Semiparametric Sieve-Type Generalized Least Squares Inference*

George Kapetanios†
Queen Mary, University of London

Zacharias Psaradakis‡
Birkbeck, University of London

December 2011; Revised June 2013

Abstract

This paper considers the problem of statistical inference in linear regression models with dependent errors. A sieve-type generalized least squares (GLS) procedure is proposed based on an autoregressive approximation to the generating mechanism of the errors. The asymptotic properties of the sieve-type GLS estimator are established under general conditions, including mixingale-type conditions as well as conditions which allow for long-range dependence in the stochastic regressors and/or the errors. A Monte Carlo study examines the finite-sample properties of the method for testing regression hypotheses.

Key Words: Autoregressive approximation; Generalized least squares; Linear regression; Long-range dependence; Short-range dependence.

JEL Classification: C22.

*The authors wish to thank an associate editor and a referee for their helpful comments and suggestions.

†School of Economics and Finance, Queen Mary, University of London, Mile End Road, London E1 4NS, U.K.; e-mail: g.kapetanios@qmul.ac.uk.

‡Department of Economics, Mathematics & Statistics, Birkbeck, University of London, Malet Street, London WC1E 7HX, U.K.; e-mail: z.psaradakis@bbk.ac.uk.

1 Introduction

Robust inference in linear regression models has attracted considerable attention in the literature. It is well known that estimators that are optimal in cross-sectional settings remain asymptotically valid in time series regressions under general regularity conditions. A leading example is the ordinary least squares (OLS) estimator of the parameters of regression models with strictly exogenous regressors and dependent errors. In such cases autocorrelation-robust inference is typically carried out by relying on the asymptotic normality of the OLS estimator and a so-called heteroskedasticity and autocorrelation consistent estimator of its asymptotic covariance matrix (see, e.g., Newey and West (1987); Andrews (1991)).

However, a number of problems remain under such a strategy. The most widely reported problem relates to the poor finite-sample properties of robust inference procedures and in particular to the large size distortions that related tests exhibit (see, e.g., den Haan and Levin (1997)). Attempts to deal with this issue, either directly or indirectly, can be found in Kiefer, Vogelsang, and Bunzel (2000), Andrews and Monahan (1992) and Jansson (2004), inter alia. A second problem that has received less attention in the literature arises from the fact that the structure of the dependence in the regression errors is not routinely exploited in order to improve the properties of robust inference procedures. Of course, the fact that the dependence structure of the errors is generally unknown poses a significant hurdle in this respect.

The present paper attempts to address the second problem and indirectly provides a means of addressing the first. In particular, we suggest using generalized least squares (GLS) as an alternative to OLS for estimating the regression coefficients. One of the advantages of this approach is that the GLS estimator is known to be the best linear unbiased estimator (BLUE) of the regression coefficients under very general conditions. The obvious difficulty, however, is that GLS requires knowledge of the covariance structure of the errors. To overcome this difficulty, we propose to follow Amemiya (1973) in employing a GLS procedure which approximates the generating mechanism of the errors by an autoregressive model (for the residuals) the order of which grows slowly with the sample size; this model is then used to obtain an estimate of the error covariance matrix that is needed for the computation of feasible GLS (FGLS) estimates. Such an approach is semiparametric in the sense that no particular finite-parameter model for the errors is assumed; instead, the infinite-parameter error process is approximated by a sequence of autoregressive models of finite but increasing order. The sequence of approximations may be viewed as a sieve, in the sense of Grenander (1981), which is why we refer to this procedure as semiparametric sieve-type GLS. Amemiya (1973) established the asymptotic normality and asymptotic Gauss–Markov efficiency of an FGLS estimator based on such an autoregressive approximation when the regressors are nonstochastic and the errors are generated by a linear process with independent and identically distributed (i.i.d.) innovations.

In this paper we extend the work of Amemiya (1973) in two important respects. Firstly, we generalize the analysis to allow for a much larger class of stochastic processes, as well as for autoregressive approximations the order of which is data-dependent and determined by means of information criteria or sequential testing. Secondly, unlike the papers cited in preceding paragraphs, all of which deal with models with regressors and errors that are short-range dependent (in the sense of having absolutely summable autocovariances), we also consider models in which the regressors and/or the errors may exhibit long-range dependence (in the sense of having autocovariances which are not absolutely summable). In models of the latter type, the OLS estimator of the regression coefficients not only fails to attain the Gauss–Markov efficiency bound but may also have a slow rate of convergence and a non-Gaussian asymptotic distribution (see, e.g., Robinson (1994); Chung (2002)). As a result, the use of robust OLS-based inferential procedures cannot be justified. Furthermore, even in models with nonstochastic regressors, where the OLS estimator is asymptotically normal for certain designs, asymptotic efficiency is generally unattainable and the rate of convergence may be slow (cf. Yajima (1988, 1991); Dahlhaus (1995)). By contrast, the GLS estimator, and suitable approximations thereof, are known to have the desirable properties of asymptotic normality and Gauss–Markov efficiency even under circumstances in which the OLS estimator has a non-Gaussian asymptotic distribution or a slow rate of convergence (cf. Robinson and Hidalgo (1997); Choy and Taniguchi (2001)).

GLS-based inference in the semiparametric context of autocorrelation of unknown form, although less studied than OLS-based inference, has been analyzed in a number of papers. Almost all of the available work, Amemiya (1973) being the obvious exception, address the problem by using semiparametric estimates of the error spectral density. In a short-range dependent environment, examples of this approach include

Hannan (1963, 1970) and Robinson (1991), among others. It is worth noting, however, that the small-sample properties of inference procedures based on such frequency-domain GLS estimators tend to be comparably poor (and, for very small samples, considerably so) compared to those obtained via OLS (cf. Robinson (1991)). In the presence of long-range-dependence, a related approach was investigated by Hidalgo and Robinson (2002), who demonstrated that the unknown spectral density of the errors may be replaced by a suitable smoothed nonparametric estimator without any effect on the first-order asymptotic distribution of their (approximate) GLS estimator. If the spectral density of the errors is a known function of finitely many unknown parameters, Robinson and Hidalgo (1997) showed that it is also possible to replace the latter by suitable estimates and employ a frequency-domain FGLS procedure. Nielsen (2005) considered an alternative frequency-domain FGLS estimator which uses only periodogram ordinates in a degenerating band around the origin in conjunction with a consistent estimator of the memory parameter of the errors.

A related strand of the literature deals with inference on the parameters of the trend function of a time series when deviations from the trend may or may not be integrated of order one, or I(1). Notable papers in this area include Perron and Yabu (2009, 2012). In these papers the authors postulate an autoregressive structure for the deviations from a deterministic trend function and use FGLS to carry out inference. In the case of I(1) deviations from the trend, inference is improved by the use of 'super-efficient' estimation of the parameters associated with the autoregressive structure of the detrended series. Perron and Yabu (2012) also consider the implications of allowing for deterministic breaks in the trend function. While the aforementioned papers make use of autoregressive approximations, as we do, and allow for I(1) components in the data-generating process, they are specifically concerned with inference on the trend component of a time series or, more generally, on deterministic components, and do not allow for the general stochastic-regressor setting considered here. In that sense, they are complementary to our analysis. A related set of contributions have been made by Harris, Harvey, Leybourne, and Taylor (2009) and Harvey, Leybourne, and Taylor (2007), who consider the problem of unit-root testing in the presence of breaks and inference on the parameters of a deterministic linear trend function in the presence of possibly I(1) noise, respectively. The former paper extends the work on GLS-based unit-root testing by Elliott, Rothenberg, and Stock (1996) and others, while the latter is based on a variant of the procedure of Newey and West (1987), suitably modified to allow for the possibility that the detrended series is I(1).

The remainder of the paper is organized as follows. Section 2 introduces the model and describes the sieve-type FGLS estimation procedure. Section 3 establishes the asymptotic properties of the FGLS procedure under conditions that allow for short-range or long-range dependence in the errors and/or regressors. Section 4 reports the results of a simulation study of the small-sample performance of the method in the context of testing regression hypotheses. Section 5 summarizes and concludes. Mathematical proofs and supporting lemmata are collected in the Appendix.

2 Model and Estimation Procedure

Consider a regression model of the form

$$y_t = x_t'\beta + u_t, \qquad t = 1, \ldots, T, \tag{1}$$

where y_t is the observable dependent variable, $x_t = (x_{1,t}, \ldots, x_{k,t})'$ is a k-dimensional vector of observable explanatory variables (with $x_{k,t} = 1$ for all t), $\beta = (\beta_1, \ldots, \beta_k)'$ is a k-dimensional vector of unknown parameters, and u_t is an unobservable random error term. The aim is inference on β.

Letting B denote the backward shift operator, we make the following assumption about the errors in (1).

Assumption 1 (i) $\{u_t\}$ *is a second-order stationary process satisfying*

$$u_t = (1-B)^{-d_u}\xi_t, \qquad \xi_t = \sum_{j=0}^{\infty} \delta_j \varepsilon_{t-j}, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{2}$$

for some $0 \le d_u < 1/2$*;* (ii) $\{\delta_j; j \ge 0\}$ *is an absolutely summable sequence of constants* (*with* $\delta_0 = 1$) *satisfying* $\sum_{j=0}^{\infty} \delta_j z^j \neq 0$ *for all complex* z *with* $|z| \le 1$*;* (iii) $\{\varepsilon_t\}$ *is an ergodic sequence of random variables such that* $E(\varepsilon_t | \mathcal{F}_{t-1}^{\varepsilon}) = 0$ *a.s.,* $E(\varepsilon_t^2 | \mathcal{F}_{t-1}^{\varepsilon}) = \sigma_\varepsilon^2 > 0$ *a.s., and* $\sup_t E(|\varepsilon_t|^4) < \infty$*,* $\mathcal{F}_t^{\varepsilon}$ *being the* σ*-field generated by* $\{\varepsilon_s; s \le t\}$*.*

As usual, for any non-integer real d, the operator $(1-B)^{-d}$ is defined by the series

$$(1-B)^{-d} = 1 + \sum_{j=1}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} B^j,$$

where $\Gamma(\cdot)$ is the gamma function (with the convention $1/\Gamma(0) = 0$). Assumption 1 requires $\{u_t\}$ to be a fractionally integrated process with memory (or long-range dependence, or fractional differencing) parameter d_u. Under this assumption, $\{u_t\}$ admits the AR(∞) representation

$$u_t = \sum_{j=1}^{\infty} \phi_j u_{t-j} + \varepsilon_t, \qquad t = 0, \pm 1, \pm 2, \ldots. \tag{3}$$

for some absolutely summable sequence of constants $\{\phi_j; j \geq 1\}$ satisfying $\phi_j \sim j^{-d_u - 1}$ as $j \to \infty$ ($d_u \neq 0$), where '$\sim$' signifies asymptotic proportionality. Similarly, it can be easily seen that $\{u_t\}$ admits the causal MA(∞) representation

$$u_t = \varepsilon_t + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j}, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{4}$$

for some square-summable sequence of constants $\{\psi_j; j \geq 1\}$ satisfying $\psi_j \sim j^{d_u - 1}$ as $j \to \infty$ ($d_u \neq 0$).

Assumption 1 covers many important families of long-range dependent processes, including autoregressive fractionally integrated moving average (ARFIMA) processes (Granger and Joyeux (1980); Hosking (1981)). When $d_u = 0$ in (2), the regression errors are a short-range dependent linear processes with martingale-difference innovations and a spectral density that is continuous and positive at the origin. We note that, for $0 < d_u < 1/2$, Assumption 1 is stronger than the corresponding assumption of Robinson and Hidalgo (1997), who only require the existence of the MA(∞) representation in (4) with square-summable coefficients and innovations which satisfy the same conditions as ours. Hidalgo and Robinson (2002), on the other hand, essentially require the innovations in (4) to behave like an i.i.d. sequence up to the 12th moment. The analysis in Amemiya (1973) requires $\{u_t\}$ to be a short-range dependent process that admits an AR(∞) representation like (3) with $\sum_{j=1}^{\infty} |\phi_{n+j}| = O(\varrho^n)$ for some $0 < \varrho < 1$ and $\{\varepsilon_t\}$ being an i.i.d. sequence.

Letting $y = (y_1, \ldots, y_T)'$, $X = (x_1, \ldots, x_T)'$ and $u = (u_1, \ldots, u_T)'$, the BLUE of the parameter β in (1) is

$$\widetilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \tag{5}$$

where $\Omega = E(uu')$. When Ω is unknown, an FGLS estimator must be used instead of the BLUE. Following Amemiya (1973), we propose to construct such an estimator by using an approximation to the AR(∞) representation of $\{u_t\}$ given in (3).

