

**HIGHER-ORDER APPROXIMATIONS FOR
FREQUENCY DOMAIN TIME SERIES REGRESSION**

BY

**ZHIJIE XIAO
AND
PETER C.B. PHILLIPS**

COWLES FOUNDATION PAPER NO. 968



**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY**

Box 208281
New Haven, Connecticut 06520-8281
1999



ELSEVIER

Journal of Econometrics 86 (1998) 297–336

JOURNAL OF
Econometrics

Higher-order approximations for frequency domain time series regression

Zhijie Xiao^{a,*}, Peter C.B. Phillips^b

^a Department of Economics, 185 Commerce west, 1206 South Sixth Street, Champaign IL 61820, USA

^b Cowles Foundation for Research in Economics, Yale University, New Haven, CT06520-8281, USA

Received 1 January 1997; received in revised form 1 September 1997; accepted 22 September 1997

Abstract

Second-order expansions and mean squared error approximations are given for efficient frequency domain regression estimators. While bandwidth choices do not figure in first order asymptotics for these estimators, they do influence second-order terms and it is shown how suitable choices will enhance second-order efficiency. Data-based bandwidth selection rules are given for practical implementation of these procedures. Two commonly used and asymptotically equivalent spectral regression estimators are studied and shown to differ in their second-order asymptotic behavior. Some Monte Carlo evidence is reported. © 1998 Elsevier Science S.A. All rights reserved.

JEL classification: C14; C22

Keywords: Data-based bandwidth selection; Higher-order approximation; Moment expansion; Second-order efficient estimation; Semiparametric estimation; Spectral regression

1. Introduction

We shall consider the problem of efficiently estimating the coefficients of the following time series regression

$$y_t = \beta' x_t + u_t, \quad t = 1, \dots, T \quad (1.1)$$

* Corresponding author. E-mail: zxiao@uiuc.edu

where it is assumed that u_t and x_s are stationary processes independent for all t , s , and have absolutely continuous spectral functions. When the autocorrelation structure of the unobservable disturbances is not parameterized, it is not possible to estimate the full covariance matrix of (u_1, \dots, u_T) and conventional time domain generalized least square regression is infeasible, although parametric approximations are possible (cf. Amemiya, 1973). Since the discrete Fourier transform (DFT) of Eq. (1.1) has residuals that are asymptotically independent, efficient methods of estimating β by spectral methods are possible and have been used in econometric applications (see, inter alia, Engle, 1974; Robinson, 1991; Corbae et al., 1994). These methods were introduced by Hannan (1963a,b) who showed that a frequency domain GLS estimator achieves asymptotically the Gauss–Markov efficiency bound under general smoothness conditions on the residual spectral density. This technique is semiparametric since it relies upon a nonparametric treatment of the regression errors. It has the advantage that it is not necessary to be explicit about the generating mechanism for the errors other than to assume stationarity.

The nonparametric spectral density estimates that are used in the efficient estimation of β entail a choice of bandwidth M . Unsuitable bandwidth selection can produce poor estimates. However, as long as the spectral density estimates are consistent, all such frequency domain GLS estimators are (first order) asymptotically equivalent. The asymptotic approximation can be quite accurate for small samples, but it is also easy to find examples where the first order asymptotic distributions provide poor approximations even when the sample contains hundreds of observations, and estimates can vary considerably with bandwidth choice. For these and other reasons, it is useful to have automated rules for bandwidth selection.

Robinson (1991) discussed frequency domain inference for time series in which the bandwidth selection for the nonparametric spectral density estimate is determined from the data. Robinson's method, like that of other automated density estimates, is based on minimizing the (integrated) mean squared error of spectral density estimates not the mean squared error of the estimator $\hat{\beta}$ itself. For many popular kernels (whose characteristic exponents equal 2), Robinson found that the optimal order of expansion for the bandwidth M is $T^{1/5}$.

We believe that higher-order approximation for the coefficient estimates can provide another way to distinguish among asymptotically equivalent procedures in this semiparametric model. Higher-order expansions have a long history of applications to econometric problems (see, inter alia, Sargan, 1976; Phillips, 1977). An application of second-order approximations to bandwidth selection in partially linear regression models has been made recently by Linton (1995a). This paper uses a similar approach in studying the higher-order properties of estimation procedures in frequency domain time series regression. We derive a stochastic expansion for two commonly used frequency domain semiparametric estimators, and approximations to their (standardized) mean squared errors

(MSE). This facilitates a comparison of different estimators at the second-order level, where differences do occur and depend on the bandwidth choice. A method of bandwidth selection is defined by minimizing the second-order effect in the mean squared error of estimates of β . Comparisons are made at the second-order level between the original Hannan efficient estimator and the commonly used spectral estimator considered by Robinson (1991).

The plan of the paper is as follows. The model and the estimators are described in the next section. Section 3 discusses the approximation for the nonparametric spectral density estimators. Formal expansions for the GLS estimator and its mean squared error are contained in Section 4, and distributional approximations are developed in Section 5. In Section 6, we study a second estimator and compare it with the estimator in Section 2. Extensions to the multivariate case are given in Section 7. Section 8 considers a small Monte Carlo study of finite sample behavior. Section 9 concludes. The detailed analysis of all these expansions and proofs of theorems are given in Appendices A–C.

2. The model

Consider the scalar version¹ of regression equation (1.1)

$$y_t = \beta x_t + u_t, \quad t = 1, \dots, T, \quad (2.1)$$

where $\{x_t\}$ and $\{u_t\}$ are stationary time series with continuous spectral density functions and $\{u_t\}_1^T$ has mean zero and covariance matrix V . Our analysis may easily be extended to multiple regressions in which case x_t and β are vectors and this extension is briefly outlined in Section 7. Distributional assumptions on u_t are generally not required, but are useful in the development of distributional approximations, as in Section 5, in which case we will employ a normality assumption and then (2.1) can be written in regression format as

$$y = X\beta + u \quad \text{where } u \sim N(0, V). \quad (2.2)$$

We assume that T is even for simplicity of exposition. Premultiplying Eq. (2.2) by the matrix $U_T = [e^{2\pi jki/T}/\sqrt{2\pi T}]$, we get the discrete Fourier transform of Eq. (2.1),

$$w_y(\lambda_t) = \beta w_x(\lambda_t) + w_u(\lambda_t), \quad \lambda_t = 2\pi t/T, \quad t = -T/2 + 1, \dots, T/2. \quad (2.3)$$

The residuals in regression (2.3), $w_u(\lambda_t)$, are approximately independent but heteroskedastic for large T . The matrix representation of (2.3) can be written

¹ An intercept term in the regression can be accommodated simply by omitting the zero frequency in the estimation.

as follows:

$$W_y = W_x \beta + W_u \quad (2.4)$$

where W_u has a complex normal distribution $N(0, \Sigma)$, with $\Sigma = U_T V U_T^*$. Here, the affix * indicates transposition combined with complex conjugation. However, since * will be used for some other notation in later sections, we hereafter use simply ', with the understanding that it means complex conjugate transpose whenever the quantities are complex, as, for example, in $\Sigma = U_T V U_T'$, $I_{uu}(\lambda) = w_u(\lambda) w_u(\lambda)'$, and $W_x' \Sigma^{-1} W_x$.

The GLS estimator of β in the frequency domain is given by $\hat{\beta} = [W_x' \Sigma^{-1} W_x]^{-1} W_x' \Sigma^{-1} W_y$. Since the covariance matrix Σ is unknown, the estimator $\hat{\beta}$ is infeasible, motivating feasible estimators of the form

$$\hat{\beta} = [W_x' \hat{\Sigma}^{-1} W_x]^{-1} W_x' \hat{\Sigma}^{-1} W_y \quad (2.5)$$

The properties of $\hat{\beta}$ depend, of course, on the choice of estimate for Σ . We shall first consider the following estimate of Σ :

$$\hat{\Sigma} = \text{diag}[\hat{f}_{uu}(\lambda_{-T/2+1}), \dots, \hat{f}_{uu}(\lambda_{T/2})], \quad (2.6)$$

where $\hat{f}_{uu}(\lambda_t)$, $t = -T/2 + 1, \dots, T/2$, are nonparametric spectral density estimators. In this paper, we use the 'leave-one-out' type estimator, which has been widely used in the existing literature (e.g., Robinson, 1987; Linton, 1995a,b), for estimating $f_{uu}(\lambda_t)$, viz.,

$$\hat{f}_{uu}(\lambda_t) = m^{-1} \sum_{\lambda_s \in B_t, s \neq t} K(\lambda_s - \lambda_t) \hat{I}_{uu}(\lambda_s) = \sum_{s \neq t} \omega_{ts} \hat{I}_{uu}(\lambda_s) \quad (2.7)$$

where

$$B_t = \left\{ \omega: \lambda_t - \frac{\pi}{2M} < \omega \leq \lambda_t + \frac{\pi}{2M} \right\}$$

is a frequency band of width π/M centered on $\lambda_t = 2\pi t/T$. Let $m = [T/2M]$, where $[\cdot]$ signifies integer part. Then each band B_t contains m fundamental frequencies λ_s . In (2.7), $K(\cdot)$ is the spectral window satisfying the properties that it is a real, even function with $(1/m) \sum_{\lambda_s \in B(\omega)} K(\lambda_s - \omega) = 1$. We denote the corresponding lag window as $k(h/M) = (1/m) \sum_{\lambda_s \in B(\omega)} K(\lambda_s - \omega) e^{-ih(\lambda_s - \omega)}$. Candidate kernel functions can be found in standard texts (e.g., Hannan, 1970; Brillinger, 1980; Priestley, 1981). $\hat{I}_{uu}(\lambda_s)$ is calculated from $\hat{w}_u(\lambda_s) = w_y(\lambda_s) - \hat{\beta}_{OLS} w_x(\lambda_s)$. Consequently, the GLS estimator for β has the following form:

$$\hat{\beta} = \left[\sum_{t=-T/2+1}^{T/2} I_{xx}(\lambda_t) \hat{f}_{uu}(\lambda_t)^{-1} \right]^{-1} \left[\sum_{t=-T/2+1}^{T/2} I_{xy}(\lambda_t) \hat{f}_{uu}(\lambda_t)^{-1} \right]. \quad (2.8)$$

Under very general conditions, $\hat{\beta}$ in (2.8) is first-order asymptotically equivalent to the infeasible GLS estimator $\tilde{\beta}$. This first-order asymptotic equivalence holds as long as the spectral density estimators in (2.8) are consistent. No guidance concerning bandwidth selection based on the mean squared error of $\hat{\beta}$ is available since the bandwidth parameter M does not show up in the first-order asymptotics. However, the performance of the estimator $\hat{\beta}$ can depend greatly on this choice.

One of the objectives of this paper is to derive higher-order expansions for the estimator $\hat{\beta}$ to compare alternative implementations of the spectral regression procedure and to define a method of bandwidth selection by minimizing the second-order effect in the mean squared error of $\hat{\beta}$. Denote

$$\begin{aligned} \hat{\Omega}_T &= T^{-1} \sum_t I_{xx}(\lambda_t) \hat{f}_{uu}(\lambda_t)^{-1}, \\ \Omega_T &= T^{-1} \sum_t I_{xx}(\lambda_t) f_{uu}(\lambda_t)^{-1}, \\ W_N &= T^{-1/2} \sum_t I_{xu}(\lambda_t) \hat{f}_{uu}(\lambda_t)^{-1}, \\ W_D &= \hat{\Omega}_T - \Omega_T = T^{-1} \sum_t I_{xx}(\lambda_t) [\hat{f}_{uu}(\lambda_t)^{-1} - f_{uu}(\lambda_t)^{-1}]. \end{aligned}$$

We then have $\sqrt{T}(\hat{\beta} - \beta) = \hat{\Omega}_T^{-1} W_N$. Expanding $\hat{\Omega}_T^{-1}$ about Ω_T to the third term and decomposing W_N into the sum of W_{N0} and W_{N1} , where $W_{N0} = T^{-1/2} \sum_t I_{xu}(\lambda_t) f_{uu}(\lambda_t)^{-1}$ and $W_{N1} = T^{-1/2} \sum_t I_{xu}(\lambda_t) [\hat{f}_{uu}(\lambda_t)^{-1} - f_{uu}(\lambda_t)^{-1}]$, we get

$$\sqrt{T}(\hat{\beta} - \beta) = (\Omega_T^{-1} - \Omega_T^{-2} W_D + \Omega_T^{-3} W_D^2 - R_1)(W_{N0} + W_{N1}), \tag{2.9}$$

where $R_1 = \hat{\Omega}_T^{-1} \Omega_T^{-3} W_D^3$. In the above expansion for $\hat{\beta}$, the leading term, $\Omega_T^{-1} W_{N0}$, is of order $O_p(1)$, and other terms are $o_p(1)$. The key elements, W_D and W_{N1} , are functions of the spectral density estimate $\hat{f}_{uu}(\lambda_t)^{-1}$. If we further expand $\hat{f}_{uu}(\lambda_t)^{-1}$ around $f_{uu}(\lambda_t)^{-1}$ to the third term and substitute the corresponding truncation into the expression of W_D and W_{N1} , we obtain an approximation of these two terms. After rearranging terms, under certain assumptions on the time series x_t and u_t , we get an expansion of the following type:

$$\begin{aligned} \sqrt{T}(\hat{\beta} - \beta) &= \Omega_T^{-1} W_{N0} + \frac{1}{\sqrt{m}} \mathcal{A}_1 + \frac{1}{M^2} \mathcal{B}_1 + \frac{1}{m} \mathcal{A}_2 + \frac{1}{M^4} \mathcal{B}_2 \\ &\quad + \text{higher-order terms} \\ &= \Psi + \text{higher-order terms} \end{aligned} \tag{2.10}$$

where $\mathcal{A}_1, \mathcal{B}_1, \mathcal{A}_2, \mathcal{B}_2$ are functions of the bias and variance terms in the nonparametric spectral density estimates (Appendix A provides explicit formulae for these terms) and where the expression ‘higher-order terms’ indicates terms of $o_p(m^{-1})$ or $o_p(M^{-4})$. A detailed analysis of this expansion is given in Appendix A. Our analysis is formal and further regularity conditions are likely to be needed to justify expansions like (2.10) as valid stochastic asymptotic expansions. The (normalized) mean squared error of the truncated expansion, $E[\Omega_T^{1/2}\Psi]^2$, can then be calculated from the above expansion, giving

$$MSE(\hat{\beta}) \approx 1 + \frac{1}{m}\mathcal{A} + \frac{1}{M^4}\mathcal{B}. \quad (2.11)$$

The second-order effect in the mean squared error, $(1/m)\mathcal{A} + (1/M^4)\mathcal{B}$, is a function of the bandwidth choice, the kernel function and the spectral densities (see Theorem 1, Section 4 for the definition of \mathcal{A} and \mathcal{B}). An optimal bandwidth for the estimation of β can be determined by minimizing the second-order effect on the $MSE(\hat{\beta})$.

It will be convenient for our development to assume at various points in our analysis some of the following properties for the time series $\{x_t\}$ and $\{u_t\}$.

