

Department of Economics

Semiparametric Sieve-Type GLS Inference in Regressions
with Long-Range Dependence

George Kapetanios and Zacharias Psaradakis

Working Paper No. 587

March 2007

ISSN 1473-0278



Semiparametric Sieve-Type GLS Inference in Regressions with Long-Range Dependence

George Kapetanios*

Queen Mary, University of London

Zacharias Psaradakis†

Birkbeck, University of London

February 2007

Abstract

This paper considers the problem of statistical inference in linear regression models whose stochastic regressors and errors may exhibit long-range dependence. A time-domain 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. A Monte Carlo study examines the finite-sample properties of the method for testing regression hypotheses.

Keywords: Autoregressive approximation; Generalized least squares; Linear regression; Long-range dependence; Spectral density.

JEL classification: C12; C13; C22.

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

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

1 Introduction

This paper considers the problem of statistical inference in linear regression models in which both the stochastic regressors and the errors may exhibit long-range dependence (in the sense of having autocorrelations which are not absolutely summable). In such models, the ordinary least squares (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 conventional inferential procedures that rely 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., Andrews (1991)) cannot be justified in these circumstances. Furthermore, even in models with nonstochastic regressors, where the OLS estimator can be asymptotically normal for certain designs, asymptotic efficiency is generally unattainable and the rate of convergence may be slow (cf. Yajima (1988, 1991); Dahlhaus (1995)).

An alternative way to conduct inference is by relying on the generalized least squares (GLS) estimator, which is the best linear unbiased estimator (BLUE) of the regression coefficients under very general conditions. 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)).

The obvious difficulty with the GLS-based approach is that it 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 utilizes an autoregressive approximation to the generating mechanism of the errors. Amemiya (1973) established the asymptotic normality and asymptotic Gauss–Markov efficiency of a feasible (or two-step) GLS estimator based on such an autoregressive approximation when the regressors are nonstochastic and the errors are generated by a short-range dependent linear process with independent and identically distributed (i.i.d.) innovations. Kapetanios (2006) extended the analysis to allow for stochastic regressors and errors which satisfy quite general near-epoch dependence conditions. What is more, the simulation evidence in Kapetanios (2006) reveals that GLS-based hypothesis tests generally have better finite-sample size and power properties than OLS-based robust

tests that rely on methods such as those discussed in Andrews and Monahan (1992) and Kiefer, Vogelsang, and Bunzel (2000).

The present paper extends the GLS approach based on autoregressive approximations to models in which the regressors and errors may exhibit arbitrarily strong long-range dependence. The basic idea is to approximate 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 estimates. Our 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. Such a sequence of approximations may be viewed as a sieve, in the sense of Grenander (1981), which is why we refer to our procedure as semiparametric sieve-type GLS.

Autoregressive sieve-type GLS is, to the best of our knowledge, the only currently available efficient estimation procedure that can be implemented in the time domain. A related approach in the frequency domain 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 feasible GLS procedure. Nielsen (2005) considered an alternative frequency-domain feasible GLS 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.

The paper proceeds as follows. Section 2 introduces the model and regularity conditions, describes the sieve-type GLS estimation procedure, and establishes its asymptotic properties. Section 3 reports the results of a simulation study of the small-sample performance of the method in the context of testing regression hypotheses. Section 4 summarizes and concludes. Mathematical proofs are presented in an Appendix.

2 Assumptions, Estimation Method, and Asymptotic Results

Consider a regression model of the form

$$y_t = \alpha + x_t' \beta + u_t, \quad t = 1, \dots, T, \quad (1)$$

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

Letting L denote the conventional lag operator ($Lu_t = u_{t-1}$), we make the following assumptions about the errors and regressors in (1).

Assumption 1 $\{u_t\}$ is a second-order stationary process satisfying

$$u_t = (1 - L)^{-d} \xi_t, \quad \xi_t = \sum_{j=0}^{\infty} \delta_j \varepsilon_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots, \quad (2)$$

for some $0 \leq d < \frac{1}{2}$. In (2), $\{\varepsilon_t\}$ is an ergodic sequence of random variables such that $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ a.s., $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2 > 0$ a.s., and $\sup_t E(|\varepsilon_t|^4) < \infty$, \mathcal{F}_t being the σ -field generated by $\{\varepsilon_s; s \leq t\}$, and $\{\delta_j; j \geq 0\}$ is a non-random sequence (with $\delta_0 = 1$) satisfying $\sum_{j=0}^{\infty} |\delta_j| < \infty$ and $\delta(z) = \sum_{j=0}^{\infty} \delta_j z^j \neq 0$ for all complex z with $|z| \leq 1$.

Assumption 2 $\{x_t\}$ is a fourth-order stationary process such that

$$\lim_{|h| \rightarrow \infty} E[(x_t - \mu_x)(x_{t+h} - \mu_x)'] = 0,$$

and

$$\lim_{|t_1| \rightarrow \infty} \sup_{|t_2|, |t_3| < \infty} |\kappa_{abcq}(0, t_1, t_2, t_3)| = 0, \quad 1 \leq a, b, c, q \leq k,$$

where $\mu_x = E(x_t)$ and $\kappa_{abcq}(0, t_1, t_2, t_3)$ is the fourth cumulant of $(x_{a,0}, x_{b,t_1}, x_{c,t_2}, x_{q,t_3})$.

Assumption 3 $\{\varepsilon_t\}$ and $\{x_t\}$ are mutually independent.