To describe the method in detail, let

$$\bar{u}_t = y_t - x_t'\bar{\beta}, \qquad t = 1, \ldots, T, \tag{6}$$

where $\bar{\beta}$ is a preliminary estimator of β. Further, for some positive integer h, chosen as a function of T so that $h \to \infty$ and $h/T \to 0$ as $T \to \infty$, let $\widehat{\phi}_h = (\widehat{\phi}_{h,1}, \ldots, \widehat{\phi}_{h,h})'$ be the hth-order OLS estimator of the autoregressive coefficients for $\{\bar{u}_t\}$, obtained as the solution to the minimization of

$$(T-h)^{-1} \sum_{t=h+1}^{T} (\bar{u}_t - \phi_{h,1}\bar{u}_{t-1} - \cdots - \phi_{h,h}\bar{u}_{t-h})^2 \tag{7}$$

over $\phi_h = (\phi_{h,1}, \ldots, \phi_{h,h})' \in \mathbb{R}^h$. A computationally attractive FGLS estimator of β may then be obtained as

$$\widehat{\beta} = (X'\widehat{\Phi}'\widehat{\Phi}X)^{-1}X'\widehat{\Phi}'\widehat{\Phi}y, \tag{8}$$

where $\widehat{\Phi}$ is the $(T-h) \times T$ matrix defined as

$$\widehat{\Phi} = \begin{pmatrix} -\widehat{\phi}_{h,h} & -\widehat{\phi}_{h,h-1} & & \cdots & & -\widehat{\phi}_{h,1} & 1 & & & \\ & -\widehat{\phi}_{h,h} & -\widehat{\phi}_{h,h-1} & & \cdots & & -\widehat{\phi}_{h,1} & 1 & 0 & \\ & & & & & \vdots & & & & \\ & 0 & & & \cdots & & & & 1 & \\ & & & & -\widehat{\phi}_{h,h} & -\widehat{\phi}_{h,h-1} & & \cdots & -\widehat{\phi}_{h,1} & 1 \end{pmatrix}.$$

This, of course, is equivalent to applying OLS to the regression of $(1 - \sum_{j=1}^{h} \widehat{\phi}_{h,j}B^j)y_t$ on $(1 - \sum_{j=1}^{h} \widehat{\phi}_{h,j}B^j)x_t$.[1]

Remark 1 An alternative FGLS estimator of β, with the same asymptotic properties as $\widehat{\beta}$, can be obtained by replacing $\widehat{\Phi}'\widehat{\Phi}$ in (8) by $\widehat{\Omega}^{-1}$, where

$$\widehat{\Omega} = \left\{ \int_{-\pi}^{\pi} e^{\imath(t-s)\omega} \widehat{f}_{\bar{u}}(\omega) \mathrm{d}\omega; t, s = 1, \ldots, T \right\}.$$

Here $\widehat{f}_{\bar{u}}(\cdot)$ is the estimator of the spectral density of $\{u_t\}$ associated with the autoregressive approximation based on $\bar{u}_t$, i.e.,

$$\widehat{f}_{\bar{u}}(\omega) = \frac{\widehat{\sigma}^2_{\bar{u},h}}{2\pi} \left| 1 - \sum_{j=1}^{h} \widehat{\phi}_{h,j} e^{\imath j\omega} \right|^{-2}, \qquad -\pi < \omega \leq \pi,$$

where $\widehat{\sigma}^2_{\bar{u},h}$ is the minimum of (7) and $\imath = \sqrt{-1}$.

[1] We note that Yule–Walker or Burg-type estimates may be used instead of the OLS estimates $\widehat{\phi}_h$ in the construction of $\widehat{\Phi}$ without changing the first-order asymptotic properties of $\widehat{\beta}$.

Remark 2 Using the discrete Fourier transforms of X and y, the BLUE $\widetilde{\beta}$ may be expressed as

$$\widetilde{\beta} = (\ddot{X}^* W \Omega^{-1} W^* \ddot{X})^{-1} \ddot{X}^* W \Omega^{-1} W^* \ddot{y},$$

where $W = \{T^{-1/2} \exp(2\pi\imath ts/T); t, s = 0, 1, \ldots, T-1\}$, $\ddot{X} = WX$, $\ddot{y} = Wy$, and the asterisk denotes matrix transposition combined with complex conjugation. It is well known that, when T is large, the matrix $W\Omega^{-1}W^*$ is approximately diagonal with elements $f_u(2\pi j/T)^{-1}$, $0 \leq j \leq T-1$, where $f_u(\cdot)$ is the spectral density of $\{u_t\}$ (cf. Grenander and Szegö (1958, p. 62)). Hence, the time-domain GLS estimator given in (5) can be shown to be asymptotically equivalent to the frequency-domain estimator

$$\widetilde{\beta}_f = (\ddot{X}^* Q^{-1} \ddot{X})^{-1} \ddot{X}^* Q^{-1} \ddot{y}, \tag{9}$$

where $Q = \text{diag}\{f_u(2\pi j/T); j = 0, 1, \ldots, T-1\}$. The approximate GLS estimator in (9) is a member of the family of estimators considered by Robinson and Hidalgo (1997), whose frequency-domain weighted least-squares estimator of $(\beta_1, \ldots, \beta_{k-1})'$ may be expressed as

$$\widetilde{\beta}_g = \left(\sum_{j=1}^{T-1} \mathcal{I}_{xx}(2\pi j/T) g(2\pi j/T) \right)^{-1} \left(\sum_{j=1}^{T-1} \mathcal{I}_{xy}(2\pi j/T) g(2\pi j/T) \right). \tag{10}$$

Here, $\mathcal{I}_{xx}(\cdot)$ and $\mathcal{I}_{xy}(\cdot)$ stand for the periodogram of $\{(x_{1,t}, \ldots, x_{k-1,t})'\}$ and the cross-periodogram of $\{(x_{1,t}, \ldots, x_{k-1,t})'\}$ and $\{y_t\}$, respectively, and $g(\cdot)$ is a real-valued, integrable, even, and periodic function on $[-\pi, \pi]$ with period 2π. If $g(\cdot) = f_u(\cdot)^{-1}$, then $\widetilde{\beta}_g = (\widetilde{\beta}_{1,f}, \ldots, \widetilde{\beta}_{k-1,f})'$, where $\widetilde{\beta}_{i,f}$ denotes the ith element of $\widetilde{\beta}_f$.

3 Asymptotic Results

3.1 Short-Range Dependence

We first consider the asymptotic properties of the FGLS estimator in the case where the errors and regressors in (1) are both short-range dependent. More specifically, we assume that the regression errors satisfy (2) with $d_u = 0$, and strengthen the assumptions about the MA(∞) innovations and weights as follows.

Assumption 2 (i) $\{\varepsilon_t\}$ *is φ-mixing of size $-\eta$, for some $\eta > 1$;* (ii) $\sup_t E(|\varepsilon_t|^{2\kappa}) < \infty$ *for some $\kappa > 2$;* (iii) $\sum_{j=0}^{\infty} j^{\zeta} |\delta_j| < \infty$ *for some $\zeta \geq 1/2$ such that $2\eta \geq \zeta$ and $\zeta(\kappa-2)/\{2(\kappa-1)\} \geq 1/2$.*

Regarding the regressors, we make the following assumption.

Assumption 3 (i) $\{x_t\}$ *is* L_{2r}*-bounded,* L_2*-near-epoch-dependent* (L_2*-NED*) *of size* $-\zeta$ *on a* k*-dimensional* φ*-mixing process of size* $-\eta$*,* $\eta > 1$*, such that* $2\eta \geq \zeta$ *and* $\zeta(r-2)/\{2(r-1)\} \geq 1/2$ *for some* $r > 2$*;* (ii) $\{x_t\}$ *and* $\{\varepsilon_t\}$ *are mutually independent;* (iii) $E(x_t x_t')$ *and* $E(T^{-1}X'\Omega^{-1}X)$ *are nonsingular.*

Assumption 3(i) is quite general and, together with Assumptions 1 and 2, implies that $\{(u_t, x_t')'\}$ is NED in L_2-norm on a φ-mixing base process.[2] This requirement is relatively mild compared to other conditions that are commonly used in the robust inference literature (such as mixing), and allows for a wide variety of linear and non-linear short-range dependent processes with bounded fourth moments (cf. Davidson (1994, 2002)). The level of detail is needed since the properties of $\{u_t\}$ are important for deriving all the results and there exists a degree of trade-off on the stringency of the conditions applied to $\{x_t\}$ and $\{u_t\}$ in order to obtain the central limit theorem that is needed in the proof of Theorem 1 below. Of course, there exist alternative sets of dependence and heterogeneity conditions that can be used to establish our results. An example are the conditions considered in Section 3.2; unlike Assumption 3(i), however, these impose conditional homoskedasticity on $\{x_t\}$. Assumption 3(i) is a substantially weakening of the assumptions in Amemiya (1973), which require $\{x_t\}$ to be a nonstochastic sequence satisfying the so-called Grenander conditions.

Assumption 3(ii) is a rather restrictive exogeneity condition. As remarked by Robinson (1991), Robinson and Hidalgo (1997) and Hidalgo and Robinson (2002), who use it as well, the assumption could probably be relaxed to a milder orthogonality condition, albeit at the cost of greater structure on $\{x_t\}$. The leading case where this assumption is violated is when lagged values of the dependent variable appear as regressors in (1). Whereas the exogeneity requirement may appear restrictive, it is implicit in some form in all work on robust inference as error autocorrelation combined with the presence of lagged dependent variables among the regressors leads to inconsistency of the OLS estimator. Assumption 3(iii) is a conventional full-rank condition that precludes multicollinearity.

Before stating our main result, we introduce two more assumptions about the preliminary estimator $\bar{\beta}$ of the regression coefficients and the order h of the autoregressive

[2] For a definition and more information on the φ-mixing and NED concepts, and the related size terminology, the reader is referred to Davidson (1994).

approximation used to obtain $\widehat{\Phi}$.

Assumption 4 $\bar{\beta} = \beta + O_p(T^{-1/2})$ *as* $T \to \infty$.

Assumption 5 (i) $h = h_T \xrightarrow{p} \infty$ *as* $T \to \infty$;(ii) $h_T = o_p\left(\{T/\ln T\}^{1/4}\right)$ *as* $T \to \infty$.

Assumption 4 is a mild requirement that is satisfied by many estimators, including the OLS estimator. Assumption 5 defines probabilistic conditions on the rate of growth of the sequence h_T. It does not necessarily specify whether the sequence is deterministic (in which case the probabilistic conditions degenerate to their deterministic equivalents) or data-dependent and, therefore, stochastic. Later results will provide conditions under which a data-dependent sequence h_T satisfies Assumption 5. The assumption is consistent with the view that $\{u_t\}$ is neither white noise nor, more generally, a finite-order autoregressive process. In the latter case, a finite h at least equal to the true autoregressive order would suffice, and existing standard results would provide justification for our FGLS approach. As the case of an AR(∞) process is more interesting and realistic, we focus on this and disallow the simpler case of finite-order autoregressive dynamics for $\{u_t\}$. This is formalized in the following assumption.

Assumption 6 $\{u_t\}$ *does not degenerate to a finite-order autoregressive process.*

The following theorem establishes the asymptotic equivalence of the FGLS estimator $\widehat{\beta}$ and the BLUE $\widetilde{\beta}$.

Theorem 1 *Suppose Assumptions* 1, 2, 3, 4 *and* 5 *hold and* $d_u = 0$. *Then*

$$\sqrt{T}(\widehat{\beta} - \widetilde{\beta}) = o_p(1) \qquad \text{as } T \to \infty,$$

and hence

$$\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V) \qquad \text{as } T \to \infty,$$

where $V = \operatorname*{plim}_{T\to\infty} \left(T^{-1}X'\Omega^{-1}X\right)^{-1}$.

An important practical issue in sieve-type GLS inference is the choice of the approximating autoregressive order h in (7). Automatic data-dependent choices for h using either information criteria or sequential testing are natural strategies to consider.

Information criteria take the general form

$$\ln \widehat{\sigma}^2_{\bar{u},h} + hC_T/T, \tag{11}$$

where $\widehat{\sigma}^2_{\bar{u},h}$ is the minimum of (7) and C_T is an information criterion specific penalty term. For instance, in the case of the familiar Akaike, Bayesian and Hannan–Quinn information criteria the penalty terms are $C_T = 2$, $C_T = \ln T$ and $C_T = 2b \ln\ln T$ ($b > 1$), respectively. The following theorem provides the theoretical justification for the use of information criteria in our setting.

Theorem 2 *Suppose Assumptions* 1, 2, 3, 4 *and* 6 *hold and* $d_u = 0$. *Let* h_T *be chosen by minimizing* (11) *over* $h \in \mathcal{H}_T$, *where* $\mathcal{H}_T = \{1, 2, \ldots, H^*_T\}$, $H^*_T = o\left(\{T/\ln T\}^{1/4}\right)$ *as* $T \to \infty$, $C_T > 1$ *and* $C_T = o\left(\{T^3 \ln T\}^{1/4}\right)$ *as* $T \to \infty$. *Then* h_T *satisfies Assumption* 5.

As an alternative to information criteria, one may use sequential testing. This amounts to starting with a general autoregressive model of order H^*_T and sequentially testing the significance of, and removing, the highest-order lag if its coefficient is found to be insignificantly different from zero. The sequence of tests stops if the highest-order lag considered is found to be significant. More specifically the following algorithm may be adopted.

Algorithm 1 (Selection of h using sequential testing)

Step 1: Start with an AR(H^*_T) model for the residual $\{\bar{u}_t\}$. Set $h^* = H^*_T$. Go to Step 2.

Step 2: Using a conventional t-test with significance level α, test the significance of the coefficient on the h^*th-order lag in an AR(h^*) model for $\bar{u}_t$. If found significant, stop and set $h_T = h^*$. Otherwise, go to Step 3.

Step 3: Set $h^* = h^* - 1$. If $h^* = 0$, set $h_T = 1$. Otherwise, go to to Step 2.

We make the following assumption about the maximum allowable autoregressive order H^*_T in Algorithm 1.

Assumption 7 $H^*_T = o\left(\{T/\ln T\}^{1/4}\right)$ *and* $\sum_{j=H^*_T+1}^{\infty} |\delta_j| = o(T^{-1/2})$ *as* $T \to \infty$.

Our next theorem establishes the validity of the procedure based on sequential testing.

Theorem 3 *Suppose Assumptions* 1, 2, 3, 4, 6 *and* 7 *hold and* $d_u = 0$. *Let* h_T *be chosen using Algorithm* 1 *with* $\alpha \in (0, 1)$. *Then* h_T *satisfies Assumption* 5.