Assumption A.1. The time series u_t and x_t are independent and stationary with $E(u_t) = 0$, $\text{cov}(x_t, x_{t+h}) = \gamma_x(h)$, $\text{cov}(u_t, u_{t+h}) = \gamma_u(h)$, and

$$\sum_h |h|^q |\gamma_x(h)| < \infty, \quad \sum_h |h|^q |\gamma_u(h)| < \infty$$

where q is the characteristic exponent of the kernel function defined as $\lim_{x \rightarrow 0} \{1 - k(x)\}/|x|^q = k_q < \infty$.

Assumption A.2. The spectral density of u_t , $f_{uu}(\cdot)$, is bounded away from the origin and $\sup_T T^{-4/5} \|\Sigma - \Sigma_0\| = o(1)$, where $\Sigma_0 = \text{diag}[f_{uu}(\lambda_{-T/2+1}), \dots, f_{uu}(\lambda_{T/2})]$, and $\|\cdot\|$ is the matrix norm defined by $\|B\| = \sup \{\|Bx\|: \|x\| < 1, \|x\| = (x'x)^{1/2}\}$.

Assumption A.3. u_t is normally distributed.

Assumptions A.1 and A.2 are sufficient for the moment expansion, and A.3 is not needed, but is used for the distributional approximation in Section 5. The conditions in Assumption A.1 not only imply that the spectral densities are continuous and bounded, but also imply the uniform boundedness of $f_q(\lambda)$, $\sup_\lambda |f_q(\lambda)| < \infty$, where $f_q(\lambda) = (1/2\pi) \sum_{h=-\infty}^{\infty} |h|^q \gamma_u(h) e^{-ih\lambda}$. For many popular kernels, where $q = 2$, the conditions in Assumption A.1 imply that the second derivatives are bounded.

3. Expansions for the spectral density estimates

In expansion (2.9), the components W_D and W_{N1} are functions of the spectral density estimates. The first step in extracting the expansion is to develop approximations for $\hat{f}_{uu}(\lambda_t)$ and $\hat{f}_{uu}(\lambda_t)^{-1}$, which we do in this section. First, we decompose the error term in the nonparametric spectral density estimator $\hat{f}_{uu}(\lambda_t)$ into three parts: B_t , the bias term due to smoothing; V_t , the variance term comes from the periodogram; and P_t , an error term comes from preliminary estimation of $w_u(\lambda_t)$. The last term, P_t , is usually of smaller order of magnitude than the first two terms, so that it can be dropped. Specifically, we have

$$\hat{f}_{uu}(\lambda_t) = f_{uu}(\lambda_t) + B_t + V_t + P_t \tag{3.1}$$

where

$$B_t = f_{uu}^*(\lambda_t) - f_{uu}(\lambda_t) = \sum_{s \neq t} \omega_{ts} [f_{uu}(\lambda_s) - f_{uu}(\lambda_t)],$$

$$V_t = \bar{f}_{uu}(\lambda_t) - f_{uu}^*(\lambda_t) = \sum_{s \neq t} \omega_{ts} [I_{uu}(\lambda_s) - f_{uu}(\lambda_s)],$$

$$P_t = \hat{f}_{uu}(\lambda_t) - \bar{f}_{uu}(\lambda_t) = \sum_{s \neq t} \omega_{ts} [\hat{I}_{uu}(\lambda_s) - I_{uu}(\lambda_s)],$$

$$\bar{f}_{uu}(\lambda_t) = m^{-1} \sum_{\lambda_s \in B_t, s \neq t} K(\lambda_s - \lambda_t) I_{uu}(\lambda_s) = \sum_{s \neq t} \omega_{ts} I_{uu}(\lambda_s),$$

$$f_{uu}^*(\lambda_t) = m^{-1} \sum_{\lambda_s \in B_t, s \neq t} K(\lambda_s - \lambda_t) f_{uu}(\lambda_s) = \sum_{s \neq t} \omega_{ts} f_{uu}(\lambda_s).$$

The order of magnitude for each of these terms in our decomposition is given by the following lemma.

Lemma 1.

$$B_t \sim -M^{-q} k_q f_q(\lambda_t) = O(M^{-q}),$$

$$V_t = O_p(m^{-1/2}),$$

$$\sqrt{m} V_t \xrightarrow{d} N\left(0, \frac{1}{2} \int_{-\infty}^{\infty} k(x)^2 dx f_{uu}^2(\lambda_t)\right), \text{ for } |\lambda_t| \geq \pi/2M,$$

$$P_t = o_p(m^{-1}).$$

where \sim denotes asymptotic equivalence.

We are principally interested in those terms that are of order $O_p(m^{-1})$ or $O_p(M^{-2q})$, and will later refer to the $o_p(m^{-1})$ and $o_p(M^{-2q})$ terms as ‘higher-order’ terms.

Next consider the expansion for $\hat{f}_{uu}(\lambda_t)^{-1}$. In a similar way, we decompose the error term for $\hat{f}_{uu}(\lambda_t)^{-1}$ into three parts, which are functions of B_t , V_t , and P_t , respectively. The order of magnitude of these terms can then be determined from Lemma 1. Write $\hat{f}_{uu}(\lambda_t)^{-1}$ as follows:

$$\begin{aligned} \hat{f}_{uu}(\lambda_t)^{-1} &= f_{uu}(\lambda_t)^{-1} + [f_{uu}^*(\lambda_t)^{-1} - f_{uu}(\lambda_t)^{-1}] \\ &\quad + [\bar{f}_{uu}(\lambda_t)^{-1} - f_{uu}^*(\lambda_t)^{-1}] + [\hat{f}_{uu}(\lambda_t)^{-1} - \bar{f}_{uu}(\lambda_t)^{-1}]. \end{aligned} \tag{3.2}$$

Expand $\bar{f}_{uu}(\lambda_t)^{-1}$ about $f_{uu}^*(\lambda_t)^{-1}$, and $f_{uu}^*(\lambda_t)^{-1}$ about $f_{uu}(\lambda_t)^{-1}$, giving

$$\begin{aligned} \bar{f}_{uu}(\lambda_t)^{-1} &= f_{uu}^*(\lambda_t)^{-1} - f_{uu}^*(\lambda_t)^{-2}V_t + f_{uu}^*(\lambda_t)^{-3}V_t^2 + R_{3a}, \\ f_{uu}^*(\lambda_t)^{-1} &= f_{uu}(\lambda_t)^{-1} - f_{uu}(\lambda_t)^{-2}B_t + f_{uu}(\lambda_t)^{-3}B_t^2 + R_{3b}. \end{aligned}$$

Here, R_{3a} , R_{3b} and the errors that come from preliminary estimation can be dropped, according to the following lemma.

Lemma 2. $\hat{f}_{uu}(\lambda_t)^{-1} - \bar{f}_{uu}(\lambda_t)^{-1} = o_p(m^{-1})$, $R_{3a} = o_p(m^{-1})$, $R_{3b} = o_p(M^{-2q})$.

We thus obtain the following expansion for $\hat{f}_{uu}(\lambda_t)^{-1}$,

$$\begin{aligned} \hat{f}_{uu}(\lambda_t)^{-1} &= f_{uu}(\lambda_t)^{-1} - f_{uu}^*(\lambda_t)^{-2}V_t + f_{uu}^*(\lambda_t)^{-3}V_t^2 - f_{uu}(\lambda_t)^{-2}B_t + f_{uu}(\lambda_t)^{-3}B_t^2 \\ &\quad + \text{higher-order terms.} \end{aligned} \tag{3.3}$$

4. The expansion of $\hat{\beta}$

With the preliminary results in Section 3, we can now calculate the stochastic expansion for the estimator $\hat{\beta}$. Let W_D^* and W_{N1}^* be the corresponding approximations of W_D and W_{N1} with $\hat{f}_{uu}(\lambda_t)^{-1}$ replaced by the truncated expansion in Eq. (3.3). Substituting W_D and W_{N1} with W_D^* and W_{N1}^* in Eq. (2.9) then gives the expansion (2.10) for the estimator $\hat{\beta}$. The remainder term in Eq. (2.10) includes R_1 as well as the remainders from replacing W_D and W_{N1} by W_D^* and W_{N1}^* . Let $MSE(\hat{\beta})$ be the standardized mean squared error of $\hat{\beta}$ (multiplied by T)

$$MSE(\hat{\beta}) = E[\Omega^{1/2}\Psi]^2, \tag{4.1}$$

where Ω is the limit of Ω_T defined in Eq. (4.3) below. The asymptotic expansion for $MSE(\hat{\beta})$ is given by the following theorem.

Theorem 1. Under Assumptions A.1 and A.2,

$$\begin{aligned} MSE(\hat{\beta}) &= 1 + \frac{1}{2m} \left[\int_{-\infty}^{\infty} k(x)^2 dx \right] + \frac{k_q^2}{M^{2q}} [\Omega^{-1}\psi_1(f_q) - \Omega^{-2}\psi_2(f_q)^2] \\ &\quad + \text{higher-order terms,} \end{aligned} \tag{4.2}$$

where

$$\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{xx}(\omega) f_{uu}(\omega)^{-1} d\omega, \quad (4.3)$$

$$\mathcal{V}_1(f_q) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{xx}(\omega) f_{uu}(\omega)^{-3} f_q(\omega)^2 d\omega, \quad (4.4)$$

$$\mathcal{V}_2(f_q) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{xx}(\omega) f_{uu}(\omega)^{-2} f_q(\omega) d\omega. \quad (4.5)$$

Remark 1. The term

$$\mathcal{K} = \frac{1}{2m} \left[\int_{-\infty}^{\infty} k(x)^2 dx \right] + \frac{k_q^2}{M^{2q}} [\Omega^{-1} \mathcal{V}_1(f_q) - \Omega^{-2} \mathcal{V}_2(f_q)^2]$$

measures the second-order effect on the asymptotic mean squared error of $\hat{\beta}$. To minimize the order of magnitude of the second-order effect, we need to balance the impact of the two terms in \mathcal{K} and this requires m and M^{2q} to be of the same order. If $M = O(T^\mu)$, ($0 < \mu < 1/2$), then $m = O(T^{1-\mu})$, and thus $1 - \mu = 2q\mu$, i.e., $\mu = (1 + 2q)^{-1}$. To simplify notation, set $2\tau = 1 - \mu = 2q\mu = 2q/(1 + 2q)$, and then the second-order effect in the mean squared error is of order $T^{-2\tau}$. When $q = 2$, we get $\mu = 1/5$, and $\tau = 2/5$. Since nonparametric techniques are used in estimating the spectral density, the second-order effect is larger than the $O_p(T^{-1})$ effect in parametric models. However, as q increases, $2\tau = 2q/(2q + 1)$ approaches 1. Thus the second-order term in the mean squared error gets close as q increases to the $O(T^{-1})$ effect that applies in the parametric case. As a result, we should be able to accelerate the convergence rate of the spectral estimates by the use of higher-order kernels so that the second-order correction in the mean squared error approaches the $O(T^{-1})$ rate in parametric models.

Remark 2. The second-order term \mathcal{K} is positive since

$$\begin{aligned} & \int_{-\pi}^{\pi} f_{xx}(\omega) f_{uu}(\omega)^{-1} d\omega \int_{-\pi}^{\pi} f_{xx}(\omega) f_{uu}(\omega)^{-3} f_q(\omega)^2 d\omega \\ & \geq \left[\int_{-\pi}^{\pi} f_{xx}(\omega) f_{uu}(\omega)^{-2} f_q(\omega) d\omega \right]^2 \end{aligned}$$

by the Cauchy–Schwarz inequality. As a result, the (second-order) variance of $\hat{\beta}$ is generally greater than that of $\tilde{\beta}$, indicating that the feasible GLS estimator has a larger second-order asymptotic variance than the GLS estimator $\tilde{\beta}$.

Remark 3. An optimal bandwidth can be found by minimizing the second-order effect in the mean squared error, i.e. we can choose M such that it minimizes

$$\frac{1}{2m} \left[\int_{-\infty}^{\infty} k(x)^2 dx \right] + \frac{k_q^2}{M^{2q}} [\Omega^{-1} \psi_1 - \Omega^{-2} \psi_2^2].$$

The optimal bandwidth is then given by

$$M = \left[\frac{2qk_q^2 [\Omega^{-1} \psi_1 - \Omega^{-2} \psi_2^2]}{\int_{-\infty}^{\infty} k(x)^2 dx} \right]^{1/(2q+1)} T^{1/(2q+1)} \\ = \delta(k, f) T^{1/(2q+1)}, \text{ say} \quad (4.6)$$

where the coefficient $\delta(k, f)$ is a function of the kernel function k and the spectral density f . To make the above bandwidth selection criterion feasible, the plug-in method can be used to obtain an estimator of δ . That is, we specify a parametric model for the error structure and estimators of these parameters are used to obtain preliminary estimates of the spectral density functions and these are then plugged into Eq. (4.6). For example, suppose x_t is generated by MA(1) process

$$x_t = \eta_t + \theta \eta_{t-1} \quad (4.7)$$

and u_t is generated by AR(1) process

$$u_t = \alpha u_{t-1} + \varepsilon_t, \quad (4.8)$$

where η and ε are both iid $N(0,1)$ variates and are independent of each other. If the Daniell window is used so that $K(\lambda_s - \omega_j) = 1$ for $\lambda_s \in B_j$ and $q = 2$, the optimal bandwidth is calculated to be

$$M = \left[\frac{2\delta_1(\alpha, \theta)}{1 + \alpha^2 \theta^2 + (\alpha - \theta)^2} - \frac{2\delta_2(\alpha, \theta)^2}{[1 + \alpha^2 \theta^2 + (\alpha - \theta)^2]^2} \right]^{1/5} T^{1/5}, \quad (4.9)$$

where δ_1 and δ_2 are defined by the formulae

$$\delta_1(\alpha, \theta) = \frac{\alpha^2 \pi^3}{18} \int_{-\pi}^{\pi} \frac{\sin^4 \omega (1 + \theta^2 + 2\theta \cos \omega)}{(1 + \alpha^2 - 2\alpha \cos \omega)^3} d\omega \\ + \frac{\alpha^2 \pi^3}{288} \int_{-\pi}^{\pi} \frac{\cos^2 \omega (1 + \theta^2 + 2\theta \cos \omega)}{1 + \alpha^2 - 2\alpha \cos \omega} d\omega \\ - \frac{\alpha^3 \pi^3}{36} \int_{-\pi}^{\pi} \frac{\sin^2 \omega \cos \omega (1 + \theta^2 + 2\theta \cos \omega)}{(1 + \alpha^2 - 2\alpha \cos \omega)^2} d\omega,$$

and

$$\delta_2(\alpha, \theta) = \frac{\alpha^2 \pi}{6} \int_{-\pi}^{\pi} \frac{\sin^2 \omega (1 + \theta^2 + 2\theta \cos \omega)}{1 + \alpha^2 - 2\alpha \cos \omega} d\omega - \frac{\alpha \pi}{24} \int_{-\pi}^{\pi} \cos \omega (1 + \theta^2 + 2\theta \cos \omega) d\omega.$$

Estimates of α and θ may then be plugged into these formulae to give a data-based optimal bandwidth formula (4.9).