As usual, the operator $(1 - L)^{-d}$ in (2) is defined by using the power series expansion of $(1 - z)^{-d}$ ($|z| < 1$), i.e.,

$$(1 - L)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} L^j,$$

where $\Gamma(\cdot)$ denotes the gamma function. Assumption 1 thus requires $\{u_t\}$ to be a fractional process with memory (or long-range dependence, or fractional differencing) parameter d . Under this assumption, $\phi(z) = (1 - z)^d/\delta(z)$ admits an absolutely convergent power series expansion $\phi(z) = \sum_{j=0}^{\infty} \phi_j z^j$ ($|z| < 1$), with $\phi_0 = 1$, implying that $\{u_t\}$ has the autoregressive representation

$$\sum_{j=0}^{\infty} \phi_j u_{t-j} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \quad (3)$$

Similarly, letting $\psi(z) = (1 - z)^{-d}\delta(z) = \sum_{j=0}^{\infty} \psi_j z^j$ ($|z| < 1$), it can be easily seen that $\{u_t\}$ admits the causal moving-average representation

$$u_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots, \quad (4)$$

with $\psi_0 = 1$, $\psi_j \sim \{\delta(1)/\Gamma(d)\} j^{d-1}$ as $j \rightarrow \infty$, and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Furthermore, the spectral density $f_u(\cdot)$ of $\{u_t\}$, satisfying

$$E(u_t u_{t+h}) = \int_{-\pi}^{\pi} e^{ih\lambda} f_u(\lambda) d\lambda, \quad h = 0, \pm 1, \pm 2, \dots$$

(with $i = \sqrt{-1}$), is of the form

$$f_u(\lambda) = (2\pi)^{-1} \sigma_{\varepsilon}^2 |\delta(e^{i\lambda})|^2 |1 - e^{i\lambda}|^{-2d}, \quad -\pi < \lambda \leq \pi,$$

and hence has a pole at the origin. 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)) and the fractional Gaussian noise (Mandelbrot and Van Ness (1968)). If $d = 0$, $\{u_t\}$ is evidently a short-range dependent linear process with $0 < f_u(0) < \infty$. We note that Assumption 1 is stronger than the corresponding assumption of Robinson and Hidalgo (1997), who only require the existence of the linear 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.

Assumptions 2–3 are the same as those employed by Robinson and Hidalgo (1997) and Hidalgo and Robinson (2002). Assumption 2 is fairly mild and allows for stochastic regressors which may exhibit short-range or long-range dependence. An example of a random process satisfying this assumption is the k -variate ARFIMA process or, more generally, the fractional process defined by

$$x_t - \mu_x = D(L)v_t, \quad v_t = \sum_{j=0}^{\infty} \Psi_j w_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots$$

Here, $D(L) = \text{diag}\{(1 - L)^{-d_1}, \dots, (1 - L)^{-d_k}\}$ for some $0 \leq d_j < \frac{1}{2}$ ($j = 1, \dots, k$), $\{\Psi_j\}$ is an absolutely summable sequence of nonstochastic $k \times k$ matrices, and $\{w_t\}$ is an ergodic k -variate martingale difference sequence (with respect to the σ -fields generated by $\{w_s; s \leq t\}$) satisfying suitable stationarity and moment conditions. It is worth pointing out that Assumptions 1–2 permit the errors and regressors to exhibit strong collective long-range dependence, in the sense that $d + d_j$ is allowed to be arbitrarily close to unity for some $1 \leq j \leq k$. This is a case of special interest because the OLS estimator of β is known to have a non-Gaussian limiting distribution and a rate of convergence slower than $O_p(1/\sqrt{T})$ when at least one of the regressors in (1) has memory parameter d_j satisfying $d_j + d > \frac{1}{2}$ (cf. Robinson (1994); Chung (2002)). The cumulant condition of Assumption 2 is weaker than the summability conditions on cumulants that are frequently used in the time-series literature and is satisfied trivially when $\{x_t\}$ is Gaussian. By way of comparison, the maintained assumption in Nielsen (2005) is that $\{(x'_t, u_t)'\}$ is a $(k + 1)$ -variate linear process with square-summable coefficients, fourth-order stationary martingale difference innovations, and spectral density matrix which satisfies certain long-range dependence conditions.

The strict exogeneity of the explanatory variables imposed by Assumption 3 is admittedly restrictive. As remarked by Robinson and Hidalgo (1997) and Hidalgo and Robinson (2002), 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. Even such a condition, however, 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 neighbourhood of the zero frequency) version of the usual orthogonality condition.

It is well known that, under Assumption 2, there exists a continuous matrix-valued spectral distribution function $F_x(\cdot)$ with Hermitian positive semidefinite increments

$F_x(\lambda_2) - F_x(\lambda_1)$, $\pi \geq \lambda_2 \geq \lambda_1 > -\pi$, such that

$$E[(x_t - \mu_x)(x_{t+h} - \mu_x)'] = \int_{-\pi}^{\pi} e^{ih\lambda} dF_x(\lambda), \quad h = 0, \pm 1, \pm 2, \dots$$

We make the following additional identifiability assumption, which guarantees non-degenerate asymptotic distributional results.

Assumption 4 *The matrix $\Sigma = \int_{-\pi}^{\pi} \{2\pi f_u(\lambda)\}^{-1} dF_x(\lambda)$ is positive definite.*

Letting $y = (y_1, \dots, y_T)'$, $X = (\check{x}_1, \dots, \check{x}_T)'$, with $\check{x}_t = (x'_t, 1)'$, and $u = (u_1, \dots, u_T)'$, the BLUE of the parameter vector $\theta = (\beta', \alpha)'$ is given by

$$\tilde{\theta} = (\tilde{\beta}', \tilde{\alpha})' = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y, \quad (5)$$

where

$$\Omega = E(uu') = \left\{ \int_{-\pi}^{\pi} e^{i(t-s)\lambda} f_u(\lambda) d\lambda; t, s = 1, \dots, T \right\}.$$

In the case where Ω is unknown, a feasible GLS estimator might be used instead of the BLUE. Such an estimator is of the form

$$\hat{\theta} = (\hat{\beta}', \hat{\alpha})' = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y, \quad (6)$$

where $\hat{\Omega}$ is a suitable estimator of Ω .