Remark 3 The conditions given in Assumption 7 clearly depend on the nature of the data-generating mechanism for $\{u_t\}$. Hence, in order to minimize dependence on these conditions, it is reasonable in practice to set the upper bound H_T^* to values close to $T^{1/4}$, e.g., $H_T^* = \lfloor T^{1/4} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer-part function. Also note that the assumption imposes both an upper and a lower bound on the rate of growth of h. This makes the selection procedure based on information criteria to be valid under slightly less restrictive conditions for the problem considered in this paper.

We end this section by considering an important question relating to the power of a conventional hypothesis test based on the FGLS estimator of β compared to the corresponding test based on the OLS estimator. Clearly both tests are consistent. Although the inefficiency of the OLS-based test is obvious, the result becomes clearer in terms of local power, on which we briefly comment.

In the case of a single regressor (which, for the purposes of the discussion on local power only, may be assumed to be different from a constant), the null hypothesis $\mathrm{H}_0 : \beta = \beta^0$ may be tested using the GLS-based t-statistic $(\widehat{V}/T)^{-1/2}(\widehat{\beta} - \beta^0)$, where $\widehat{V}$ is a consistent estimator of V.[3] It is then straightforward to show that, under a sequence of local alternatives of the form $\mathrm{H}_{1T} : \beta = \beta_T = \beta^0 + cT^{-1/2}$, where c is a constant, $(\widehat{V}/T)^{-1/2}(\widehat{\beta} - \beta^0) \xrightarrow{\mathcal{D}} \mathcal{N}(cV^{-1/2}, 1)$ as $T \to \infty$. Similarly, for the OLS-based t-statistic we have $(\widehat{V}_{\mathrm{OLS}}/T)^{-1/2}(\widehat{\beta}_{\mathrm{OLS}} - \beta^0) \xrightarrow{\mathcal{D}} \mathcal{N}(cV_{\mathrm{OLS}}^{-1/2}, 1)$ as $T \to \infty$ under H_{1T}, where $\widehat{\beta}_{\mathrm{OLS}}$, V_{OLS}, and $\widehat{V}_{\mathrm{OLS}}$ denote, respectively, the OLS estimator of β, its asymptotic variance and a consistent estimator of V_{OLS}. Since, $V \leq V_{\mathrm{OLS}}$, the local power of the GLS-based test dominates that of the OLS-based test.

3.2 Long-Range Dependence

We now consider the case where the errors and/or regressors in (1) may exhibit long-range dependence. In order to allow for this possibility, we make the following

[3] The estimator $\widehat{V}$ can be easily obtained from the OLS regression of $(1 - \sum_{j=1}^{h} \widehat{\phi}_{h,j} B^j) y_t$ on $(1 - \sum_{j=1}^{h} \widehat{\phi}_{h,j} B^j) x_t$.

assumptions about the regressors (here, and in the sequel, $\|\cdot\|$ denotes the Euclidean norm for vectors and matrices, and I_k is the identity matrix of order k).

Assumption 8 (i) $\{x_t\}$ *is a fourth-order stationary process satisfying*

$$x_t - \mu_x = D(B)v_t, \qquad v_t = \sum_{j=0}^{\infty} \Psi_j w_{t-j}, \qquad t = 0, \pm 1, \pm 2, \ldots. \tag{12}$$

where $D(B) = \text{diag}\{(1-B)^{-d_1}, \ldots, (1-B)^{-d_k}\}$ *for some* $0 \leq d_i < 1/2$ $(i = 1, \ldots, k)$ *and* $\mu_x \in \mathbb{R}^k$ *is a constant;* (ii) $\{\Psi_j\}$ *is an absolutely summable sequence of non-stochastic* $k \times k$ *matrices* (*with* $\Psi_0 = I_k$)*;* (iii) $\{w_t\}$ *is an ergodic sequence of* k*-dimensional random vectors such that* $E(w_t|\mathcal{F}_{t-1}^w) = 0$ *a.s.,* $E(w_t w_t'|\mathcal{F}_{t-1}^w) = \Sigma_w$ *a.s., with* $|\Sigma_w| > 0$ *and* $\|\Sigma_w\| < \infty$*, and* $\sup_t E(\|w_t\|^4) < \infty$*,* $\mathcal{F}_t^w$ *being the* σ*-field generated by* $\{w_s; s \leq t\}$*.*

Assumption 9 $\{w_t\}$ *and* $\{\varepsilon_t\}$ *are mutually independent.*

Assumption 8, which is a multivariate variant of Assumption 1, is fairly mild and allows for stochastic regressors which may exhibit long-range dependence. Under this assumption, the k-variate fractionally integrated process defined by (12) admits a causal MA(∞) representation with square-summable coefficients which decay at slow hyperbolic rates. Needless to say, Assumption 8 also caters for the possibility that the regressors are short-range dependent, albeit under stricter conditions than those in Assumption 3; if $d_1 = \cdots = d_k = 0$, then $\{x_t\}$ is a linear process with martingale-difference innovations and a continuous spectral density matrix. By way of comparison, we note that Robinson and Hidalgo (1997) and Hidalgo and Robinson (2002) assume that $\{x_t\}$ is an ergodic, fourth-order stationary process satisfying a suitable cumulant condition. The maintained assumption in Nielsen (2005) is that $\{(x_t', u_t)'\}$ is a linear process with square-summable weights, fourth-order stationary martingale-difference innovations, and a spectral density matrix which satisfies certain long-range dependence conditions.

The strict exogeneity of the explanatory variables imposed by Assumption 9 is admittedly restrictive. We again note the remark by Robinson and Hidalgo (1997) and Hidalgo and Robinson (2002), that the assumption could probably be relaxed to a milder orthogonality condition, albeit at the cost of greater structure on $\{x_t\}$ and greater technical complexity. However, as mentioned before, even such a condition would rule out dynamic specifications in which the regressors include lagged values of the regressand. The assumption used by Nielsen (2005) is essentially a local (in the neighborhood of the zero frequency) version of the usual orthogonality condition.

In order to establish the asymptotic distribution of the FGLS estimator, Assumption 4 will be needed, which requires the preliminary estimator $\bar{\beta}$ to be $\sqrt{T}$-consistent. In the presence of long-range dependence, this generally rules out the OLS estimator, which has a rate of convergence slower than $O_p(T^{-1/2})$ when at least one of the regressors in (1) has memory parameter d_i such that $d_i + d_u > 1/2$ (cf. Robinson (1994); Chung (2002)). However, one may use the weighted least-squares estimator in (10) with a weight function $g(\cdot)$ which satisfies the conditions of Robinson and Hidalgo (1997); an example of such a function is $g(\omega) = |1 - e^{\imath\omega}|$.

We shall also make use of the following assumption, which requires h to increase with the sample size at a suitable rate.

Assumption 10 *$h = h_T \to \infty$ as $T \to \infty$ so that $h_T = O(\{\ln T\}^r)$ for some $0 < r < \infty$.*

The theorem below gives sufficient conditions for the FGLS estimator $\widehat{\beta}$ to have the same asymptotic distribution as the BLUE $\widetilde{\beta}$ in the presence of long-range dependence. The asymptotic distribution of the BLUE can be found in Robinson and Hidalgo (1997).

Theorem 4 *Suppose Assumptions* 1, 4, 8, 9 *and* 10 *hold. If, in addition,*

$$(5/2)d_u + (1/2)\max\{d_i; 1 \leq i \leq k\} < 1, \tag{13}$$

then

$$\sqrt{T}(\widehat{\beta} - \widetilde{\beta}) = o_p(1) \qquad \text{as } T \to \infty.$$

Remark 4 We conjecture that the result of the above theorem will hold for all linear, long-range dependent, stationary processes $\{x_t\}$ and $\{u_t\}$, rather than just those satisfying (13). Our simulation results suggest that this is the case. Note also that condition (13) is more flexible than conditions usually invoked in the literature to derive standard (short-range dependence type) asymptotics for long-range dependent processes. These are usually of the form $d_u < 1/4$ or $\max\{d_i; 1 \leq i \leq k\} < 1/4$. By contrast, $d_u < 2/5$ and $\max\{d_i; 1 \leq i \leq k\} < 1/2$ are cases that can be potentially accommodated by our results.

As in the short-range dependent case discussed in Section 3.1, data-dependent choices of the approximating autoregressive order $h = h_T$ can be allowed for. Any choice h_T obtained by using a data-driven selection procedure which guarantees that h_T satisfies Assumption 10 with probability approaching one as $T \to \infty$ is sufficient for Theorem 4 to hold. Using the results in Poskitt (2007), it can be shown, in a similar way to Theorem 2, that information criteria such as the familiar Akaike, Bayesian and Hannan–Quinn criteria return a lag order h_T which is asymptotically acceptable, provided the maximum allowable order is allowed to grow to infinity with T at a rate $O(\{\ln T\}^r)$ for some $0 < r < \infty$.

4 Monte Carlo Experiments

The theoretical part of the paper has argued that the approach of Amemiya (1973) can be extended to cover a rich class of short-range dependent and long-range dependent processes, be made automatic through the use of information criteria and/or sequential tests, and be applied to the problem of robust inference. Nevertheless, it is not clear to what extent the asymptotic results provide good small-sample approximations. The aim of the Monte Carlo study of this section is to provide some answers to this question by examining the small-sample size and power properties of hypothesis tests.

4.1 Experimental Design

In our numerical experiments, artificial data $\{y_t\}$ are generated according to the model

$$y_t = \beta_1 x_{1,t} + \beta_2 + u_t, \qquad t = 1, \ldots, T,$$

with $\beta_1 \in \{0, 0.05, 0.1, 0.2, 0.5\}$ and $\beta_2 = 0$. The data-generating mechanism for $\{x_{1,t}\}$ is either the AR(1) model $(1 - 0.5B)x_{1,t} = w_t$ or the ARFIMA$(0, d_x, 0)$ model $x_{1,t} = (1-B)^{-d_x} w_t$ with $d_x \in \{0.2, 0.4\}$; in either case, $\{w_t\}$ are i.i.d. $\mathcal{N}(0, 1)$ random variables. Similarly, the errors are allowed to exhibit either long-range or short-range dependence. In the former case, $\{u_t\}$ is generated as the ARFIMA$(1, d_u, 0)$ process $(1-\phi B)u_t = (1-B)^{-d_u}\varepsilon_t$ with $\phi \in \{0, 0.5, 0.9, 0.98\}$ and $d_u \in \{0.2, 0.4\}$. In the latter case, we consider the following data-generating processes ($\mathbb{I}(\cdot)$ denotes the indicator function):

1. $u_t = 0.3u_{t-1} + \varepsilon_t$ (AR1)

2. $u_t = 0.95u_{t-1} + \varepsilon_t$ (AR2)

3. $u_t = \varepsilon_t + 0.5\varepsilon_{t-1}$ (MA)

4. $u_t = 0.5u_{t-1}\varepsilon_{t-1} + \varepsilon_t$ (BIL)

5. $u_t = 0.8\varepsilon_{t-1}^2 + \varepsilon_t$ (NMA)

6. $u_t = 0.5u_{t-1}\mathbb{I}(u_{t-1} \leq 1) + 0.4u_{t-1}\mathbb{I}(u_{t-1} > 1) + \varepsilon_t$ (TAR1)

7. $u_t = 0.95\sqrt{|u_{t-1}|} + \varepsilon_t$ (SQRT)

8. $u_t = -\mathbb{I}(u_{t-1} \leq 0) + \mathbb{I}(u_{t-1} > 0) + \varepsilon_t$ (SGN)

9. $u_t = 0.95u_{t-1}\mathbb{I}(u_{t-1} \leq 0) - 0.2u_{t-1}\mathbb{I}(u_{t-1} > 0) + \varepsilon_t$ (TAR2)

10. $u_t = \sqrt{\nu_t}\varepsilon_t, \quad \nu_t = 0.25 + 0.5\nu_{t-1} + 0.5y_{t-1}^2\mathbb{I}(\varepsilon_t \leq 0) + 0.2y_{t-1}^2\mathbb{I}(\varepsilon_t > 0)$ (NARCH)

For all the designs, $\{\varepsilon_t\}$ are i.i.d. $\mathcal{N}(0,1)$ random variables independent of $\{w_t\}$. AR1, AR2 and MA are linear processes, and hence the sieve-type GLS is expected to work best for these. In order to investigate the robustness of the method to failure of the linearity assumption, the remaining eight processes under consideration are nonlinear. We feel they represent a reasonable sample of nonlinear processes used in the literature and have taken many of them from Hong and White (2005). BIL is a bilinear AR(1) process, NMA is a nonlinear MA(1) process, TAR1 and TAR2 are threshold AR(1) processes, SQRT is a fractional AR(1) process, SGN is a sign AR(1) process, and NARCH is a nonlinear GARCH process. If the sieve-type GLS procedure were found to perform well in these cases, then it would be reasonable to claim that linear autoregressive approximations are worth considering more generally. It is worth pointing out that nonlinear short-range dependent processes have not been widely used in Monte Carlo studies in the robust inference literature. This seems surprising, especially in the case of methods that utilize autoregressive prewhitening (such as those discussed in Andrews and Monahan (1992) and den Haan and Levin (2000), among others), since such methods may be reasonably expected to work better for linear processes compared to other nonparametric approaches.

In the experiments, the objective is to test the null hypothesis $\mathrm{H}_0 : \beta_1 = 0$ against the alternative $\mathrm{H}_1 : \beta_1 \neq 0$ using a t-type statistic; the latter is constructed using either the FGLS coefficient estimator and its variance or the OLS coefficient estimator and a heteroskedasticity and autocorrelation robust variance estimator. The value $\beta_1 = 0$ is thus used to compute Type I error probabilities, while values $\beta_1 > 0$ are used for power calculations. To ensure meaningful power comparisons, power calculations are carried out using size-adjusted critical values computed from the experiments in which $\beta_1 = 0$. The number of Monte Carlo replications is 20,000 in the case of size calculations and 5,000 otherwise. The nominal significance level of tests is 0.05.