Remark 4. Asymptotic expansions for other regression statistics like standard error estimates, t -ratios and Wald statistics can all be developed in a similar way. To conserve space, we do not provide formulae here but details of these expansions are available in Xiao (1997).

5. Distributional approximation

Under Assumption A.3, $W_u \sim N(0, \Sigma)$, and so, conditional on $\{x_t\}$,

$$\sqrt{T}(\tilde{\beta} - \beta) = \Omega_{\tilde{\Sigma}}^{-1} W_{N_0}^* \sim N(0, \Omega_{\tilde{\Sigma}}^{-1}), \tag{5.1}$$

where $\Omega_{\tilde{\Sigma}} = T^{-1} W_x' \Sigma^{-1} W_x$, $W_{N_0}^* = T^{-1/2} W_x' \Sigma^{-1} W_u$. The probability distribution of $\tilde{\beta}$ is then given by

$$\Pr(\Omega_{\tilde{\Sigma}}^{1/2} \sqrt{T}(\hat{\beta} - \beta) \leq r) = \Pr(\Omega_{\tilde{\Sigma}}^{1/2} \sqrt{T}(\tilde{\beta} - \beta) \leq r - \Omega_{\tilde{\Sigma}}^{1/2} \sqrt{T}(\hat{\beta} - \tilde{\beta})).$$

For the estimators considered in this paper, $\hat{\Sigma}$ and $\hat{\beta} - \tilde{\beta}$ do not depend on β as long as x_t and u_t do not. Since, for given Σ , $\tilde{\beta}$ is a complete sufficient statistic for β , both $\hat{\Sigma}$ and $\hat{\beta} - \tilde{\beta}$ are distributed independent of $\tilde{\beta}$. Thus

$$\begin{aligned} \Pr[\Omega_{\tilde{\Sigma}}^{1/2} \sqrt{T}(\tilde{\beta} - \beta) \leq r - \Omega_{\tilde{\Sigma}}^{1/2} \sqrt{T}(\hat{\beta} - \tilde{\beta}) | \{x_t\}] \\ = E[\Phi(r - \Omega_{\tilde{\Sigma}}^{1/2} \sqrt{T}(\hat{\beta} - \tilde{\beta})) | \{x_t\}], \end{aligned}$$

where $\Phi(\cdot)$ is the c.d.f. of a standard normal variate. To develop a distributional expansion for $\hat{\beta}$, we now just need to approximate the expectation of $\Phi[r - \Omega_{\tilde{\Sigma}}^{1/2} \sqrt{T}(\hat{\beta} - \tilde{\beta})]$. Our arguments in this section are all conditional on $\{x_t\}$, and so we drop the conditioning symbolism in what follows to simplify notation.

If $\Omega_{\tilde{\Sigma}}^{1/2} \sqrt{T}(\hat{\beta} - \tilde{\beta}) = Z_T + R_T$, where $T^r Z_T$ possesses bounded moments in the limit and R_T has well behaved tail probabilities in the sense that $\Pr[T^{2r} \log T \|R_T\| > c] = o(T^{-2r})$ for some constant c , then restricting our attention to the region where $|R_T| < c/T^{2r} \log T$ yields an error of order $o(T^{-2r})$ – see

Sargan and Mikhail (1971) and Rothenberg (1984a,b). It follows that

$$\begin{aligned} \Pr(\Omega_{\varepsilon}^{1/2} \sqrt{T}(\hat{\beta} - \beta) \leq r) &= E(\Phi(r - Z_T)) + o_p(T^{-2\tau}) \\ &= \Phi(r) - E(Z_T)\varphi(r) - \frac{1}{2}E(Z_T^2)r\varphi(r) + o_p(T^{-2\tau}) \\ &= \Phi(r) - \frac{1}{2}E(Z_T^2)r\varphi(r) + o_p(T^{-2\tau}) \\ &= \Phi[r(1 - EZ_T^2/2)] + o_p(T^{-2\tau}), \end{aligned} \quad (5.2)$$

where φ is the standard normal p.d.f. It is verified in Appendix B that

$$\Pr[T^{2\tau} \log T \|R_T\| > c] = o(T^{-2\tau}). \quad (5.3)$$

Thus, we get the following theorem.

Theorem 2. Under Assumptions A.1–A.3, the probability distribution of the GLS estimator (2.8) can be expanded in the form

$$\Pr[(1 - EZ_T^2/2)\Omega_{\varepsilon}^{1/2} \sqrt{T}(\hat{\beta} - \beta) \leq r] = \Phi(r) + o(T^{-2\tau}), \quad (5.4)$$

where

$$EZ_T^2 = \frac{1}{2m} \int_{-\infty}^{\infty} k(x)^2 dx + \frac{k_q^2}{M^{2q}} [\Omega^{-1} \psi_1(f_q) - \Omega^{-2} \psi_2(f_q)^2] + o(T^{-2\tau}). \quad (5.5)$$

Remark 5. The theorem shows that $\hat{\beta}$ is asymptotically normal to the second-order with an error of $o(T^{-2\tau})$ and where $\tau = q/(1 + 2q)$. Thus, the effect of feasible frequency domain GLS estimation on second-order asymptotics is simply to scale the limiting normal distribution of the GLS estimator by a factor which depends on the spectral estimates and the bandwidth choice.

6. The expansion for Hannan's estimator

Another version of the frequency domain GLS estimator that has been widely used in the literature is the following one suggested by Hannan (1963a,b):

$$\hat{\beta}_{\text{H0}} = \left[\sum_{j=-M+1}^M \hat{f}_{xx}(\omega_j) \hat{f}_{uu}(\omega_j)^{-1} \right]^{-1} \left[\sum_{j=-M+1}^M \hat{f}_{xy}(\omega_j) \hat{f}_{uu}(\omega_j)^{-1} \right]. \quad (6.1)$$

This estimator of β is based on consistent spectral estimates $\hat{f}_{xx}(\omega_j)$, $\hat{f}_{uu}(\omega_j)$ and $\hat{f}_{xy}(\omega_j)$ such as

$$\begin{aligned} \hat{f}_{uu}(\omega_j) &= m^{-1} \sum_{\lambda_s \in B_j} K(\lambda_s - \omega_j) \hat{I}_{uu}(\lambda_s) = \sum_s \omega_{js} \hat{I}_{uu}(\lambda_s), \\ \hat{f}_{xx}(\omega_j) &= m^{-1} \sum_{\lambda_s \in B_j} K(\lambda_s - \omega_j) I_{xx}(\lambda_s) = \sum_s \omega_{js} I_{xx}(\lambda_s), \\ \hat{f}_{xy}(\omega_j) &= m^{-1} \sum_{\lambda_s \in B_j} K(\lambda_s - \omega_j) I_{xy}(\lambda_s) = \sum_s \omega_{js} I_{xy}(\lambda_s). \end{aligned}$$

Each frequency band $B_j = \{\omega: \omega_j - \pi/2M < \omega \leq \omega_j + \pi/2M\}$ is centered on $\omega_j = \pi j/M$, $j = -M + 1, \dots, M$. In this section, we consider the following ‘leave-one-out’ estimator for $f_{uu}(\omega_j)$

$$\tilde{f}_{uu}(\omega_j) = \frac{1}{2m} \sum_{\lambda_s \in B_j^*} W(\lambda_s - \omega_j) \hat{I}_{uu}(\lambda_s), \tag{6.2}$$

where $B_j^* = B_{j-1} \cup B_{j+1}$, and $W(\cdot)$ works as a spectral window with the property that $(1/2m) \sum_s W(\lambda_s - \omega_j) = 1$. The use of such an estimator is primarily motivated by technical convenience. In particular, it simplifies the calculation of the second-order effect without changing the order of magnitude. A simple example of this type estimator is

$$\tilde{f}_{uu}(\omega_j) = \frac{1}{2} [\hat{f}_{uu}(\omega_{j-1}) + \hat{f}_{uu}(\omega_{j+1})]. \tag{6.3}$$

For simplicity, we will use this estimator in our analysis although the more complicated estimator (6.2) with the spectral window $W(\cdot)$ could be used without changing the results below in a significant way. (In particular, the orders of magnitude of the correction terms stay the same.) Let

$$\hat{\beta}_H = \left[\sum_{j=-M+1}^M \hat{f}_{xx}(\omega_j) \tilde{f}_{uu}(\omega_j)^{-1} \right]^{-1} \left[\sum_{j=-M+1}^M \hat{f}_{xy}(\omega_j) \tilde{f}_{uu}(\omega_j)^{-1} \right]. \tag{6.4}$$

The estimator $\hat{\beta}_H$ is first-order asymptotically equivalent to the estimator $\hat{\beta}$ in Eq. (2.8), studied in the previous section. However, we see here that these two estimators do differ at the second-order level. Denote

$$\Sigma_T = \frac{1}{2M} \sum_{j=-M+1}^M f_{xx}(\omega_j) f_{uu}(\omega_j)^{-1}, \quad \hat{\Sigma}_T = \frac{1}{2M} \sum_{j=-M+1}^M \hat{f}_{xx}(\omega_j) \tilde{f}_{uu}(\omega_j)^{-1},$$

$$X_N = \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) \tilde{f}_{uu}(\omega_j)^{-1}, \quad X_{N0} = \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) f_{uu}(\omega_j)^{-1}.$$

Then

$$\sqrt{T}(\hat{\beta}_H - \beta) = \hat{\Sigma}_T^{-1} X_N = \left(\Sigma_T^{-1} - \frac{\Sigma_T^{-2} X_D}{\sqrt{m}} + \frac{\Sigma_T^{-3} X_D^2}{m} + R_T \right) X_N, \tag{6.5}$$

where

$$X_D = \sqrt{m}(\hat{\Sigma}_T - \Sigma_T),$$

and

$$R_T = -\frac{1}{m^{3/2}} \hat{\Sigma}_T^{-1} \Sigma_T^{-3} X_D^3.$$

Without much loss of generality, we will consider the most common case $q = 2$ in the following analysis. Notice that under Assumption A.1, the second derivatives of the spectral density functions are bounded. Proceeding in the same way as for the estimator $\hat{\beta}$, we obtain the following stochastic expansion for $\hat{\beta}_H$:

$$\begin{aligned} \sqrt{T}(\hat{\beta}_H - \beta) &= \Sigma_T^{-1} X_{N0} + \frac{1}{\sqrt{m}} A_1 + \frac{1}{m} A_2 + \frac{1}{M^2} B_1 + \frac{1}{M^4} B_2 \\ &+ \text{higher-order terms.} \\ &= \Psi_H + \text{higher-order terms.} \end{aligned} \quad (6.6)$$

The terms, A_1, A_2, B_1, B_2 , in Eq. (6.6) are functions of $B_{xj}, V_{xj}, B_{uj}, V_{uj}$, which are the bias and variance terms in the nonparametric estimators of $f_{xx}(\omega_j)$ and $f_{uu}(\omega_j)$. A detailed analysis of the expansion is given in Appendix C of the paper, which provides explicit formulae for the expansion of $\hat{\beta}_H$ and its asymptotic mean square error. Here we will give only the main results and discuss their implications.

The mean squared error of $\sqrt{T}(\hat{\beta}_H - \beta)$ can be approximated by the second moment of the truncated expansion $\Sigma_T^{-1} X_{N0} + (1/\sqrt{m})A_1 + (1/m)A_2 + (1/M^2)B_1 + (1/M^4)B_2$ from Eq. (6.6). It turns out that the calculation of the mean squared error expansion of $\hat{\beta}_H$ involves the approximation of $E\hat{f}_{xu}(\omega_j)^2$. Conditional on $\{x_i\}$, we have

$$\sqrt{m}\hat{f}_{xu}(\omega_j) \sim N(0, f_{xx}(\omega_j)f_{uu}(\omega_j)) + \text{higher-order terms.}$$

We can show that the conditional variance $E[\sqrt{m}\hat{f}_{xu}(\omega_j)]^2$ can be approximated by $f_{xx}(\omega_j)f_{uu}(\omega_j) + O_p(M^{-2})$, where the $O_p(M^{-2})$ term is the cumulated bias term due to approximating $f_{uu}(\lambda_s)$ by $f_{uu}(\omega_j)$ in each frequency band $B_j = \{\omega: \omega_j - \pi/2M < \omega \leq \omega_j + \pi/2M\}$. This term plays an important role in the second-order effects of the mean squared error of $\hat{\beta}_H$. It turns out that this term dominates (in order of magnitude) the squared bias terms which come from B_{xj} and B_{uj} , and, as a result, the normalized mean squared error of $\hat{\beta}_H$ can be approximated by an expression of the form

$$1 + \frac{1}{m}A + \frac{1}{M^2}B + \text{higher-order terms.}$$

The second-order effect, $(1/m)A + (1/M^2)B$ (see Appendix C for the definition of A and B), differs from that of $MSE(\hat{\beta})$ in previous sections of the paper. Let $MSE(\hat{\beta}_H)$ be the standardized mean squared error of $\hat{\beta}_H$, i.e. $E[a^{-1/2}\Omega^{1/2}\Psi_H]^2$, and let $a = \frac{1}{2}\int_{-\infty}^{\infty} k(x)^2 dx$ be the limit of $(1/m)\sum_s K(\lambda_s - \omega)^2$ as $T \rightarrow \infty$, and $\bar{K}(\theta) = K(\theta/2M)$. The expansion for $MSE(\hat{\beta}_H)$ is given in the following result.

Theorem 3. Under Assumptions A.1, A.2

$$MSE(\hat{\beta}_H) \approx 1 + \frac{a}{2m} + \frac{1}{M^2} \Omega^{-1} \left[\frac{1}{a} \mathcal{V}_3 + 2k_2 \mathcal{V}_4 \right], \tag{6.7}$$

where

$$\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{xx}(\omega) f_{uu}(\omega)^{-1} d\omega, \tag{6.8}$$

$$\begin{aligned} \mathcal{V}_3 = & \frac{1}{16\pi} \int_{-\pi}^{\pi} K(\theta)^2 \theta^2 d\theta \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} f_{xx}''(\omega) f_{uu}(\omega)^{-1} d\omega \right. \\ & + \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{xx}(\omega) f_{uu}(\omega)^{-2} f_{uu}''(\omega)^2 d\omega \\ & \left. + \frac{1}{\pi} \int_{-\pi}^{\pi} f_{xx}'(\omega) f_{uu}'(\omega) f_{uu}(\omega)^{-2} d\omega \right], \tag{6.9} \end{aligned}$$

$$\mathcal{V}_4 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{uu}(\omega)^{-1} f_{x2}(\omega) d\omega. \tag{6.10}$$

Remark 6. Among the second-order effects, the $O_p(m^{-1})$ term, $a/2m$, comes from the variance of the nonparametric spectral density estimate, and the coefficient, $a/2$, reflects a scaling effect which depends on the kernel function. The $O_p(M^{-2})$ term, $(1/M^2)\Omega^{-1}[(1/a)\mathcal{V}_3 + 2k_2\mathcal{V}_4]$, is a bias term that comes from $E\hat{f}_{xu}(\omega_j)^2$ when we estimate $f_{xx}(\omega_j)$ by $E\hat{I}_{xx}(\lambda_j)$ in the frequency band B_j . This $O_p(M^{-2})$ effect depends on the kernel function and on the slope of the spectral density.