Following Amemiya (1973), we propose to construct $\hat{\Omega}$ by using an approximation to the autoregressive representation of $\{u_t\}$ given in (3). To describe the method in detail, let

$$\bar{u}_t = y_t - \bar{y} - (x_t - \bar{x})'\bar{\beta}, \quad t = 1, \dots, T,$$

where $\bar{y} = T^{-1} \sum_{t=1}^T y_t$, $\bar{x} = T^{-1} \sum_{t=1}^T x_t$, and $\bar{\beta}$ is a preliminary estimator of β . Further, for some positive integer p chosen as a function of T so that $p \rightarrow \infty$ and $p/T \rightarrow 0$ as $T \rightarrow \infty$, let $\hat{\phi}_p = (\hat{\phi}_{p,1}, \dots, \hat{\phi}_{p,p})'$ be the p -th order OLS estimator of the autoregressive coefficients for $\{\bar{u}_t\}$, obtained as the solution to the minimization of

$$(T-p)^{-1} \sum_{t=p+1}^T (\bar{u}_t + \phi_{p,1}\bar{u}_{t-1} + \dots + \phi_{p,p}\bar{u}_{t-p})^2 \quad (7)$$

over $\phi_p = (\phi_{p,1}, \dots, \phi_{p,p})' \in \mathbb{R}^p$. The estimator of $f_u(\cdot)$ associated with this autoregressive approximation is given by

$$\hat{f}_{\bar{u}}(\lambda) = \frac{\hat{\sigma}_{\bar{u},p}^2}{2\pi} \left| 1 + \sum_{j=1}^p \hat{\phi}_{p,j} e^{ij\lambda} \right|^{-2}, \quad -\pi < \lambda \leq \pi, \quad (8)$$

where $\widehat{\sigma}_{\bar{u},p}^2$ is the minimum of (7). Our feasible GLS estimator is then defined as in (6) with

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

The covariance matrix of $\widehat{\theta}$ can be estimated by $(X' \widehat{\Omega}^{-1} X)^{-1}$.

Remark 1 It is worth noting that a computationally more attractive feasible GLS estimator of θ , with the same asymptotic properties as $\widehat{\theta}$, can be obtained by replacing $\widehat{\Omega}^{-1}$ in (6) with $\widehat{\Phi}' \widehat{\Phi}$, where $\widehat{\Phi}$ is the $(T - p) \times T$ matrix defined as

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

This, of course, is equivalent to applying OLS to the regression of $(1 + \sum_{j=1}^p \widehat{\phi}_{p,j} L^j) y_t$ on $(1 + \sum_{j=1}^p \widehat{\phi}_{p,j} L^j) \bar{x}_t$. We also note that Yule–Walker or Burg-type estimates may be used instead of the OLS estimates $\widehat{\phi}_p$ and $\widehat{\sigma}_{\bar{u},p}^2$; this will not change the first-order asymptotic properties of $\widehat{f}_{\bar{u}}(\lambda)$ and $\widehat{\theta}$.

Remark 2 Using the discrete Fourier transforms of X and y , the GLS estimator $\widetilde{\theta}$ given in (5) may be written as

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

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

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

where $Q = \text{diag}\{f_u(2\pi j/T); j = 0, 1, \dots, T-1\}$. 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 β can be written as

$$\tilde{\beta}_g = \left(\sum_{j=1}^{T-1} I_{xx}(2\pi j/T) g(2\pi j/T) \right)^{-1} \left(\sum_{j=1}^{T-1} I_{xy}(2\pi j/T) g(2\pi j/T) \right). \quad (10)$$

Here, $I_{xx}(\cdot)$ and $I_{xy}(\cdot)$ stand for the periodogram of $\{x_t\}$ and the cross-periodogram of $\{x_t\}$ and $\{y_t\}$, respectively, and $g(\cdot)$ is a real-valued, integrable, even, and periodic function on $[-\pi, \pi]$ with period 2π . We note that the omission of $j = 0$ from the summations in (10) is equivalent to sample-mean correction of the data, and $\tilde{\beta}_g = \tilde{\beta}_f$ if $g(\cdot) = f_u(\cdot)^{-1}$.

To establish the asymptotic distribution of our feasible GLS estimator, we make the following assumptions about the preliminary estimator $\bar{\beta}$ of the slope coefficients and the order p of the autoregressive approximation used to obtain $\hat{\Omega}$ (or $\hat{\Phi}$) .

Assumption 5 $\bar{\beta} = \beta + O_p(1/\sqrt{T})$.

Assumption 6 $p = p(T) \rightarrow \infty$ and $p(T) = O(\{\log T\}^r)$ as $T \rightarrow \infty$ for some $0 < r < \infty$.

Assumption 5 requires the estimator $\bar{\beta}$ to be \sqrt{T} -consistent. This generally rules out the OLS estimator, but 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 $g(\lambda) = |1 - e^{i\lambda}|$. Assumption 6 requires p to increase with the sample size at a suitably slow rate, and is similar to assumptions that are commonly employed in the theory of autoregressive approximations (cf. An, Chen, and Hannan (1982); Hannan and Kavalieris (1986)).

The theorem that follows shows that the autoregressive spectral estimator given in (8) is uniformly consistent.

Theorem 1 *If Assumptions 1–6 hold, then*

$$\sup_{\lambda} \left| \hat{f}_{\bar{u}}(\lambda) - f_u(\lambda) \right| = o_p(1) \quad \text{as } T \rightarrow \infty. \quad (11)$$

In view of the result in (11) and Theorem 1 of Robinson and Hidalgo (1997), it can now be shown that the feasible GLS estimator $\hat{\beta}$ has the same Gaussian asymptotic distribution as the BLUE $\tilde{\beta}$.

Theorem 2 *If Assumptions 1–6 hold, then*

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^{-1}) \quad \text{as } T \rightarrow \infty,$$

where $\xrightarrow{\mathcal{D}}$ signifies convergence in distribution.

We conclude this section by noting that a practical issue in sieve-type GLS inference is the choice of the approximating autoregressive order p in (7). Any data-dependent choice \hat{p}_T , say, obtained by using a data-driven selection procedure which guarantees that \hat{p}_T satisfies Assumption 6 with probability approaching one as $T \rightarrow \infty$ is sufficient for Theorems 1–2 to hold. Using the results in Poskitt (2005), it can be shown that information criteria such as the familiar Akaike, Bayesian and Hannan–Quinn criteria return a lag order \hat{p}_T which is asymptotically acceptable; what is required is that the maximum allowable order p_{\max} , say, be set so as to satisfy $p_{\max} = O(\{\log T\}^r)$ as $T \rightarrow \infty$ for some $0 < r < \infty$.