When the errors and regressors are short-range dependent, we consider two versions of the GLS-based test depending on the strategy used to choose the order h of the autoregressive approximation. For the first version, h is selected by minimizing the Bayesian information criterion of Schwarz (1978) over the range $0 \leq h \leq H_T^*$. We set $H_T^* = 10 \lfloor T^{1/4} \rfloor$, which is an appropriate choice in view of Theorem 2, and label the resulting test GLS_B. The second version uses the sequential testing procedure described in Algorithm 1 with $\alpha = 0.01$. We set $H_T^* = 10 \lfloor T^{1/4} \rfloor$, which is an appropriate choice in view of Theorem 3 and Remark 3, and label the resulting test $\mathrm{GLS}_{1\%}$. In either case, the preliminary estimator $\bar{\beta}$ used to compute the residuals in (6) is OLS.

We consider three OLS-based competitors to the tests constructed using FGLS estimates. The first is a test based on the Newey and West (1987) robust covariance estimator with bandwidth set equal to $\lfloor T^{1/5} \rfloor$, which is labeled NW. The second is based on the approach advocated by Kiefer, Vogelsang, and Bunzel (2000), which is labeled KVB. The third test is based on the approach of Andrews and Monahan (1992) using the quadratic spectral kernel, AR(1) prewhitening, and AR(1)-based automatic bandwidth selection, as discussed in that paper; this test is labeled QS.

When long-range dependence is considered, we examine regression tests based on our time-domain FGLS estimator $\widehat{\beta}$ and the frequency-domain GLS estimator of Hidalgo and Robinson (2002).[4] In our implementation of the FGLS procedure, the preliminary estimator used to compute the residuals in (6) is OLS. This, of course, is

[4] We are grateful to Štěpána Lazarová for providing us with the computer code used in the simulation study of Hidalgo and Robinson (2002).

not a theoretically attractive choice because the OLS estimator of the slope coefficient is $\sqrt{T}$-consistent only when $d_x + d_u < 1/2$. Nevertheless, we use the OLS estimator because of its numerical convenience and familiarity, and because we wish to examine whether this choice has deleterious effects on the small-sample properties of the FGLS procedure. The order of the autoregressive approximation h is selected by minimizing the Bayesian information criterion over the range $0 \leq h \leq \lfloor 2 \ln T \rfloor$. Frequency-domain GLS estimates are computed using the weight function $c(u) = 1 - |u|$ and bandwidth parameter $m = \lfloor T/16 \rfloor$, in the notation of Hidalgo and Robinson (2002). The preliminary estimator employed is again OLS; interestingly, Hidalgo and Robinson (2002) found this choice of preliminary estimator to yield satisfactory results even in situations where it is theoretically unjustifiable owing to the reason mentioned in the previous sentence. The tests based on the sieve-type GLS and the frequency-domain GLS are labeled GLS-TD and GLS-FD, respectively.

4.2 Simulation Results

Tables 1–3 report Monte Carlo estimates of the rejection probabilities of tests in the case where the stochastic regressor and the errors exhibit short-range dependence. The null rejection probabilities shown in Table 1 reveal that both GLS-based tests have very good performance. $\text{GLS}_{1\%}$ tends to have an advantage over GLS_B in small samples ($T = 20, 30$), but otherwise the performance of the two tests is quite similar. For this reason, we focus on the $\text{GLS}_{1\%}$ test. As expected, in the case of linear data-generating processes for the errors, $\text{GLS}_{1\%}$ dominates all other tests by substantial margins (see, e.g., the case of AR2). More surprisingly perhaps, $\text{GLS}_{1\%}$ is also the superior test when $\{u_t\}$ is a nonlinear process. More specifically, it dominates NW and QS for all design points. KVB performs better than $\text{GLS}_{1\%}$ in some cases. However, even in these cases the difference between KVB and $\text{GLS}_{1\%}$ is small and becomes negligible when $T \geq 30$. What is more, in cases where $\text{GLS}_{1\%}$ outperforms KVB, it does so by a considerable margin.

These results are backed up by a more detailed study of the null rejection probabilities shown in Figure 1, where T varies between 10 and 100 at steps of 1 observation. The panels in Figure 1 group the performance of the tests for a given process for $\{u_t\}$. To gain a different insight, Figure 2 presents the same results but here each panel groups the results for one test across error processes. Performance across error processes is much less variable for GLS-based tests than for OLS-based tests. The GLS-based tests perform impressively regardless of the structure of the process generating the errors. This is not the case for OLS-based tests. NW is the least successful and, while QS improves upon NW, it never dominates either KVB or $\text{GLS}_{1\%}$. As a final point, we comment on the difference in performance between $\text{GLS}_{1\%}$ and GLS_B. Although they perform equally well for $T > 30$, the performance for small T is different enough to warrant some explanation. Closer examination suggests that $\text{GLS}_{1\%}$ performs better because it relies on a more parsimonious and less variable order selection procedure, which results in better size control.

Rejection probabilities for the power experiments are reported in Tables 2 and 3. Since our computations use size-adjusted critical values, it is not surprising that $\text{GLS}_{1\%}$ and GLS_B perform very similarly. The two GLS-based tests are never dominated by the OLS-based tests. In many cases the performance of $\text{GLS}_{1\%}$ and GLS_B is similar to that of NW, and to a lesser extent QS. $\text{GLS}_{1\%}$ and GLS_B dominate KVB in most cases by substantial margins.

Overall, it is clear that GLS-based tests, and especially $\text{GLS}_{1\%}$, suffer little of the size distortions widely reported in the literature on OLS-based procedures. Their power performance in small samples reflects their theoretical asymptotic properties by being superior to that of OLS-based tests. Importantly, these results seem to extend to cases where the generating mechanism of the errors is nonlinear, implying that autoregressive approximations have wide applicability, as indeed is suggested by the work of Bickel and Bühlmann (1997).[5]

Turning to designs with long-range dependence, Table 4 contains estimates of the rejection probabilities of the GLS-TD and GLS-FD tests in the case where both the stochastic regressor and the errors are long-range dependent. For the most part, the discrepancy between the empirical and nominal Type I error probabilities of the GLS-TD test is smaller than that of the GLS-FD test, especially in the smaller samples. The GLS-TD test is also more robust than the GLS-FD test with respect to strong collective long-range dependence (although it is possible that the performance of the GLS-FD test would improve in cases where $d_x + d_u > 1/2$ if a $\sqrt{T}$-consistent

[5] Bickel and Bühlmann (1997) show that the closure (with respect to certain metrics) of the class of MA(∞) or AR(∞) processes is fairly large. Roughly speaking, for any stationary nonlinear process, there exists another process in the closure of linear processes having identical sample paths with probability exceeding $1/e \approx 0.368$.

preliminary estimator of the slope parameter was used instead of OLS). As far as power is concerned, the GLS-TD test has a clear advantage for most parameter configurations. This advantage is particularly prominent in cases where there is strong long-range dependence in the errors.

In additional experiments not reported here, we considered situations where either the regressor or the error (but not both) is long-range dependent. The case where the regressor is long-range dependent but the error is not, included experiments where the error follows all of the nonlinear processes presented earlier. Results under the null hypothesis exhibit similar patterns to those reported in Table 4. More specifically, in the presence of long-range dependent errors, the GLS-TD test dominates GLS-FD in terms of size distortion, especially in the smaller samples. The GLS-TD test generally has a power advantage too over the GLS-FD test, the advantage being most prominent for small T and large d_u. When the regressor exhibits long-range dependence, the empirical size of the GLS-TD test is closer to the nominal level than the GLS-FD test, the difference between the two tests being especially noticeable for the smaller sample sizes. The performance of the GLS-TD test is particularly impressive in light of the fact that the sieve-type GLS procedure relies explicitly on a linear autoregressive approximation to the generating mechanism of the errors, a mechanism which is linear in the minority of cases considered in the experiments. In fact, as observed in the case of the GLS_B and $\text{GLS}_{1\%}$ tests, nonlinearity in the errors does not seem to have any noticeable effects on the size properties of the GLS-TD test compared to cases where errors are generated by linear processes. The GLS-TD test also dominates the GLS-FD test in terms of power for almost all designs, although the differences between the two tests are not substantial.[6]

In sum, the sieve-type GLS-based tests are found to have the best overall performance in our experiments. They generally exhibit smaller size distortions than other GLS-based or OLS-based tests, especially when the sample size is small, and they also tend to be superior in terms of size-adjusted power. Furthermore, they are robust with respect to the presence of neglected nonlinearity in the regression errors.

[6] The full set of simulation results is available from the authors upon request.

5 Conclusion

This paper has considered the use of time-domain sieve-type GLS based on autoregressive approximations for inference in linear regression models. By allowing the order of the autoregressive approximation to increase with the sample size at an appropriate rate, it has been shown that the sieve-type GLS estimator of the regression coefficients is $\sqrt{T}$-consistent, asymptotically normal and asymptotically Gauss–Markov efficient under general conditions, including mixingale-type conditions as well as conditions which permit long-range dependence in the stochastic regressors and the errors. A Monte Carlo study has revealed that hypothesis tests based on sieve-type GLS have better finite-sample size and power properties than tests based on an alternative frequency-domain GLS procedure or OLS combined with a robust estimator of the OLS asymptotic covariance matrix.

6 Appendix

This Appendix presents the proofs of the theorems stated in the main text. We begin by introducing two lemmata which are used in our proofs. Throughout, limits in order symbols are taken as $T \to \infty$, unless stated otherwise, and $c, c_1, c_2, \ldots$ denote finite constants that may assume different values upon each appearance.

Lemma 1 *Suppose Assumptions* 1*,* 2 *and* 3 *hold and* $d_u = 0$*. Then* $\{x_t u_t\}$ *is an* L_2*-mixingale of size* $-1/2$*.*

Proof. If $\{x_t\}$ and $\{u_t\}$ are L_r-bounded ($r \geq 2$), L_2-NED processes of size $-\zeta$ on a φ-mixing process of size $-\eta$ ($\eta > 1$), then, by Example 17.17 of Davidson (1994), $\{x_t u_t\}$ is L_2-NED of size $-\{\zeta(\kappa - 2)\}/\{2(\kappa - 1)\} \leq -1/2$ on a φ-mixing process of size $-\eta$. In view of Theorem 17.5(ii) of Davidson (1994), this in turn implies that $\{x_t u_t\}$ is an L_2-mixingale of size $-1/2$ if $2\eta > \zeta$. Hence it is enough to show that $\{u_t\}$ is L_r-bounded ($r \geq 2$), L_2-NED of the required size.

By Burkholder's inequality for martingale-difference sequences (e.g., Davidson (1994, Theorem 15.18)) and Hölder's inequality,

$$E(|u_t|^r) \leq cE\left(\left\{\sum_{j=0}^{\infty} |\delta_j|^2 |\varepsilon_{t-j}|^2\right\}^{r/2}\right) \leq c\left(\sum_{j=0}^{\infty} |\delta_j|^2\right)^{r/2-1}\left(\sum_{j=0}^{\infty} |\delta_j|^2 E(|\varepsilon_{t-j}|^r)\right),$$

so $\{u_t\}$ is L_r-bounded if $\sup_t E(|\varepsilon_t|^r) < \infty$. Moreover, writing $\|\cdot\|_r$ for the L_r-norm, we have, by Minkowski's inequality,

$$\left\|u_t - E(u_t|\mathcal{F}^{\varepsilon}_{t-m})\right\|_2 = \left\|\sum_{j=m+1}^{\infty} \delta_j \varepsilon_{t-j}\right\|_2 \leq \sup_t \|\varepsilon_t\|_2 \sum_{j=m+1}^{\infty} |\delta_j|,$$

for any integer $m > 0$. But, when $d_u = 0$, Assumption 2(iii) implies that

$$\lim_{m\to\infty} m^{\zeta} \sum_{j=m+1}^{\infty} |\delta_j| = 0,$$

and consequently $\{u_t\}$ is L_2-NED of size $-\zeta$, which completes the proof. ■

Lemma 2 *Let $\phi_h = (\phi_{h,1}, .., \phi_{h,h})'$ satisfy the equations $\Gamma_h \phi_h = \gamma_h$, where $\Gamma_h = \{\gamma_u(s-j); j, s = 1, \ldots, h\}$, $\gamma_h = (\gamma_u(1), \ldots, \gamma_u(h))'$ and $\gamma_u(s) = E(u_t u_{t+s})$. If the assumptions of either Theorem 1 or Theorem 4 hold, then*

$$\sum_{j=1}^{h} \left|\widehat{\phi}_{h,j} - \phi_{h,j}\right|^2 = O_p\left(h\left(\frac{\ln T}{T}\right)^{1-2d_u}\right),$$

uniformly in h.