Remark 7. When we calculate the mean squared error of a spectral density estimator, we have a squared bias term which is of order $O(M^{-4})$. However, in the expression for $MSE(\hat{\beta}_H)$, not only are there $O(M^{-4})$ terms from the squared bias, but there are also $O(M^{-2})$ bias terms that originate in the second-order bias effect in $E\hat{f}_{xu}(\omega_j)^2$ and these dominate the $O(M^{-4})$ terms. When this term is positive, in order to minimize the order of magnitude of the second-order effect

$$\frac{a}{2m} + \frac{1}{M^2} \Omega^{-1} \left[\frac{1}{a} \mathcal{V}_3 + 2k_2 \mathcal{V}_4 \right],$$

we have to set m and M^2 to be the same order. If $M = O(T^\mu)$, ($0 < \mu < 1/2$), then $m = O(T^{1-\mu})$, and thus $1 - \mu = 2\mu$, i.e., $\mu = 1/3$, and $MSE(\hat{\beta}_H) = 1 + O_p(T^{-2/3})$. As a result, the trade-off between bias and variance yields an optimal rate of $M \sim T^{1/3}$ for the bandwidth to minimize the second-order effect of the estimator $\hat{\beta}_H$. This order of magnitude differs from the optimal order for the bandwidth obtained for the estimator (2.8). It also differs from the optimal order

in estimating a spectral density at a single point (in which case $M \sim T^{1/5}$). Specifically, we let the frequency band shrink more quickly (at a rate of $T^{-1/3}$) here than in the case of estimator (2.8). As a result, the second-order effect in the mean squared error of the estimator $\hat{\beta}_H$ is of order $T^{-2/3}$.

Remark 8. We can choose the optimal bandwidth by minimizing the second-order effect in the MSE, i.e., we choose M such that it minimizes

$$\frac{a}{2m} + \frac{1}{M^2} \left[\frac{\mathcal{V}_3}{a} + 2k_2 \mathcal{V}_4 \right] \Omega^{-1}.$$

The optimal bandwidth is then

$$M = \left[\frac{2\mathcal{V}_3 + 4ak_2\mathcal{V}_4}{a^2\Omega} \right]^{1/3} T^{1/3} = \delta'(k, f) T^{1/3}. \quad (6.11)$$

Like formula (4.6), the coefficient $\delta'(k, f)$ depends on the kernel function k and the spectral density. A feasible procedure is again obtained by using plug-in estimates based on a simple parametric model like an AR(1). For the example considered in Remark 3, i.e., processes (4.7) and (4.8), the corresponding formula for the optimal bandwidth (6.11) is

$$M = \left\{ \frac{\pi^2 [\delta_3(\alpha, \theta) + 3\alpha\theta]}{12[1 + \alpha^2\theta^2 + (\alpha - \theta)^2]} \right\}^{1/3} T^{1/3},$$

where δ_3 is defined by

$$\delta_3(\alpha, \theta) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \frac{[8\alpha^2 \sin^2 \omega - 2\alpha \cos \omega (1 + \alpha^2 - 2\alpha \cos \omega)]^2 (1 + \theta^2 + 2\theta \cos \omega)}{(1 + \alpha^2 - 2\alpha \cos \omega)^4} d\omega.$$

Fig. 1 plots the second-order adjusted asymptotic MSE of the two estimators for certain AR and MA processes. We call $\hat{\beta}$ in (2.8) estimator 1 and $\hat{\beta}_H$ in (6.4) estimator 2 in these graphs. The two curves in each graph depict the value of

$$1 + \frac{1}{2m} \left[\int_{-\infty}^{\infty} k(x)^2 dx \right] + \frac{k_q^2}{M^{2q}} [\Omega^{-1} \mathcal{V}_1(f_q) - \Omega^{-2} \mathcal{V}_2(f_q)^2]$$

for $\hat{\beta}$ with (4.6) plugged in, and the value of

$$1 + \frac{a}{2m} + \frac{1}{M^2} \Omega^{-1} \left[\frac{1}{a} \mathcal{V}_3 + 2k_2 \mathcal{V}_4 \right]$$

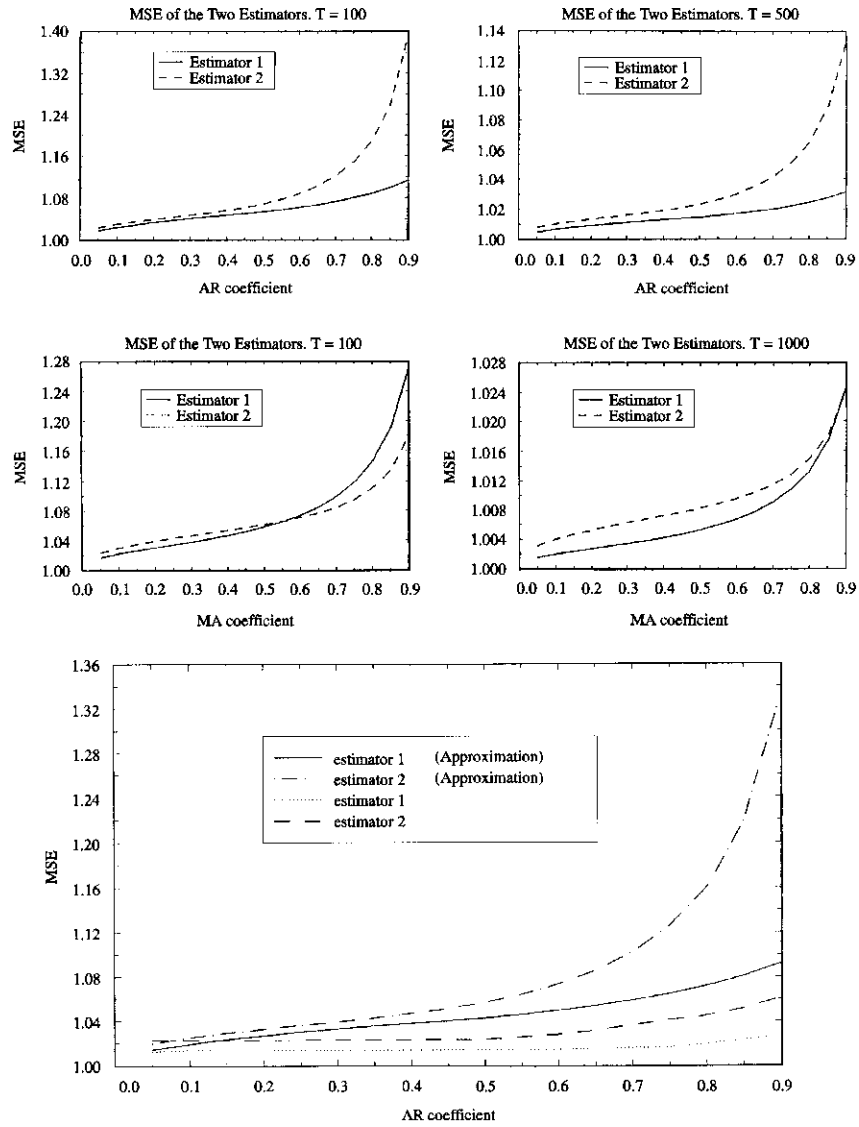


Fig. 1. Mean squared errors of the two estimators.

with (6.11) plugged in. The models considered are

$$x_t = \eta_t + 0.5\eta_{t-1}, \tag{6.12}$$

$$u_t = \alpha u_{t-1} + \varepsilon_t, \tag{6.13}$$

and

$$x_t = \eta_t - 0.5\eta_{t-1}, \quad (6.14)$$

$$u_t = \varepsilon_t - \theta\varepsilon_{t-1}. \quad (6.15)$$

We let α and θ take values from 0.05 to 0.9, and Fig. 1 shows how the adjusted MSE changes with these parameters settings.

Remark 9. These graphs show that, generally speaking, estimator 1 is superior at the second-order level. Intuitively, $\hat{\beta}$ uses T unsmoothed regressors, $\sum_{t=1}^T I_{xx}(\lambda_t) f_{uu}(\lambda_t)^{-1}$, in estimation, while $\hat{\beta}_H$ uses $2M$ smoothed regressors, $\sum_{j=-M+1}^M \hat{f}_{xx}(\omega_j) \hat{f}_{uu}(\lambda_t)^{-1}$. As a result, the first estimator has a stronger signal than the second since the smoothing reduces the strength of the signal. However, these differences do not show up in the first-order asymptotics but play an important role in second-order effects, including the order of magnitude of the error.

7. Multivariate extension

Our analysis can be extended to the multivariable case where x_t and β are $p \times 1$ vectors. In this section, I_{xx} and f_{xx} are $p \times p$ matrices, and I_{xy} , I_{xu} are $p \times 1$ vectors. For convenience, we work with the scalar standardized quantities of these two estimators,

$$[c' \Omega^{-1} c]^{-1/2} \sqrt{T} c' (\hat{\beta} - \beta),$$

and

$$[c' \Omega^{-1} c]^{-1/2} \sqrt{T} c' (\hat{\beta}_H - \beta),$$

where c is any $p \times 1$ vector. By a geometric series expansion, we get the representation

$$\begin{aligned} \sqrt{T} c' (\hat{\beta} - \beta) &= c' [\Omega_T^{-1} - \Omega_T^{-1} W_D \Omega_T^{-1} \\ &\quad + \Omega_T^{-1} W_D \Omega_T^{-1} W_D \Omega_T^{-1}] W_N - R_T = \Phi - R_T, \end{aligned}$$

where $R_T = c' \hat{\Omega}_T^{-1} W_D \Omega_T^{-1} W_D \Omega_T^{-1} W_D \Omega_T^{-1} W_N$. Let $MSE_c(\hat{\beta})$ denote the standardized mean squared error of $\hat{\beta}$, i.e., $MSE_c(\hat{\beta}) = E\{[c' \Omega^{-1} c]^{-1/2} \Phi\}^2$. The expansion of $MSE_c(\hat{\beta})$ is then given by the following theorem.

Theorem 4. Under Assumptions A.1 and A.2,

$$\begin{aligned} MSE_c(\hat{\beta}) &\approx 1 + \frac{1}{m} \left[\frac{1}{2} \int_{-\infty}^{\infty} k(x)^2 dx \right] \\ &\quad + \frac{k_a^2}{M^{2a}} \left\{ \frac{c' \Omega^{-1} [\mathcal{V}_1 - \mathcal{V}_2 \Omega^{-1} \mathcal{V}_2] \Omega^{-1} c}{c' \Omega^{-1} c} \right\}, \end{aligned}$$

where Ω , \mathcal{V}_1 , \mathcal{V}_2 are $p \times p$ matrices defined by the formulae (4.3)–(4.5).

Remark 10. We continue to use the notation $\Omega_T, \Omega, \mathcal{V}_1, \mathcal{V}_2, W_D, W_N$ here (and similar formulae to that of previous sections), even though they are now matrices or vectors rather than scalars in this section.

Similar results to those of Section 6 also apply for the Hannan estimator $\hat{\beta}_H$. Let $MSE_c(\hat{\beta}_H)$ be the standardized mean squared error of the truncated expansion of $\hat{\beta}_H$. Theorem 5 gives the expansion of $MSE_c(\hat{\beta}_H)$ for $\hat{\beta}_H$.

Theorem 5. Under Assumptions A.1, A.2,

$$MSE_c(\hat{\beta}_H) \approx 1 + \frac{a}{2m} + \frac{1}{M^2} \left\{ \frac{c' \Omega^{-1} [\mathcal{V}_3 + 2ak_2 \mathcal{V}_4] \Omega^{-1} c}{ac' \Omega^{-1} c} \right\},$$

where $\Omega, \mathcal{V}_3, \mathcal{V}_4$ are $p \times p$ matrices defined by matrix analogues of the formulae (6.8)–(6.10).

8. Monte Carlo results

A small simulation experiment was conducted to evaluate the bandwidth selection procedures and the sampling performance of estimators that use these automated bandwidth choices. The model used for data generation was the following:

$$y_t = \beta x_t + u_t, \quad \beta = 1,$$

$$x_t = \eta_t + \theta \eta_{t-1}, \quad \theta = 0.5.$$

Two different specifications of u_t were considered, and they are denoted *DGP(1)* and *DGP(2)*:

$$DGP(1): \text{AR}(1), \quad u_t = \alpha u_{t-1} + \varepsilon_t, \quad \alpha = 0.9$$

$$DGP(2): \text{MA}(1), \quad u_t = \varepsilon_t + \alpha \varepsilon_{t-1}, \quad \alpha = 0.5$$

In each case, η_t and ε_t are both iid $N(0,1)$ variates and are independent of each other. The sampling performance of the frequency domain estimators $\hat{\beta}, \hat{\beta}_H$ and the simple OLS estimator were examined for the case of each *DGP* and for different sample sizes. We use the following notation in our discussion.

1. OLS: OLS estimator for β .
2. GLS1: Frequency domain GLS estimator $\hat{\beta}$ in (2.8) by using the bandwidth (4.6).
3. GLS2: Frequency domain GLS estimator $\hat{\beta}$ in (2.8) using the simple ‘rule-of-thumb’ bandwidth $M = T^{1/3}$.
4. GLS3: Frequency domain GLS estimator $\hat{\beta}$ in (2.8) using the simple ‘rule-of-thumb’ bandwidth $M = T^{1/5}$.

5. GLS1H: Frequency domain GLS estimator (6.4) using the bandwidth (6.11).
6. GLS2H: Frequency domain GLS estimator (6.4) using the simple ‘rule-of-thumb’ bandwidth $M = T^{1/3}$.
7. GLS3H: Frequency domain GLS estimator (6.4) using the simple ‘rule-of-thumb’ bandwidth $M = T^{1/5}$.

We tried the sample sizes $T = 2^6 = 64$, $2^7 = 128$, $2^8 = 256$. These highly composite sample sizes were chosen, as in Robinson (1991), to take advantage of fast discrete Fourier transform routines. The number of replications was 4000 for each case. Since our interest is primarily in bandwidth selection, we just used the Daniell window $K(\lambda_s - \omega_j) = 1$ for $\lambda_s \in B_j$. In Figs. 2–7, true formulae were used for approximations (4.6) and (6.11) in estimates GLS1 and GLS1H. A comparison between the results of using true bandwidth formulae and the plug-in method is shown in Figs. 8 and 9. Figs. 2 and 3 plot the empirical distributions of the estimator $\hat{\beta}$ in (2.8) (GLS1, GLS2) and the OLS estimator for the sample sizes $T = 2^7$ and $T = 2^8$ when the data were generated by $DGP(1)$. We can see that as the sample size increases, the performance of the frequency domain *GLS* estimator is improved. Figs. 4 and 5 give the empirical distributions of these estimators for the sample sizes $T = 2^7$ and $T = 2^8$ when the data were generated by $DGP(2)$. The sampling performance differs across data generating processes. The frequency domain *GLS* estimator has better small sample performance in the case of $DGP(1)$

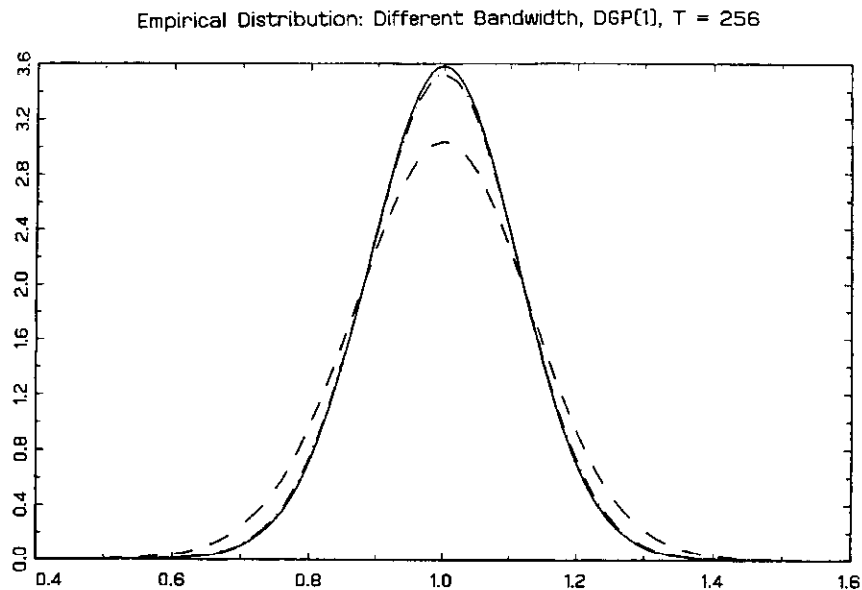


Fig. 2. GLS1 (solid), GLS2 (dots and dashes), OLS (dotted).

Empirical Distribution: Different Bandwidth, DGP(1), T = 128

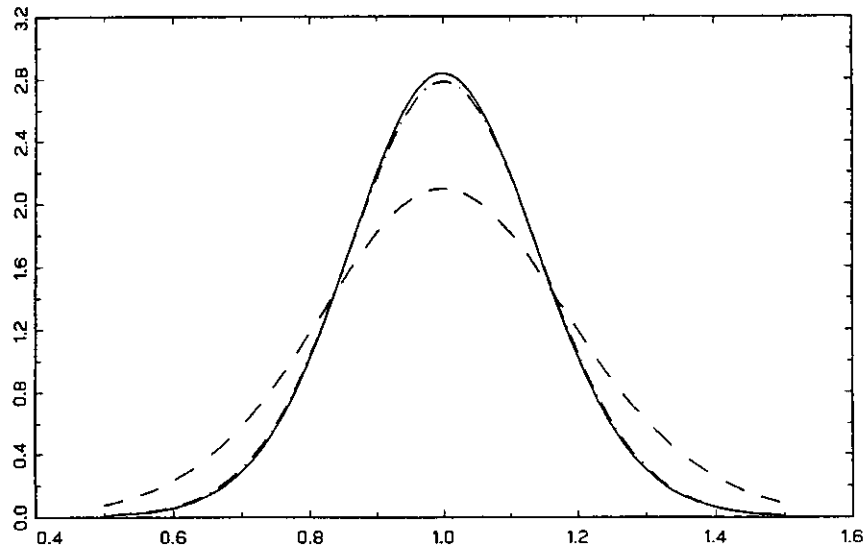


Fig. 3. GLS1 (solid), GLS2 (dots and dashes), OLS (dashed).

Empirical Distribution: Different Bandwidth, DGP(2), T = 128

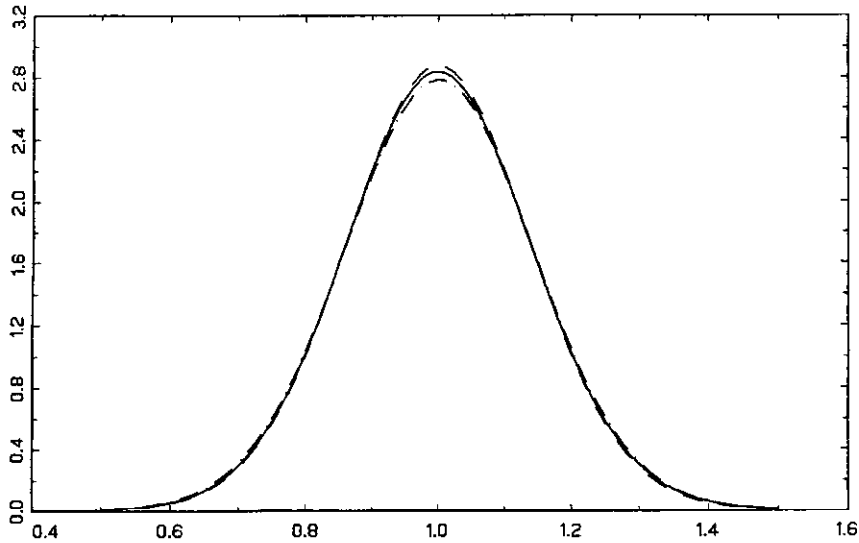
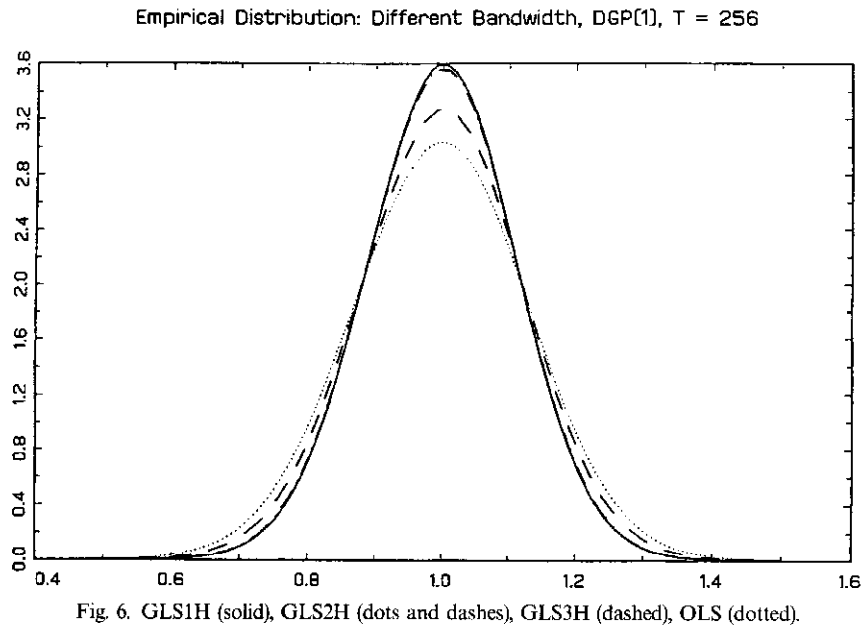
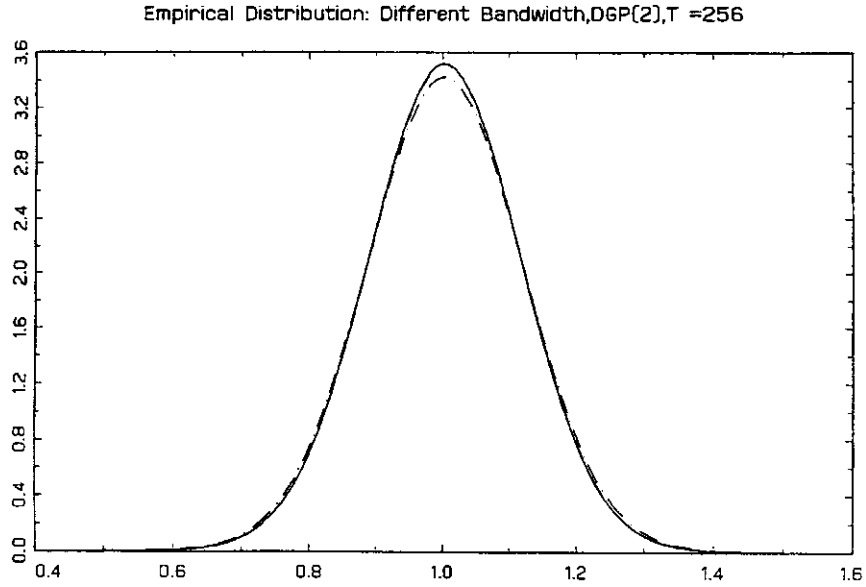


Fig. 4. GLS1 (solid), GLS2 (dots and dashes), OLS (dashed).



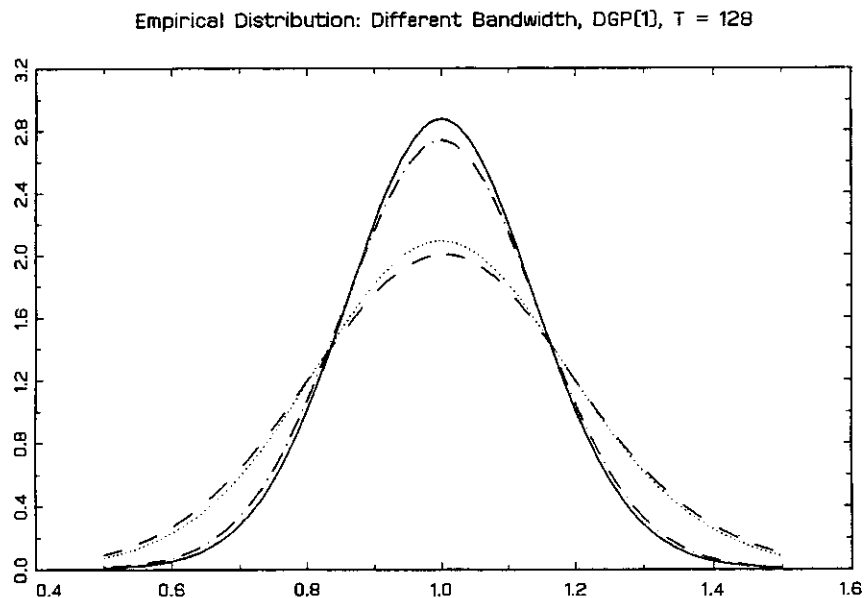


Fig. 7. GLS1H (solid), GLS2H (dots and dashes), GLS3H (dashed), OLS (dotted).

than in *DGP(2)*. Similar phenomenon can be found for the Hannan estimator (Figs. 6 and 7).

These simulation results emphasize the importance of bandwidth selection. Variation across M is apparent in the figures. For the Hannan estimator $\hat{\beta}_H$, a bandwidth selection for M of order $T^{1/3}$ generally provides better estimation for β than M of order $T^{1/5}$, corroborating second-order asymptotic theory. An unsuitable choice of bandwidth can lead to poor estimation for β . In the case of *DGP(1)*, substantial efficiency gains are achieved by choosing the bandwidth to be of order $M^{1/3}$ for the Hannan's estimator. However, less favorable conclusions are found when the error term is generated by *DGP(2)*.

The parametric plug-in method of using an AR(1) formula for the error process has also been tried. Figs. 8 and 9 graph the empirical distributions for estimators *GLS1H* using the true bandwidth formula and the plug-in formula when the sample size is 2^6 . Fig. 8 compares the distributions for estimator *GLS1H* when the error process was generated by AR(1) model and an AR(1) structure was used in the plug-in formula. Fig. 9 graphs these distributions when the error process was generated by MA(1) model and we used an AR(1) plug-in formula. When the prespecified model is close to the true error process, there is no big difference between the empirical distribution of the estimator that uses the true optimal bandwidth formula and the estimator using a plug-in formula.

Feasible GLS and infeasible GLS are also compared. Fig. 10 reports the results for the case of *DGP(2)* with sample size $T = 2^7$, other cases being similar. The

Effect of the Plug in Method, AR(1) Case

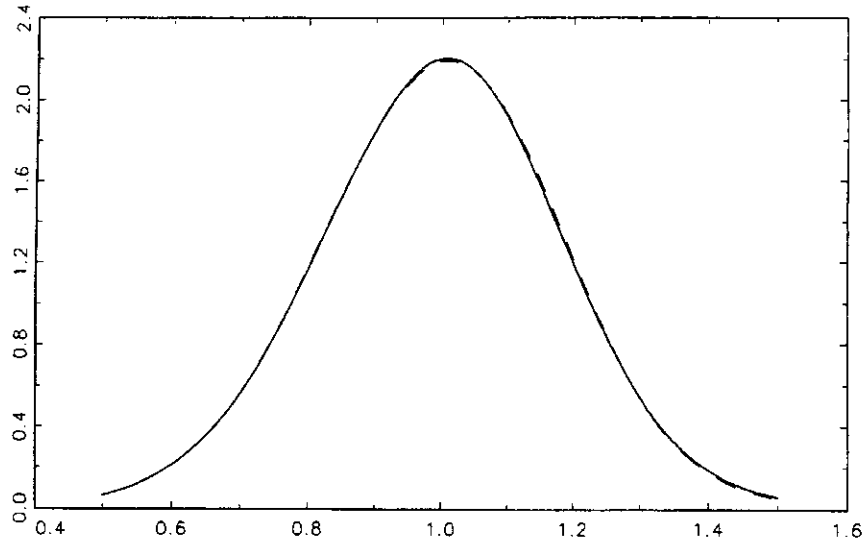


Fig. 8. GLS1H (solid), GLS1H-plug in (dashed).

Effect of the Plug in Method, MA(1) Case

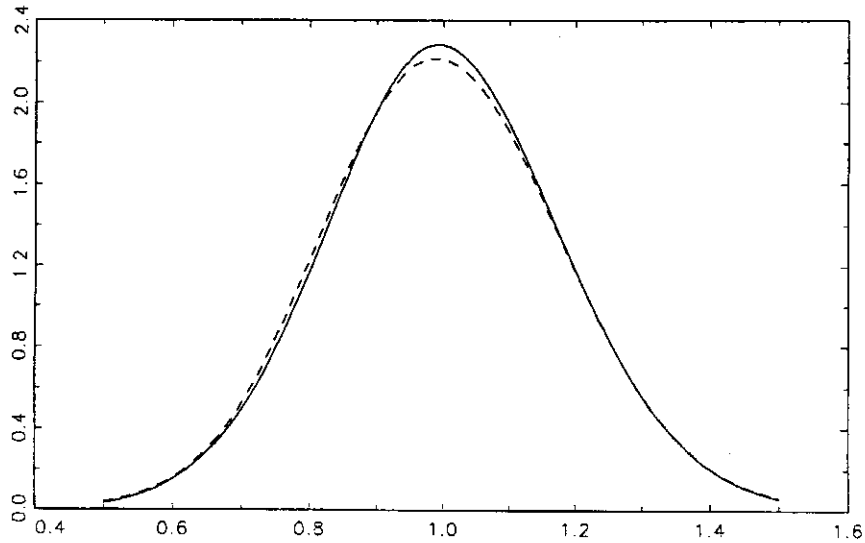


Fig. 9. GLS1H (solid), GLS1H-plug in (dashed).