3 Simulation Evidence

The theoretical part of the paper has argued that the approach of Amemiya (1973) can be extended to cover a rich class of long-range dependent processes, be made automatic through the use of information criteria, 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.

In our numerical experiments, artificial data $\{y_t\}$ are generated according to the model in (1) with $k = 1$, $\alpha = 0$ and $\beta_1 \in \{0, 0.2, 0.5\}$. The generating mechanism of the regressor is either the AR(1) model $x_{1,t} = 0.5x_{1,t-1} + w_t$ or the ARFIMA($0, d_u, 0$) model $x_{1,t} = (1 - L)^{-d_x}w_t$ with $d_x \in \{0.1, 0.2, 0.3, 0.4, 0.49\}$; in both cases, the innovations $\{w_t\}$ are i.i.d. $\mathcal{N}(0, 1)$ random variables. Similarly, the regression 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 $u_t = \phi u_{t-1} + (1 - L)^{-d_u} \varepsilon_t$ with $\phi \in \{0, 0.5, 0.9, 0.98\}$ and $d_u \in \{0.1, 0.2, 0.3, 0.4, 0.49\}$. In the latter case, we consider the following data-generating processes (with $I_{\{A\}}$ denoting the indicator of the event A):

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. $u_t = 0.5I_{\{u_{t-1} \leq 1\}}u_{t-1} + 0.4I_{\{u_{t-1} > 1\}}u_{t-1} + \varepsilon_t$ (TAR1)
7. $u_t = 0.95\sqrt{|u_{t-1}|} + \varepsilon_t$ (SQRT)
8. $u_t = -I_{\{u_{t-1} \leq 0\}} + I_{\{u_{t-1} > 0\}} + \varepsilon_t$ (SGN)
9. $u_t = 0.95I_{\{u_{t-1} \leq 0\}}u_{t-1} - 0.2I_{\{u_{t-1} > 0\}}u_{t-1} + \varepsilon_t$ (TAR2)
10. $u_t = \sqrt{\eta_t}\varepsilon_t, \quad \eta_t = 1 + 0.8\varepsilon_t^2$ (NARCH1)
11. $u_t = \sqrt{\eta_t}\varepsilon_t, \quad \eta_t = 0.25 + 0.5\eta_{t-1} + 0.5y_{t-1}^2I_{\{\varepsilon_t \leq 0\}} + 0.2y_{t-1}^2I_{\{\varepsilon_t > 0\}}$ (NARCH2)

For all the designs, the innovations $\{\varepsilon_t\}$ 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 NARCH1 and NARCH2 are nonlinear GARCH processes. 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 our experiments, the objective is to test the null hypothesis $\mathcal{H}_0 : \beta_1 = 0$ against the alternative $\mathcal{H}_1 : \beta_1 \neq 0$ using a t -type statistic. The parameter value $\beta_1 = 0$ is thus used to compute Type I error probabilities, while the values $\beta_1 = 0.2$ and $\beta_1 = 0.5$ 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 $\beta_1 = 0$. We consider sample sizes $T \in \{16, 32, 64, 128, 256, 512\}$. 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.

In our implementation of the sieve-type GLS procedure, the preliminary estimator used to compute the regression residuals is OLS. This, of course, is not a theoretically attractive choice because the OLS estimator of the slope coefficients is \sqrt{T} -consistent only when $d_x + d_u < \frac{1}{2}$. 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 GLS procedure. The order of the autoregressive approximation is selected by minimizing the Bayesian information criterion of ? over the range $0 \leq p \leq \lfloor 2 \log T \rfloor$ (where $\lfloor \cdot \rfloor$ denotes the integer-part function).

As a competitor to tests based on sieve-type GLS, we consider the frequency-domain GLS procedure of Hidalgo and Robinson (2002).¹ Frequency-domain GLS estimates are computed using the weight function $c(u) = 1 - |u|$ and bandwidth parameter $m = \lfloor T/16 \rfloor$. 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 paragraph.

Table 1 contains Monte Carlo estimates of the rejection probabilities of tests in the case where the stochastic regressor exhibits short-range dependence but the errors are long-range dependent. Here and in the other tables, tests based on time-domain sieve-type GLS are labelled GLS–TD, while tests based on frequency-domain GLS are labelled GLS–FD. 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

¹We would like to thank Štěpána Lazarová for providing us with the computer code used in the simulation study of Hidalgo and Robinson (2002).

GLS–FD test, the difference being particularly prominent in the case of the two smallest sample sizes considered. Both tests exhibit little size distortion for $T \geq 128$; the most noticeable exceptions occur when $d_u > 0.4$, suggesting that Gaussian asymptotic approximations become less accurate as the memory parameter of the errors approaches the boundary of the stationary region. Where power is concerned, the GLS–TD test has an advantage over the GLS–FD test for most design points; this advantage is most prominent for small T and large d_u . As expected, for given d_u and ϕ , the power of both tests tends to approach unity as the sample size increases.

The simulation results for a model with a long-range dependent regressor and short-range dependent errors are presented in Table 2. In nearly every case (with $T \leq 256$), the empirical size of the GLS–TD test is closer to the nominal 0.05 level than the GLS–FD test, the difference between the two tests being especially noticeable for the smaller sample sizes. The empirical size of the GLS–TD test never exceeds 0.076 when $T \geq 64$, the corresponding figure for GLS–FD being 0.127. The performance of the GLS–TD test is particularly impressive in the 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 only three of the eleven cases considered in the experiments. In fact, 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. It is not surprising that test power increases with both the sample size and the difference between the true value of the slope coefficient and its null value.

The estimated rejection probabilities of tests in the case where both the regressor and the errors exhibit long-range dependence are shown in Table 3. Results under the null hypothesis exhibit similar patterns to those reported before. More specifically, the GLS–TD test dominates GLS–FD in terms of size distortion, 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 $d_x + d_u > 0.5$ if a \sqrt{T} -consistent preliminary estimator of the slope parameters was used instead of OLS). The GLS–TD test, and to a lesser extent the GLS–FD test, tend to be somewhat

conservative in the larger samples if the memory parameter of the errors exceeds 0.4. 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 ($d_u > 0.4$).

