac{1}{T}\sum X_{tp}x_{t-1}^{**}\right)^{-1}.$$

We can also denote the estimated values of the true matrices in a more compact way

$$\begin{aligned}\sqrt{T}\widetilde{\phi} &= \left[\widehat{\gamma}_0^* - \widehat{\Sigma}_{*p}'\widehat{\Sigma}_{xp}^{-1}\widehat{\Sigma}_{*p}\right]^{-1}\left(\frac{1}{\sqrt{T}}\sum x_{t-1}^{**}\varepsilon_{tp}\right) \\ &- \left[\widehat{\gamma}_0^* - \widehat{\Sigma}_{*p}'\widehat{\Sigma}_{xp}^{-1}\widehat{\Sigma}_{*p}\right]^{-1}\widehat{\Sigma}_{*p}'\widehat{\Sigma}_{xp}^{-1}\left(\frac{1}{\sqrt{T}}\sum X_{tp}\varepsilon_{tp}\right).\end{aligned}$$

Further, we define the asymptotic counterpart of $\sqrt{T}\widetilde{\phi}$ as follows

$$\begin{aligned}\sqrt{T}\phi^* &= \left[\gamma_0^* - \Sigma_{*p}'\Sigma_{xp}^{-1}\Sigma_{*p}\right]^{-1}\left(\frac{1}{\sqrt{T}}\sum x_{t-1}^{**}\varepsilon_{t}\right) \\ &- \left[\gamma_0^* - \Sigma_{*p}'\Sigma_{xp}^{-1}\Sigma_{*p}\right]^{-1}\Sigma_{*p}'\Sigma_{xp}^{-1}\left(\frac{1}{\sqrt{T}}\sum X_{tp}\varepsilon_{t}\right).\end{aligned}$$

Now, we need to show that

$$\sqrt{T}(\widetilde{\phi} - \phi^*) \xrightarrow{p} 0. \tag{29}$$

In order to prove (29), denote $\widehat{A} = \left[\widehat{\gamma}_0^* - \widehat{\Sigma}_{*p}'\widehat{\Sigma}_{xp}^{-1}\widehat{\Sigma}_{*p}\right]^{-1}$ and $A = \left[\gamma_0^* - \Sigma_{*p}'\Sigma_{xp}^{-1}\Sigma_{*p}\right]^{-1} = \sigma_{\omega p}^{-2}$. We have

$$\widehat{A}\left(\frac{1}{\sqrt{T}}\sum x_{t-1}^{**}\varepsilon_{tp}\right) - A\left(\frac{1}{\sqrt{T}}\sum x_{t-1}^{**}\varepsilon_{t}\right) =$$
$$\widehat{A}\left(\frac{1}{\sqrt{T}}\sum x_{t-1}^{**}(\varepsilon_{tp} - \varepsilon_t)\right) + \left[\widehat{A} - A\right]\left(\frac{1}{\sqrt{T}}\sum x_{t-1}^{**}\varepsilon_{t}\right),$$

and will show that this expression converges in probability to zero. Consider the first term in the last line and conclude according to (25) that

$$\left|\frac{1}{\sqrt{T}}\sum x^{**}_{t-1}(\varepsilon_{tp}-\varepsilon_t)\right| = o_p(1), \tag{30}$$

and $\widehat{A} = O_p(1)$ due to (19). Furthermore, it is easy to see that $\left|T^{-1/2}\sum x^{**}_{t-1}\varepsilon_t\right| = O_p(1)$, and $\left|\widehat{A}-A\right| = o_p(1)$. The desired result is now obvious

$$\left|\widehat{A}\left(\frac{1}{\sqrt{T}}\sum x^{**}_{t-1}\varepsilon_{tp}\right) - A\left(\frac{1}{\sqrt{T}}\sum x^{**}_{t-1}\varepsilon_t\right)\right| = o_p(1).$$

Denote $\widehat{B} = \left[\widehat{\gamma}^*_0 - \widehat{\Sigma}'_{*p}\widehat{\Sigma}^{-1}_{xp}\widehat{\Sigma}_{*p}\right]^{-1}\widehat{\Sigma}'_{*p}\widehat{\Sigma}^{-1}_{xp}$ and $B = \left[\gamma^*_0 - \Sigma'_{*p}\Sigma^{-1}_{xp}\Sigma_{*p}\right]^{-1}\Sigma'_{*p}\Sigma^{-1}_{xp} = A\Sigma'_{*p}\Sigma^{-1}_{xp}$. Then we have

$$\begin{aligned}
&\widehat{B}\left(\frac{1}{\sqrt{T}}\sum X_{tp}\varepsilon_{tp}\right) - B\left(\frac{1}{\sqrt{T}}\sum X_{tp}\varepsilon_t)\right) \\
&= \left[\widehat{B}-B\right]\left(\frac{1}{\sqrt{T}}\sum X_{tp}(\varepsilon_{tp}-\varepsilon_t)\right) + \left[\widehat{B}-B\right]\left(\frac{1}{\sqrt{T}}\sum X_{tp}\varepsilon_t\right) + \\
&\quad + B\left(\frac{1}{\sqrt{T}}\sum X_{tp}(\varepsilon_{tp}-\varepsilon_t)\right).
\end{aligned}$$

Consider the term on the third line. We proceed with the following transformation

$$B\left(\frac{1}{\sqrt{T}}\sum X_{tp}(\varepsilon_{tp}-\varepsilon_t)\right) = A\left(\frac{1}{\sqrt{T}}\sum u_{tp}(\varepsilon_{tp}-\varepsilon_t)\right),$$

where $u_{tp} = \Sigma'_{*p}\Sigma^{-1}_{xp}X_{tp}$ and $\mathbb{E}(u^2_{tp}) = \Sigma'_{*p}\Sigma^{-1}_{xp}\Sigma_{*p}$ is bounded, because, according to Lemma 6, it converges to a constant as p grows. From Lemma 7, we know that $u_{tp} = u_t + o_p(1)$, where u_t is a square summable process with the same innovations ε_t. It follows

$$\left(\frac{1}{\sqrt{T}}\sum u_{tp}(\varepsilon_{tp}-\varepsilon_t)\right) = \left(\frac{1}{\sqrt{T}}\sum u_t(\varepsilon_{tp}-\varepsilon_t)\right) + o_p(1)$$

and similarly to (30) we conclude that

$$\left|\frac{1}{\sqrt{T}}\sum u_t(\varepsilon_{tp}-\varepsilon_t)\right| = o_p(1).$$

It is rather tedious, but straightforward to show that $\left\|\widehat{B}-B\right\| = o_p(p^{-1})$, so following relationship holds:

$$\left|\left[\widehat{B}-B\right]\left(\frac{1}{\sqrt{T}}\sum X_{tp}\varepsilon_t\right)\right| \le \left\|\widehat{B}-B\right\|\left\|\frac{1}{\sqrt{T}}\sum X_{tp}\varepsilon_t\right\| = o_p(p^{-1})O_p(p^{1/2}) = o_p(1).$$