Proof. We examine the general case with long-range dependence; for the case of short-range dependence the result follows upon setting $d_u = 0$. Let $\widehat{\phi}^u_h = (\widehat{\phi}^u_{h,1}, .., \widehat{\phi}^u_{h,h})'$ satisfy the equations $\widehat{\Gamma}^u_h \widehat{\phi}^u_h = \widehat{\gamma}^u_h$, where $\widehat{\Gamma}^u_h = \{\widehat{\gamma}_u(j,s); j, s = 1, \ldots, h\}$, $\widehat{\gamma}^u_h = (\widehat{\gamma}_u(1), \ldots, \widehat{\gamma}_u(h))'$, $\widehat{\gamma}_u(s) = \widehat{\gamma}_u(0,s)$, and $\widehat{\gamma}_u(j,s) = T^{-1}\sum_{t=1+\max\{j,s\}}^{T} u_{t-j}u_{t-s}$. Letting $\widehat{\Gamma}_h$ and $\widehat{\gamma}_h$ be the quantities corresponding to $\widehat{\Gamma}^u_h$ and $\widehat{\gamma}^u_h$, respectively, with $\bar{u}_t$ in place of u_t, and noting that each element of $\widehat{\Gamma}_h - \widehat{\Gamma}^u_h$ and $\widehat{\gamma}_h - \widehat{\gamma}^u_h$ can be shown to be $O_p(T^{-1})$, we have

$$\begin{aligned}
\sum_{j=1}^{h} \left|\widehat{\phi}_{h,j} - \widehat{\phi}^u_{h,j}\right|^2 &\leq \left\|\widehat{\phi}_h - \widehat{\phi}^u_h\right\|^2 = \left\|\widehat{\Gamma}^{-1}_h \widehat{\gamma}_h - \left(\widehat{\Gamma}^u_h\right)^{-1}\widehat{\gamma}^u_h\right\|^2 \\
&\leq \left(\left\|\widehat{\Gamma}_h - \widehat{\Gamma}^u_h\right\| \|\widehat{\gamma}^u_h\| + \|\widehat{\gamma}_h - \widehat{\gamma}^u_h\| \left\|\left(\widehat{\Gamma}^u_h\right)^{-1}\right\|\right)^2 \\
&= \left\|\widehat{\Gamma}_h - \widehat{\Gamma}^u_h\right\|^2 \|\widehat{\gamma}^u_h\|^2 + 2\left\|\widehat{\Gamma}_h - \widehat{\Gamma}^u_h\right\| \|\widehat{\gamma}^u_h\| \|\widehat{\gamma}_h - \widehat{\gamma}^u_h\| \left\|\left(\widehat{\Gamma}^u_h\right)^{-1}\right\| \\
&+ \|\widehat{\gamma}_h - \widehat{\gamma}^u_h\|^2 \left\|\left(\widehat{\Gamma}^u_h\right)^{-1}\right\|^2 \\
&= O_p(h^2T^{-2})O_p(h) + 2O_p(hT^{-1})O_p(h^{1/2})O_p(h^{1/2}T^{-1})O_p(h) \\
&+ O_p(hT^{-2})O_p(h^2) \\
&= O_p(h^3T^{-2}).
\end{aligned}$$

Furthermore, by Theorem 5 of Poskitt (2007),

$$\sum_{j=1}^{h} \left| \widehat{\phi}_{h,j}^{u} - \phi_{h,j} \right|^2 = O_p \left(\frac{h}{\sqrt{\lambda_{\min}(\Gamma_h^2)}} \left(\frac{\ln T}{T} \right)^{1-2d_u} \right)$$

uniformly in h, where $\lambda_{\min}(\Gamma_h^2)$ is the smallest eigenvalue of Γ_h^2. In consequence, there exist constants $c_1 > 0$ and $c_2 > 0$ such that

$$\begin{aligned} \sum_{j=1}^{h} \left| \widehat{\phi}_{h,j} - \phi_{h,j} \right|^2 &\leq c_1 \sum_{j=1}^{h} \left| \widehat{\phi}_{h,j} - \widehat{\phi}_{h,j}^{u} \right|^2 + c_2 \sum_{j=1}^{h} \left| \widehat{\phi}_{h,j}^{u} - \phi_{h,j} \right|^2 \\ &= O_p \left(\frac{h}{\sqrt{\lambda_{\min}(\Gamma_h^2)}} \left(\frac{\ln T}{T} \right)^{1-2d_u} \right), \end{aligned}$$

and the desired result follows. ■

Proof of Theorem 1. Observe that

$$\begin{aligned} \sqrt{T}\left(\widehat{\beta} - \beta\right) = &\left[\frac{1}{T} \sum_{t=1}^{T} \left(x_t - \sum_{j=1}^{h} \widehat{\phi}_{h,j} x_{t-j} \right) \left(x_t - \sum_{j=1}^{h} \widehat{\phi}_{h,j} x_{t-j} \right)' \right]^{-1} \\ &\times \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(x_t - \sum_{j=1}^{h} \widehat{\phi}_{h,j} x_{t-j} \right) \left(u_t - \sum_{j=1}^{h} \widehat{\phi}_{h,j} u_{t-j} \right) \right]. \end{aligned}$$

Hence, the first part of the theorem follows if

$$\begin{aligned} P_1 = &\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(x_t - \sum_{j=1}^{h} \widehat{\phi}_{h,j} x_{t-j} \right) \left(u_t - \sum_{j=1}^{h} \widehat{\phi}_{h,j} u_{t-j} \right) \\ &- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(x_t - \sum_{j=1}^{\infty} \phi_j x_{t-j} \right) \left(u_t - \sum_{j=1}^{\infty} \phi_j u_{t-j} \right) \\ = &\, o_p(1) \end{aligned} \tag{14}$$

and

$$\begin{aligned} P_2 = &\frac{1}{T} \sum_{t=1}^{T} \left(x_t - \sum_{j=1}^{h} \widehat{\phi}_{h,j} x_{t-j} \right) \left(x_t - \sum_{j=1}^{h} \widehat{\phi}_{h,j} x_{t-j} \right)' \\ &- \frac{1}{T} \sum_{t=1}^{T} \left(x_t - \sum_{j=1}^{\infty} \phi_j x_{t-j} \right) \left(x_t - \sum_{j=1}^{\infty} \phi_j x_{t-j} \right)' \\ = &\, o_p(1). \end{aligned} \tag{15}$$

Considering (14) first, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(x_t-\sum_{j=1}^{h}\widehat{\phi}_{h,j}x_{t-j}\right)\left(u_t-\sum_{j=1}^{h}\widehat{\phi}_{h,j}u_{t-j}\right)$$
$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left(x_t-\sum_{j=1}^{\infty}\phi_j x_{t-j}\right)-\sum_{j=1}^{\infty}\left(\widehat{\phi}_{h,j}-\phi_j\right)x_{t-j}\right]$$
$$\times\left[\left(u_t-\sum_{j=1}^{\infty}\phi_j u_{t-j}\right)-\sum_{j=1}^{\infty}\left(\widehat{\phi}_{h,j}-\phi_j\right)u_{t-j}\right],$$

with $\widehat{\phi}_{h,j}=0$ for $j>h$, and so

$$P_1=-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{\infty}\left(\widehat{\phi}_{h,j}-\phi_j\right)x_{t-j}\right]\left(u_t-\sum_{j=1}^{\infty}\phi_j u_{t-j}\right)$$
$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(x_t-\sum_{j=1}^{\infty}\phi_j x_{t-j}\right)\left[\sum_{j=1}^{\infty}\left(\widehat{\phi}_{h,j}-\phi_j\right)u_{t-j}\right]$$
$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{\infty}\left(\widehat{\phi}_{h,j}-\phi_j\right)x_{t-j}\right]\left[\sum_{j=1}^{\infty}\left(\widehat{\phi}_{h,j}-\phi_j\right)u_{t-j}\right]$$
$$=-S_1-S_2+S_3.$$

Regarding S_1, we have

$$S_1=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{\infty}\left(\widehat{\phi}_{h,j}-\phi_j\right)x_{t-j}\right]\varepsilon_t$$
$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j}-\phi_{h,j}\right)x_{t-j}\right]\varepsilon_t+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)x_{t-j}\right]\varepsilon_t$$
$$=J_1+J_2, \tag{16}$$

with $\phi_{h,j}=0$ for $j>h$. By the Cauchy–Schwarz inequality,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j}-\phi_{h,j}\right)x_{t-j}\right]\varepsilon_t$$
$$=\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j}-\phi_{h,j}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_{t-j}\varepsilon_t\right)$$
$$\leq\left[\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j}-\phi_{h,j}\right)^2\right]^{1/2}\left[\sum_{j=1}^{h}\left(\frac{1}{T}\sum_{t=1}^{T}x'_{t-j}\varepsilon_t^2x_{t-j}\right)\right]^{1/2}.$$

But, by Assumptions 1(iii) and 3, and Theorem 17.9 of Davidson (1994), $\{x'_{t-j}\varepsilon_t^2 x_{t-j}\}$ is an L_1-NED process which, by Theorem 19.11 of Davidson (1994), obeys a law of large numbers. Since

$$\sup_{1\le j\le h}\left|\frac{1}{T}\sum_{t=1}^{T} x'_{t-j}\varepsilon_t^2 x_{t-j}\right| \le \sum_{j=1}^{h}\frac{1}{T}\sum_{t=1}^{T} x'_{t-j}\varepsilon_t^2 x_{t-j},$$

it follows that

$$\left[\sum_{j=1}^{h}\left(\frac{1}{T}\sum_{t=1}^{T} x'_{t-j}\varepsilon_t^2 x_{t-j}\right)\right]^{1/2} = O_p(h^{1/2}).$$

Since, in addition,

$$\left[\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j}-\phi_{h,j}\right)^2\right]^{1/2} = O_p\left(\left\{hT^{-1}\ln T\right\}^{1/2}\right)$$

by Lemma 2, we conclude that

$$J_1 = o_p(1), \tag{17}$$

as long as $h = o\left(\{T/\ln T\}^{1/2}\right)$. Moreover, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)x_{t-j}\right]\varepsilon_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Q_T\varepsilon_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} J_{tT},$$

where $J_{tT} = Q_T\varepsilon_t$. By Theorem 7.6.6 of Anderson (1971), $\text{var}(J_{tT}) = o(1)$. Hence, by a triangular array central limit theorem, it follows that

$$J_2 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} J_{tT} = o_p(1),$$

which, together with (17), shows that $S_1 = o_p(1)$.

We now examine S_2. It is easy to see that, under the assumptions of the theorem, $x_t - \sum_{j=1}^{\infty}\phi_j x_{t-j}$ is an NED process. Then, a similar treatment to that used for S_1 leads to $S_2 = o_p(1)$.

Finally, we consider S_3. We have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{\infty}\left(\widehat{\phi}_{h,j}-\phi_j\right)x_{t-j}\right]\left[\sum_{j=1}^{\infty}\left(\widehat{\phi}_{h,j}-\phi_j\right)u_{t-j}\right]$$

$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j}-\phi_{h,j}\right)x_{t-j}+\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)x_{t-j}\right]$$

$$\times\left[\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j}-\phi_{h,j}\right)u_{t-j}+\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)u_{t-j}\right]$$

$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)x_{t-j}\right]\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)u_{t-j}\right]$$

$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j}-\phi_{h,j}\right)x_{t-j}\right]\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)u_{t-j}\right]$$

$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)x_{t-j}\right]\left[\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j}-\phi_{h,j}\right)u_{t-j}\right]$$

$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j}-\phi_{h,j}\right)x_{t-j}\right]\left[\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j}-\phi_{h,j}\right)u_{t-j}\right]$$

$$=K_1+K_2+K_3+K_4. \tag{18}$$

We first examine K_1. We have that

$$\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)x_{t-j}\right]\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)u_{t-j}\right]\right\|$$

$$\leq\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\sum_{j=1}^{\infty}\phi_{h,j}x_{t-j}\right)\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)u_{t-j}\right]\right\|$$

$$+\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\sum_{j=1}^{\infty}\phi_{j}x_{t-j}\right)\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)u_{t-j}\right]\right\|.$$

Noting that Theorem 7.6.6. of Anderson (1971) holds uniformly over t, and using that theorem, we have

$$\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\sum_{j=1}^{\infty}\phi_{h,j}x_{t-j}\right)\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)u_{t-j}\right]\right\|$$

$$+\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\sum_{j=1}^{\infty}\phi_{j}x_{t-j}\right)\left[\sum_{j=1}^{\infty}\left(\phi_{h,j}-\phi_j\right)u_{t-j}\right]\right\|$$

$$=o_p(1), \tag{19}$$

provided that $h \to \infty$ as $T \to \infty$. To see this, we note that

$$\sum_{j=1}^{\infty} (\phi_{h,j} - \phi_j) u_{t-j} = \left(u_t - \sum_{j=1}^{\infty} \phi_j u_{t-j} \right) - \left(u_t - \sum_{j=1}^{\infty} \phi_{h,j} u_{t-j} \right) = \varepsilon_t - \varepsilon_{h,t}.$$

Letting $E_{t-1}(\cdot)$ denote conditional expectation given the the σ-field generated by $\{(\varepsilon_s, x_s); s < t\}$, it can be seen that $\{\varepsilon_t - \varepsilon_{h,t}\}$ is a triangular array such that $\lim_{T\to\infty} E_{t-1}(\varepsilon_t - \varepsilon_{h,t}) = 0$ uniformly over t and $\lim_{T\to\infty} E_{t-1}(\{\varepsilon_t - \varepsilon_{h,t}\}^2)$ does not depend on t. Further, $\varepsilon_t - \varepsilon_{h,t}$ is independent of x_{t-j} for all integers j. Since

$$\sum_{j=1}^{\infty} \phi_j x_{t-j} = x_t - \left(x_t - \sum_{j=1}^{\infty} \phi_j x_{t-j} \right),$$

it follows that $\sum_{j=1}^{\infty} \phi_j x_{t-j}$ is a stationary NED process of the same size as $\{x_t\}$. This is also the case for $\sum_{j=1}^{\infty} \phi_{h,j} x_{t-j}$. It is then easy to see that

$$\lim_{T\to\infty} E_{t-1} \left(\sum_{j=1}^{\infty} \phi_j x_{t-j} (\varepsilon_t - \varepsilon_{h,t}) \right) = 0 \tag{20}$$

uniformly over t, and

$$E_{t-1} \left(\left[\sum_{j=1}^{\infty} \phi_j x_{t-j} (\varepsilon_t - \varepsilon_{h,t}) \right]^2 \right) = c_T, \tag{21}$$

where $c_T \geq 0$ does not depend on t. This implies that a martingale-difference central limit theorem holds for $\sum_{j=1}^{\infty} \phi_j x_{t-j} (\varepsilon_t - \varepsilon_{h,t})$.[7] Since, in addition, the variance of $\sum_{j=1}^{\infty} \phi_j x_{t-j} (\varepsilon_t - \varepsilon_{h,t})$ approaches zero as $T \to \infty$, uniformly over t, (19) holds true.

Moving on to K_2, $K_2 = o_p(1)$ follows immediately from Theorem 7.6.6. of Anderson (1971) and (17). Arguing as in the case of S_2, it can also be deduced that $K_3 = o_p(1)$. Finally, we consider K_4. We have

$$\begin{aligned} &\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\sum_{j=1}^{h} \left(\widehat{\phi}_{h,j} - \phi_{h,j} \right) x_{t-j} \right] \left[\sum_{j=1}^{h} \left(\widehat{\phi}_{h,j} - \phi_{h,j} \right) u_{t-j} \right] \\ &= \sum_{j=1}^{h} \sum_{s=1}^{h} \left(\widehat{\phi}_{h,j} - \phi_{h,j} \right) \left(\widehat{\phi}_{h,s} - \phi_{h,s} \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{t-j} u_{t-s} \right). \end{aligned} \tag{22}$$

[7] Note that the standard central limit theorem for martingale-difference sequences specifies that (20) and (21) hold for all T and not only in the limit, but a casual examination of the proof of Theorem 24.3 of Davidson (1994) shows that (20) and (21) are sufficient for the theorem to hold.

By Lemma 1, $\{x_t u_t\}$ is an L_2-NED process of size $-1/2$ which, in view of Theorem 1 of De Jong (1997), obeys a central limit theorem. Thus

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{t-j} u_{t-s} = O_p\left(1\right).$$

But

$$\sum_{j=1}^{h} \sum_{s=1}^{h} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{t-j} u_{t-s} \right) = O_p\left(h^2\right)$$

uniformly in j and s. Then, by Lemma 2,

$$\sum_{j=1}^{h} \sum_{s=1}^{h} \left(\widehat{\phi}_{h,j} - \phi_{h,j}\right) \left(\widehat{\phi}_{h,s} - \phi_{h,s}\right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{t-j} u_{t-s} \right) = O_p\left(h^4 T^{-1} \ln T\right),$$

which is $o_p(1)$ as long as $h = o\left(\{T/\ln T\}^{1/4}\right)$, and thus $K_4 = o_p(1)$. This completes the proof of $P_1 = o_p(1)$. By reasoning in a similar way, it can be shown that $P_2 = o_p(1)$, and so the first part of the theorem is established.

The second part of the theorem follows in a straightforward manner using the fact that, in view of Lemma 1, $\{x_t u_t\}$ obeys a central limit theorem (cf. De Jong (1997, Theorem 1)) and $\{x_t x_t'\}$ obeys a mixingale law of large numbers (cf. Davidson (1994, Theorem 19.11))). ■

Proof of Theorem 2. Part (ii) of Assumption 5 holds on account of the fact that $H_T^* = o\left(\{T/\ln T\}^{1/4}\right)$. Thus, it remains to prove that $h_T \to \infty$ as $T \to \infty$. By Theorem 7.4.7(b) of Hannan and Deistler (1988),

$$\ln \widehat{\sigma}^2_{\bar{u},h} + \frac{hC_T}{T} = c + \frac{h(C_T - 1)}{T} + \frac{\widehat{\sigma}^2_{\bar{u},h} - \sigma^2_{\varepsilon}}{\sigma^2_{\varepsilon}} \{1 + o_p(1)\}, \tag{23}$$

where c is a constant not depending on h. But (23) implies that $h_T \to \infty$ as $T \to \infty$. To see this, note that, under Assumption 6, $(\widehat{\sigma}^2_{\bar{u},h} - \sigma^2_{\varepsilon})/\sigma^2_{\varepsilon}$ is bounded away from zero almost surely for any finite value of h. Since $h(C_T - 1)/T \to 0$ as $T \to \infty$ when $C_T = o\left(\{T^3 \ln T\}^{1/4}\right)$, it follows that (11) cannot be minimized for any finite value of h. ■

Proof of Theorem 3. To establish the assertion of the theorem, it is sufficient to show that Lemma 5.1 of Ng and Perron (1995) holds for $h_T = H_T^*$. The desired result follows then by Lemma 5.2 of Ng and Perron (1995). Under the conditions of the theorem, Lemma 5.1 of Ng and Perron (1995) holds for $h_T = H_T^*$ if Theorem 4 of Lewis and Reinsel (1985) or, equivalently, Theorem 7.4.8 of Hannan and Deistler (1988) holds when $h_T = H_T^*$. But the latter theorem holds provided $\sqrt{T}\sum_{j=H_T^*+1}^{\infty} |\delta_j| \to 0$ as $T \to \infty$, and the desired result is obtained.[8] ■

Proof of Theorem 4. The assertion of the theorem can be proved by showing that (14) and (15) hold under the conditions of the theorem. Thus, using the same notation as in the proof of Theorem 1, we begin by showing that S_1, S_2 and S_3 are $o_p(1)$.

As in (16), put $S_1 = J_1 + J_2$. Now note that

$$\sum_{j=1}^{h}\left(\frac{1}{T}\sum_{t=1}^{T} x'_{t-j}\varepsilon_t^2 x_{t-j}\right) = \sum_{j=1}^{h}\left(\frac{1}{T}\sum_{t=1}^{T} x'_{t-j}\left(\varepsilon_t^2 - \sigma_\varepsilon^2\right) x_{t-j}\right) + \sum_{j=1}^{h}\left(\frac{\sigma_\varepsilon^2}{T}\sum_{t=1}^{T} x'_{t-j} x_{t-j}\right).$$

By Assumptions 1, 8 and 9, it follows that $\{x'_{t-j}\left(\varepsilon_t^2 - \sigma_\varepsilon^2\right) x_{t-j}\}$ is a square-integrable martingale-difference sequence, which implies that $(1/T)\sum_{t=1}^{T} x'_{t-j}\left(\varepsilon_t^2 - \sigma_\varepsilon^2\right) x_{t-j} = O_p(T^{-1/2})$. Further, by Assumption 8, $(1/T)\sum_{t=1}^{T} x'_{t-j} x_{t-j}$ obeys a law of large numbers and is, therefore, bounded in probability. Since

$$\sup_{1\le j\le h}\left|\frac{\sigma_\varepsilon^2}{T}\sum_{t=1}^{T} x'_{t-j} x_{t-j}\right| \le \sum_{j=1}^{h}\frac{\sigma_\varepsilon^2}{T}\sum_{t=1}^{T} x'_{t-j} x_{t-j},$$

it follows that

$$\left(\sum_{j=1}^{h}\left[\frac{1}{T}\sum_{t=1}^{T} x'_{t-j}\varepsilon_t^2 x_{t-j}\right]\right)^{1/2} = O_p(h).$$

Since, in addition,

$$\left(\sum_{j=1}^{h}\left(\widehat{\phi}_{h,j} - \phi_{h,j}\right)^2\right)^{1/2} = O_p\left(\left\{h(T/\ln T)^{2d_u-1}\right\}^{1/2}\right)$$

by Lemma 2, we conclude that

$$J_1 = o_p(1), \tag{24}$$

as long as $h = O(\{\ln T\}^a)$ for some $0 < a < \infty$. Moreover, by the same argument used in the proof of Theorem 1, we deduce that $J_2 = o_p(1)$, thus showing that $S_1 = o_p(1)$.

[8] It is worth noting that Kuersteiner (2005) argues that the chi-square approximation in Lemma 5.1 of Ng and Perron (1995) is not valid. However, irrespectively of the validity of the result in question, it is straightforward to establish that, as long as the probabilities with which the null hypothesis of the test used in the sequential testing procedure of Definition 3.1 of Ng and Perron (1995) is accepted when true and rejected when false are bounded away from 1 and 0, respectively, Lemma 5.2 of Ng and Perron (1995) is valid, thereby establishing our result.

Regarding S_2, it is easy to see that, under the assumptions of the theorem, the ith component of $x_t - \sum_{j=1}^{\infty} \phi_j x_{t-j}$ has memory parameter $d_i - d_u$. Then, a similar treatment to that used for S_1 leads to $S_2 = o_p(1)$.

Next, as in (18), put $S_3 = K_1 + K_2 + K_3 + K_4$. To see that the result in (19) holds true under the conditions of the theorem, note that $\sum_{j=1}^{\infty} \phi_j x_{t-j}$ is a stationary process with the same memory parameters as x_t, and this is also the case for $\sum_{j=1}^{\infty} \phi_{h,j} x_{t-j}$. It is then easy to see that (20) and (21) hold, and (19) follows by the argument used in the proof of Theorem 1, thus showing that $K_1 = o_p(1)$. The result $K_2 = o_p(1)$ follows from Theorem 7.6.6 of Anderson (1971) and (24). Arguing as in the case of S_1 and S_2, it can also be deduced that $K_3 = o_p(1)$. Finally, an examination of the proof of Theorem 2 of Chung (2002) suggests that, uniformly in j and s,

$$\frac{1}{T} \sum_{t=1}^{T} x_{t-j} u_{t-s} = O_p\left(T^{-d^*}\right),$$

where

$$d^* = (1/2)\mathbb{I}(d_x + d_u < 1/2) - (1/2)(1 + d_x + d_u)\mathbb{I}(d_x + d_u > 1/2)$$

and $d_x = \max\{d_i; 1 \leq i \leq k\}$. Focusing on the 'worst case' scenario, under which $d_x + d_u > 1/2$, we have

$$\frac{1}{T} \sum_{t=1}^{T} x_{t-j} u_{t-s} = O_p\left(T^{-(1+d_x+d_u)/2}\right).$$

Hence, (22) is of the same order in probability as

$$T^{(d_x+d_u)/2} \sum_{j=1}^{h} \sum_{s=1}^{h} \left(\widehat{\phi}_{h,j} - \phi_{h,j}\right) \left(\widehat{\phi}_{h,s} - \phi_{h,s}\right).$$

Then, by Lemma 2,

$$T^{(d_x+d_u)/2} \sum_{j=1}^{h} \sum_{s=1}^{h} \left(\widehat{\phi}_{h,j} - \phi_{h,j}\right) \left(\widehat{\phi}_{h,s} - \phi_{h,s}\right) = O_p\left(h^2 T^{(d_x+d_u)/2} \left(T^{-1} \ln T\right)^{1-2d_u}\right),$$

which is $o_p(1)$ as long as $(5d_u/2) + (d_x/2) - 1 < 0$, thus deducing that $K_4 = o_p(1)$ and concluding that $P_1 = o_p(1)$. By reasoning in a similar way, it can be shown that $P_2 = o_p(1)$, which completes the proof of the theorem. ■

References

Amemiya, T. (1973): "Generalized Least Squares with an Estimated Autocovariance Matrix," *Econometrica*, 41, 723–732.

Anderson, T. W. (1971): *The Statistical Analysis of Time Series.* Wiley, New York.

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Andrews, D. W. K., and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

Bickel, P. J., and P. Bühlmann (1997): "Closure of Linear Processes," *Journal of Theoretical Probability*, 10, 445–479.

Choy, K., and M. Taniguchi (2001): "Stochastic Regression Model with Dependent Disturbances," *Journal of Time Series Analysis*, 22, 175–196.

Chung, C.-F. (2002): "Sample Means, Sample Autocovariances, and Linear Regression of Stationary Multivariate Long Memory Processes," *Econometric Theory*, 18, 51–78.

Dahlhaus, R. (1995): "Efficient Location and Regression Estimation for Long Range Dependent Regression Models," *Annals of Statistics*, 23, 1029–1047.

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford University Press, Oxford.

——— (2002): "Establishing Conditions for the Functional Central Limit Theorem in Nonlinear and Semiparametric Time Series Processes," *Journal of Econometrics*, 106, 243–269.

De Jong, R. M. (1997): "Central Limit Theorems for Dependent Heterogeneous Random Variables," *Econometric Theory*, 13, 353–367.

den Haan, W. J., and A. Levin (1997): "A Practitioner's Guide to Robust Covariance Matrix Estimation," in *Handbook of Statistics: Robust Inference*, ed. by G. S. Maddala, and C. R. Rao, vol. 15, pp. 299–342. North-Holland, Amsterdam.

——— (2000): "Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order," Technical Working Paper No. 255, National Bureau of Economic Research.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

Granger, C. W. J., and R. Joyeux (1980): "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15–29.

Grenander, U. (1981): *Abstract Inference*. Wiley, New York.

Grenander, U., and G. Szegö (1958): *Toeplitz Forms and Their Applications*. University of California Press, Berkeley.

Hannan, E. J. (1963): "Regression for Time Series," in *Time Series Analysis*, ed. by M. Rosenblatt, pp. 17–37. Wiley, New York.

——— (1970): *Multiple Time Series*. Wiley, New York.

Hannan, E. J., and M. Deistler (1988): *The Statistical Theory of Linear Systems*. Wiley, New York.

Harris, D., D. I. Harvey, S. J. Leybourne, and A. M. R. Taylor (2009): "Testing for a Unit Root in the Presence of a Possible Break in Trend," *Econometric Theory*, 25, 1545–1588.

Harvey, D. I., S. J. Leybourne, and A. M. R. Taylor (2007): "A Simple, Robust and Powerful Test of the Trend Hypothesis," *Journal of Econometrics*, 141, 1302–1330.

Hidalgo, J., and P. M. Robinson (2002): "Adapting to Unknown Disturbance Autocorrelation in Regression with Long Memory," *Econometrica*, 70, 1545–1581.

Hong, Y., and H. White (2005): "Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence," *Econometrica*, 73, 837–901.

Hosking, J. R. M. (1981): "Fractional Differencing," *Biometrika*, 65, 165–176.

Jansson, M. (2004): "The Error in Rejection Probability of Simple Autocorrelation-Robust Tests," *Econometrica*, 72, 937–946.

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68, 695–714.

Kuersteiner, G. (2005): "Automatic Inference for Infinite Order Vector Autoregressions," *Econometric Theory*, 20, 85–115.

Lewis, R., and G. C. Reinsel (1985): "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of the Multivariate Analysis*, 16, 393–411.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Ng, S., and P. Perron (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.

Nielsen, M. Ø. (2005): "Semiparametric Estimation in Time-Series Regression with Long-Range Dependence," *Journal of Time Series Analysis*, 26, 279–304.

Perron, P., and T. Yabu (2009): "Estimating Deterministic Trends with an Integrated or Stationary Noise Component," *Journal of Econometrics*, 151, 56–69.

——— (2012): "Testing for Shifts in Trend With an Integrated or Stationary Noise Component," *Journal of Business and Economic Statistics*, 27, 369–396.

Poskitt, D. S. (2007): "Autoregressive Approximation in Nonstandard Situations: The Fractionally Integrated and Non-Invertible Cases," *Annals of the Institute of Statistical Mathematics*, 59, 697–725.

Robinson, P. M. (1991): "Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models," *Econometrica*, 59, 1329–1363.

——— (1994): "Time Series with Strong Dependence," in *Advances in Econometrics: Sixth World Congress*, ed. by C. A. Sims, vol. 1, pp. 47–95. Cambridge University Press, Cambridge.

Robinson, P. M., and F. J. Hidalgo (1997): "Time Series Regression with Long-Range Dependence," *Annals of Statistics*, 25, 77–104.

Schwarz, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Yajima, Y. (1988): "On Estimation of a Regression Model with Long-Memory Stationary Errors," *Annals of Statistics*, 16, 791–807.

——— (1991): "Asymptotic Properties of the LSE in a Regression Model with Long-Memory Stationary Errors," *Annals of Statistics*, 19, 158–177.

Table 1: Rejection probabilities for $\beta_1 = 0$

	$T = 20$					$T = 30$					$T = 50$				
Model	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$
AR1	.169	.121	.112	.144	.097	.136	.090	.094	.124	.080	.114	.070	.076	.100	.066
AR2	.241	.110	.150	.176	.087	.215	.076	.126	.158	.065	.174	.056	.099	.135	.052
MA	.164	.121	.105	.139	.093	.135	.096	.090	.121	.083	.116	.078	.077	.097	.070
BILIN	.136	.114	.090	.116	.087	.103	.083	.072	.094	.072	.091	.070	.064	.076	.064
NMA	.124	.107	.080	.107	.081	.096	.082	.069	.092	.072	.087	.068	.064	.080	.064
TAR1	.127	.106	.083	.108	.082	.102	.081	.075	.095	.070	.090	.064	.064	.080	.061
SQRT	.161	.111	.105	.143	.089	.130	.085	.088	.122	.075	.110	.067	.075	.099	.064
SGN	.200	.114	.130	.159	.094	.179	.089	.117	.154	.081	.149	.069	.097	.131	.066
TAR2	.218	.110	.136	.167	.086	.190	.080	.118	.155	.070	.159	.058	.096	.132	.055
NARCH	.127	.112	.079	.115	.086	.098	.085	.068	.096	.074	.082	.070	.057	.080	.066
	$T = 100$					$T = 200$					$T = 400$				
Model	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$
AR1	.088	.058	.065	.079	.058	.075	.055	.058	.065	.055	.066	.053	.053	.058	.053
AR2	.140	.050	.079	.113	.050	.125	.049	.065	.094	.051	.102	.051	.059	.081	.052
MA	.086	.063	.061	.070	.062	.076	.060	.056	.059	.060	.064	.055	.051	.048	.056
BILIN	.067	.054	.051	.059	.053	.061	.052	.051	.056	.053	.058	.052	.052	.052	.052
NMA	.066	.056	.054	.061	.056	.056	.051	.050	.052	.051	.057	.055	.052	.054	.055
TAR1	.073	.061	.059	.068	.061	.062	.058	.056	.059	.058	.056	.053	.052	.055	.053
SQRT	.087	.058	.064	.077	.058	.074	.053	.055	.065	.053	.066	.055	.055	.060	.056
SGN	.117	.057	.074	.102	.057	.098	.056	.061	.081	.056	.082	.053	.057	.067	.054
TAR2	.127	.051	.072	.104	.050	.115	.051	.061	.087	.052	.096	.053	.056	.070	.054
NARCH	.062	.057	.051	.061	.057	.060	.059	.052	.060	.060	.054	.052	.047	.052	.054

Table 2: Rejection probabilities for $\beta_1 \in \{0.05, 0.1\}$

$\beta_1 = .05$															
	$T = 20$					$T = 30$					$T = 50$				
Model	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$
AR1	.052	.048	.050	.051	.050	.058	.063	.050	.060	.064	.067	.064	.064	.061	.065
AR2	.054	.049	.055	.051	.050	.051	.057	.049	.056	.057	.051	.063	.053	.054	.066
MA	.056	.055	.055	.057	.053	.056	.053	.058	.053	.052	.061	.056	.060	.062	.057
BILIN	.052	.055	.054	.052	.054	.061	.056	.059	.058	.057	.059	.061	.059	.063	.061
NMA	.047	.047	.048	.049	.047	.053	.055	.054	.056	.055	.060	.055	.057	.058	.054
TAR1	.059	.051	.060	.059	.055	.061	.057	.057	.061	.060	.067	.067	.066	.067	.066
SQRT	.059	.052	.055	.056	.057	.053	.059	.053	.057	.058	.062	.065	.060	.063	.065
SGN	.049	.051	.050	.049	.051	.059	.062	.058	.059	.062	.053	.056	.054	.051	.058
TAR2	.043	.049	.047	.048	.051	.054	.057	.052	.056	.060	.057	.067	.054	.054	.067
NARCH	.058	.051	.056	.056	.054	.056	.058	.055	.056	.055	.062	.061	.066	.063	.061
	$T = 100$					$T = 200$					$T = 400$				
Model	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$
AR1	.075	.076	.068	.072	.077	.093	.103	.079	.094	.103	.162	.174	.131	.158	.173
AR2	.056	.086	.050	.052	.086	.060	.133	.055	.059	.131	.054	.193	.053	.051	.193
MA	.065	.075	.063	.063	.072	.090	.108	.082	.090	.109	.127	.165	.111	.127	.163
BILIN	.071	.073	.070	.074	.073	.102	.103	.082	.101	.102	.139	.137	.105	.136	.137
NMA	.072	.073	.067	.071	.072	.095	.093	.082	.097	.092	.114	.114	.096	.116	.114
TAR1	.073	.076	.072	.076	.076	.114	.110	.088	.116	.109	.188	.190	.135	.190	.189
SQRT	.075	.080	.069	.075	.081	.098	.110	.087	.097	.110	.155	.167	.122	.154	.167
SGN	.058	.069	.051	.057	.070	.058	.086	.056	.057	.087	.084	.144	.075	.086	.144
TAR2	.055	.082	.054	.056	.081	.054	.109	.054	.052	.110	.064	.179	.061	.066	.178
NARCH	.080	.073	.074	.079	.073	.106	.098	.095	.105	.098	.153	.151	.136	.153	.151
$\beta_1 = .1$															
	$T = 20$					$T = 30$					$T = 50$				
Model	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$
AR1	.063	.064	.058	.066	.064	.072	.080	.065	.072	.080	.093	.106	.081	.092	.107
AR2	.057	.062	.055	.050	.061	.056	.078	.056	.057	.081	.059	.111	.054	.062	.113
MA	.058	.063	.062	.059	.059	.071	.075	.073	.068	.071	.087	.090	.077	.091	.093
BILIN	.053	.058	.056	.058	.059	.075	.068	.068	.070	.073	.088	.089	.079	.090	.091
NMA	.059	.057	.057	.058	.058	.070	.072	.071	.068	.071	.080	.075	.074	.077	.076
TAR1	.070	.063	.067	.067	.065	.078	.073	.067	.076	.076	.107	.110	.090	.107	.112
SQRT	.063	.064	.057	.057	.061	.076	.074	.069	.071	.075	.093	.103	.084	.092	.104
SGN	.060	.056	.056	.056	.061	.054	.067	.053	.054	.069	.063	.081	.064	.065	.084
TAR2	.060	.064	.056	.056	.068	.054	.070	.049	.055	.075	.069	.105	.063	.069	.105
NARCH	.075	.075	.078	.080	.077	.090	.089	.088	.083	.087	.112	.112	.102	.110	.110
	$T = 100$					$T = 200$					$T = 400$				
Model	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$	NW	GLS$_B$	KVB	QS	GLS$_{1\%}$
AR1	.138	.158	.109	.131	.159	.253	.301	.191	.254	.302	.458	.531	.346	.455	.529
AR2	.055	.190	.049	.052	.191	.068	.347	.060	.062	.343	.072	.600	.065	.070	.600
MA	.138	.168	.116	.135	.165	.212	.277	.170	.209	.277	.420	.524	.310	.413	.522
BILIN	.143	.135	.123	.146	.135	.222	.219	.172	.221	.217	.391	.382	.279	.390	.383
NMA	.122	.122	.099	.119	.123	.211	.205	.166	.215	.205	.330	.322	.238	.331	.322
TAR1	.162	.165	.126	.164	.166	.302	.305	.226	.304	.303	.566	.571	.411	.569	.570
SQRT	.135	.155	.118	.133	.155	.253	.295	.202	.248	.295	.433	.497	.322	.438	.496
SGN	.080	.135	.075	.077	.136	.112	.224	.104	.112	.224	.202	.407	.156	.202	.406
TAR2	.066	.176	.064	.063	.176	.090	.325	.078	.083	.325	.115	.541	.101	.114	.538
NARCH	.182	.169	.148	.172	.168	.277	.258	.212	.268	.259	.481	.470	.359	.475	.471

Table 3: Rejection probabilities for $\beta_1 \in \{0.2, 0.5\}$															
$\beta_1 = .2$															
	$T = 20$					$T = 30$					$T = 50$				
Model	NW	GLS_B	KVB	QS	$GLS_{1\%}$	NW	GLS_B	KVB	QS	$GLS_{1\%}$	NW	GLS_B	KVB	QS	$GLS_{1\%}$
AR1	.110	.104	.102	.109	.113	.146	.157	.127	.141	.167	.236	.274	.192	.227	.279
AR2	.070	.100	.064	.065	.108	.066	.167	.066	.067	.177	.087	.293	.078	.077	.307
MA	.096	.102	.086	.090	.101	.128	.140	.122	.116	.149	.191	.222	.164	.183	.231
BILIN	.106	.094	.102	.101	.102	.153	.141	.133	.146	.147	.218	.215	.187	.225	.224
NMA	.100	.084	.093	.094	.096	.129	.119	.114	.121	.120	.181	.170	.147	.179	.172
TAR1	.127	.114	.117	.129	.128	.184	.172	.157	.178	.182	.287	.298	.231	.285	.299
SQRT	.102	.113	.093	.099	.114	.149	.151	.129	.142	.155	.213	.262	.178	.213	.267
SGN	.081	.083	.075	.073	.089	.088	.125	.079	.085	.128	.108	.187	.097	.101	.189
TAR2	.080	.097	.074	.071	.107	.084	.157	.079	.081	.166	.095	.279	.088	.092	.285
NARCH	.155	.137	.150	.140	.150	.214	.193	.185	.195	.196	.303	.284	.255	.288	.287
	$T = 100$					$T = 200$					$T = 400$				
Model	NW	GLS_B	KVB	QS	$GLS_{1\%}$	NW	GLS_B	KVB	QS	$GLS_{1\%}$	NW	GLS_B	KVB	QS	$GLS_{1\%}$
AR1	.429	.500	.311	.412	.501	.732	.814	.553	.731	.814	.957	.983	.803	.955	.981
AR2	.091	.580	.082	.078	.581	.111	.870	.094	.108	.869	.151	.994	.131	.141	.994
MA	.377	.478	.291	.361	.480	.658	.797	.497	.654	.795	.928	.979	.763	.928	.979
BILIN	.392	.389	.304	.393	.389	.644	.652	.477	.641	.648	.888	.902	.706	.883	.901
NMA	.345	.334	.252	.340	.334	.599	.589	.438	.603	.589	.839	.846	.653	.840	.846
TAR1	.502	.527	.387	.509	.527	.818	.827	.624	.819	.825	.986	.988	.872	.987	.988
SQRT	.412	.486	.310	.398	.488	.713	.789	.530	.706	.789	.955	.973	.802	.955	.973
SGN	.184	.399	.158	.175	.400	.355	.674	.272	.344	.674	.625	.937	.464	.623	.938
TAR2	.129	.510	.109	.123	.515	.190	.805	.160	.175	.802	.305	.982	.238	.294	.981
NARCH	.509	.496	.403	.494	.498	.754	.738	.587	.746	.737	.925	.932	.799	.923	.931
$\beta_1 = .5$															
	$T = 20$					$T = 30$					$T = 50$				
Model	NW	GLS_B	KVB	QS	$GLS_{1\%}$	NW	GLS_B	KVB	QS	$GLS_{1\%}$	NW	GLS_B	KVB	QS	$GLS_{1\%}$
AR1	.377	.374	.321	.355	.417	.587	.650	.466	.539	.665	.815	.895	.654	.792	.899
AR2	.169	.427	.159	.153	.476	.195	.702	.171	.174	.729	.236	.939	.214	.201	.944
MA	.339	.376	.300	.318	.399	.515	.614	.430	.466	.632	.765	.868	.612	.742	.872
BILIN	.376	.334	.325	.341	.381	.559	.545	.464	.513	.565	.765	.772	.633	.750	.784
NMA	.331	.287	.282	.305	.326	.492	.462	.404	.451	.481	.694	.698	.557	.670	.705
TAR1	.463	.431	.415	.439	.477	.684	.682	.565	.643	.701	.892	.911	.761	.881	.914
SQRT	.382	.394	.319	.336	.432	.569	.617	.462	.524	.636	.788	.881	.636	.770	.884
SGN	.216	.305	.193	.191	.332	.282	.504	.242	.258	.518	.436	.790	.354	.406	.793
TAR2	.213	.389	.186	.189	.441	.265	.630	.223	.241	.659	.336	.898	.285	.296	.903
NARCH	.561	.508	.509	.503	.558	.710	.697	.609	.655	.708	.857	.864	.758	.832	.867
	$T = 100$					$T = 200$					$T = 400$				
Model	NW	GLS_B	KVB	QS	$GLS_{1\%}$	NW	GLS_B	KVB	QS	$GLS_{1\%}$	NW	GLS_B	KVB	QS	$GLS_{1\%}$
AR1	.990	.997	.900	.986	.997	1.00	1.00	.986	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR2	.309	1.00	.257	.258	1.00	.426	1.00	.351	.394	1.00	.630	1.00	.490	.596	1.00
MA	.977	.997	.864	.970	.997	1.00	1.00	.983	1.00	1.00	1.00	1.00	.999	1.00	1.00
BILIN	.955	.966	.857	.946	.965	.997	.999	.970	.995	.999	1.00	1.00	.997	1.00	1.00
NMA	.927	.942	.811	.922	.941	.998	.999	.958	.998	.999	1.00	1.00	.997	1.00	1.00
TAR1	.996	.997	.942	.996	.997	1.00	1.00	.996	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SQRT	.988	.996	.902	.985	.996	1.00	1.00	.986	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SGN	.772	.983	.601	.748	.983	.988	1.00	.871	.986	1.00	1.00	1.00	.976	1.00	1.00
TAR2	.506	.997	.404	.471	.997	.722	1.00	.591	.687	1.00	.919	1.00	.773	.911	1.00
NARCH	.970	.977	.906	.967	.977	.997	.997	.978	.997	.997	.999	1.00	.997	1.00	1.00

Table 4: Rejection probabilities under long-range dependence in x_t and u_t

			GLS-TD						GLS-FD					
d_x	d_u	ϕ/T	16	32	64	128	256	512	16	32	64	128	256	512
							$\beta_1 = 0$							
0.2	0.2	0	0.143	0.084	0.071	0.062	0.060	0.059	0.411	0.163	0.100	0.067	0.058	0.054
0.2	0.2	0.5	0.136	0.078	0.061	0.051	0.052	0.049	0.416	0.164	0.097	0.067	0.058	0.052
0.2	0.2	0.9	0.140	0.077	0.058	0.052	0.055	0.051	0.421	0.169	0.096	0.068	0.062	0.055
0.2	0.2	0.98	0.137	0.074	0.056	0.051	0.050	0.049	0.428	0.166	0.099	0.068	0.056	0.051
0.2	0.4	0	0.131	0.059	0.043	0.041	0.042	0.043	0.468	0.183	0.084	0.053	0.041	0.035
0.2	0.4	0.5	0.128	0.055	0.041	0.037	0.037	0.040	0.477	0.176	0.084	0.048	0.034	0.030
0.2	0.4	0.9	0.123	0.056	0.036	0.032	0.033	0.032	0.473	0.184	0.084	0.048	0.033	0.027
0.2	0.4	0.98	0.131	0.062	0.048	0.039	0.041	0.041	0.457	0.193	0.096	0.054	0.038	0.030
0.4	0.2	0	0.163	0.103	0.085	0.078	0.069	0.064	0.418	0.183	0.114	0.084	0.068	0.064
0.4	0.2	0.5	0.159	0.090	0.069	0.059	0.053	0.055	0.428	0.177	0.111	0.075	0.062	0.059
0.4	0.2	0.9	0.164	0.089	0.064	0.058	0.056	0.052	0.430	0.179	0.108	0.078	0.065	0.056
0.4	0.2	0.98	0.163	0.088	0.064	0.055	0.053	0.054	0.445	0.188	0.112	0.076	0.063	0.058
0.4	0.4	0	0.157	0.077	0.047	0.042	0.044	0.044	0.493	0.216	0.104	0.062	0.046	0.035
0.4	0.4	0.5	0.150	0.073	0.048	0.036	0.035	0.034	0.512	0.224	0.115	0.061	0.038	0.027
0.4	0.4	0.9	0.153	0.069	0.045	0.035	0.031	0.029	0.523	0.241	0.119	0.067	0.042	0.028
0.4	0.4	0.98	0.152	0.080	0.053	0.043	0.040	0.039	0.495	0.236	0.129	0.069	0.044	0.032
							$\beta_1 = 0.2$							
0.2	0.2	0	0.089	0.193	0.348	0.619	0.894	0.996	0.073	0.170	0.316	0.606	0.894	0.995
0.2	0.2	0.5	0.100	0.192	0.378	0.655	0.925	0.997	0.077	0.171	0.341	0.634	0.914	0.997
0.2	0.2	0.9	0.101	0.213	0.412	0.690	0.948	0.999	0.074	0.172	0.366	0.640	0.938	0.999
0.2	0.2	0.98	0.102	0.219	0.431	0.730	0.959	1.000	0.064	0.168	0.375	0.690	0.947	0.999
0.2	0.4	0	0.106	0.269	0.525	0.838	0.987	1.000	0.059	0.138	0.320	0.620	0.936	0.999
0.2	0.4	0.5	0.106	0.258	0.554	0.844	0.983	1.000	0.057	0.124	0.279	0.542	0.855	0.990
0.2	0.4	0.9	0.126	0.279	0.522	0.820	0.980	1.000	0.052	0.107	0.222	0.456	0.750	0.954
0.2	0.4	0.98	0.107	0.253	0.501	0.806	0.979	1.000	0.065	0.136	0.273	0.558	0.842	0.988
0.4	0.2	0	0.105	0.188	0.344	0.584	0.862	0.991	0.069	0.167	0.318	0.576	0.859	0.988
0.4	0.2	0.5	0.105	0.192	0.359	0.622	0.888	0.995	0.078	0.167	0.325	0.603	0.875	0.995
0.4	0.2	0.9	0.103	0.201	0.372	0.646	0.909	0.996	0.069	0.166	0.332	0.601	0.897	0.995
0.4	0.2	0.98	0.111	0.213	0.378	0.660	0.925	0.998	0.075	0.179	0.329	0.619	0.908	0.997
0.4	0.4	0	0.111	0.233	0.446	0.766	0.965	1.000	0.070	0.152	0.277	0.522	0.838	0.986
0.4	0.4	0.5	0.107	0.245	0.473	0.742	0.957	0.998	0.063	0.142	0.252	0.441	0.739	0.948
0.4	0.4	0.9	0.121	0.245	0.448	0.736	0.940	0.995	0.061	0.128	0.206	0.380	0.623	0.861
0.4	0.4	0.98	0.116	0.237	0.417	0.722	0.946	0.998	0.066	0.151	0.237	0.455	0.734	0.932
							$\beta_1 = 0.5$							
0.2	0.2	0	0.335	0.725	0.961	1.000	1.000	1.000	0.188	0.651	0.954	0.999	1.000	1.000
0.2	0.2	0.5	0.337	0.742	0.976	1.000	1.000	1.000	0.182	0.660	0.962	1.000	1.000	1.000
0.2	0.2	0.9	0.345	0.777	0.980	1.000	1.000	1.000	0.167	0.677	0.959	1.000	1.000	1.000
0.2	0.2	0.98	0.359	0.804	0.986	1.000	1.000	1.000	0.162	0.682	0.963	1.000	1.000	1.000
0.2	0.4	0	0.418	0.874	0.992	1.000	1.000	1.000	0.118	0.503	0.860	0.993	1.000	1.000
0.2	0.4	0.5	0.422	0.867	0.989	1.000	1.000	1.000	0.094	0.430	0.784	0.968	0.999	1.000
0.2	0.4	0.9	0.435	0.863	0.983	1.000	1.000	1.000	0.088	0.358	0.660	0.916	0.995	1.000
0.2	0.4	0.98	0.421	0.858	0.990	1.000	1.000	1.000	0.120	0.490	0.801	0.970	0.999	1.000
0.4	0.2	0	0.318	0.697	0.943	0.998	1.000	1.000	0.165	0.630	0.917	0.998	1.000	1.000
0.4	0.2	0.5	0.328	0.712	0.961	1.000	1.000	1.000	0.169	0.643	0.946	0.999	1.000	1.000
0.4	0.2	0.9	0.336	0.732	0.967	1.000	1.000	1.000	0.147	0.617	0.940	0.999	1.000	1.000
0.4	0.2	0.98	0.347	0.758	0.971	1.000	1.000	1.000	0.150	0.629	0.942	1.000	1.000	1.000
0.4	0.4	0	0.369	0.811	0.977	0.999	1.000	1.000	0.113	0.465	0.787	0.971	0.998	1.000
0.4	0.4	0.5	0.385	0.793	0.974	0.997	1.000	1.000	0.094	0.389	0.707	0.917	0.993	1.000
0.4	0.4	0.9	0.390	0.798	0.961	0.996	1.000	1.000	0.090	0.349	0.596	0.845	0.970	0.997
0.4	0.4	0.98	0.384	0.798	0.971	0.999	1.000	1.000	0.110	0.459	0.727	0.924	0.991	1.000

Figure 1: Null rejection probabilities grouped according to u_t process for $T = 10, 11, \ldots, 100$

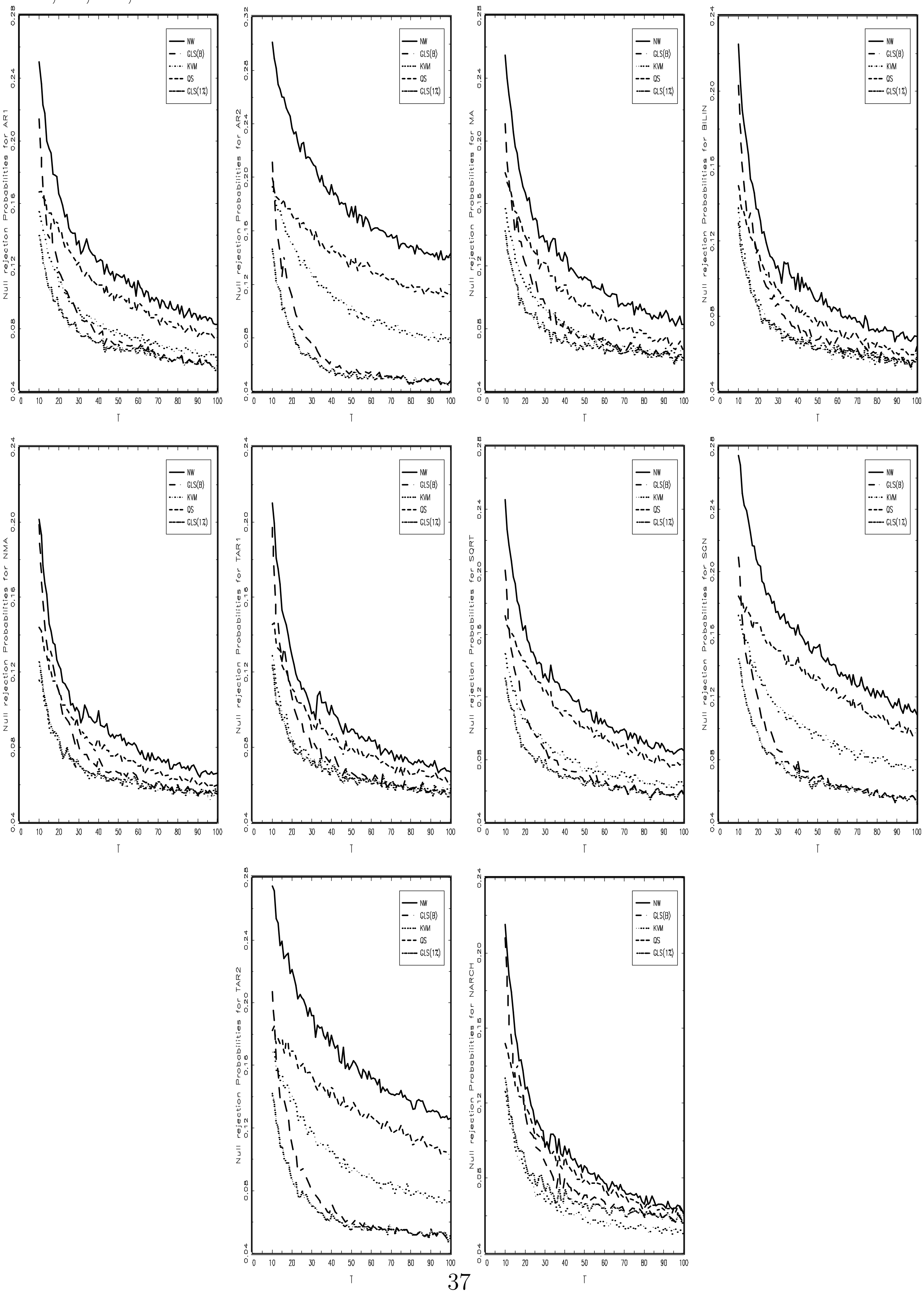

Figure 2: Null rejection probabilities grouped according to test for $T = 10, 11, \ldots, 100$