In sum, the GLS–TD test is found to have the best overall performance in our experiments. It generally exhibits smaller size distortions than the GLS–FD test, especially when the sample size is small, and is also superior in terms of size-adjusted power. Furthermore, it is robust with respect to the presence of neglected nonlinearity in the regression errors.

4 Conclusion

This paper has suggested the use of time-domain sieve-type GLS based on autoregressive approximations for inference in regression models with long-range dependent data. 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 slope parameters is \sqrt{T} -consistent, asymptotically normal and asymptotically Gauss–Markov efficient under general conditions which permit long-range dependence in both 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.

5 Appendix

This Appendix provides proofs of Theorems 1 and 2. Throughout, limits in order symbols are taken as T tends to infinity. C , C_1 and C_2 denote finite constants whose value may change upon each appearance.

Proof of Theorem 1: Let $\phi_p^u = (\phi_{p,1}^u, \dots, \phi_{p,p}^u)'$ and $\sigma_{u,p}^2$ be the solution of the equations

$$\Gamma_p \phi_p^u = -\gamma_p,$$

$$\sigma_{u,p}^2 = \gamma_u(0) + (\phi_p^u)' \gamma_p,$$

where $\Gamma_p = \{\gamma_u(j-s); j, s = 1, \dots, p\}$, $\gamma_p = (\gamma_u(1), \dots, \gamma_u(p))'$ and $\gamma_u(j) = E(u_t u_{t+j})$. Further, let $\widehat{\phi}_p^u = (\widehat{\phi}_{p,1}^u, \dots, \widehat{\phi}_{p,p}^u)'$ be the solution of the equations

$$\widehat{\Gamma}_p \widehat{\phi}_p^u = -\widehat{\gamma}_p,$$

$$\widehat{\sigma}_{u,p}^2 = \widehat{\gamma}_u(0) + (\widehat{\phi}_p^u)' \widehat{\gamma}_p,$$

where $\widehat{\Gamma}_p = \{\widehat{\gamma}_u(j, s); j, s = 1, \dots, p\}$, $\widehat{\gamma}_p = (\widehat{\gamma}_u(1), \dots, \widehat{\gamma}_u(p))'$, $\widehat{\gamma}_u(j) = \widehat{\gamma}_u(0, j)$, and $\widehat{\gamma}_u(j, s) = T^{-1} \sum_{t=\max\{j,s\}+1}^T u_{t-j} u_{t-s}$. Also, for $-\pi < \lambda \leq \pi$, define the following autoregressive spectral densities:

$$\begin{aligned}\widehat{f}_u(\lambda) &= \frac{\widehat{\sigma}_{u,p}^2}{2\pi} \left| 1 + \sum_{j=1}^p \widehat{\phi}_{p,j}^u e^{ij\lambda} \right|^{-2}, \\ \bar{f}_{p,u}(\lambda) &= \frac{\sigma_{u,p}^2}{2\pi} \left| 1 + \sum_{j=1}^p \phi_{p,j}^u e^{ij\lambda} \right|^{-2}, \\ f_{p,u}(\lambda) &= \frac{\sigma_{u,p}^2}{2\pi} \left| 1 + \sum_{j=1}^p \phi_j e^{ij\lambda} \right|^{-2}.\end{aligned}$$

By the triangle inequality, we have

$$\begin{aligned}\sup_{\lambda} \left| \widehat{f}_{\bar{u}}(\lambda) - f_u(\lambda) \right| &\leq \sup_{\lambda} \left| \widehat{f}_{\bar{u}}(\lambda) - \widehat{f}_u(\lambda) \right| + \sup_{\lambda} \left| \widehat{f}_u(\lambda) - \bar{f}_{p,u}(\lambda) \right| \\ &\quad + \sup_{\lambda} \left| \bar{f}_{p,u}(\lambda) - f_{p,u}(\lambda) \right| + \sup_{\lambda} |f_{p,u}(\lambda) - f_u(\lambda)| \\ &= S_1 + S_2 + S_3 + S_4,\end{aligned}$$

with an obvious notation for S_1, \dots, S_4 . The theorem may, therefore, be proved by showing that S_1 and S_2 are $o_p(1)$ and that S_3 and S_4 are $o(1)$.

Consider the term S_1 first. It is easy to see that

$$\left| \widehat{f}_{\bar{u}}(\lambda) - \widehat{f}_u(\lambda) \right| \leq C \sum_{j=1}^p \left| \widehat{\phi}_{p,j} - \widehat{\phi}_{p,j}^u \right| e^{ij\lambda}.$$

However, in view of Assumption 5, we have

$$\begin{aligned}
\sum_{j=1}^p \left| \widehat{\phi}_{p,j} - \widehat{\phi}_{p,j}^u \right| &\leq C \sum_{h=1}^p \left(T^{-1} \sum_{t=h+1}^T \bar{u}_t \bar{u}_{t-h} - T^{-1} \sum_{t=h+1}^T u_t u_{t-h} \right) \\
&\leq C_1 p \left(T^{-1} \sum_{t=1}^T \bar{u}_t^2 - T^{-1} \sum_{t=1}^T u_t^2 \right) \\
&\leq C_2 p \left(T^{-1} \sum_{t=1}^T u_t (\bar{u}_t - u_t) \right) \\
&= C_2 p \left((\bar{\beta} - \beta) T^{-1} \sum_{t=1}^T (x_t - \bar{x}) u_t \right) \\
&= o_p(1),
\end{aligned}$$

and hence $S_1 = o_p(1)$.