For the remaining term, it holds

$$\begin{aligned}\left|\left[\widehat{B}-B\right]\left(\frac{1}{\sqrt{T}}\sum X_{tp}(\varepsilon_{tp}-\varepsilon_t)\right)\right| &\leq \left\|\widehat{B}-B\right\|\left\|\frac{1}{\sqrt{T}}\sum X_{tp}(\varepsilon_{tp}-\varepsilon_t)\right\| \\ &= o_p(p^{-1})O_p(p^{1/2})o_p(1)=o_p(1),\end{aligned}$$

due to (25). Statement (29) is now proven.
Having shown that the distribution of $\sqrt{T}\widetilde{\phi}$ can be studied employing an expression with the same asymptotic distribution, we now investigate the distribution of this expression. We have

$$\sqrt{T}\phi^* = \left[\gamma_0^* - \Sigma'_{*p}\Sigma_{xp}^{-1}\Sigma_{*p}\right]^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}(x_{t-1}^{**}-\Sigma'_{*p}\Sigma_{xp}^{-1}X_{tp})\varepsilon_t\right).$$

Denote π_p^2 as the variance of the term in the sum and calculate it:

$$\begin{aligned}\pi_p^2 &= \mathbb{E}\left[(x_{t-1}^{**}-\Sigma'_{*p}\Sigma_{xp}^{-1}X_{tp})\varepsilon_t\right]^2 \\ &= \mathbb{E}\left[(x_{t-1}^{**}\varepsilon_t)^2 - 2\Sigma'_{*p}\Sigma_{xp}^{-1}X_{tp}x_{t-1}^{**}\varepsilon_t^2 + \Sigma'_{*p}\Sigma_{xp}^{-1}X_{tp}X'_{tp}\varepsilon_t^2\Sigma_{xp}^{-1}\Sigma_{*p}\right] \\ &= \mathbb{E}(x_{t-1}^{**}\varepsilon_t)^2 - 2\Sigma'_{*p}\Sigma_{xp}^{-1}\mathbb{E}(X_{tp}x_{t-1}^{**}\varepsilon_t^2) + \Sigma'_{*p}\Sigma_{xp}^{-1}\mathbb{E}(X_{tp}X'_{tp}\varepsilon_t^2)\Sigma_{xp}^{-1}\Sigma_{*p},\end{aligned}$$

so we can define the following process with zero mean and unit variance

$$Z_{Tt} = \pi_p^{-1}(x_{t-1}^{**}-\Sigma'_{*p}\Sigma_{xp}^{-1}X_{tp})\varepsilon_t\ .$$

Since the third term was shown in Gonçalves and Kilian (2003) to converge, it follows with Lemma 9 that π_p^2 is bounded and converges to a positive constant as $T\to\infty$. We have to prove that

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} Z_{Tt} \xrightarrow{d} \mathbf{N}(0,1). \tag{31}$$

To that end, we apply a CLT for martingale differences (cf. Davidson, 1994, p. 383). Hence we need to show that

(a) $T^{-1}\sum Z_{Tt}^2 - 1 \xrightarrow{p} 0$, and

(b) $\max_{p+1\leq t\leq T} T^{-1/2}\left|Z_{Tt}\right| \xrightarrow{p} 0$.

We start with (a) and have

$$
\begin{aligned}
T^{-1}\sum Z_{Tt}^2 - 1 &= \pi_p^{-2}T^{-1}\sum\left[(x_{t-1}^{**} - \Sigma_{*p}'\Sigma_{xp}^{-1}X_{tp})^2\varepsilon_t^2 - \pi_p^2\right] \\
&= \pi_p^{-2}T^{-1}\sum\left[(x_{t-1}^{**}\varepsilon_t)^2 - \mathbb{E}(x_{t-1}^{**}\varepsilon_t)^2\right] \\
&\quad -2\pi_p^{-2}\Sigma_{*p}'\Sigma_{xp}^{-1}T^{-1}\sum\left[X_{tp}x_{t-1}^{**}\varepsilon_t^2 - \mathbb{E}(X_{tp}x_{t-1}^{**}\varepsilon_t^2)\right] \\
&\quad +\pi_p^{-2}\Sigma_{*p}'\Sigma_{xp}^{-1}T^{-1}\sum\left[X_{tp}X_{tp}'\varepsilon_t^2 - \mathbb{E}(X_{tp}X_{tp}'\varepsilon_t^2)\right]\Sigma_{xp}^{-1}\Sigma_{*p} \\
&= \pi_p^{-2}S_{1T} - 2\pi_p^{-2}\Sigma_{*p}'\Sigma_{xp}^{-1}S_{2T} + \pi_p^{-2}\Sigma_{*p}'\Sigma_{xp}^{-1}S_{3T}\Sigma_{xp}^{-1}\Sigma_{*p},
\end{aligned}
$$

where S_{1T}, S_{2T}, S_{3T} are defined implicitly and are examined next. Consider

$$
\mathbb{E}\left(S_{1T}^2\right) = \mathbb{E}\left((T-p)^{-1}\sum_{t=p+1}^{T}\left[(x_{t-1}^{**}\varepsilon_t)^2 - \mathbb{E}(x_{t-1}^{**}\varepsilon_t)^2\right]\right)^2.
$$

Since $p = o(T)$, we may write

$$
\begin{aligned}
\mathbb{E}\left(S_{1T}^2\right) &= T^{-2}\sum_{t=p+1}^{T}\sum_{s=p+1}^{T} Cov((x_{t-1}^{**}\varepsilon_t)^2, (x_{s-1}^{**}\varepsilon_s)^2) + o(1) \\
&= T^{-2}\sum_{i,j,k,l=0}^{\infty}\psi_i\psi_j\psi_k\psi_l \cdot \\
&\quad \cdot\sum_{t=p+1}^{T}\sum_{s=p+1}^{T} Cov(\varepsilon_{t-1-i}\varepsilon_{t-1-j}\varepsilon_t^2, \varepsilon_{t-1-k}\varepsilon_{t-1-l}\varepsilon_s^2) + o(1).
\end{aligned}
$$

Due to stationarity, we employ a change of variable, $\tau = s - t$, and have the following transformation neglecting the term $o(1)$

$$
\begin{aligned}
\mathbb{E}\left(S_{1T}^2\right) &= T^{-1}\sum_{i,j,k,l=0}^{\infty}\psi_i\psi_j\psi_k\psi_l \cdot \\
&\quad \cdot\sum_{\tau=-T}^{T}\left(1 - \frac{|\tau|}{T}\right) Cov(\varepsilon_0\varepsilon_{i-j}\varepsilon_{1+i}^2, \varepsilon_{\tau-k+i}\varepsilon_{\tau-l+i}\varepsilon_{\tau+1+i}^2) \\
&\le T^{-1}\sum_{i,j,k,l=0}^{\infty}\psi_i\psi_j\psi_k\psi_l\sum_{\tau=-\infty}^{\infty} Cov(\varepsilon_0\varepsilon_{i-j}\varepsilon_{1+i}^2, \varepsilon_{\tau-k+i}\varepsilon_{\tau-l+i}\varepsilon_{\tau+1+i}^2) \\
&= T^{-1}\sum_{\tau=-\infty}^{\infty}\sum_{i,j,k,l=0}^{\infty}\psi_i\psi_j\psi_k\psi_l Cov(\varepsilon_0\varepsilon_{i-j}\varepsilon_{1+i}^2, \varepsilon_{\tau-k+i}\varepsilon_{\tau-l+i}\varepsilon_{\tau+1+i}^2).
\end{aligned}
$$