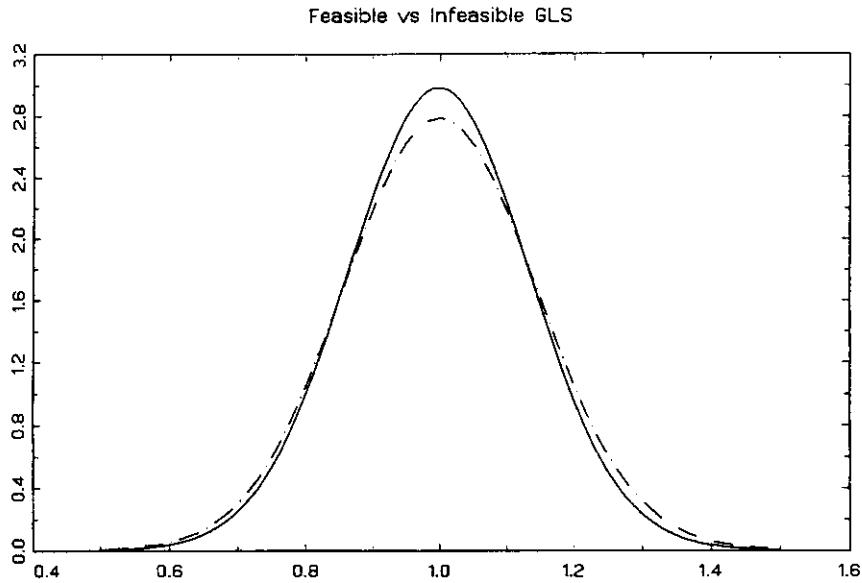


Fig. 10. GLS1 (dots and dashes), infeasible GLS (solid).

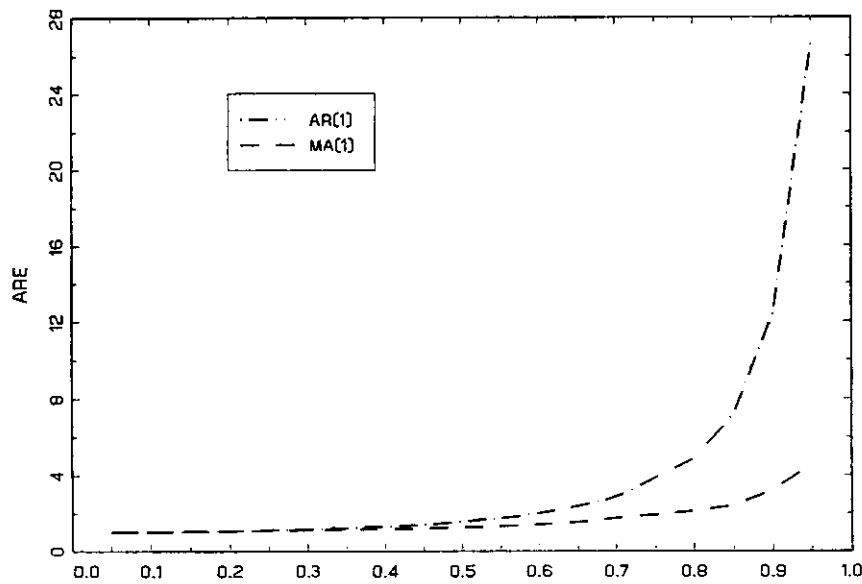


Fig. 11. Relative efficiency of GLS over OLS.

relative performance of the frequency domain estimator and the OLS estimator depends on the exact distribution of the errors in small samples. Generally speaking, the frequency domain estimators are more efficient than the OLS estimator for *DGP(1)*. But OLS does do better than the GLS estimators in the case of *DGP(2)* for relatively small samples. The main reason for this phenomenon is that the asymptotic relative efficiency of GLS over OLS ($ARE = Var(\hat{\beta}_{GLS})/Var(\hat{\beta}_{OLS})$) is much higher in the case of *DGP(1)* than in *DGP(2)*. Fig. 11 plots the *ARE* of GLS over OLS for AR and MA errors. The solid line corresponds to an AR(1) error process $u_t = \alpha u_{t-1} + \varepsilon_t$ with α taking values from 0.1 to 0.9. The dashed line is for an MA error $u_t = \varepsilon_t + \alpha \varepsilon_{t-1}$, with α taking values from 0.1 to 0.9. The greater potential gains from GLS estimation in the AR(1) case are apparent in the figure.

9. Conclusion

This paper develops second-order expansions for efficient, frequency domain semiparametric estimators and gives second-order approximations to their mean squared errors. While choice of the bandwidth parameter does not figure in the first order asymptotics for these estimators, it does influence the second-order terms. It can therefore be chosen in such a way as to minimize second-order effects and thereby enhance second-order efficiency. Second-order formulae also provide a mechanism for data-based bandwidth selection rules that are useful for the practical implementation of these procedures. Under normality assumptions, a more specific distributional approximation for the estimator is possible and is given in Section 5. Two commonly used versions of the frequency domain estimator are studied. One of these ($\hat{\beta}_H$) is due originally to Hannan (1963a,b) and involves more smoothing over frequencies than the other ($\hat{\beta}$). It is shown that while these two estimators are asymptotically equivalent, they do differ at the second-order level. As a result, different bandwidths should be set for these two estimation procedures to minimize the second-order effect in the asymptotic mean squared error. For the commonly occurring quadratic kernel functions, the optimal expansion rate for the bandwidth parameter M for these two estimators ($\hat{\beta}_H$, $\hat{\beta}$) are $T^{1/3}$ and $T^{1/5}$ respectively. Monte Carlo evidence emphasizes the importance of bandwidth selection in practical applications. Although the relative performance of different estimation procedures depends on the form of the error process, we find that the bandwidth selection procedures suggested by the second-order asymptotics perform reasonably well for both of the frequency domain estimators $\hat{\beta}$ and $\hat{\beta}_H$.

Acknowledgements

Our thanks go to the Co-Editor, Peter Robinson, Oliver Linton, and a referee for helpful comments on an earlier version of this paper. The paper was typed by the

authors in SW2.5 and computations were performed in GAUSS. The paper was completed while Phillips was visiting the University of Auckland. Xiao thanks the Cowles Foundation for support under an Anderson Fellowship and Phillips thanks the NSF for support under Grant No. SBR 94-22922.

Appendix A. Expansion of estimator $\hat{\beta}$: Lemmas and proofs

The following lemma shows that, in nonparametric spectral density estimation, the bias term coming from the periodogram itself is of order T^{-1} .

Lemma A.1. Under Assumption A.2, $E I_{uu}(\lambda_t) - f_{uu}(\lambda_t) = O(T^{-1})$, the $O(T^{-1})$ term is uniform in λ_s , (cf. Brillinger, 1980, Theorem 5.2.2)

Proof of Lemma 1. The proofs for the results of B_t and V_t are the same as the standard argument for the bias and variance of kernel spectral density estimates. For P_t , note that

$$\begin{aligned}
 P_t &= \sum_{s \neq t} \omega_{ts} [\hat{I}_{uu}(\lambda_s) - I_{uu}(\lambda_s)] \\
 &= -2(\hat{\beta} - \beta) \sum_{s \neq t} \omega_{ts} I_{xu}(\lambda_s) + (\hat{\beta} - \beta)^2 \sum_{s \neq t} \omega_{ts} I_{xx}(\lambda_s).
 \end{aligned}
 \tag{A.1'}$$

Notice that $\hat{\beta} - \beta = O_p(T^{-1/2})$ and it can be verified that $\sum_{s \neq t} \omega_{ts} I_{xx}(\lambda_s) = O_p(1)$ and $\sum_{s \neq t} \omega_{ts} I_{xu}(\lambda_s) = O_p(m^{-1/2})$, thus the first term in Eq. (A.1') is of order $O_p(T^{-1/2} m^{-1/2})$ and the second term is $O_p(T^{-1})$. As a result, $P_t = o_p(m^{-1})$. \square

Proof of Lemma 2. The results follow from Lemma 1. \square

In view of Lemma 2, we see that R_1 in Eq. (2.9) is of higher-order of magnitude. Substituting Eq. (3.3) into the expressions for W_D and W_{N1} , we get the following:

$$W_D = -W_{D0} + W_{D1} - b_{D1} + b_{D2} + R_{D1},
 \tag{A.1}$$

$$W_{N1} = -Q_{N1} - L_{N1} + L_{N2} + L_{N3} + Q_{N2} + C_{N1},
 \tag{A.2}$$

where

$$W_{D0} = T^{-1} \sum_{t=-T/2+1}^{T/2} I_{xx}(\lambda_t) f_{uu}^*(\lambda_t)^{-2} V_t,$$

$$W_{D1} = T^{-1} \sum_t I_{xx}(\lambda_t) f_{uu}^*(\lambda_t)^{-3} E V_t^2,$$

$$b_{D1} = T^{-1} \sum_t I_{xx}(\lambda_t) f_{uu}(\lambda_t)^{-2} B_t,$$

$$\begin{aligned}
b_{D2} &= T^{-1} \sum_t I_{xx}(\lambda_t) f_{uu}(\lambda_t)^{-3} B_t^2, \\
R_{D1} &= T^{-1} \sum_t I_{xx}(\lambda_t) f_{uu}^*(\lambda_t)^{-3} [V_t^2 - EV_t^2], \\
L_{N1} &= T^{-1/2} \sum_t w_x(\lambda_t) w_u(\lambda_t) f_{uu}(\lambda_t)^{-2} B_t, \\
L_{N2} &= T^{-1/2} \sum_t w_x(\lambda_t) w_u(\lambda_t) f_{uu}^*(\lambda_t)^{-3} EV_t^2, \\
L_{N3} &= T^{-1/2} \sum_t w_x(\lambda_t) w_u(\lambda_t) f_{uu}(\lambda_t)^{-3} B_t^2, \\
Q_{N1} &= T^{-1/2} \sum_t w_x(\lambda_t) w_u(\lambda_t) f_{uu}^*(\lambda_t)^{-2} V_t, \\
Q_{N2} &= T^{-1/2} \sum_t w_x(\lambda_t) w_u(\lambda_t) f_{uu}^*(\lambda_t)^{-3} \left[\sum_s \omega_{ss}^2 (\zeta_s^2 - E\zeta_s^2) \right], \\
C_{N1} &= T^{-1/2} \sum_t w_x(\lambda_t) w_u(\lambda_t) f_{uu}^*(\lambda_t)^{-3} \left[\sum_{s \neq j} \omega_{is} \omega_{tj} \zeta_s \zeta_j \right],
\end{aligned}$$

where $\zeta_s = I_{uu}(\lambda_s) - f_{uu}(\lambda_s)$. The order of magnitude for each term in W_{N1} and W_D is given in the next lemma.

Lemma A.2. $W_{D0} = O_p(T^{-1/2})$, $W_{D1} = O_p(m^{-1})$, $b_{D1} = O_p(M^{-q})$, $b_{D2} = O_p(M^{-2q})$, $R_{D1} = O_p(T^{-1/2}m^{-1})$, $L_{N1} = O_p(M^{-q})$, $L_{N2} = O_p(m^{-1})$, $L_{N3} = O_p(M^{-2q})$, $Q_{N1} = O_p(m^{-1/2})$, $Q_{N2} = O_p(m^{-1})$, $C_{N1} = O_p(m^{-1})$

Proof. The orders of $W_{D1}, b_{D1}, b_{D2}, L_{N2}$ follow from the results of Lemma 2. The proofs of $R_{D1}, L_{N1}, L_{N3}, Q_{N1}, Q_{N2}, C_{N1}$ follow directly by calculating the second-order moments. \square

Substituting (A.1) and (A.2) into expansion (2.9), and dropping those terms that are of order $O_p(M^{-2q})$ or $O_p(m^{-1})$ according to the results of Lemma A.2, we find

$$\begin{aligned}
\sqrt{T}(\hat{\beta} - \beta) &= \Omega_T^{-1} \{ W_{N0} - [L_{N1} - b_{D1} \Omega_T^{-1} W_{N0}] \\
&\quad + [L_{N2} - W_{D1} \Omega_T^{-1} W_{N0}] \\
&\quad + [L_{N3} - b_{D2} \Omega_T^{-1} W_{N0} - b_{D1} \Omega_T^{-1} L_{N1} + b_{D1}^2 M^{-2} W_{N0}] \\
&\quad - Q_{N1} + Q_{N2} + C_{N1} - W_{D0} \Omega_T^{-1} W_{N0} - b_{D1} \Omega_T^{-1} Q_{N1} \} \\
&\quad + \text{higher-order terms} \tag{A.3}
\end{aligned}$$

$$\begin{aligned}
 &= \Omega_T^{-1}W_{N0} + \frac{1}{\sqrt{m}}\mathcal{A}_1 + \frac{1}{M^2}\mathcal{B}_1 + \frac{1}{m}\mathcal{A}_2 + \frac{1}{M^4}\mathcal{B}_2 \\
 &\quad + \text{higher-order terms,} \tag{A.4}
 \end{aligned}$$

where $\mathcal{A}_1 = \Omega_T^{-1}Q_{N1}$, $\mathcal{B}_1 = \Omega_T^{-1}(L_{N1} - b_{D1}\Omega_T^{-1}W_{N0})$, $\mathcal{A}_2 = \Omega_T^{-1}(L_{N2} - W_{D1}\Omega_T^{-1}W_{N0} + Q_{N2} + C_{N1})$, $\mathcal{B}_2 = \Omega_T^{-1}(L_{N3} - b_{D2}\Omega_T^{-1}W_{N0} - b_{D1}\Omega_T^{-1}L_{N1} + b_{D1}^2M^{-2}W_{N0})$. The following lemma, which gives correlations among terms in Eq. (A.3), helps us to eliminate higher-order terms in the expansion of the mean squared error.

Lemma A.3. The following expectations are either of order $o(m^{-1})$ or $o(M^{-2q})$

$$E[W_{N0}(L_{N1} - b_{D1}\Omega_T^{-1}W_{N0})], \tag{A.5}$$

$$E[W_{N0}(L_{N2} - W_{D1}\Omega_T^{-1}W_{N0})], \tag{A.6}$$

$$E[W_{N0}(L_{N3} - b_{D2}\Omega_T^{-1}W_{N0})], \tag{A.7}$$

$$E[W_{N0}(-b_{D1}\Omega_T^{-1}L_{N1} + b_{D1}^2M^{-2}W_{N0})], \tag{A.8}$$

$$E[Q_{N1}(L_{N1} - b_{D1}\Omega_T^{-1}W_{N0})], \tag{A.9}$$

$$E(W_{N0}Q_{N1}), \tag{A.10}$$

$$E(W_{N0}Q_{N2}), \tag{A.11}$$

$$E(W_{N0}C_{N1}), \tag{A.12}$$

$$E(W_{N0}W_{D0}\Omega_T^{-1}W_{N0}). \tag{A.13}$$

Proof. The proofs are similar as those of Lemma A.2. \square

As a result of the above lemma, we get

$$\begin{aligned}
 \text{MSE}(\hat{\beta}) &= \text{Var}(\Omega_T^{-1/2}W_{N0}) + \text{Var}[\Omega_T^{-1/2}(L_{N1} - b_{D1}\Omega_T^{-1}W_{N0})] \\
 &\quad + \text{Var}(\Omega_T^{-1/2}Q_{N1}) + \text{higher-order terms} \tag{A.14}
 \end{aligned}$$

$$= 1 + \frac{1}{m}\mathcal{A} + \frac{1}{M^4}\mathcal{B} + \text{higher-order terms,} \tag{A.15}$$

where $\mathcal{A} = \frac{1}{2}\int k(x)^2 dx$, $\mathcal{B} = k_q^2[\Omega^{-1}\mathcal{V}_1 - \Omega^{-2}\mathcal{V}_2^2]$.

Appendix B. Lemmas for the distributional approximation

Lemma B.1. $\hat{\Sigma}$ and $\hat{\beta} - \tilde{\beta}$ do not depend on β .

Proof. Notice that $\hat{\Sigma} = \text{diag}[\hat{f}_{uu}(\lambda_1), \dots, \hat{f}_{uu}(\lambda_T)]$, and $\hat{f}_{uu}(\lambda_s)$ is calculated from $\hat{w}_u(\lambda_s) = w_y(\lambda_s) - \hat{\beta}_{OLS}w_x(\lambda_s)$ where $\hat{\beta}_{OLS} = [W'_xW_x]^{-1}W'_xW_y$. Thus \hat{W}_u does not

depend on β since neither x nor u do. As a result, $\hat{f}_{uu}(\lambda_t)$ and $\hat{\Sigma}$ do not depend on β , and $\hat{\beta} - \tilde{\beta}$ is independent of β . \square

The following lemmas gives some uniform convergence results which are useful in controlling higher-order terms and help us to establish condition (5.3) under which Theorem 2 holds.

Lemma B.2.

$$\begin{aligned} \sup_t |B_t| &= O(M^{-q}), \\ \sup_t |V_t| &= O_p(m^{-1/2+\varepsilon}) \quad \text{for any } \varepsilon > 0, \\ \sup_t |P_t| &= o_p(T^{-2\tau}). \end{aligned}$$

Proof. The results for B_t and V_t follow from Brillinger (1980), (Theorem 7.7.4). The third result follows from the fact that the moments of $w_u(\lambda)$ are uniformly bounded under normality assumption. \square

Since $|\hat{f}_u(\lambda_t) - f_u(\lambda_t)|^p \leq c_0\{|B_t|^p + |V_t|^p + |P_t|^p\}$ for some constant c_0 , we have

$$\sup_t |\hat{f}_u(\lambda_t) - f_u(\lambda_t)|^p = O(T^{-p\tau+\varepsilon}).$$

Let $A_t = \{|\hat{f}_u(\lambda_t) - f_u(\lambda_t)| > f_u(\lambda_t)/2\}$, $A = \cup_{t=1}^T A_t$, then

Lemma B.3. $\Pr(A) = o(T^{-2\tau})$

Proof. Omitted. \square

Restricting our attention to the region of the complement of A , A^c , we get

Lemma B.4.

$$\begin{aligned} \sup_T E|W_N|^p &< \infty, \\ \sup_T E|\tilde{W}_D|^p &< \infty, \quad \tilde{W}_D = M^q W_D. \end{aligned}$$

Proof. Omitted. \square

Notice that $\Omega_{\Sigma}^{1/2}\sqrt{T}(\hat{\beta} - \tilde{\beta}) = \Omega_{\Sigma}^{1/2}\sqrt{T}(\hat{\beta} - \beta) - \Omega_{\Sigma}^{1/2}\sqrt{T}(\tilde{\beta} - \beta) = Z_T + R_T$, where $Z_T = (\Omega_{\Sigma}^{-1} - \Omega_{\Sigma}^{-2}W_D + \Omega_{\Sigma}^{-3}W_D^2)W_N$, and $R_T = \hat{\Omega}_{\Sigma}^{-1}\Omega_{\Sigma}^{-3}W_D^3W_N$. Let

$G = \{T^{2\tau} \log T \|R_T\| > c\}$ and $H = \{\hat{\Omega}_T > \Omega/2\}$, we have

$$\Pr(G) \leq \Pr(G \cap H) + \Pr(H^c),$$

and since that $\Omega_T \rightarrow \Omega$, and $\hat{\Omega}_T \rightarrow \Omega > 0$ in probability,

$$\begin{aligned} \Pr(H^c) &= \Pr\{|\hat{\Omega}_T - \Omega| > c_*\} \\ &= \Pr\{M^{-q} |\tilde{W}_D| > c_*\} \\ &\leq c_* \frac{E|\tilde{W}_D|^p}{M^{pq}}, \end{aligned}$$

where c_* is a constant whose value is not always the same. By Lemma B.3, we focus on the region of A^c since the error is of order $o(T^{-2\tau})$. Thus, $\Pr(H^c) = o(T^{-2\tau})$ provided $p > 2$ and $\sup_T E|\tilde{W}_D|^p < \infty$. For $\Pr(G \cap H)$, we have

$$\begin{aligned} G \cap H &\subset \left\{ |2\Omega^{-1} \Omega_T^{-3} W_D^3 W_N| > \frac{c}{T^{2\tau} \log T} \right\} \\ &= \left\{ |\tilde{W}_D|^3 |W_N| > \frac{c_* M^{3q}}{T^{2\tau} \log T} \right\}. \end{aligned}$$

Then

$$\begin{aligned} \Pr(G \cap H) &\leq c_* \frac{E|\tilde{W}_D|^{3p} |W_N|^p T^{2p\tau} \log^p T}{M^{3pq}} \\ &= c_* \frac{E|\tilde{W}_D|^{3p} |W_N|^p \log^p T}{T^{pq\mu}}. \end{aligned}$$

Therefore, provided $\sup_T E|\tilde{W}_D|^{3p} |W_N|^p < \infty$, and $p > 2$, $\Pr(G \cap H) = o(T^{-2\tau})$. It follows that $\Pr(G) = o(T^{-2\tau})$, which justifies the error in (5.3) of Theorem 2.

Appendix C. The expansion of Hannan's estimator: Lemmas and proofs

C.1. The approximation for the spectral density estimator $\hat{f}_{uu}(\omega_j)$

Let

$$f_{xx}^*(\omega_j) = m^{-1} \sum_{\lambda_s \in B_j} K(\lambda_s - \omega_j) f_{xx}(\lambda_s) = \sum_s \omega_{js} f_{xx}(\lambda_s),$$

$$B_{xj} = f_{xx}^*(\omega_j) - f_{xx}(\omega_j) = \sum_s \omega_{js} [f_{xx}(\lambda_s) - f_{xx}(\omega_j)],$$

and

$$V_{xj} = \hat{f}_{xx}(\omega_j) - f_{xx}^*(\omega_j) = \sum_s \omega_{js} [I_{xx}(\lambda_s) - f_{xx}(\lambda_s)].$$

Then

$$\hat{f}_{xx}(\omega_j) = f_{xx}(\omega_j) + B_{xj} + V_{xj}. \quad (\text{C.1})$$

Lemma C.1. If $\lim_{x \rightarrow 0} \{1 - k(x)\}/|x|^q = k_q < \infty$, and

$$f_{xq}(\lambda_j) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} |h|^{q\gamma_x} e^{-ih\lambda_j},$$

then

$$B_{xj} \sim -M^{-q} k_q f_{xq}(\omega_j) = O(M^{-q})$$

$$\sqrt{m} V_{xj} \xrightarrow{d} N(0, a f_{xx}^2(\omega_j)) \quad \text{for } \omega_j \neq 0,$$

where $a = \frac{1}{2} \int_{-\infty}^{\infty} k(x)^2 dx$ is the limit of $(1/m) \sum_s K(\lambda_s - \omega_j)^2$.

Proof. Similar to that of Lemma 1. \square

Define

$$\bar{f}_{uu}(\omega_j) = m^{-1} \sum_{\lambda_s \in B_j} K(\lambda_s - \omega_j) I_{uu}(\lambda_s) = \sum_s \omega_{js} I_{uu}(\lambda_s),$$

$$f_{uu}^*(\omega_j) = m^{-1} \sum_{\lambda_s \in B_j} K(\lambda_s - \omega_j) f_{uu}(\lambda_s) = \sum_s \omega_{js} f_{uu}(\lambda_s),$$

$$\bar{f}_{uu}(\omega_j) = \frac{1}{2} [\bar{f}_{uu}(\omega_{j-1}) + \bar{f}_{uu}(\omega_{j+1})],$$

$$f_{uu}^{**}(\omega_j) = \frac{1}{2} [f_{uu}^*(\omega_{j-1}) + f_{uu}^*(\omega_{j+1})],$$

$$B_{uj} = f_{uu}^{**}(\omega_j) - f_{uu}(\omega_j),$$

$$V_{uj} = \bar{f}_{uu}(\omega_j) - f_{uu}^{**}(\omega_j),$$

$$P_{uj} = \bar{f}_{uu}(\omega_j) - \bar{f}_{uu}(\omega_j).$$

Then

$$\bar{f}_{uu}(\omega_j) = f_{uu}(\omega_j) + B_{uj} + V_{uj} + P_{uj}.$$

Lemma C.2.

$$B_{uj} = \frac{1}{M^2} \left[\frac{\pi^2}{2} f''_{uu}(\omega_j) - \frac{k_2}{2} f_{u2}(\omega_{j-1}) - \frac{k_2}{2} f_{u2}(\omega_{j+1}) \right] + o(M^{-2})$$

$$= \frac{1}{M^2} \left[\frac{\pi^2}{2} + \frac{1}{16\pi} \int_{-\pi}^{\pi} \bar{K}(\theta) \theta^2 d\theta \right] f''_{uu}(\omega_j) + o(M^{-2}),$$

$$\sqrt{2m} V_{uj} \xrightarrow{d} N(0, af''_{uu}(\omega_j)) \quad \text{for } \omega_j \neq 0,$$

$$P_{uj} = o_p(m^{-1}).$$

Proof. Similar to that of Lemma 1. \square

By the same argument as in Section 3, we obtain the following expansion for the reciprocal density:

$$\begin{aligned} \tilde{f}_{uu}(\omega_j)^{-1} - f_{uu}(\omega_j)^{-1} &= -f_{uu}^{**}(\omega_j)^{-2} V_{uj} + f_{uu}^{**}(\omega_j)^{-3} V_{uj}^2 \\ &\quad - f_{uu}(\omega_j)^{-2} B_{uj} + f_{uu}(\omega_j)^{-3} B_{uj}^2 \\ &\quad + o_p(m^{-1} + M^{-4}). \end{aligned} \tag{C.2}$$

C.2. Approximation of X_D and X_N

$$\begin{aligned} \hat{\Sigma}_T - \Sigma_T &= \frac{1}{2M} \sum_{j=-M+1}^M \{ f_{xx}(\omega_j) \\ &\quad \times [-f_{uu}^*(\omega_j)^{-2} V_{uj} + f_{uu}^*(\omega_j)^{-3} V_{uj}^2 - f_{uu}(\omega_j)^{-2} B_{uj} + f_{uu}(\omega_j)^{-3} B_{uj}^2] \\ &\quad + V_{xj} [f_{uu}(\omega_j)^{-1} - f_{uu}^*(\omega_j)^{-2} V_{uj} + f_{uu}^*(\omega_j)^{-3} V_{uj}^2 \\ &\quad - f_{uu}(\omega_j)^{-2} B_{uj} + f_{uu}(\omega_j)^{-3} B_{uj}^2] \\ &\quad + B_{xj} [f_{uu}(\omega_j)^{-1} - f_{uu}^*(\omega_j)^{-2} V_{uj} + f_{uu}^*(\omega_j)^{-3} V_{uj}^2 \\ &\quad - f_{uu}(\omega_j)^{-2} B_{uj} + f_{uu}(\omega_j)^{-3} B_{uj}^2] \} \\ &\quad + \text{higher-order terms.} \end{aligned}$$

Dropping those terms that are of order $o(m^{-1})$ and $o(M^{-4})$, we get

$$\begin{aligned} X_D &= \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M \{ f_{xx}(\omega_j) \\ &\quad \times [-f_{uu}^*(\omega_j)^{-2} V_{uj} + f_{uu}^*(\omega_j)^{-3} V_{uj}^2 - f_{uu}(\omega_j)^{-2} B_{uj} + f_{uu}(\omega_j)^{-3} B_{uj}^2] \\ &\quad + V_{xj} [f_{uu}(\omega_j)^{-1} - f_{uu}^*(\omega_j)^{-2} V_{uj} - f_{uu}(\omega_j)^{-2} B_{uj}] \\ &\quad + B_{xj} [f_{uu}(\omega_j)^{-1} - f_{uu}^*(\omega_j)^{-2} V_{uj} - f_{uu}(\omega_j)^{-2} B_{uj}] \} \\ &\quad + \text{higher-order terms} \end{aligned}$$

$$\begin{aligned}
&= -\frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{xx}(\omega_j) f_{uu}^*(\omega_j)^{-2} V_{uj} \\
&\quad - \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{xx}(\omega_j) f_{uu}(\omega_j)^{-2} B_{uj} \\
&\quad + \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{xx}(\omega_j) f_{uu}(\omega_j)^{-3} B_{uj}^2 \\
&\quad + \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{xx}(\omega_j) f_{uu}^*(\omega_j)^{-3} E V_{uj}^2 \\
&\quad + \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{xx}(\omega_j) f_{uu}^*(\omega_j)^{-3} [V_{uj}^2 - E V_{uj}^2] \\
&\quad + \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}(\omega_j)^{-1} V_{xj} \\
&\quad - \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}^*(\omega_j)^{-2} V_{xj} V_{uj} \\
&\quad - \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}(\omega_j)^{-2} B_{uj} V_{xj} \\
&\quad + \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}(\omega_j)^{-1} B_{xj} \\
&\quad - \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}(\omega_j)^{-2} B_{xj} B_{uj} \\
&\quad - \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}^*(\omega_j)^{-2} B_{xj} V_{uj} \\
&\quad + \text{higher-order terms,}
\end{aligned}$$

and

$$\begin{aligned}
X_N &= \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) \hat{f}_{uu}(\omega_j)^{-1} \\
&= \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) \\
&\quad \times [f_{uu}(\omega_j)^{-1} - f_{uu}^*(\omega_j)^{-2} V_{uj} + f_{uu}^*(\omega_j)^{-3} V_{uj}^2 \\
&\quad - f_{uu}(\omega_j)^{-2} B_{uj} + f_{uu}(\omega_j)^{-3} B_{uj}^2] \\
&\quad + \text{higher-order terms}
\end{aligned}$$

$$\begin{aligned}
&= \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) f_{uu}(\omega_j)^{-1} \\
&\quad - \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) f_{uu}^*(\omega_j)^{-2} V_{uj} \\
&\quad + \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) f_{uu}^*(\omega_j)^{-3} V_{uj}^2 \\
&\quad - \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) f_{uu}(\omega_j)^{-2} B_{uj} \\
&\quad + \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) f_{uu}(\omega_j)^{-3} B_{uj}^2.
\end{aligned}$$

The order of magnitude for each term in X_D and X_N is given in the following lemma.

Lemma C.3.

$$\begin{aligned}
L_{u1} &= \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{xx}(\omega_j) f_{uu}^*(\omega_j)^{-2} V_{uj} = O_p(M^{-1/2}), \\
b_{u1} &= \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{xx}(\omega_j) f_{uu}(\omega_j)^{-2} B_{uj} = O(M^{-2}m^{1/2}), \\
b_{u2} &= \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{xx}(\omega_j) f_{uu}(\omega_j)^{-3} B_{uj}^2 = O(M^{-4}m^{1/2}), \\
b_{u3} &= \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{xx}(\omega_j) f_{uu}^*(\omega_j)^{-3} E V_{uj}^2 = O(m^{-1/2}), \\
R_{u3} &= \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{xx}(\omega_j) f_{uu}^*(\omega_j)^{-3} [V_{uj}^2 - E V_{uj}^2] = O_p(m^{-1/2}M^{-1/2}), \\
L_{x1} &= \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}(\omega_j)^{-1} V_{xj} = O_p(M^{-1/2}), \\
L_{xu} &= \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}^*(\omega_j)^{-2} V_{xj} V_{uj} = O_p(m^{-1/2}M^{-1/2}),
\end{aligned}$$

$$L_{x2} = \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}(\omega_j)^{-2} B_{uj} V_{xj} = O_p(M^{-5/2}),$$

$$b_{x1} = \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}(\omega_j)^{-1} B_{xj} = O(m^{1/2} M^{-2}),$$

$$b_{xu} = \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}(\omega_j)^{-2} B_{xj} B_{uj} = O(M^{-4} m^{1/2}),$$

$$L_{u2} = \frac{\sqrt{m}}{2M} \sum_{j=-M+1}^M f_{uu}^*(\omega_j)^{-2} B_{xj} V_{uj} = O_p(M^{-5/2}),$$

$$X_{N0} = \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) f_{uu}(\omega_j)^{-1} = O_p(1),$$

$$Q_{N1}^H = \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) f_{uu}^*(\omega_j)^{-2} V_{uj} = O_p(m^{-1/2}),$$

$$C_{u1} = \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) f_{uu}^*(\omega_j)^{-3} V_{uj}^2 = O_p(m^{-1}),$$

$$L_{N1}^H = \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) f_{uu}(\omega_j)^{-2} B_{uj} = O_p(M^{-2}),$$

$$L_{N2}^H = \sqrt{\frac{m}{2M}} \sum_{j=-M+1}^M \hat{f}_{xu}(\omega_j) f_{uu}(\omega_j)^{-3} B_{uj}^2 = O_p(M^{-4}).$$

Proof. Similar to that of Lemma A.2. \square

Thus

$$\begin{aligned} X_D &= (b_{x1} - b_{u1}) + (L_{x1} - L_{u1}) + b_{u2} + b_{u3} \\ &\quad - b_{xu} + R_{u3} - L_{xu} - L_{x2} - L_{u2} \end{aligned} \quad (C.3)$$

$$X_N = X_{N0} - Q_{N1}^H + C_{u1} - L_{N1}^H + L_{N2}^H \quad (C.4)$$

C.3. Expansion for the moments $E\hat{f}_{xu}(\omega)^2$

The square of $\sqrt{T}(\hat{\beta} - \beta)$ is approximated by the moment of the truncated expansion

$$\left[\left(\Sigma_T^{-1} - \frac{\Sigma_T^{-2} X_D}{\sqrt{m}} + \frac{\Sigma_T^{-3} X_D^2}{m} \right) X_N \right],$$

and the calculation of EW_{N0}^2 involves $E\hat{f}_{xu}(\omega_j)^2$. Notice that

$$E\hat{f}_{xu}(\omega_j)^2 \sim \frac{1}{m^2} \sum_{\lambda_s \in B_j} K(\lambda_s - \omega_j)^2 EI_{xx}(\lambda_s) EI_{uu}(\lambda_s).$$

We decompose $EI_{xx}(\lambda_s)$ for $\lambda_s \in B_j$ as follows:

$$EI_{xx}(\lambda_s) = f_{xx}(\omega_j) + [f_{xx}(\lambda_s) - f_{xx}(\omega_j)] + [EI_{xx}(\lambda_s) - f_{xx}(\lambda_s)].$$

Under very general conditions (e.g. Assumption A.1), the last term, $[EI_{xx}(\lambda_s) - f_{xx}(\lambda_s)]$, is of smaller order of magnitude, $o_p(m^{-1})$, than the bias term $[f_{xx}(\lambda_s) - f_{xx}(\omega_j)]$, which represents the bias due to smoothing. Thus,

$$\begin{aligned} & \frac{1}{m} \sum_{\lambda_s} K(\lambda_s - \omega_j)^2 EI_{xx}(\lambda_s) EI_{uu}(\lambda_s) \\ &= \frac{1}{m} \sum_{\lambda_s} K(\lambda_s - \omega_j)^2 f_{xx}(\omega_j) f_{uu}(\omega_j) \\ & \quad + \frac{1}{m} \sum_{\lambda_s} K(\lambda_s - \omega_j)^2 f_{xx}(\omega_j) [f_{uu}(\lambda_s) - f_{uu}(\omega_j)] \\ & \quad + \frac{1}{m} \sum_{\lambda_s} K(\lambda_s - \omega_j)^2 f_{uu}(\omega_j) [f_{xx}(\lambda_s) - f_{xx}(\omega_j)] \\ & \quad + \frac{1}{m} \sum_{\lambda_s} K(\lambda_s - \omega_j)^2 [f_{xx}(\lambda_s) - f_{xx}(\omega_j)] [f_{uu}(\lambda_s) - f_{uu}(\omega_j)] \\ & \quad + o(m^{-1}). \end{aligned}$$

For the second term, notice that, under our assumptions,

$$f_{uu}(\lambda_s) - f_{uu}(\omega_j) = f_{uu}'(\omega_j)(\lambda_s - \omega_j) + \frac{1}{2} f_{uu}''(\omega_j)(\lambda_s - \omega_j)^2 + o_p((\lambda_s - \omega_j)^2).$$

Thus

$$\begin{aligned} & \frac{1}{m} \sum_{\lambda_s} K(\lambda_s - \omega_j)^2 [f_{uu}(\lambda_s) - f_{uu}(\omega_j)] \\ & \sim f_{uu}''(\omega_j) \frac{1}{2m} \sum_{\lambda_s} K(\lambda_s - \omega_j)^2 (\lambda_s - \omega_j)^2. \end{aligned}$$

Let $\lambda_s - \omega_j = \theta_p/2M$, then $\theta_p = -\pi + 2\pi p/m$, $p = 1, \dots, m$, and $K(\lambda_s - \omega_j) = K(\theta_p/2M) = \bar{K}(\theta_p) = (1/2M) \sum_{h=-T/2+1}^{T/2} k(h/M) e^{-i(h/2M)\theta_p}$. Thus,

$$\begin{aligned} & \frac{1}{m} \sum_{\lambda_s} K(\lambda_s - \omega_j)^2 f_{xx}(\omega_j) [f_{uu}(\lambda_s) - f_{uu}(\omega_j)] \\ & \sim \frac{1}{16M^2\pi} \frac{2\pi}{m} \sum_{p=1}^m \bar{K}(\theta_p)^2 \theta_p^2 f_{uu}''(\omega_j) f_{xx}(\omega_j) \\ & \sim \frac{1}{16M^2\pi} \int_{-\pi}^{\pi} \bar{K}(\theta)^2 \theta^2 d\theta f_{xx}(\omega_j) f_{uu}''(\omega_j). \end{aligned}$$

Similarly,

$$\begin{aligned} & \frac{1}{m} \sum_{\lambda_s} K(\lambda_s - \omega_j)^2 f_{uu}(\omega_j) [f_{xx}(\lambda_s) - f_{xx}(\omega_j)] \\ & \sim \frac{1}{16M^2\pi} \int_{-\pi}^{\pi} \bar{K}(\theta)^2 \theta^2 d\theta f_{uu}(\omega_j) f_{xx}'(\omega_j), \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{m} \sum_{\lambda_s} K(\lambda_s - \omega_j)^2 [f_{xx}(\lambda_s) - f_{xx}(\omega_j)] [f_{uu}(\lambda_s) - f_{uu}(\omega_j)] \\ & \sim \frac{1}{8M^2\pi} \int_{-\pi}^{\pi} \bar{K}(\theta)^2 \theta^2 d\theta f_{xx}(\omega_j) f_{uu}'(\omega_j) \\ & + \frac{1}{128M^4\pi} \int_{-\pi}^{\pi} \bar{K}(\theta)^2 \theta^4 d\theta f_{xx}''(\omega_j) f_{uu}''(\omega_j). \end{aligned}$$

Thus

$$\begin{aligned} E\hat{f}_{xu}(\omega_j)^2 & \sim \frac{1}{m} [df_{xx}(\omega_j) f_{uu}(\omega_j) \\ & + \frac{1}{16M^2\pi} \int_{-\pi}^{\pi} \bar{K}(\theta)^2 \theta^2 d\theta f_{uu}(\omega_j) f_{xx}'(\omega_j) \\ & + \frac{1}{16M^2\pi} \int_{-\pi}^{\pi} \bar{K}(\theta)^2 \theta^2 d\theta f_{uu}''(\omega_j) f_{xx}(\omega_j) \\ & + \frac{1}{8M^2\pi} \int_{-\pi}^{\pi} \bar{K}(\theta)^2 \theta^2 d\theta f_{xx}'(\omega_j) f_{uu}'(\omega_j)]. \end{aligned}$$

C.4. Second-order expansion of $\sqrt{T}(\hat{\beta}_H - \beta)$

Plugging Eqs. (C.3) and (C.4) in expression (6.5), and dropping those terms that are $o_p(m^{-1})$ and $o_p(M^{-2})$, we get,

$$\begin{aligned} \sqrt{T}(\hat{\beta}_H - \beta) & = \Sigma_T^{-1} \left\{ X_{N0} - Q_{N1}^H - \left[L_{N1}^H - \frac{1}{\sqrt{m}} (b_{u1} - b_{x1}) \Sigma_T^{-1} X_{N0} \right] \right. \\ & \quad \left. + C_{u1} + \frac{1}{\sqrt{m}} (L_{u1} - L_{x1}) \Sigma_T^{-1} X_{N0} - \frac{1}{\sqrt{m}} b_{u3} \Sigma_T^{-1} X_{N0} \right\} \\ & \quad + \text{higher-order terms} \\ & \sim \Sigma_T^{-1} X_{N0} + \frac{1}{\sqrt{m}} A_1 + \frac{1}{M^2} B_1, \end{aligned} \tag{C.5}$$

where $A_1 = \Sigma_T^{-1} Q_{N1}^H$, and $B_1 = \Sigma_T^{-1} [L_{N1}^H - (1/\sqrt{m})(b_{u1} - b_{x1})\Sigma_T^{-1} X_{N0}]$. The following lemma gives the correlations among the terms in Eq. (C.22).

Lemma C.4.

$$\begin{aligned} EX_{N0}^2 &= a\Omega + \frac{1}{M^2}\mathcal{V}_3, \\ E(Q_{N1}^H)^2 &= \frac{a^2}{2m}\Omega, \\ EX_{N0}Q_{N1}^H &= 0, \\ E\frac{1}{\sqrt{m}}b_{u3}\Sigma_T^{-1}X_{N0}^2 &= EX_{N0}C_{u1} = \frac{a^2}{2m}\Omega, \\ E\Sigma_T^{-2}(L_{x1} - L_{u1})X_{N0}^2 &= 0, \\ EX_{N0}\left[L_{N1}^H + \frac{1}{\sqrt{m}}(b_{x1} - b_{u1})\Sigma_T^{-1}X_{N0}\right] &= -\frac{ak_2}{M^2}\mathcal{V}_4. \end{aligned}$$

Proof. Omitted. \square

Proof of Theorem 3

$$\begin{aligned} \text{MSE}(\hat{\beta}) &= Ea^{-1}\Omega_T^{-1} \\ &\times \left\{ X_{N0}^2 + Q_{N1}^2 - 2X_{N0}Q_{N1} - 2X_{N0}\left[L_{N1} - \frac{1}{\sqrt{m}}(b_{u1} - b_{x1})\Sigma_T^{-1}X_{N0}\right] \right. \\ &\quad \left. + 2X_{N0}\left[C_{u1} - \frac{1}{\sqrt{m}}b_{u3}\Sigma_T^{-1}X_{N0}\right] + \frac{2}{\sqrt{m}}(L_{u1} - L_{x1})\Sigma_T^{-1}X_{N0}^2 \right\} + o_p(\cdot) \\ &= 1 + \frac{a}{2m} + \frac{1}{M^2}\left[\frac{\mathcal{V}_1}{a} + 2k_2\mathcal{V}_2\right]\Omega^{-1}. \quad \square \end{aligned}$$

References

- Brillinger, D.R., 1980. *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston, New York.
- Corbae, D., Ouliaris, S., Phillips, P.C.B., 1994. A reexamination of the consumption function using frequency domain regressions. *Empirical Economics* 19, 595–609.
- Engle, R., 1974. Band spectrum regression. *International Economic Review* 15 (1), 1–11.
- Hannan, E.J., 1963a. Regression for time series. In: Rosenblatt, M. (Ed.), *Time Series Analysis*. Wiley, New York.
- Hannan, E.J., 1963b. Regression for time series with errors of measurement. *Biometrika* 50, 293–302.
- Hannan, E.J., 1970. *Multiple Time Series*. Wiley, New York.

- Linton, O.B., 1995a. Second order approximation in the partially linear regression model. *Econometrica* 63, 1079–1112.
- Linton, O.B., 1995b. Second order approximation in a linear regression with heteroskedasticity of unknown form. *Econometric Reviews* 15, 1–32.
- Phillips, P.C.B., 1977. A general theorem in the theory of asymptotic expansion as approximations to the finite sample distributions of econometric Estimators. *Econometrica* 45, 463–485.
- Priestley, M.B., 1981. *Spectral Analysis and Time Series*. Academic Press, New York.
- Robinson, P.M., 1987. Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 56, 875–891.
- Robinson, P.M., 1991. Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica* 59, 755–786.
- Rothenberg, T.J., 1984a. Approximating the distributions of econometric estimators and test statistics. In Griliches, Z., Intriligator, M. (Eds.), *Handbook of Econometrics*, Chap. 4, vol. 2, North-Holland, New York.
- Rothenberg, T.J., 1984b. Approximate normality of generalized least squares estimators. *Econometrica* 52, 811–825.
- Sargan, J.D., 1976. Econometric estimators and the Edgeworth expansion. *Econometrica* 44, 421–448.
- Sargan, J.D., Mikhail, W.M., 1971. A general approximation to the distribution of instrumental variables estimates. *Econometrica* 39, 131–169.
- Xiao, Z., 1997, Efficiency Issues in stationary and nonstationary time series regression, Ph. D. Thesis, Department of Economics, Yale University.