Consider next the term S_2 . It is not difficult to see that $S_2 = o_p(1)$ provided that

$$\sup_{\lambda} \left| \sum_{j=1}^p \left(\widehat{\phi}_{p,j}^u - \phi_{p,j}^u \right) e^{ij\lambda} \right| = o_p(1) \quad (12)$$

and

$$|\widehat{\sigma}_{u,p}^2 - \sigma_{u,p}^2| = o_p(1). \quad (13)$$

But

$$\sup_{\lambda} \left| \sum_{j=1}^p \left(\widehat{\phi}_{p,j}^u - \phi_{p,j}^u \right) e^{ij\lambda} \right| \leq C p \sum_{j=1}^p \left(\widehat{\phi}_{p,j}^u - \phi_{p,j}^u \right)^2$$

and

$$|\widehat{\sigma}_{u,p}^2 - \sigma_{u,p}^2| = |\widehat{\gamma}_u(0) - \gamma_u(0) + (\widehat{\phi}_p^u)' \widehat{\gamma}_p - (\phi_p^u)' \gamma_p|.$$

Therefore, (12) and (13) follow if

$$\sum_{j=1}^p \left(\widehat{\phi}_{p,j}^u - \phi_{p,j}^u \right)^2 = O_p(T^{-r}) \text{ for some } r > 0 \quad (14)$$

and

$$\sum_{j=1}^p |\widehat{\gamma}_u(j) - \gamma_u(j)| = O_p(T^{-r}) \text{ for some } r > 0. \quad (15)$$

In view of Theorem 5.1 of Poskitt (2005), (14) holds for all $0 \leq d < \frac{1}{2}$. Furthermore, by Theorem 4.1 of Poskitt (2005),

$$\max_{0 \leq j \leq p} |\widehat{\gamma}_u(j) - \gamma_u(j)| = O_p(T^{-r}) \text{ for some } r > 0.$$

Consequently, since

$$\sum_{j=1}^p |\hat{\gamma}_u(j) - \gamma_u(j)| \leq p \max_{0 \leq j \leq p} |\hat{\gamma}_u(j) - \gamma_u(j)|,$$

(15) holds for all $0 \leq d < \frac{1}{2}$.

Moving on to the term S_3 , we have

$$\begin{aligned} & |\bar{f}_{p,u}(\lambda) - f_{p,u}(\lambda)| \\ &= \frac{\left| \left| 1 + \sum_{j=1}^p \phi_{p,j}^u e^{ij\lambda} \right|^2 - \left| 1 + \sum_{j=1}^p \phi_j e^{ij\lambda} \right|^2 \right|}{2\pi \left| 1 + \sum_{j=1}^p \phi_{p,j}^u e^{ij\lambda} \right|^2 \left| 1 + \sum_{j=1}^p \phi_j e^{ij\lambda} \right|^2} \\ &\leq \frac{\left| \sum_{j=1}^p (\phi_{p,j}^u - \phi_j) e^{ij\lambda} \right|^2 - \left| 1 + \sum_{j=1}^p \phi_{p,j}^u e^{ij\lambda} \right| \left| \sum_{j=1}^p (\phi_{p,j}^u - \phi_j) e^{ij\lambda} \right|}{2\pi \left| 1 + \sum_{j=1}^p \phi_{p,j}^u e^{ij\lambda} \right|^2 \left| 1 + \sum_{j=1}^p \phi_j e^{ij\lambda} \right|^2}. \end{aligned}$$

Since

$$\left| \sum_{j=1}^p (\phi_{p,j}^u - \phi_j) e^{ij\lambda} \right|^2 \leq \sum_{j=1}^p |\phi_{p,j}^u - \phi_j|^2 \leq \sqrt{p} \left\{ \sum_{j=1}^p (\phi_{p,j}^u - \phi_j)^2 \right\}^{\frac{1}{2}},$$

we examine $\sum_{j=1}^p (\phi_{p,j}^u - \phi_j)^2$. Letting $\phi^{(p)} = (\phi_1, \dots, \phi_p)'$, we have

$$\phi_p^u - \phi^{(p)} = -\Gamma_p^{-1} \zeta_p,$$

where $\zeta_p = (\zeta_p(1), \dots, \zeta_p(p))'$ with $\zeta_p(j) = \sum_{h=1}^{\infty} \phi_{p+h} \gamma_u(p+h-j)$, $1 \leq j \leq p$. But

$$|\zeta_p(j)| \leq \gamma_u(0) \sum_{j=p+1}^{\infty} |\phi_j| = o(\{\log T\}^r), \quad \text{for all } r > 0,$$

and so

$$\begin{aligned} & p (\phi_p^u - \phi^{(p)})' (\phi_p^u - \phi^{(p)}) \\ &\leq p^3 \left(\sum_{j=p+1}^{\infty} |\phi_j| \right) \{\rho_{\min}(\Gamma_p^2)\}^{-1} \\ &= o(\{\log T\}^r), \quad \text{for all } r > 0, \end{aligned}$$

where $\rho_{\min}(\cdot)$ denotes the minimum eigenvalue of its argument. Hence, $S_3 = o(1)$.

Finally, consider the term S_4 . We have

$$|f_{p,u}(\lambda) - f_u(\lambda)| \leq \frac{\left| \sum_{j=p+1}^{\infty} \phi_j e^{ij\lambda} \right|^2 - \left| 1 + \sum_{j=1}^p \phi_j e^{ij\lambda} \right| \left| \sum_{j=p+1}^{\infty} \phi_j e^{ij\lambda} \right|}{2\pi \left| 1 + \sum_{j=1}^p \phi_j e^{ij\lambda} \right|^2 \left| 1 + \sum_{j=1}^{\infty} \phi_j e^{ij\lambda} \right|^2}.$$

But since

$$\left| \sum_{j=p+1}^{\infty} \phi_j e^{ij\lambda} \right| \leq \sum_{j=p+1}^{\infty} |\phi_j| = o(\{\log T\}^r), \text{ for all } r > 0,$$

we conclude that $S_4 = o(1)$, which completes the proof of the theorem. ■

Proof of Theorem 2: To establish the claim of the theorem, it is sufficient to show that

$$\sqrt{T}(\tilde{\beta} - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^{-1}) \quad \text{as } T \rightarrow \infty, \quad (16)$$

and

$$\sqrt{T}(\tilde{\beta} - \beta) - \sqrt{T}(\hat{\beta} - \beta) = o_p(1). \quad (17)$$