Employing Theorem 2.3.2 of Brillinger (1975, p. 21) we can express the covariances in the last line as the sum of products of cumulants of ε_t of order eight of lower, see also Gonçalves and Kilian (2003). We only consider the component of the sum including cumulants of order eight. Lemma 10 shows the following sum to be finite:

$$\sum_{\tau}\sum_{i,j,k,l}\psi_i\psi_j\psi_k\psi_l\kappa_\varepsilon(0,i-j,1+i,1+i,\tau-k+i,\tau-l+i,\tau+1+i,\tau+1+i).$$

The sums over other indecomposable partitions can be shown to be finite in the same way. This result implies that $\exists K>0$ so that

$$\mathbb{E}\left(S_{1T}^2\right)\leq K/T\rightarrow 0. \tag{32}$$

Next we consider the second term,

$$\mathbb{E}\left\|S_{2T}\right\|^2 \;=\; \sum_{m=1}^{p}\mathbb{E}\left(S_{2Tm}\right)^2,$$

where

$$S_{2Tm}=T^{-1}\sum_{t=p+1}^{T}\left[x_{t-m}x_{t-1}^{**}\varepsilon_t^2-\mathbb{E}(x_{t-m}x_{t-1}^{**}\varepsilon_t^2)\right].$$

We have

$$\begin{aligned}\mathbb{E}\left(S_{2Tm}\right)^2 &= \mathbb{E}\left(T^{-1}\sum_{t=p+1}^{T}\left[x_{t-m}x_{t-1}^{**}\varepsilon_t^2-\mathbb{E}(x_{t-m}x_{t-1}^{**}\varepsilon_t^2)\right]\right)^2\\ &= T^{-2}\sum_{t=p+1}^{T}\sum_{s=p+1}^{T}Cov(x_{t-m}x_{t-1}^{**}\varepsilon_t^2,x_{s-m}x_{s-1}^{**}\varepsilon_s^2),\end{aligned}$$

which equals

$$T^{-2}\sum_{i,j,k,l=-\infty}^{\infty}b_i\psi_jb_k\psi_l\sum_{t=p+1}^{T}\sum_{s=p+1}^{T}Cov(\varepsilon_{t-m-i}\varepsilon_{t-1-j}\varepsilon_t^2,\varepsilon_{s-m-k}\varepsilon_{s-1-l}\varepsilon_s^2)$$

bounded by

$$\frac{1}{T}\sum_{i,j,k,l=-\infty}^{\infty}b_i\psi_jb_k\psi_l\sum_{\tau=-\infty}^{\infty}Cov(\varepsilon_0\varepsilon_{m+i-j-1}\varepsilon_{m+i}^2,\varepsilon_{s-t+i-k}\varepsilon_{s-t-1-l+m+i}\varepsilon_{s-t+m+i}^2),$$

again considering that $p = o(T)$. The proof of boundedness of the covariance sum is similar to the case of S_{1T}, and we have

$$\mathbb{E}\,\|S_{2T}\|^2 \leq Kp/T \to 0 \tag{33}$$

because of $p^4/T \to 0$. The convergence of S_{3T} is proven in Gonçalves and Kilian (2003), so that

$$\mathbb{E}\,\|S_{3T}\|^2 \leq Kp^2/T \to 0,$$

which proves (a), given that π_p^{-2} and $\|\Sigma'_{*p}\Sigma_{xp}^{-1}\|$ are bounded, see Lemma 9 and Lemma 5.

To prove (b) we note that for any $\eta > 0$ and some $r > 1$

$$\mathbb{P}\left(\max_{p+1\leq t\leq T} |Z_{Tt}| > \eta\sqrt{T}\right) \leq \sum \mathbb{P}\left(|Z_{Tt}| > \eta\sqrt{T}\right) \leq \sum \frac{\mathbb{E}|Z_{Tt}|^r}{T^{r/2}\eta^r}.$$

By the Cauchy-Schwarz inequality we have

$$\begin{aligned}
\mathbb{E}|Z_{Tt}|^r &= \mathbb{E}\left|\pi_p^{-1}(x_{t-1}^{**} - \Sigma'_{*p}\Sigma_{xp}^{-1}X_{tp})\varepsilon_t\right|^r \\
&\leq \left|\pi_p^{-1}\right|^r \left(\mathbb{E}\left|x_{t-1}^{**} - \Sigma'_{*p}\Sigma_{xp}^{-1}X_{tp}\right|^{2r}\right)^{1/2} \left(\mathbb{E}\,|\varepsilon_t|^{2r}\right)^{1/2}
\end{aligned}$$

where $\pi_p^{-1} = O(1)$ and $\mathbb{E}\,|\varepsilon_t|^{2r} < \infty$ for all $r \leq 4$ by our assumptions. Now consider

$$\begin{aligned}
\mathbb{E}\left|x_{t-1}^{**} - \Sigma'_{*p}\Sigma_{xp}^{-1}X_{tp}\right|^{2r} &\leq \left[\|x_{t-1}^{**}\|_{2r} + \|\Sigma'_{*p}\Sigma_{xp}^{-1}X_{tp}\|_{2r}\right]^{2r} \\
&\leq \left[\|x_{t-1}^{**}\|_{2r} + \left\|\Sigma'_{*p}\Sigma_{xp}^{-1}\right\| \left[\mathbb{E}\|X_{tp}\|^{2r}\right]^{1/2r}\right]^{2r}
\end{aligned}$$

where $\|\cdot\|_r = [\mathbb{E}|\cdot|^r]^{1/r}$ and further $\mathbb{E}|x_{t-1}^{**}|^{2r} < \infty$ for all $r \leq 4$ by the assumption about cumulants of ε_t and Lemma 8, which can be easily extended to the case of 8th cumulants of x_t given the summability of 8th cumulants of ε_t, as well as $\left\|\Sigma'_{*p}\Sigma_{xp}^{-1}\right\| < \infty$ according to Lemma 5. Finally,

$$\mathbb{E}\|X_{tp}\|^{2r} = \mathbb{E}\left|\sum_{i=1}^{p} x_{t-i}^2\right|^r \leq \left(\sum_{i=1}^{p} \|x_{t-i}^2\|_r\right)^r = O(p^r)$$

because $\mathbb{E}|x_{t-i}|^{2r} < \infty$ for some $r \leq 4$ and $i = 1, \ldots, p$ and therefore we have $\mathbb{E}\,|Z_{Tt}|^r = O(p^{r/2})$, leading to

$$\mathbb{P}\left(\max_{p+1\leq t\leq T} |Z_{Tt}| > \eta\sqrt{T}\right) \leq \sum_{t=p+1}^{T} \frac{\mathbb{E}|Z_{Tt}|^r}{T^{r/2}\eta^r} = O\left(\frac{p^{r/2}}{T^{r/2-1}}\right) = o(1)$$

for $r = 3$ because of $p^4/T \to 0$. The proof of (31) is now complete.
Now we define the matrix

$$\Gamma_p = \Sigma_p^{-1} \mathbb{E}\left(W_{tp} W_{tp}' \varepsilon_t^2\right) \Sigma_p^{-1}$$

whereas it is easy to see that

$$[\Gamma_p]_{11} = \left[\gamma_0^* - \Sigma_{*p}' \Sigma_{xp}^{-1} \Sigma_{*p}\right]^{-2} \pi_p^2 \, .$$

Obviously, (31) can be represented as

$$\frac{\sqrt{T}\phi^*}{\left([\Gamma_p]_{11}\right)^{0.5}} \xrightarrow{d} \mathbf{N}(0,1).$$

However, in practical applications Γ_p is of course unknown, and therefore needs to be estimated consistently. We proceed in the manner of Gonçalves and Kilian (2003, Theorem 2.3) and estimate the matrix of fourth moments in the following way

$$\widehat{\mathbb{E}}\left(W_{tp} W_{tp}' \varepsilon_t^2\right) = T^{-1} \sum_{t=p+1}^{T} V_{tp} V_{tp}' \widehat{\varepsilon}_{tp}^{\,2} \tag{34}$$

which is a version of the Eicker-White heteroskedasticity-robust estimator, where $V_{tp} = (x_{t-1}^*, X_{tp}')'$ and $\widehat{\varepsilon}_{tp} = x_t - V_{tp}' \widehat{\beta}_p$. Using Lemma 2 and Proposition 1 it is easy to show that

$$T^{-1} \sum_{t=p+1}^{T} V_{tp} V_{tp}' \widehat{\varepsilon}_{tp}^{\,2} = T^{-1} \sum_{t=p+1}^{T} W_{tp} W_{tp}' \varepsilon_{tp}^{\,2} + o_p(1)$$

and employing the arguments of Theorem 2.3 in Gonçalves and Kilian (2003), results (32) and (33), as well as the assumption $p^4/T \to 0$, see Gonçalves and Kilian (2003, Proof of Theorem 2.3) for reference, we conclude that $\widehat{\mathbb{E}}\left(W_{tp} W_{tp}' \varepsilon_t^2\right)$ in (34) is a consistent estimator of the matrix of fourth moments. It also follows with slight modifications that $\sqrt{T}\ \hat{s}(\widehat{\phi})$, with $\hat{s}(\widehat{\phi})$ defined in (9), has a proper positive probability limit as $T \to \infty$. The desired result now follows. ■

Proof of Proposition 3

Consider the OLS estimator under the local alternative, $\widehat{\underline{\beta}}_p^{\,\delta}$. Correspondingly, denote $\widetilde{V}_{tp} = (\widetilde{x}_{t-1}^*, \widetilde{x}_{t-1}, \ldots, \widetilde{x}_{t-p})'$; therefore, we have

$$\widehat{\underline{\beta}}_p^{\,\delta} = \left[\frac{1}{T} \sum \widetilde{V}_{tp} \widetilde{V}_{tp}'\right]^{-1} \left(\frac{1}{T} \sum \widetilde{V}_{tp} \widetilde{x}_t\right).$$

Let us consider the properties of the process x^{δ}_{t-2} defined in (11). Just like in Lemma 2, define asymptotic counterpart

$$x^{\delta\delta}_{t-2} = \sum_{j=1}^{\infty} \frac{x^{**}_{t-1-j}}{j}, \tag{35}$$

and express it as an infinite MA process, $x^{\delta\delta}_{t-2} = \sum_{j=1}^{\infty} \eta_{j-1}\varepsilon_{t-1-j}$, with η_j the convolution of ψ_j and $\frac{1}{j}$. Tedious, but straightforward calculations lead to the conclusion $\eta_j = O(\ln j\,/j)$, which ensures stationarity and ergodicity of the process $x^{\delta\delta}_{t-2}$. The difference of the processes x^{δ}_{t-2} and $x^{\delta\delta}_{t-2}$ has itself an MA representation,

$$x^{\delta\delta}_{t-2} - x^{\delta}_{t-2} = \sum_{j=t-1}^{\infty} \widetilde{\eta}_{j-1}\varepsilon_{t-1-j},$$

where the coefficients $\widetilde{\eta}_{j-1}$, similarly to $\widetilde{\psi}_j$ in Lemma 2, converge for fixed t and $j \to \infty$ to the theoretical coefficients η_{j-1}, so, following Lemma 4, we may write

$$\widetilde{\eta}_{j-1} = O\left(\ln j\,/j\right),$$

which, together with MDS property of ε_t, leads to

$$\mathbb{E}\left(x^{\delta\delta}_{t-2} - x^{\delta}_{t-2}\right)^2 = O\left(\sum_{j=t-1}^{\infty} \frac{\ln^2 j}{j^2}\right).$$

Using the integral $\int \frac{\ln^2 x}{x^2}dx = -\frac{\ln^2 x}{x} - \frac{2\ln x}{x} - \frac{2}{x}$ to approximate the sum, it follows

$$x^{\delta}_{t-2} = x^{\delta\delta}_{t-2} + O_p\left(\frac{\ln t}{\sqrt{t}}\right).$$

It is straightforward to show, along the lines of Proposition 4, that

$$\left\|\frac{1}{T}\sum \widetilde{V}_{tp}\widetilde{V}'_{tp} - \frac{1}{T}\sum W_{tp}W'_{tp}\right\| = O_p\left(\frac{p}{\sqrt{T}}\right) = o_p(1),$$

so $\frac{1}{T}\sum \widetilde{V}_{tp}\widetilde{V}'_{tp}$ "converges" to Σ_p. Now, consider the expression $\sqrt{T}(\underline{\widehat{\beta}}_p^{\,\delta} - \underline{\beta}_p)$. It can be easily seen that

$$\sqrt{T}(\underline{\widehat{\beta}}_p^{\,\delta} - \underline{\beta}_p) = \sqrt{T}(\underline{\beta}_p - \underline{\beta}_p) + \delta\Sigma_p^{-1}\Xi_p + O_p(T^{-1}),$$

where

$$\Xi_p = \begin{pmatrix} \mathbb{E}[x^{\delta\delta}_{t-2}x_t] + \mathbb{E}[(x^{**}_{t-1})^2] \\ \mathbb{E}[x^{**}_{t-2}x_t] + \mathbb{E}[x^{**}_{t-1}x_{t-1}] \\ \mathbb{E}[x^{**}_{t-3}x_t] + \mathbb{E}[x^{**}_{t-1}x_{t-2}] \\ \vdots \\ \mathbb{E}[x^{**}_{t-p-1}x_t] + \mathbb{E}[x^{**}_{t-1}x_{t-p}] \end{pmatrix} = \begin{pmatrix} \mathbb{E}[x^{\delta\delta}_{t-2}x_t] + \gamma^*_0 \\ \gamma^{*x}_{-2} + \gamma^{*x}_0 \\ \gamma^{*x}_{-3} + \gamma^{*x}_1 \\ \vdots \\ \gamma^{*x}_{-p-1} + \gamma^{*x}_{p-1} \end{pmatrix}.$$

Write

$$\Sigma_p^{-1}\Xi_p = \begin{bmatrix} \gamma^*_0 & \Sigma'_{*p} \\ \Sigma_{*p} & \Sigma_{xp} \end{bmatrix}^{-1} \begin{pmatrix} \mathbb{E}[x^{\delta\delta}_{t-2}x_t] + \gamma^*_0 \\ \overset{\bullet}{\Sigma}_{*p} + \Sigma_{*p} \end{pmatrix},$$

with $\overset{\bullet}{\Sigma}_{*p} = (\gamma^{*x}_{-2}, \ldots, \gamma^{*x}_{-p-1})'$, or

$$\Sigma_p^{-1}\Xi_p = \begin{bmatrix} \gamma^*_0 & \Sigma'_{*p} \\ \Sigma_{*p} & \Sigma_{xp} \end{bmatrix}^{-1} \begin{pmatrix} \mathbb{E}[x^{\delta\delta}_{t-2}x_t] \\ \overset{\bullet}{\Sigma}_{*p} \end{pmatrix} + \begin{bmatrix} \gamma^*_0 & \Sigma'_{*p} \\ \Sigma_{*p} & \Sigma_{xp} \end{bmatrix}^{-1} \begin{pmatrix} \gamma^*_0 \\ \Sigma_{*p} \end{pmatrix}.$$

Consider the first element of this vector,

$$\begin{aligned} \left[\Sigma_p^{-1}\Xi_p\right]_1 &= \left(\gamma^*_0 - \Sigma'_{*p}\Sigma^{-1}_{xp}\Sigma_{*p}\right)^{-1} \mathbb{E}[x^{\delta\delta}_{t-2}x_t] - \left(\gamma^*_0 - \Sigma'_{*p}\Sigma^{-1}_{xp}\Sigma_{*p}\right)^{-1} \Sigma'_{*p}\Sigma^{-1}_{xp}\overset{\bullet}{\Sigma}_{*p} \\ &\quad + \left(\gamma^*_0 - \Sigma'_{*p}\Sigma^{-1}_{xp}\Sigma_{*p}\right)^{-1} \gamma^*_0 - \left(\gamma^*_0 - \Sigma'_{*p}\Sigma^{-1}_{xp}\Sigma_{*p}\right)^{-1} \Sigma'_{*p}\Sigma^{-1}_{xp}\Sigma_{*p} \\ &= \left(\gamma^*_0 - \Sigma'_{*p}\Sigma^{-1}_{xp}\Sigma_{*p}\right)^{-1} \left(\mathbb{E}[x^{\delta\delta}_{t-2}x_t] - \Sigma'_{*p}\Sigma^{-1}_{xp}\overset{\bullet}{\Sigma}_{*p}\right) + 1. \end{aligned}$$

Let $A_p = Q_p - \Sigma'_{*p}\Sigma^{-1}_{xp}\overset{\bullet}{\Sigma}_{*p}$, with $Q_p = \left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{p}\right)\overset{\bullet}{\Sigma}_{*p}$, and note that $Q_p \to \mathbb{E}[x^{\delta\delta}_{t-2}x_t]$ as $p \to \infty$. Then, write $A_p = B_p\overset{\bullet}{\Sigma}_{*p}$, with $B_p = \left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{p}\right) - \Sigma'_{*p}\Sigma^{-1}_{xp}$. Post-multiply B_p with Σ_{xp} to obtain

$$B_p\Sigma_{xp} = \{b_j\}_{1\le j\le p} = \left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{p}\right)\Sigma_{xp} - \Sigma'_{*p}.$$

We now have

$$\begin{aligned} \gamma^{*x}_j &= \mathbb{E}\left[x^{**}_{t-1}x_{t-1-j}\right] = \mathbb{E}\left[\left(\sum_{k\ge1}\frac{x_{t-k}}{k}\right)x_{t-1-j}\right] = \sum_{k\ge1}\frac{\gamma_{|k-1-j|}}{k} \\ &= \left(1, \frac{1}{2}, \frac{1}{3}, \ldots\right)(\gamma_j, \ldots, \gamma_0, \ldots, \gamma_p, \ldots)' \\ &= \left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{p}\right)\Sigma^{[j]}_{xp} + \sum_{k\ge p+1}\frac{\gamma_{|k-1-j|}}{k}, \end{aligned}$$

where $\Sigma_{xp}^{[j]}$ is the j^{th} column of the matrix Σ_{xp}. Further,

$$\sum_{k\geq p+1}\frac{\gamma_{|k-1-j|}}{k} = o\left(\sum_{k\geq p+1}\frac{1}{k\,(k-j)}\right) = o\left(\frac{1}{j}\sum_{k\geq p+1}\left(\frac{1}{k-j}-\frac{1}{k}\right)\right)$$
$$= o\left(\frac{1}{j\,(p-j)}\right).$$

It follows

$$b_j = o\left(\frac{1}{j\,(p-j)}\right) = o\left(\frac{1}{p}\left(\frac{1}{p-j}+\frac{1}{j}\right)\right) = o\left(\frac{1}{p}\right),$$

and hence

$$\|B_p\Sigma_{xp}\| = \sum_{j=1}^{p} b_j^2 = o\left(\frac{p}{p^2}\right) = o\,(1)\,.$$

Since the norms of the matrix Σ_{xp} and the vector $\dot{\Sigma}_{*p}$ are bounded, it follows that the norm of A_p vanishes as $p\to\infty$, and, due to the convergence of $\gamma_0^* - \Sigma_{*p}'\Sigma_{xp}^{-1}\Sigma_{*p}$ and of Q_p, it holds

$$\left[\Sigma_p^{-1}\Xi_p\right]_1 = 1 + o\,(1)\,.$$

It follows directly from the distribution of the estimators that $\left\|\widehat{\underline{\beta}}_p^{\delta} - \underline{\beta}_p\right\|$ is of order $o_p\left(p^{-1/2}\right)$, and, due to this consistency, it is straightforward to show that the standard deviation of $\widehat{\phi}^{\delta}$ behaves just as under the null hypothesis, which completes the result. ■

Proof of Proposition 4

The proof consists of three steps. First, we show $\widehat{x}_t = x_t + O_p\left(t^{-0.5}\right)$.
To this end, recall that $d_i = O\left(i^{-d-1}\right)$ and $d_0 = 1$. Then,

$$(1-L)_t^d\, f(t) = O\left(\sum_{i=1}^{t-1} i^{-d-1}\,(t-i)^{\alpha}\right).$$

Now write

$$\sum_{i=0}^{t-1} i^{-d-1}\,(t-i)^{\alpha} = t^{\alpha-d}\sum_{i=0}^{t-1}\left(\frac{i}{t}\right)^{-d-1}\left(\frac{t-i}{t}\right)^{\alpha}\frac{1}{t} \tag{36}$$

For $t\to\infty$, it obviously holds that $\sum_{i=0}^{t-1}\left(\frac{i}{t}\right)^{-d-1}\left(\frac{t-i}{t}\right)^{\alpha}\frac{1}{t} \to \int_0^1 s^{-d-1}\,(1-s)^{\alpha}\,ds$, which is finite as long as $\alpha > -1$ and $d > 0$. Thus the sum on the right-hand of Equation (36) is bounded and

$$(1-L)_t^d\, f(t) = O\left(t^{\alpha-d}\right).$$

OLS regression of $x_{t,obs}$ on $(1-L)_t^d\ f(t)$ leads to

$$\widehat{\eta}-\eta = O_p\left(\frac{\sum_{t=1}^T t^{\alpha-d}x_t}{\sum_{t=1}^T t^{2(\alpha-d)}}\right),\text{ and } \widehat{x}_t = x_t - (\widehat{\eta}-\eta)\,O\left(t^{\alpha-d}\right).$$

Two cases arise. On the one hand, $\alpha - d \geq 0$. Then, write

$$\frac{\sum_{t=1}^T t^{\alpha-d}x_t}{\sum_{t=1}^T t^{2(\alpha-d)}} = T^{-0.5-(\alpha-d)}\frac{T^{-0.5-(\alpha-d)}\sum_{t=1}^T t^{\alpha-d}x_t}{T^{-1-2(\alpha-d)}\sum_{t=1}^T t^{2(\alpha-d)}}$$

and following result is easily shown to hold for a Brownian motion W:

$$\frac{T^{-0.5-(\alpha-d)}\sum_{t=1}^T t^{\alpha-d}x_t}{T^{-1-2(\alpha-d)}\sum_{t=1}^T t^{2(\alpha-d)}} \xrightarrow{d} \frac{\int_0^1 s^{\alpha-d}dW(s)}{\int_0^1 s^{2(\alpha-d)}ds}.$$

It follows

$$\widehat{\eta}-\eta = O_p\left(T^{-0.5-(\alpha-d)}\right)$$

and, since $(t/T)^{\alpha-d} = O(1)$,

$$\widehat{x}_t = x_t + (\widehat{\eta}-\eta)\,O\left(t^{\alpha-d}\right) = x_t + O_p\left(T^{-0.5}\right) = x_t + O_p\left(t^{-0.5}\right). \tag{37}$$

On the other hand, $\alpha - d < 0$, where, we distinguish 3 subcases: $-0.5 < \alpha - d < 0$, $\alpha - d = -0.5$ and $\alpha - d < -0.5$. For $-0.5 < \alpha - d < 0$, we obtain just like before,

$$\widehat{\eta}-\eta = O_p\left(T^{-0.5-(\alpha-d)}\right),$$

from which it follows

$$\widehat{x}_t = x_t + (\widehat{\eta}-\eta)\,O\left(t^{\alpha-d}\right) = x_t + O_p\left(t^{-0.5}\right), \tag{38}$$

since $T^{-0.5-(\alpha-d)}t^{\alpha-d} = T^{-0.5-(\alpha-d)}t^{0.5+\alpha-d}t^{-0.5}$ and $(T/t)^{-0.5-(\alpha-d)} = O(1)$. For $\alpha - d = -0.5$, the denominator of $\widehat{\eta}-\eta$ is obviously $O(\ln T)$. With $\mathbb{E}(x_t) = 0$, the variance of the numerator is

$$\mathbb{E}\left(\sum_{t=1}^T \frac{1}{\sqrt{t}}x_t\right)^2 = \sum_{t=1}^T \frac{1}{t}\gamma_0^x + \sum_{i=1}^{T-1}\frac{1}{\sqrt{i}}\sum_{j=i+1}^T \frac{1}{\sqrt{j}}\gamma_{j-i}^x.$$

Due to Lemma 3, $\gamma_{j-i}^x = o\left((j-i)^{-1}\right)$. Then,

$$\sum_{j=i+1}^T \frac{1}{\sqrt{j}}\gamma_{j-i}^x = o\left(\frac{1}{i}\sum_{j=i+1}^T\left(\frac{\sqrt{j}}{j-i} - \frac{1}{\sqrt{j}}\right)\right).$$

The first summand on the right-hand side dominates the second, and, since $j \leq T$, we obtain

$$\sum_{j=i+1}^{T} \frac{1}{\sqrt{j}} \gamma_{j-i}^{x} = o\left(\frac{\sqrt{T} \ln (T-i)}{i}\right).$$

Hence, since $T > T - i$,

$$\sum_{i=1}^{T-1} \frac{1}{\sqrt{i}} \sum_{j=i+1}^{T} \frac{1}{\sqrt{j}} \gamma_{j-i}^{x} = o\left(\sqrt{T} \ln T \sum_{i=1}^{T-1} \frac{1}{i^{1.5}}\right) = o\left(\ln T\right).$$

The numerator is thus of order $\sqrt{\ln T}$, and $\widehat{\eta} - \eta = o_p\left(1\right)$, leading to

$$\widehat{x}_t = x_t + \left(\widehat{\eta} - \eta\right) O\left(t^{-0.5}\right) = x_t + o_p\left(t^{-0.5}\right). \tag{39}$$

For $\alpha - d < -0.5$, the denominator of $\widehat{\eta} - \eta$ converges to a constant, while the numerator can be shown to be $O_p\left(1\right)$, so

$$\widehat{x}_t = x_t + O_p\left(t^{\alpha - d}\right) = x_t + o_p\left(t^{-0.5}\right). \tag{40}$$

Note that the $O_p\left(t^{-0.5}\right)$ terms in Equations (37), (38), (39) and (40) can be written as $O\left(t^{-0.5}\right) \cdot O_p\left(1\right)$, where the $O_p\left(1\right)$ term is the same for all t.
In the second step, the term $\widehat{x}_{t-1}^{*}$ is examined:

$$\widehat{x}_{t-1}^{*} = \sum_{j=1}^{t-1} \frac{\widehat{x}_{t-j}}{j} = \sum_{j=1}^{t-1} \frac{x_{t-j}}{j} + O_p\left(\sum_{j=1}^{t-1} \frac{1}{j\sqrt{t-j}}\right).$$

Further,

$$\begin{aligned} \sum_{j=1}^{t-1} \frac{1}{j\sqrt{t-j}} &= \sum_{j=1}^{t-1} \frac{\sqrt{t-j}}{j\left(t-j\right)} = \sum_{j=1}^{t-1} \frac{\sqrt{t-j}}{t}\left(\frac{1}{j} + \frac{1}{\left(t-j\right)}\right) \\ &= \frac{1}{\sqrt{t}} \sum_{j=1}^{t-1} \frac{1}{j}\sqrt{1 - \frac{j}{t}} + \frac{1}{t} \sum_{j=1}^{t-1} \frac{1}{\sqrt{t-j}}. \end{aligned}$$

Since $0 < \sqrt{1 - j/t} < 1$, $\sum_{j=1}^{t-1} j^{-1} = O\left(\ln t\right)$, $\sum_{j=1}^{t-1} (t-j)^{-0.5} = O\left(\sqrt{t}\right)$ and, from Lemma 2, $x_{t-1}^{*} = x_{t-1}^{**} + O_p\left(t^{-0.5}\right)$, it follows

$$\widehat{x}_{t-1}^{*} = x_{t-1}^{**} + O_p\left(\ln t / \sqrt{t}\right). \tag{41}$$

Again, the stochastic component of this $O_p\left(\ln t / \sqrt{t}\right)$ term is the same for all t.
The third step is concerned with the effect of these approximations on the estimators. Let $\mathbf{1}_{p+1} = (1, 1, \ldots, 1)' \in \mathbb{R}^{p+1}$ and consider $\widehat{W_{tp}} = W_{tp} + \mathbf{1}_{p+1} \cdot O_p\left(\ln t / \sqrt{t}\right)$

and $\widehat{x}_t = x_t + O_p\left(\ln t/\sqrt{t}\right)$, thus accounting for Equations (37), (38), (39) and (40), as well as (41). The OLS estimator $\underline{\widehat{\beta}}_{pr}$ of $\underline{\beta}_p$ is given by

$$\underline{\widehat{\beta}}_{pr} = \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}'_{tp}\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}\widehat{x}_t\right).$$

First, we need to show that

$$\left\|\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}'_{tp} - \frac{1}{T}\sum W_{tp}W'_{tp}\right\| = o_p\left(p^{-1/2}\right).$$

and

$$\left\|\frac{1}{T}\sum \widehat{W}_{tp}\widehat{x}_t - \frac{1}{T}\sum W_{tp}x_t\right\| = o_p(1),$$

since inverting a matrix and building a matrix norm are continuous operations, and $\text{plim}T^{-1}\sum W_{tp}W'_{tp}$. The matrix $\frac{1}{T}\sum W_{tp}W'_{tp}$ is nonsingular for any fixed p, see Lemma 5. Consider the first diagonal element of $\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}'_{tp}$, which can be expressed as

$$\begin{aligned}&\frac{1}{T}\sum\left(x^{**}_{t-1}\right)^2 + \frac{2}{T}\sum x^{**}_{t-1}O_p\left(\frac{\ln t}{\sqrt{t}}\right) + O_p\left(\frac{1}{T}\sum\frac{\ln^2 t}{t}\right)\\ &= \frac{1}{T}\sum\left(x^{**}_{t-1}\right)^2 + O_p\left(\frac{\ln^{1.5} T}{\sqrt{T}}\right),\end{aligned}$$

due to $\sum \frac{\ln^2 t}{t} = O\left(\ln^3 T\right)$ and

$$\sum x^{**}_{t-1}\frac{\ln t}{\sqrt{t}} \leq \sqrt{\sum\left(x^{**}_{t-1}\right)^2\sum\frac{\ln^2 t}{t}} = O_p\left(\sqrt{T\ln^3 T}\right).$$

Similar arguments apply to the other elements of $\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}'_{tp}$ and $\frac{1}{T}\sum \widehat{W}_{tp}\widehat{x}_t$, which leads to

$$\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}'_{tp} - \frac{1}{T}\sum W_{tp}W'_{tp} = O_p\left(\frac{\ln^{1.5} T}{\sqrt{T}}\right)\cdot \mathbf{1}_{p+1}\mathbf{1}'_{p+1},$$

Hence,

$$\left\|\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}'_{tp} - \frac{1}{T}\sum W_{tp}W'_{tp}\right\| = O_p\left(\frac{\ln^{1.5} T}{\sqrt{T}}\right)\cdot\left\|\mathbf{1}_{p+1}\mathbf{1}'_{p+1}\right\| = o_p\left(p^{-1/2}\right),$$

due to $\left\|\mathbf{1}_{p+1}\mathbf{1}'_{p+1}\right\| = O(p) = o\left(\sqrt[4]{T}\right)$. Also,

$$\frac{1}{T}\sum \widehat{W}_{tp}\widehat{x}_t - \frac{1}{T}\sum W_{tp}x_t = \frac{1}{T}\sum \widehat{W}_{tp}O_p\left(\frac{\ln t}{\sqrt{t}}\right) + \frac{1}{T}\sum \mathbf{1}_{p+1}x_tO_p\left(\frac{\ln t}{\sqrt{t}}\right).$$

Since $\left\|\widehat{W}_{tp}\right\|$ and $\|1_{p+1}\|$ are both of (stochastic) order $\sqrt{p}$, it is easily derived that

$$\left\|\frac{1}{T}\sum \widehat{W}_{tp}\widehat{x}_t - \frac{1}{T}\sum W_{tp}x_t\right\| = o_p(1).$$

Then, we need to show that

$$\left\|\sqrt{T}\left(\widehat{\underline{\beta}}_{pr} - \underline{\beta}_p\right) - \sqrt{T}\left(\widetilde{\underline{\beta}}_p - \underline{\beta}_p\right)\right\| = o_p(1),$$

where $\widetilde{\underline{\beta}}_p = \left[W_{tp}W_{tp}'\right]^{-1}(W_{tp}x_t)$. Consider with ε_{tp} from (7)

$$\begin{aligned}
\widehat{\underline{\beta}}_{pr} &= \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}\widehat{x}_t\right) \\
&= \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}x_t\right) + \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}O_p\left(\frac{\ln t}{\sqrt{t}}\right)\right) \\
&= \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}\left(\varepsilon_{tp} + W_{tp}'\underline{\beta}_p\right)\right) + \\
&\quad + \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}O_p\left(\frac{\ln t}{\sqrt{t}}\right)\right) \\
&= \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}\varepsilon_{tp}\right) + \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}W_{tp}'\right)\underline{\beta}_p + \\
&\quad + \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}O_p\left(\frac{\ln t}{\sqrt{t}}\right)\right) \\
&= \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum W_{tp}\varepsilon_{tp}\right) + \\
&\quad + \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \mathbf{1}_{p+1}\varepsilon_{tp}O_p\left(\frac{\ln t}{\sqrt{t}}\right)\right) + \\
&\quad + \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right)\beta_p \\
&\quad - \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}\mathbf{1}_{p+1}'O_p\left(\frac{\ln t}{\sqrt{t}}\right)\right)\underline{\beta}_p \\
&\quad + \left[\frac{1}{T}\sum \widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{T}\sum \widehat{W}_{tp}O_p\left(\frac{\ln t}{\sqrt{t}}\right)\right)
\end{aligned}$$

leading to following expression for $\sqrt{T}\left(\widehat{\underline{\beta}}_{pr}-\underline{\beta}_{p}\right)-\sqrt{T}\left(\widetilde{\underline{\beta}}_{p}-\underline{\beta}_{p}\right)$:

$$\left[\frac{1}{T}\sum\widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{\sqrt{T}}\sum\mathbf{1}_{p+1}\varepsilon_{tp}O_p\left(\frac{\ln t}{\sqrt{t}}\right)\right) \tag{42}$$

$$-\left[\frac{1}{T}\sum\widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{\sqrt{T}}\sum\widehat{W}_{tp}\mathbf{1}_{p+1}'O_p\left(\frac{\ln t}{\sqrt{t}}\right)\right)\underline{\beta}_{p} \tag{43}$$

$$+\left[\frac{1}{T}\sum\widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}\left(\frac{1}{\sqrt{T}}\sum\widehat{W}_{tp}O_p\left(\frac{\ln t}{\sqrt{t}}\right)\right)+o_p(1). \tag{44}$$

For each term, the norm of $\left[T^{-1}\sum\widehat{W}_{tp}\widehat{W}_{tp}'\right]^{-1}$ is bounded, and so is the norm of $\underline{\beta}_{p}$. The expression (43) is of stochastic order

$$O_p\left(\frac{1}{\sqrt{T}}\sum W_{tp}\frac{\ln t}{\sqrt{t}}\mathbf{1}_{p+1}'\right)+O_p\left(\frac{1}{\sqrt{T}}\sum\mathbf{1}_{p+1}\mathbf{1}_{p+1}'\frac{\ln^2 t}{t}\right).$$

While $\left\|T^{-1/2}\sum\mathbf{1}_{p+1}\mathbf{1}_{p+1}'O_p\left(\ln^2 t/t\right)\right\|=O_p\left(p\ln^3 T/\sqrt{T}\right)=o_p(1)$, the second term is trickier. Consider $\sum(\ln t/\sqrt{t})x_{t-h}$. Since $\ln t/\sqrt{t}=o\left(t^{-0.5+\lambda}\right)$ for any $\lambda>0$, we may choose some $\lambda\in(0,0.25)$ and write

$$\sum\frac{\ln t}{\sqrt{t}}x_{t-h}=o_p\left(T^{\lambda}\sum\frac{T^{0.5-\lambda}}{t^{0.5-\lambda}}\frac{x_{t-h}}{\sqrt{T}}\right).$$

But

$$\sum\frac{T^{0.5-\lambda}}{t^{0.5-\lambda}}\frac{x_{t-h}}{\sqrt{T}}\xrightarrow{d}\int_0^1 s^{0.5-\lambda}dW(s),$$

which is $O_p(1)$, since its variance, $\int_0^1 s^{1-2\lambda}ds$ is easily checked to be finite. Hence, $\sum(\ln t/\sqrt{t})x_{t-h}$ is $o_p(T^{\lambda})$, as is $\sum(\ln t/\sqrt{t})x_{t-h}^{**}$, in spite of long range dependence of x_{t-h}^{**}. Since $pT^{\lambda}/\sqrt{T}\to 0$, we have $T^{-0.5}\sum W_{tp}(\ln t/\sqrt{t})\mathbf{1}_{p+1}'=o_p(1)$ and, obviously, so is the expression (44). Finally, (42) can be shown to be $o_p(1)$ using Equation (7) and a similar reasoning as above, thus concluding the proof. ■

References

[1] Agiakloglou, C., Newbold, P. (1994), Lagrange multiplier tests for fractional difference; *Journal of Time Series Analysis* 15, 253-262.

[2] Amemiya, T. (1985), Advanced Econometrics; Harvard University Press, Cambridge, Massachusets.

[3] Andrews, D. W. K. (1991), Heteroskedasticity and autocorrelation consistent covariance matrix estimation; *Econometrica* 59, 817-858.

[4] Berk, K. N. (1974), Consistent autoregressive spectral estimates; *The Annals of Statistics* 2, 489-502.

[5] Bollerslev, T. (1986), Generalized autoregressive conditional heteroscedasticity; *Journal of Econometrics* 31, 307-327.

[6] Breitung, J., Hassler, U. (2002), Inference on the cointegration rank in fractionally integrated processes; *Journal of Econometrics* 110, 167-185.

[7] Brillinger, D. R. (1975), Times Series: Data Analysis and Theory; Holt, Rinehart and Winston, New York.

[8] Chang, Y., Park, J. (2002), On the asymptotics of ADF tests for unit roots; *Econometric Reviews* 21, 431-448.

[9] Davidson, J. (1994), Stochastic Limit Theory: An Introduction for Econometricians; Oxford University Press, Oxford.

[10] Demetrescu, M. (2006), On the Dickey-Fuller test with White standard errors; mimeo.

[11] Dickey, D. A., Fuller, W. A. (1979), Distribution of estimators for autoregressive time series with a unit root; *Journal of American Statistical Association* 74, 427-431.

[12] Dickey, D. A., Fuller, W. A. (1981), Likelihood ratio statistics for autoresgressive time series with a unit root; *Econometrica* 49, 1057-1072.

[13] Diebold, F. X., Rudebusch, G. D. (1991), On the power of Dickey-Fuller tests against fractional alternatives; *Economic Letters* 35, 155-160.

[14] Dolado, J. J., Gonzalo, J., Mayoral, L. (2002), A fractional Dickey-Fuller test for unit roots; *Econometrica* 70, 1963-2006.

[15] Gonçalves, S., Kilian, L. (2003), Asymptotic and bootstrap inference for $AR(\infty)$ processes with conditional heteroscedasticity; *CIRANO Working Paper 2003-s28.*

[16] Granger, C. W. J. (1981), Some properties of time series data and their use in econometric model specification; *Journal of Econometrics* 16, 121-130.

[17] Granger, C. W. J., Joyeux, R. (1980), An introduction to long memory time series models and fractional differencing; *Journal of Time Series Analysis* 1, 15-39.

[18] Haldrup, N. (1994), Heteroscedasticity in non-stationary time series, some Monte Carlo evidence; *Statistical Papers* 35, 287-307.

[19] Hassler, U., Breitung, J. (2006), A residual-based LM type test against fractional cointegration; *Econometric Theory* forthcoming.

[20] Hassler, U., Wolters, J. (1994), On the power of unit roots against fractionally integrated alternatives; *Economic Letters* 45, 1-5.

[21] Henry, M. (2001), Robust automatic bandwidth for long memory; *Journal of Time Series Analysis* 22, 293-316.

[22] Henry, M., Zaffaroni, P. (2003), The long-range dependence paradigm for macroeconomics and finance; In: Doukhan, P., Oppenheim, G., Taqqu M.S. (eds.), Theory and Applications of Long-Range Dependence; Birkhäuser, Boston, 417-438.

[23] Krämer, W. (1998), Fractional integration and the augmented Dickey-Fuller test; *Economics Letters* 61, 269-272.

[24] Leeb, H., Pötscher, B. M. (2005), Model selection and inference: Facts and fiction; *Econometric Theory* 21, 21-59.

[25] Lewis, R., Reinsel, G. C. (1985), Prediction of multivariate time series by autoregressive model fitting; *Journal of Multivariate Analysis* 16, 393-411.

[26] Lobato, I. N., Velasco, C. (2005), Optimal Fractional Dickey-Fuller Tests for Unit Roots; mimeo.

[27] Lütkepohl, H. (1996), Handbook of Matrices; Wiley, New York.

[28] Milhoj, A. (1985), The moment structure of ARCH processes; *Scandinavian Journal of Statistics* 12, 281-292.

[29] Ng, S., Perron, P. (1995), Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag; *Journal of the American Statistical Association* 90, 268-281.

[30] Robinson, P. M. (1991), Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression; *Journal of Econometrics* 47, 67-84.

[31] Robinson, P. M. (1994), Efficient tests of nonstationary hypotheses; *Journal of the American Statistical Association* 89, 1420-1437.

[32] Robinson, P. M. (1995), Gaussian semiparametric estimation of long range dependence; *Annals of Statistics* 23, 1630-1661.

[33] Robinson, P. M., Henry, M. (1999), Long and short memory conditional heteroskedasticity in estimating the memory parameter of levels; *Econometric Theory* 15, 299-336.

[34] Said, S. E., Dickey, D. A. (1984), Testing for unit roots in autoregressive-moving average models of unknown order; *Biometrika* 71, 599-607.

[35] Schwert, G. W. (1989), Tests for unit roots: A Monte Carlo investigation; *Journal of Business and Economic Statistics* 7, 147-160.

[36] Sowell, F. B. (1990), The fractional unit root distribution; *Econometrica* 58, 495-505.

[37] Tanaka, K. (1999), The nonstationary fractional unit root; *Econometric Theory* 15, 549-582.

[38] Velasco, C. (1999), Gaussian semiparametric estimation of non-stationary Time Series; *Journal of Time Series Analysis* 20, 87-127.