Using classical results from Grenander and Szegö (1958) (cf. Remark 2) and Theorem 1 of Robinson and Hidalgo (1997), the asymptotic normality of $\tilde{\beta}$ in (16) can be deduced. It remains to verify that the asymptotic equivalence in (17) holds. This will follow if we show that

$$(Z'\Omega^{-1}Z)^{-1} Z'\Omega^{-1}u - (Z'\hat{\Omega}^{-1}Z)^{-1} Z'\hat{\Omega}^{-1}u = o_p(1), \quad (18)$$

where $Z = (1/\sqrt{T})X$. To this end, express the left-hand side of (18) as

$$\begin{aligned} & \left\{ (Z'\Omega^{-1}Z)^{-1} - (Z'\hat{\Omega}^{-1}Z)^{-1} \right\} Z'\Omega^{-1}u - (Z'\hat{\Omega}^{-1}Z)^{-1} (Z'\hat{\Omega}^{-1}u - Z'\Omega^{-1}u) \\ &= J_1 + J_2. \end{aligned}$$

Writing z_j for the j -th column of Z , it is not difficult to see that the (j, s) -th element of $Z'\Omega^{-1}Z - Z'\hat{\Omega}^{-1}Z$, given by $z'_j \Omega^{-1}(\Omega - \hat{\Omega})\hat{\Omega}^{-1}z_s$, is bounded from above by

$$C \rho_{\max}(\{\Omega - \hat{\Omega}\}^2) \rho_{\max}(\Omega^{-2}) \rho_{\max}(\hat{\Omega}^{-2}), \quad (19)$$

with probability approaching one as $T \rightarrow \infty$, where $\rho_{\max}(\cdot)$ denotes the maximum eigenvalue of its argument. By the assumptions of the theorem, $\rho_{\max}(\Omega^{-2})$ is bounded. Furthermore, in view of Theorem 1, $\rho_{\max}(\{\Omega - \hat{\Omega}\}^2) = o_p(1)$ and $\rho_{\max}(\hat{\Omega}^{-2})$ is bounded. In addition, the variance of $Z'\Omega^{-1}u$ is bounded by a central limit theorem for long-range dependent processes (see Robinson and Hidalgo (1997, Proposition 2)). Combining these results gives $J_1 = o_p(1)$. A similar argument leads to the conclusion that $z'_j (\hat{\Omega} - \Omega)uu'(\hat{\Omega} - \Omega)z_j$ is bounded from above by

$$C \rho_{\max}(\{\hat{\Omega} - \Omega\}^2) \rho_{\max}(\Omega^{-2}),$$

with probability approaching one as $T \rightarrow \infty$, from which it follows that $J_2 = o_p(1)$. This completes the proof of (18) and of the theorem. ■

References

- AMEMIYA, T. (1973): “Generalized Least Squares with an Estimated Autocovariance Matrix,” *Econometrica*, 41, 723–732.
- AN, H.-Z., Z.-G. CHEN, AND E. J. HANNAN (1982): “Autocorrelation, Autoregression and Autoregressive Approximation,” *Annals of Statistics*, 10, 926–936 (Correction, 11, 1018).
- 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.
- 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.
- DEN HAAN, W. J., AND A. LEVIN (2000): “Robust Covariance Matrix Estimation with Data Dependent VAR Prewitening Order,” *NBER Technical Working Paper No. 255*.
- 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., AND L. KAVALIERIS (1986): “Regression, Autoregression Models,” *Journal of Time Series Analysis*, 7, 27–49.
- 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, 837–902.
- HOSKING, J. R. M. (1981): “Fractional Differencing,” *Biometrika*, 65, 165–176.
- KAPETANIOS, G. (2006): “Sieve-Type GLS Tests of Regression Hypotheses with Autocorrelation-Robust Size and Optimal Power Properties,” Department of Economics, Queen Mary, University of London.
- KIEFER, N. M., T. J. VOGELSANG, AND H. BUNZEL (2000): “Simple Robust Testing of Regression Hypotheses,” *Econometrica*, 68, 695–714.
- MANDELBROT, B. B., AND J. W. VAN NESS (1968): “Fractional Brownian Motions, Fractional Noises and Applications,” *SIAM Review*, 10, 422–437.
- NIELSEN, M. Ø. (2005): “Semiparametric Estimation in Time-Series Regression with Long-Range Dependence,” *Journal of Time Series Analysis*, 26, 279–304.
- POSKITT, D. S. (2005): “Autoregressive Approximation in Nonstandard Situations: The Non-Invertible and Fractionally Integrated Cases,” Working Paper 16/05, Department of Econometrics and Business Statistics, Monash University.
- ROBINSON, P. M. (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.
- 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 under long-range dependence in u_t

d_u	ϕ/T	GLS-TD					GLS-FD				
		16	32	64	128	256	512	16	32	64	128
$\beta_1^0 = 0$											
0.1	0	0.157	0.096	0.072	0.063	0.057	0.053	0.377	0.159	0.100	0.074
0.1	0.5	0.159	0.098	0.074	0.063	0.056	0.054	0.382	0.164	0.100	0.072
0.1	0.9	0.163	0.098	0.072	0.062	0.059	0.056	0.391	0.159	0.097	0.069
0.1	0.98	0.167	0.099	0.071	0.062	0.059	0.057	0.403	0.167	0.096	0.069
0.2	0	0.164	0.093	0.066	0.062	0.058	0.056	0.409	0.166	0.094	0.069
0.2	0.5	0.168	0.092	0.068	0.055	0.055	0.050	0.422	0.170	0.101	0.070
0.2	0.9	0.174	0.093	0.064	0.055	0.051	0.050	0.435	0.175	0.102	0.069
0.2	0.98	0.168	0.088	0.063	0.057	0.050	0.051	0.446	0.181	0.098	0.068
0.3	0	0.165	0.087	0.062	0.053	0.052	0.050	0.453	0.184	0.099	0.065
0.3	0.5	0.170	0.085	0.057	0.053	0.051	0.048	0.455	0.180	0.099	0.069
0.3	0.9	0.164	0.078	0.055	0.049	0.049	0.050	0.461	0.187	0.092	0.060
0.3	0.98	0.164	0.077	0.053	0.047	0.048	0.050	0.468	0.185	0.089	0.057
0.4	0	0.164	0.077	0.047	0.045	0.048	0.047	0.470	0.180	0.081	0.051
0.4	0.5	0.159	0.071	0.048	0.044	0.043	0.046	0.466	0.174	0.076	0.045
0.4	0.9	0.161	0.071	0.048	0.040	0.039	0.043	0.469	0.180	0.080	0.044
0.4	0.98	0.161	0.077	0.052	0.043	0.045	0.046	0.452	0.185	0.090	0.052
0.49	0	0.159	0.073	0.047	0.043	0.039	0.045	0.455	0.179	0.083	0.047
0.49	0.5	0.159	0.071	0.047	0.036	0.037	0.037	0.447	0.183	0.080	0.044
0.49	0.9	0.150	0.070	0.040	0.032	0.030	0.029	0.445	0.187	0.084	0.044
0.49	0.98	0.152	0.063	0.036	0.030	0.024	0.023	0.454	0.188	0.083	0.049
$\beta_1^0 = 0.2$											
0.1	0	0.088	0.187	0.376	0.652	0.924	0.997	0.077	0.176	0.352	0.638
0.1	0.5	0.086	0.173	0.344	0.613	0.898	0.994	0.080	0.162	0.325	0.612
0.1	0.9	0.094	0.177	0.337	0.593	0.877	0.994	0.067	0.167	0.332	0.599
0.1	0.98	0.086	0.166	0.333	0.606	0.873	0.993	0.065	0.159	0.317	0.607
0.2	0	0.094	0.180	0.339	0.593	0.887	0.994	0.068	0.156	0.303	0.592
0.2	0.5	0.083	0.159	0.327	0.613	0.886	0.994	0.061	0.146	0.308	0.580
0.2	0.9	0.083	0.162	0.351	0.615	0.891	0.995	0.064	0.139	0.308	0.580
0.2	0.98	0.077	0.166	0.356	0.637	0.912	0.997	0.069	0.132	0.306	0.601
0.3	0	0.089	0.174	0.368	0.671	0.923	0.998	0.063	0.132	0.306	0.620
0.3	0.5	0.078	0.188	0.397	0.690	0.932	0.998	0.058	0.137	0.305	0.607
0.3	0.9	0.083	0.199	0.390	0.699	0.939	0.999	0.070	0.132	0.289	0.608
0.3	0.98	0.083	0.198	0.423	0.722	0.955	1.000	0.060	0.124	0.294	0.597
0.4	0	0.072	0.185	0.443	0.756	0.968	0.999	0.055	0.113	0.264	0.572
0.4	0.5	0.084	0.204	0.446	0.764	0.974	1.000	0.053	0.105	0.243	0.507
0.4	0.9	0.081	0.196	0.450	0.786	0.974	1.000	0.057	0.095	0.186	0.405
0.4	0.98	0.084	0.192	0.430	0.747	0.963	1.000	0.063	0.112	0.238	0.504
0.49	0	0.079	0.192	0.442	0.750	0.966	1.000	0.058	0.087	0.209	0.430
0.49	0.5	0.079	0.194	0.436	0.764	0.963	0.999	0.059	0.080	0.154	0.336
0.49	0.9	0.083	0.209	0.427	0.750	0.954	0.997	0.053	0.066	0.122	0.223
0.49	0.98	0.082	0.209	0.425	0.706	0.937	0.993	0.047	0.071	0.094	0.154

Continued on next page

Continued from last page													
d_u	ϕ/T	GLS-TD						GLS-FD					
		16	32	64	128	256	512	16	32	64	128	256	512
$\beta_1^0 = 0.5$													
0.1	0	0.308	0.754	0.972	1.000	1.000	1.000	0.218	0.714	0.964	1.000	1.000	1.000
0.1	0.5	0.307	0.704	0.965	1.000	1.000	1.000	0.211	0.668	0.959	1.000	1.000	1.000
0.1	0.9	0.287	0.698	0.955	0.999	1.000	1.000	0.186	0.658	0.953	0.999	1.000	1.000
0.1	0.98	0.277	0.676	0.952	0.999	1.000	1.000	0.172	0.628	0.943	1.000	1.000	1.000
0.2	0	0.298	0.698	0.955	0.999	1.000	1.000	0.176	0.626	0.942	1.000	1.000	1.000
0.2	0.5	0.262	0.684	0.952	1.000	1.000	1.000	0.160	0.607	0.934	1.000	1.000	1.000
0.2	0.9	0.251	0.681	0.963	1.000	1.000	1.000	0.140	0.573	0.936	1.000	1.000	1.000
0.2	0.98	0.268	0.710	0.974	1.000	1.000	1.000	0.142	0.578	0.936	0.999	1.000	1.000
0.3	0	0.268	0.717	0.980	1.000	1.000	1.000	0.131	0.575	0.942	1.000	1.000	1.000
0.3	0.5	0.280	0.738	0.980	1.000	1.000	1.000	0.128	0.554	0.929	0.999	1.000	1.000
0.3	0.9	0.283	0.748	0.982	1.000	1.000	1.000	0.127	0.501	0.905	0.999	1.000	1.000
0.3	0.98	0.284	0.768	0.987	1.000	1.000	1.000	0.100	0.466	0.882	0.997	1.000	1.000
0.4	0	0.289	0.771	0.987	1.000	1.000	1.000	0.095	0.431	0.826	0.989	1.000	1.000
0.4	0.5	0.305	0.778	0.981	1.000	1.000	1.000	0.081	0.364	0.752	0.970	1.000	1.000
0.4	0.9	0.294	0.768	0.981	1.000	1.000	1.000	0.081	0.304	0.633	0.916	0.995	1.000
0.4	0.98	0.294	0.769	0.986	1.000	1.000	1.000	0.100	0.409	0.767	0.974	0.999	1.000
0.49	0	0.289	0.776	0.982	0.999	1.000	1.000	0.081	0.328	0.667	0.926	0.996	1.000
0.49	0.5	0.278	0.765	0.979	1.000	1.000	1.000	0.081	0.251	0.524	0.805	0.968	0.999
0.49	0.9	0.295	0.758	0.973	1.000	1.000	1.000	0.069	0.174	0.371	0.621	0.884	0.987
0.49	0.98	0.282	0.728	0.956	0.998	1.000	1.000	0.067	0.156	0.270	0.455	0.694	0.919

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Copyright © 2007 George Kapetanios and Zacharias Psaradakis
All rights reserved

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm