Nonparametric Kernel Estimation of Evolutionary Autoregressive Processes

Woocheol Kim *

This paper develops a new econometric tool for evolutionary autoregressive models, where the AR coefficients change smoothly over time. To estimate the unknown functional form of time-varying coefficients, we propose a modified local linear smoother. The asymptotic normality and variance of the new estimator are derived by extending the Phillips and Solo device to the case of evolutionary linear processes. As an application for statistical inference, we show how Wald tests for stationarity and misspecification could be formulated based on the finite-dimensional distributions of kernel estimates. We also examine the finite sample performance of the method *via* numerical simulations.

Keywords: Autoregressive models, Evolutionary linear processes, Local linear fits, Locally stationary processes, Phillips and Solo device, Time-varying coefficients

JEL Classification: C14

I. Introduction

Stationarity has been a fundamental assumption in time series analysis. In a stationary system, the statistical properties of the process do not change over time, which is desirable if the data measure deviates from what is believed to be a steady-state equilibrium. However, the notion of stationarity is best considered to be a mathematical idealiza-

* Assistant Professor, Department of Taxation, University of Seoul, 163 Siripdae-ro, Dongdaemun-gu, Seoul 130-743, Korea, (Tel) +82-2-6490-6799, (Fax) +82-2-6490-6794, (E-mail) sunrise@uos.ac.kr. I would like to thank Rainer Dahlhaus, Oliver Linton, Michael Neuman, Peter Phillips, Donald Andrews, and Wolfgang Härdle for helpful discussions and comments.

tion, which is often too simple to capture the complicated dynamic structure of economic time series. The availability of a longer historical data series only serves to increase doubts about the realism of such restrictions. A more serious case occurs in practical applications when the period of interest tends to undergo frequent structural changes. For example, the long-term behavior of most economies tends to show what appears to be a slow but steady adjustment process, which cannot be properly analyzed by using the stationary approach. In this paper, we attempt to widen the empirical diversity of time series models by adopting a general class of evolutionary processes that can accommodate a variety of complicated forms of nonstationary behavior. Specifically, we extend the application of autoregressive (AR) models to a general nonstationary process by allowing the AR coefficients to change smoothly over time. An evolutionary AR(p) process, $\{y_t\}_{t=1}^{n}$ is defined to have the following data generation process (DGP):

$$y_t = \sum_{k=1}^{p} \alpha_k(t / n) y_{t-k} + \varepsilon_t, \tag{1}$$

where ε_t is i.i.d. $(0,\ \sigma_\varepsilon^2)$.

Unrestricted nonstationarity, however, may entail a large of arbitrariness in the time-dependent behavior of a process, thus making the development of a meaningful asymptotic theory impossible. When a process is evolutionary, increasing the number of observations over time does not necessarily imply an increase in information. In particular, one cannot expect an ensemble average to be consistently estimated by the corresponding temporal average.[1] To avoid pathological cases arising from extreme nonstationarity, we impose a number of restrictions on the process to control the extent of the deviations from stationarity. A natural approach for of doing so is to embed a stationary structure on the process in the vicinity of each time point. This idea is similar to the notion that underlies the nonparametric technique of fitting a line locally to a nonlinear curve. In this case, a smoothness condition on the curve is required to validate the approach. Likewise, in the present case, the imposition of local stationarity involves the use of a smoothness constraint on the evolution of the nonstationary processes. A rigorous definition of

[1] This breakdown might seem to be linked more directly to the violation of ergodicity rather than stationarity. However note that under stationarity, one still has convergence to ensemble averages conditional on the invariant algebra.

local stationarity was introduced by Dahlhaus (1996b), who imposed a smoothness condition in terms of the components in the spectral representation of the process. Heuristically, we can say that a process is locally stationary if the law of motion is smoothly time-varying. Thus, a locally stationary process behaves in a manner similar to a stationary process in the neighborhood of each instant in time, but has global nonstationary behavior. In example (1) above, the evolutionary AR model is locally stationary if the coefficients are smooth functions of time. Thus, as far as the local properties of this model are concerned, the statistical tools for stationarity can be used in deriving the asymptotics (see Section III).

The efforts to search for a framework for nonstationary processes have a long history in statistics and other applied sciences. In early empirical works, Granger and Hatanaka (1964) and Brillinger and Hatanaka (1969) advocated the spectral analysis of nonstationary processes in the frequency domain. Priestley *et al.* [(Cramér (1961), as well as Priestley (1965), Priestley and Tong (1973)] gave the first theoretical treatment of nonstationarity by defining time-dependent (or evolutionary) spectral density and estimating the spectral functions. The monograph by Priestley (1981) collected these main results. Since the early 1990s, the field has undergone a number of breakthroughs following a series of developments by Dahlhaus (1996a, 1996b), which provided a more rigorous definition and treatment of locally stationary processes. Under this framework, Neumann and Von Sachs (1997) applied wavelet methods for the adaptive estimation of evolutionary spectra.

The main contribution of this paper is the presentation of the nonparametric kernel estimation of time varying AR coefficients of an evolutionary process defined in (1). Dahlhaus (1997) takes a fully parametric approach and assumes specific functional forms for AR coefficients when constructing a local Whittle likelihood. In a practical sense, however, assuming that we have no prior information on the time dependency of the parameters is reasonable. Empirical economists often find the determination of evolution in the coefficients is itself of direct interest. Thus, the approach selected in this paper is to impose no functional restrictions on the coefficients and to estimate them as unknown functions of time by applying nonparametric kernel methods. The second contribution lies in the novelty of the statistical theory used in deriving the asymptotic properties for locally stationary processes. In Dahlhaus (1997), the asymptotic results are derived based on a somewhat complicated theory of evolutionary spectra. By contrast, in our approach, the structure of the local linear smoother makes the derivation of the limiting theory relatively easy. The intuition is that, in a limiting case, kernel methods enable us to be only concerned with local properties of locally stationary processes. Therefore, the well-established results for stationary processes can be utilized in deriving the asymptotics of the kernel estimates. To demonstrate the validity of this argument, the Phillips-Solo device (1992) is extended to the case of generalized linear representations of locally stationary processes and is used intensively as a standard machinery.

The remainder of this paper is as follows: Section II defines the local linear smoother for estimating the AR coefficients. In Section III, an asymptotic theory is derived for the time-varying coefficient estimators, and tests for stationarity and misspecification are suggested based on finite-dimensional distributions of these estimates. Section IV reports results from numerical simulations. Technical conditions and proofs are collected in Section V.

II. Kernel Estimation

Throughout this paper, we will use the following notation to represent coefficients as functions of a rescaled time index, that is, $\alpha_{k,t,n}=\alpha_k(t/n)$ with $\alpha(\cdot):[0,\ 1]\rightarrow\mathbb{R}$. To estimate $\alpha(\cdot)\equiv(\alpha_1(\cdot),\ \ldots,\ \alpha_p(\cdot))^T$, we apply the nonparametric method of local linear smoothing. If $\alpha_k(\cdot)$ is differentiable at u, $\alpha_k(u)$ can be approximated locally by

$$\alpha_k(t/n)\simeq\alpha_k(u)+\alpha'_k(u)(t/n-u).$$

Let $K_h(\cdot)=(1/h)K(\cdot/h)$ be a nonnegative weight function on a compact support. Given the observations $\{y_t\}_{t=1}^{n+p}$, we define the kernel-weighted least squares estimator of $\alpha_k(u)$'s and their first derivatives, $\alpha'(u)$'s, as

$$\begin{aligned}&\{\hat{\alpha}_k(u),\ \hat{\alpha}'_k(u)\}_{k=1}^{p}\\&=\arg\min_{a_{k0},a_{k1}}\sum_{t=p+1}^{n+p}\left\{y_t-\sum_{k-1}^{p}\left[a_{k0}+a_{k1}\left(\frac{t}{n}-u\right)\right]y_{t-k}\right\}^2K_h\left(\frac{t}{n}-u\right).\end{aligned}\tag{2}$$

Minimizing (2) with regard to the a_{k0}'s and a_{k1}'s yield $\hat{\alpha}(u)$ of the form

$$\begin{aligned}\hat{\alpha}(u)&\equiv[\hat{\alpha}_1(u),\ \ldots,\ \hat{\alpha}_p(u),\ \hat{\alpha}'_1(u),\ \ldots,\ \hat{\alpha}'_p(u)]^T\\&=(Z^TWZ)^{-1}(Z^TW_y),\end{aligned}\tag{3}$$

where

$$y=(y_{p+1}, \ldots, y_{n+p})^T,$$
$$Y_{t-1}=(y_{t-1}, \ldots, y_{t-p})^T, \; Y=(Y_p, \ldots, Y_{n-1})^T$$
$$Z=[I_n, D_n]Y \text{ with } D_n=diag[(1/n-u), \ldots, (n/n-u)],$$
$$W=diag[K_h(1/n-u), \ldots, K_h(n/n-u)].$$

The first p-elements of $\hat{\alpha}(u)$ are an estimate for the level of time-varying coefficients, and the remaining elements for their first derivatives. The latter property can be regarded as a unique benefit from local polynomial regression. By concentrating on the level of $\alpha(\cdot)$, not on its derivatives, we denote the estimates of $\alpha(u)$ by

$$\hat{\alpha}(u)=[\hat{\alpha}_1(u), \ldots, \hat{\alpha}_p(u)]^T=E_0(Z^T WZ)^{-1}(Z^T W_y), \tag{4}$$

where $E_0=[I_p, O_{p\times p}]$. If we rewrite Equation (4) in terms of sample moments, the estimator is understood exactly the same way as the weighted least squares estimator in a linear model. D_h be a $(2p\times 2p)$ diagonal matrix, the first p diagonal elements of which are one with other diagonal elements being h. Observe that

$$\hat{\alpha}(u)=E_0 D_h [(ZD_h)^T WZD_h]^{-1}[(ZD_h)^T W_y]=E_0 S_n^{-1} t_n, \tag{5}$$

where S_n is a $2p\times 2p$ matrix $[S_{n(i+j-2)}(u)]_{i,j=1,2}$, and $t_n=[t_{n0}(u), t_{n1}(u)]^T$, with

$$S_{nl}(u)=\frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(\frac{t}{n}-u\right)\left[\frac{1}{h}\left(\frac{t}{n}-u\right)\right]^l Y_{t-1}Y_{t-1}^T, \text{ for } l=0,1,2,$$
$$t_{nl}(u)=\frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(\frac{t}{n}-u\right)\left[\frac{1}{h}\left(\frac{t}{n}-u\right)\right]^l Y_{t-1}y_t, \text{ for } l=0,1.$$

Here, the estimation errors, $\hat{\alpha}(r)-\alpha(r)$, are not as simple as those associated with the usual least squares framework, given that the coefficients, $\alpha_{t,n}$, depend on the time index, t. The kernel estimate is subject to some bias as in the standard nonparametric method. The following lemma verifies this argument by decomposing the estimation error from the modified local linear fit into two parts: the bias term and the leading stochastic term.

Lemma 1. (Decomposition of Estimation Errors) Under E.1,

$$\hat{\alpha}(u)-\alpha(u)=B_n+\tilde{t}_n+o_p(h^2),\ \text{for } u \text{ in } (0,\ 1), \tag{6}$$

where

$$B_n = \frac{h^2}{2}E_0S_n^{-1}[S_{n2}, S_{n3}]^T\alpha''(u),$$
$$\tilde{t}_n = E_0S_n^{-1}\tilde{\tau}_n,$$
$$\tilde{\tau}_n = [\tilde{\tau}_{0n}, \tilde{\tau}_{1n}]^T,$$
$$\tilde{\tau}_{nl} = \frac{1}{n}\sum_{t=p+1}^{n+p}K_h\left(\frac{t}{n}-u\right)\left[\frac{1}{h}\left(\frac{t}{n}-u\right)\right]^l Y_{t-1}\varepsilon_t.$$

III. Statistical Results

The asymptotic properties of our estimator, $\hat{\alpha}_0(\cdot)$, are derived by generalizing the device of Phillips and Solo (1992) to the case of evolutionary linear processes. In the Appendix, we first show that the locally stationary AR process in (1) is a special case of evolutionary process and then develop the second-order Beveridge-Nelson (BN) decomposition for the sample moments of S_{ni} and $\tilde{\tau}_{nk}$ in (6). Let a function $\phi_k : [0,\ 1] \mapsto R$ be defined as $\phi_k(u) = \lim_{n\to\infty}\phi_k([nu]/n)$ with $\phi_k(t/n) \equiv \Sigma_{j=0}^{\infty}\varphi_{tj}\,\varphi_{(t+k)(k+j)}$. Also, let $\Gamma(u)$ be a symmetric $p\times p$ matrix with the h-th off-diagonal elements being $[\phi_h(u),\ \ldots,\ \phi_h(u)]_{1\times(p-h)}$, for $h=1,\ \ldots,\ p-1$, and the diagonal, $[\phi_0(u),\ \ldots,\ \phi_0(u)]_{1\times p}$. The results in the following lemmas give the probability limits of S_{ni} and the bias term, as well as the asymptotic distribution of the stochastic term $\tilde{\tau}_n$.

Lemma 2. Assume that E.1 through E.3 and A.2 hold. If $h\to 0$ and $nh^2 \to\infty$, then,

$$S_{nl} \xrightarrow{p} \left(\sigma_\varepsilon^2\int K(r)r^l dr\right)\Gamma(u),\ \text{for } l=0,1,2,3.$$

Lemma 3. Assume that all the conditions in Lemma 2 hold. Then,

$$B_n = \frac{h^2}{2}E_0S_n^{-1}[S_{n2}, S_{n3}]^T\alpha''(u) \xrightarrow{p} \frac{h^2}{2}\mu_K^2\alpha''(u).$$

We then have to derive asymptotic distribution of the main stochastic term, $E_0 S_n^{-1} \tilde{\tau}_n$. Given that $E_0 S_n^{-1}$ converges to $[\Gamma^{-1}(u), O_{p\times p}]$ by Lemma 2, we only have to deal with the first term of $\tilde{\tau}_n$.

Lemma 4. Assume that E.1 through E.3 and A.1 hold. If $h \to 0$ and $nh \to \infty$, then,

$$\sqrt{nh}\,\tilde{\tau}_{n0} \xrightarrow{D} N(0, \Sigma),$$

where $\Sigma = \sigma_\varepsilon^4 \left(\int K^2(r)dr\right)\Gamma(u)$.

Considering that $B_n = O_p(h^2)$ and $\tilde{\tau}_n = O_p(1/\sqrt{nh})$, the above results indicate that $\hat{\alpha}(u)$ is a consistent estimator when $h \to 0$ and $nh^2 \to \infty$. Notably, the asymptotic bias in Lemma 3 has the same form as the standard local linear fit. Lemma 3 and 4 yield the following theorem:

Theorem 5. Assume that E.1 through E.3 and A.1 through A.2 hold. If $h \to 0$ and $nh^2 \to \infty$, then,

$$\sqrt{nh}\,[\hat{\alpha}(u) - \alpha(u) - B_n] \xrightarrow{D} N(0, \Sigma_\alpha(u)),$$

where $\Sigma_\alpha(u) = \|K\|_2^2\, \Gamma^{-1}(u)$.

For a stationary AR(1) case, $\Gamma(u)$ is simplified to be $\Sigma_{j=0}^{\infty} \varphi_j^2 = \Sigma_{j=0}^{\infty} \alpha^{2j} = 1/(1-\alpha^2)$, which implies that $\Sigma_\alpha(u)$ of Theorem 5 can be interpreted as a nonparametric generalization of the asymptotic variance of ordinary least squares in a stationary AR model. Let $\hat{\varepsilon}_t = y_t - \Sigma_{k=1}^{p} \hat{\alpha}_k(t/n)\, y_{t-k}$ and $\hat{\sigma}_\varepsilon^2 = \Sigma_{t=p+1}^{n} \hat{\varepsilon}_t^2/(n-p)$. By Lemma 2, $\Gamma(u)$ is consistently estimated by

$$\hat{\Gamma}(u) \equiv S_{n0}(u)/\hat{\sigma}_\varepsilon^2 = \hat{\sigma}_\varepsilon^{-2}\, \frac{1}{n} \sum_{t=1}^{n} K_h\left(\frac{t}{n} - u\right) Y_t Y_t',$$

and $\Sigma_\alpha(u)$ by

$$\hat{\Sigma}_\alpha(u) \equiv \|K\|_2^2\, \hat{\Gamma}^{-1}(u) = \|K\|_2^2\, S_{n0}^{-1}(u)\, \hat{\sigma}_\varepsilon^2.$$

Considering that $\hat{\alpha}(u_1)$ and $\hat{\alpha}(u_2)$ are asymptotically uncorrelated for $u_1 \neq u_2$, their joint distribution is also asymptotically normal with a covariance of $\mathrm{diag}\{\Sigma_\alpha(u_1), \Sigma_\alpha(u_2)\}$. Thus, the normalized sum of squared errors over d time points follows a Chi-square distribution of degree dp.

Corollary 6. Assume that all the conditions in Theorem 5 hold. Then,

$$H_n = \sum_{i=1}^{d} nh\,[\hat{\alpha}_0(u_i) - \alpha(u_i) - B_n(u_i)]' \, \Sigma_{\alpha}^{-1}(u_i)[\hat{\alpha}_0(u_i) - \alpha(u_i) - B_n(u_i)] \overset{D}{\rightarrow} \chi^2(dp),$$

where $u_i \in [0, 1]$, for all $i=1, \ldots, d$.

Remark 7. (Tests for Misspecification and Stationarity) Corollary 6 is related to the construction of a Wald test for misspecification. Consider the null hypothesis of $H_0: \alpha(u_i) = \alpha^*(u_i)$ for all $i=1, \ldots, d$, against the general alternative. A feasible Wald test statistic is given by

$$\hat{H}_n = \sum_{i=1}^{d} nh\,[\hat{\alpha}_0(u_i) - \alpha^*(u_i) - \hat{B}_n]' \, \hat{\Sigma}_{\alpha}^{-1}[\hat{\alpha}_0(u_i) - \alpha^*(u_i) - \hat{B}_n], \tag{7}$$

and follows a $\chi^2(dp)$ asymptotically under the null hypothesis. Given that $|\alpha(u_i) - \alpha^*(u_i)| \neq 0$ under the alternative, $\hat{H}_n$ goes to infinity as $n \rightarrow \infty$, that is, the test is consistent. In a similar manner, we can set up a test for stationarity against general nonstationarity by assuming a null hypothesis, $H_0: \alpha(u_i) = \alpha^*$ for all $i=1, \ldots, d$. Considering that Corollary 6 still holds for a constant coefficient case, the average of coefficient estimates converges to the true value, α^*, at a faster rate than $\sqrt{nh}$ under H_0. The same effect can be achieved by applying least squares, the convergence rate of which is $\sqrt{n}$ under H_0. In this case, the test statistics is given by

$$\tilde{H}_n = \sum_{i=1}^{d} nh\,[\hat{\alpha}_0(u_i) - \bar{\hat{\alpha}} - \hat{B}_n]' \, \hat{\Sigma}_{\alpha}^{-1}[\hat{\alpha}_0(u_i) - \bar{\hat{\alpha}} - \hat{B}_n], \tag{8}$$

where $\bar{\hat{\alpha}} = 1/(n-p)\Sigma_{t=p+1}^{n} \hat{\alpha}_0(t/n)$, or $\bar{\hat{\alpha}} = (Y_t'Y_t)^{-1}Y_t'y_t$. $\tilde{H}_n$ weakly converges to $\chi^2(dp)$ under H_0, and the test is consistent, given that $|\hat{\alpha}_0(u_i) - \bar{\hat{\alpha}}| \overset{p}{\rightarrow} |\alpha(u_i) - (1/d)\,\Sigma_{j=1}^{d} \alpha(u_j)| \neq 0$ under H_A.

IV. Numerical Studies

Simulations. We perform a number of numerical simulations to investigate the finite sample performance of the kernel estimator defined in Section II. In the simulations, we used three different types of time-varying

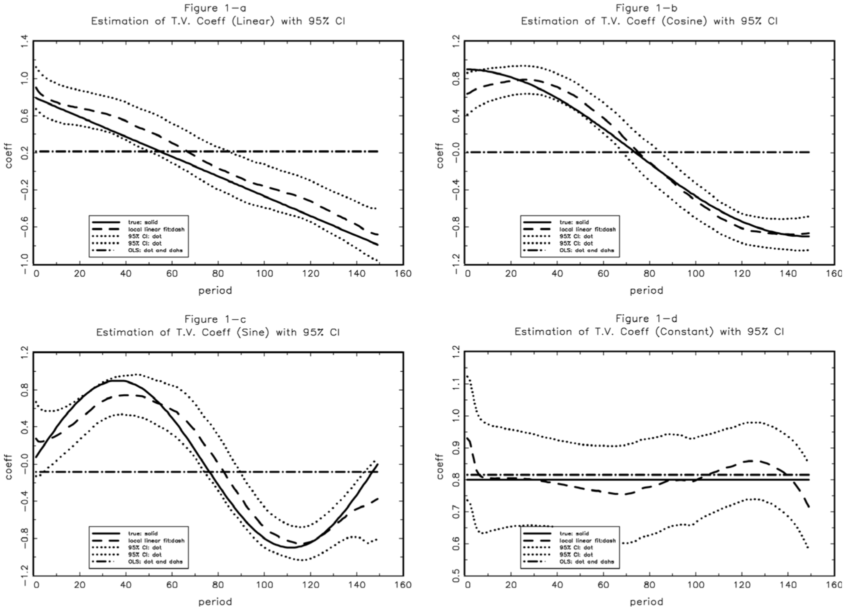

FIGURE 1
SIMULATION RESULTS

AR(1) models with $y_t = \alpha(t/n) y_{t-1} + 0.5\varepsilon_t$, $t = 1, \ldots, n$, where ε_t are i.i.d $N(0, 1)$ and

$$\text{Model I: } \alpha(r) = -1.6r + 0.8,$$
$$\text{Model II: } \alpha(r) = 0.9\cos(\pi r),$$
$$\text{Model III: } \alpha(r) = 0.9\sin(2\pi r).$$

For each model, we applied the local linear smoother to estimate the AR(1) coefficients and to report their basic statistical results. A set of simulated data with a sample size of $n = 150$ is generated from each model. We performed 2500 replications. For the kernel estimators, the Epanechnikov kernel function was used with a bandwidth, $h = b\sigma_n n^{-1/5}$, where σ_n is a standard deviation of $\{t/n\}_{t=1}^{n}$, and the constant b ranges from 1.4 to 2.5. Figure 1 shows the estimates for a typical sample along with asymptotic confidence intervals (CIs).

Considering the nonparametric nature of our smoothers, the estimators seem to work relatively well even in a sample as small as $n = 150$. Figure 1(c) indicates that the estimation of a sinusoidal trend in the coefficient involves more biases than others. The constant coefficients in Figure

TABLE 1
COVERAGE OF TRUE VALUES IN THE 95% CI (MODEL I)

	at 100 equidistant points in (0,1)	at 100 randomly-chosen points in (0,1)
Pr.	94.6%	94%

TABLE 2
AVERAGE MEAN SQUARED/ABSOLUTE ERRORS

Bandwidth	nh	AMSE	AMAE
		$p=1$	$p=1$
0.6	9.6	0.16	0.13
0.9	14.4	0.13	0.11
1.2	19.2	0.12	0.10
1.5	24.0	0.11	0.10
1.8	28.8	0.11	0.09
OLS		0.59	0.52

1(d) are efficiently estimated by the parametric least squares, but the nonparametric fits are close to the truth except at the boundaries. The asymptotic CIs cover the true functions at almost all points, but seem somewhat narrow, especially for the sinusoidal specification. This condition can be partly attributed to the disregarded biases in constructing confidence intervals. To verify with the asymptotic results of Theorem 5, we also compute the probability that the true coefficients are included in the 95% asymptotic CIs in the case of Model I. Table 1 shows that the real coverage rate is close to the value suggested by theoretical asymptotic distributions. In Table 2, we summarize the average mean squared errors of kernel estimates for various bandwidth choices when the true DGP is Model II.

V. Conditions and Proofs

A. Section II

Conditions:

E.1. The function $\{\alpha_k(\cdot)\}_{k=1}^{p}$ is twice continuously differentiable u with uniformly bounded second-order derivatives, and the roots of $\Sigma_{k=1}^{p}$ $\alpha_k(u)\ z^j$ are uniformly bounded away from the unit circle.

E.2. The kernel $K(\cdot)$ is a continuous symmetric nonnegative function on a compact support, satisfying $\sup_r |K(r)|^p = \| K \|^p_\infty < \infty$.

E.3. $\int K(r)\, dr = 1$, $\mu_K^2 = \int K(r)\, r^2 dr < \infty$, $\int K^2(r)\, dr = \| K \|_2^2 < \infty$, and $\int K^2(r)\, r^2 dr < \infty$.

Proof of Lemma 1. From the basic equations: with $E_1 \equiv [O_{p\times p},\ I_p]$

$$E_0(Z^T WZ)^{-1}(Z^T WZ)\ E_0^T = I_p,\ ZE_0^T = Y,$$
$$E_0(Z^T WZ)^{-1}(Z^T WZ)\ E_1^T = O_{p\times p},\ ZE_1^T = D_n Y,$$

it follows that

$$\alpha(u) = E_0(Z^T WZ)^{-1}(Z^T WZ)\ E_0^T \alpha(u) = E_0(Z^T WZ)^{-1} Z^T WY\alpha(u),$$

and

$$0 = E_0(Z^T WZ)^{-1}(Z^T WZ)\ E_1^T \alpha'(u) = E_0(Z^T WZ)^{-1} Z^T WD_n Y\alpha'(u).$$

The estimation error is then

$$\begin{aligned}\hat{\alpha}(u) - \alpha(u) &= E_0(Z^T WZ)^{-1}(Z^T Wy) - E_0(Z^T WZ)^{-1} Z^T WY\alpha(u)\\ &= E_0(Z^T WZ)^{-1} Z^T W[y - Y\alpha(u)] = E_0(Z^T WZ)^{-1} Z^T W[y - Y\alpha(u) - \\ &\quad D_n Y\alpha'(u)]\\ &= E_0 D_h[(ZD_h)^T WZD_h]^{-1}(ZD_h)^T W[y - Y\alpha(u) - D_n Y\alpha'(u)]\\ &= E_0[(ZD_h)^T WZD_h]^{-1}(ZD_h)^T W[y - Y\alpha(u) - D_n Y\alpha'(u)].\end{aligned}$$

Using the definition, $[b_\lambda]_{\lambda=0,1} \equiv [b_1,\ b_2]^T$, we rewrite the numerator of $\hat{\alpha}(u) - \alpha(u)$ as

$$\begin{aligned}&(ZD_h)^T W[y - Y\alpha(u) - D_n Y\alpha'(u)]\\ &= \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(\frac{t}{n} - u\right)\left[\frac{1}{h^\lambda}\left(\frac{t}{n} - u\right)^\lambda\right]_{\lambda=0,1} Y_{t-1}\left\{y_t - Y_{t-1}^T\alpha(u) - \left(\frac{t}{n} - u\right)Y_{t-1}^T\alpha'(u)\right\}\\ &= \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(\frac{t}{n} - u\right)\left[\frac{1}{h^\lambda}\left(\frac{t}{n} - u\right)^\lambda\right]_{\lambda=0,1} Y_{t-1}Y_{t-1}^T\left\{\alpha\left(\frac{t}{n}\right) - \alpha(u) - \left(\frac{t}{n} - u\right)\alpha'(u)\right\}\\ &+ \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(\frac{t}{n} - u\right)\left[\frac{1}{h^\lambda}\left(\frac{t}{n} - u\right)^\lambda\right]_{\lambda=0,1} Y_{t-1}\varepsilon_t.\end{aligned}$$

Considering the Taylor expansion of $\alpha(t/n)$ around u, the first term is approximated by

$$\frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(\frac{t}{n}-u\right)\left[\frac{1}{h^{\lambda}}\left(\frac{t}{n}-u\right)^{\lambda}\right]_{\lambda=2,3} Y_{t-1}Y_{t-1}^{T}\left[\frac{h^2}{2}\alpha''(u)\right],$$

and the estimation error is thus decomposed into two parts:

$$\hat{\alpha}(u)-\alpha(u)$$

$$= E_0[(ZD_h)^T WZD_h]^{-1}\frac{1}{n}\sum_{t=1}^{n} K_h\left(\frac{t}{n}-r\right)\left[\frac{1}{h^{\lambda}}\left(\frac{t}{n}-r\right)^{\lambda}\right]_{\lambda=2,3} Y_{t-1}Y_{t-1}^{T}\left[\frac{h^2}{2}\alpha''(u)\right]$$

$$+E_0[(ZD_h)^T WZD_h]^{-1}\frac{1}{n}\sum_{t=1}^{n} K_h\left(\frac{t}{n}-r\right)\left[\frac{1}{h^{\lambda}}\left(\frac{t}{n}-u\right)^{\lambda}\right]_{\lambda=0,1} Y_{t-1}\varepsilon_t + o_p(h^2).$$

B. Evolutionary Linear Processes and BN Decompositions

When the roots of $\Sigma_{k=1}^{p}\alpha_k(u)\,z^j$ are uniformly bounded away from the unit circle, it follows under the conditions on the bounded derivatives for $\alpha_k(\cdot)$ and $\sigma(\cdot)$ (see Mélard 1985) that the difference equations in (1) have a solution of the form

$$y_{t,n}=\sum_{j=0}^{\infty}\varphi_{j,t,n}\varepsilon_{t-j},$$

where

$$\sum_{j=0}^{\infty}|\varphi_{j,t,n}|<\infty, \text{ uniformly in } t \text{ and } n.$$

Lemma P.1. If $\alpha_k(\cdot)$'s are continuous and differentiable in u with a uniformly bounded derivative, then, for $\{y_{t,n}\}$ in (1), there exists a (unique) sequence of differentiable functions, $\{\varphi_j(\cdot)\,|\,\varphi_j:[0,1]\rightarrow R\}_{j=0}^{\infty}$, such that

$$\text{i) } \sup_t\left|y_{t,n}-\sum_{j=0}^{\infty}\varphi_j(t/n)\varepsilon_{t-j}\right|=O_p(1/n), \tag{9}$$

$$\text{ii) } \sup_t\sum_{j=0}^{\infty}|\varphi_j(t/n)|<\infty.$$

Proof of Lemma P.1. Let

$$A(u,\lambda) \equiv \frac{\sigma_\varepsilon}{\sqrt{2\pi}}\left[1-\sum_{k=0}^{p}\alpha_k(u)\exp(-i\lambda k)\right]^{-1} \text{ and } f(u,\lambda) = |A(u,\lambda)|^2 .$$

Observing that for a given u, $f(u, \lambda)$ is the spectral density function of a stationary AR(p) process, we define $\{\varphi_j(\cdot)\}_{j=0}^{\infty}$ to be a moving-average (MA) coefficient given by the MA representation of the AR process. Then, from the stability condition, ii) is satisfied, and, by construction, it holds that $(\sigma_\varepsilon/\sqrt{2\pi})\Sigma_{j=0}^{\infty}\varphi_j(u)\exp(-i\lambda j)=A(u, \lambda)$ for all u. The smoothness of $\varphi_j(\cdot)$ stems from the differentiability of $\{\alpha_k(\cdot)\}$. To show i), consider a spectral representation of (1),

$$y_{t,n} = \frac{\sigma_\varepsilon}{\sqrt{2\pi}}\int_{-\pi}^{\pi}\exp(i\lambda t)A_{t,n}^{0}(\lambda)dZ_X(\lambda),$$

where $A_{t,n}^{0}(\lambda)\equiv(\sigma_\varepsilon/\sqrt{2\pi})\Sigma_{j=0}^{\infty}\varphi_{j,t,n}(t/n)\exp(-i\lambda j)$. Given that $\{y_{t,n}\}$ in (1) is locally stationary with a time-varying spectral density of $f(u, \lambda)$ by Dahlhaus (1996b, Theorem 2.3), it follows that, for some constant K_1,

$$\sup_{t,\lambda}\left|A_{t,n}^{0}(\lambda)-A\left(\frac{t}{n},\lambda\right)\right| \le K_1 n^{-1}, \text{ for all } n,$$

which implies

$$\begin{aligned}\sup_t\left|y_{t,n}-\sum_{j=0}^{\infty}\varphi_j(t/n)\varepsilon_{t-j}\right| &= \sup_t \frac{\sigma_\varepsilon}{\sqrt{2\pi}}\left|\int_{-\pi}^{\pi}\exp(i\lambda t)[A_{t,n}^{0}(\lambda)-A(\frac{t}{n},\lambda)]dZ_X(\lambda)\right|\\ &\le K_2\sup_{t,\lambda}\left|A_{t,n}^{0}(\lambda)-A(\frac{t}{n},\lambda)\right|\\ &\le K_3 n^{-1}, \text{ for all } n,\end{aligned}$$

where $Z_X(\lambda)$ is a stochastic process of orthogonal increments on $[-\pi, \pi]$ with $\overline{Z_X(\lambda)}=Z_X(-\lambda)$.

In a simple AR(1) case, $\varphi_{j,t,n}$ is equal to $\Pi_{k=0}^{j}\alpha[(t-k)/n]$, but $\varphi_j(t/n)=\alpha(t/n)^j$. The above lemma suggests that $\Sigma_{j=0}^{\infty}[\varphi_{j,t,n}-\varphi_j(t/n)]\varepsilon_{t-j}=0$ does not hold in a finite sample, but it does asymptotically.

The approximate MA representation in Lemma P.1 now enables us to apply the Phillips-Solo device of the second-order BN decompositions to the sample moments of S_{nl} and $\tilde{\tau}_{nl}$ in (6). Recall that a function, $\phi_h : [0, 1] \mapsto R$ is such that

$$\phi_{h,j}(t/n,\ (t+h)/n) \equiv \varphi_j(t/n)\varphi_{j+h}((t+h)/n)$$

Conditions:

A.1. ε_t is i.i.d $(0,\ \sigma^2,\ \kappa_4)$, where κ_4 is a finite fourth cumulant.
A.2. (a) $\sup_{t\le n}\Sigma_{j=0}^{\infty} j^{1/2}\ \varphi_j^2(t/n) < \infty$, (b) $\sup_{t\le n}\Sigma_{j=0}^{\infty} j^{1/2}[\varphi_j'(t/n)]^2 = o(n^2)$.

Given that $\phi(\cdot)$ is defined on compact set, it is bounded and square integrable, $\int_0^1 \phi_h^2(r)dr < \infty$. The summability conditions in A.2(a) is, except for a number of generalizing modifications, of the same kind used in Phillips and Solo (1992) for the validity of the BN decomposition. A.2(b) is an additional condition required to restrict the changes in the time-varying coefficients. Notably, $\phi_h(\cdot)$ is continuously differentiable, that is, $\phi_h(\cdot) \in C^2$. We now show the validity of BN decomposition when applied to an evolutionary AR process. From Lemma P.1, it follows that

$$\begin{aligned} y_t y_{t+h} &\simeq \sum_{j=0}^{\infty} \varphi_j(t/n)\varepsilon_{t-j} \sum_{k=0}^{\infty} \varphi_k((t+h)/n)\varepsilon_{t+h-k} \\ &= \sum_{j=0}^{\infty} \varphi_j(t/n)\varphi_{j+h}((t+h)/n)\varepsilon_{t-j}^2 \\ &\quad + \sum_{j=0}^{\infty} \sum_{k=0, k\neq h+j}^{\infty} \varphi_j(t/n)\varphi_k((t+h)/n)\varepsilon_{t-j}\varepsilon_{t+h-k} \\ &= \sum_{j=0}^{\infty} \varphi_j(t/n)\varphi_{j+h}((t+h)/n)\varepsilon_{t-j}^2 \\ &\quad + \sum_{j=0}^{\infty} \sum_{r=-\infty, r\neq 0}^{\infty} \varphi_j(t/n)\varphi_{j+h+r}((t+h)/n)\varepsilon_{t-j}\,\varepsilon_{t-j-r}, \end{aligned}$$

where we assume that $\varphi_j(\cdot) = 0$ for all $s < 0$. Following the same argument by Phillips and Solo (1992), we consider the second-order BN decomposition as follows:

By defining

$$\phi_h(t/n,(t+h)/n;L)=\sum_{j=0}^{\infty}\varphi_j(t/n)\varphi_{(j+h)}((t+h)/n)L^j,$$

we obtain

$$y_t y_{t+h}=\phi_h(t/n,(t+h)/n;L)\varepsilon_t^2+\sum_{r=-\infty,r\neq 0}^{\infty}\phi_{h+r}(t/n,(t+h)/n;L)\varepsilon_t\varepsilon_{t-r}. \quad (11)$$

Observe that

$$\begin{aligned}\phi_{h+r}(t/n,(t+h)/n;L)&=\phi_{h+r}(t/n,(t+h)/n;1)-\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)(1-L)\\&=\phi_{h+r}(t/n,(t+h)/n;1)-(1-L)\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)\\&\quad+[\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)-\tilde{\phi}_{h+r}(t-1/n,(t+h-1)/n;L)]L,\end{aligned}$$

where

$$\begin{aligned}\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)&=\sum_{j=0}^{\infty}\tilde{\phi}_{h+r,j}(t/n,(t+h)/n)L^j\\&=\sum_{j=0}^{\infty}[\sum_{s=j+1}^{\infty}\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)]L^j.\end{aligned}$$

This condition implies the two-level BN decomposition:

$$\begin{aligned}&\phi_{h+r}(t/n,(t+h)/n;L)\varepsilon_t\varepsilon_{t-r}\\&=\phi_{h+r}(t/n,(t+h)/n;1)\varepsilon_t\varepsilon_{t-r}-(1-L)\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)\varepsilon_t\varepsilon_{t-r}\\&\quad+[\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)-\tilde{\phi}_{h+r}(t-1/n,(t+h-1)/n;L)]\varepsilon_{t-1}\varepsilon_{t-r-1}\\&=\phi_{h+r}(t/n,(t+h)/n;1)\varepsilon_t\varepsilon_{t-r}-(1-L)\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)\varepsilon_t\varepsilon_{t-r}+o_p(1)\end{aligned} \quad (12)$$

the validity of which depends on the condition:

(i) $\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)\varepsilon_t\varepsilon_{t-r}\in L^2$,
(ii) $[\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)-\tilde{\phi}_{h+r}(t-1/n,(t+h-1)/n;L)]\varepsilon_{t-1}\varepsilon_{t-r-1}=o_p(1)$.

To prove (i), we first consider

$$\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)\varepsilon_t\varepsilon_{t-r}=\sum_{j=0}^{\infty}[\sum_{s=j+1}^{\infty}\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)]\varepsilon_{t-j}\varepsilon_{t-r-j}.$$

Then, it suffices to show that

$$\sum_{j=0}^{\infty}[\sum_{s=j+1}^{\infty}\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)]^2$$
$$=\sum_{j=0}^{\infty}[\sum_{s=j+1}^{\infty}s^{1/4}\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)/s^{1/4}]^2$$
$$\leq\sum_{j=0}^{\infty}\left(\sum_{s=j+1}^{\infty}s^{1/2}\varphi_s^2(t/n)\right)\left(\sum_{s=j+1}^{\infty}\varphi_{s+h+r}^2((t+h)/n)/s^{1/2}\right)$$
$$\leq\sum_{s=1}^{\infty}s^{1/2}\varphi_s^2(t/n)\left(\sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty}\varphi_{s+h+r}^2((t+h)/n)/s^{1/2}\right)$$
$$=\sum_{s=1}^{\infty}s^{1/2}\varphi_s^2(t/n)\left(\sum_{s=1}^{\infty}\varphi_{s+h+r}^2((t+h)/n)/s^{1/2}\sum_{j=0}^{s-1}1\right)$$
$$=\sum_{s=1}^{\infty}s^{1/2}\varphi_s^2(t/n)\left(\sum_{s=1}^{\infty}\varphi_{s+h+r}^2((t+h)/n)s^{1/2}\right)$$
$$\leq\left(\sup_t\sum_{s=1}^{\infty}s^{1/2}\varphi_s^2(t/n)\right)^2<\infty$$

To prove (ii), noting that

$$[\tilde{\phi}_{h+r}(t/n,L)-\tilde{\phi}_{h+r}(t-1/n,L)]\varepsilon_{t-1}\varepsilon_{t-r-1}$$
$$=\sum_{j=0}^{\infty}\left\{\sum_{s=j+1}^{\infty}\Delta_t[\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)]\right\}\varepsilon_{t-1-j}\varepsilon_{t-r-1-j},$$

we only need to show that

$$\sum_{j=0}^{\infty}\left\{\sum_{s=j+1}^{\infty}\Delta_t[\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)]\right\}^2=o(1).$$

First, observe that

$$\Delta_t[\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)$$
$$=\Delta_t\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)+\varphi_s(t-1/n)\Delta_t\varphi_{s+h+r}((t+h)/n),$$

and then it holds that

$$\sum_{j=0}^{\infty}\left(\sum_{s=j+1}^{\infty}\Delta_t[\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)]\right)^2$$

$$\leq 2\left\{\sum_{j=0}^{\infty}\left(\sum_{s=j+1}^{\infty}\Delta_t\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)\right)^2\right.$$

$$\left.+\sum_{j=0}^{\infty}\left(\sum_{s=j+1}^{\infty}\varphi_s(t-1/n)\Delta_t\varphi_{s+h+r}((t+h)/n)\right)^2\right\}$$

$$\leq 2\left\{\sum_{s=1}^{\infty}s^{1/2}[\Delta_t\varphi_s(t/n)]^2\left(\sum_{s=1}^{\infty}\varphi_{s+h+r}^2((t+h)/n)s^{1/2}\right)\right.$$

$$\left.+\sum_{s=1}^{\infty}s^{1/2}\varphi_s^2(t-1/n)\left(\sum_{s=1}^{\infty}[\Delta_t\varphi_{s+h+r}((t+h)/n)]^2 s^{1/2}\right)\right\}$$

$$\leq 2\sup_t\sum_{s=1}^{\infty}s^{1/2}\varphi_s^2(t/n)\left\{\sum_{s=1}^{\infty}s^{1/2}[\Delta_t\varphi_s(t/n)]^2+\sum_{s=1}^{\infty}[\Delta_t\varphi_{s+h+r}((t+h)/n)]^2 s^{1/2}\right\}$$

$$\leq 4\left(\sup_t\sum_{s=1}^{\infty}s^{1/2}\varphi_s^2(t/n)\right)\left(\sup_t\sum_{s=1}^{\infty}s^{1/2}[\Delta_t\varphi_s(t/n)]^2\right)$$

$$\to 4\left(\sup_t\sum_{s=1}^{\infty}s^{1/2}\varphi_s^2(r)\right)\frac{1}{n^2}\left(\sup_t\sum_{s=1}^{\infty}s^{1/2}[\varphi_s'(r)]^2\right)$$

$$\to 0,$$

as n goes to ∞(with $t=[nr]$), given that

$$\left[\frac{\Delta_t\varphi_s(t/n)}{n}\right]^2\to[\varphi_s'(r)]^2$$

and $\sup_t\Sigma_{s=0}^{\infty}s^{1/2}[\varphi_s'(r)]^2=o(n^2)$.

Now, (11) and (12) imply

$$\begin{aligned}y_t y_{t+h} &= \phi_h(t/n,(t+h)/n;1)\varepsilon_t^2+\sum_{r=-\infty,r\neq 0}^{\infty}\phi_{h+r}(t/n,(t+h)/n;1)\varepsilon_t\varepsilon_{t-r}\\ &\quad-(1-L)\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)\varepsilon_t^2-(1-L)\\ &\quad\sum_{r=-\infty,r\neq 0}^{\infty}\tilde{\phi}_{h+r}(t/n,(t+h)/n;L)\varepsilon_t\varepsilon_{t-r}+o_p(1),\end{aligned}\tag{13}$$

Lemma P.2. (Validity of second-order BN decomposition) Under E.1, A.1, and A.2, the BN decomposition in (13) is valid, that is,

$$\sum_{j=0}^{\infty} \tilde{\phi}_{h+r,j}^2(t/n,(t+h)/n) = \sum_{j=0}^{\infty} [\sum_{s=j+1}^{\infty} \varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)]^2 < \infty$$

$$\sum_{j=0}^{\infty} \left\{ \sum_{s=j+1}^{\infty} \Delta_t[\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)] \right\}^2 = o(1).$$

C. Section III

Proof of Lemma 2. We only prove the case for the representative element of S_{nl},

$$S_{nl,d} = \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(\frac{t}{n}-u\right)\frac{1}{h^l}\left(\frac{t}{n}-u\right)^l y_{t-1}y_{t-1+d}$$

$$= \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l y_t y_{t+d}.$$

BN decomposition in Lemma P.1, when applied to $S_{nl,d}$, yields

$$S_{nl,d}=M_{1n}+M_{2n}+M_{3n},$$

where

$$M_{1n} = \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l \phi_d(t/n,(t+d)/n;1)\varepsilon_t^2,$$

$$M_{2n} = \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l \varepsilon_t\varepsilon_t^{\phi},$$

$$\varepsilon_t^{\phi} = \sum_{r=-\infty,r\neq 0}^{\infty} \phi_{d+r}(t/n,(t+d)/n;1)\varepsilon_{t-r}$$

$$M_{3n} = -M_{31n} - M_{32n},$$

$$M_{31n} = \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l (1-L)\tilde{\phi}_{d+r}(t/n,(t+d)/n;L)\varepsilon_t^2,$$

$$M_{32n} = \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l (1-L)$$

$$\sum_{r=-\infty,r\neq 0}^{\infty} \tilde{\phi}_{d+r}(t/n,(t+d)/n;L)\varepsilon_t\varepsilon_{t-r}.$$

(i) Considering that ε_t is i.i.d., the standard argument of Law of large numbers implies

$$M_{1n} \overset{p}{\to} \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l \phi_d(t/n,(t+d)/n;1)\sigma_\varepsilon^2$$

$$\to \sigma_\varepsilon^2\int_0^1 K_h(r-u)\left[\frac{1}{h}(r-u)\right]^l \phi_d(r,r;1)dr$$

$$= \sigma_\varepsilon^2\int_{-\infty}^{\infty} K(s)s^l\phi_h(u-hs,u-hs;1)ds$$

$$\to \sigma_\varepsilon^2\phi_h(u,u;1)\int_{-\infty}^{\infty} K(s)s^l ds,$$

where the last equation is given by Dominated Convergence Theorem.

(ii) Considering that $E(M_{2n})=0$ (from $E(\varepsilon_t\varepsilon_{t-r})=0,\ \forall r\neq 0$), we show $E(M_{2n}^2)\to 0$ for $M_{2n}=o_p(1)$. First, observe that

$$\sigma_{\phi,d}^2(t/n) \equiv E(\varepsilon_t^{\phi 2}) = \sigma_\varepsilon^2 \sum_{r=-\infty, r\neq 0}^{\infty} \phi_{d+r}^2(t/n,(t+d)/n;1)$$

$$= \sigma_\varepsilon^2 \sum_{r=0(r\neq d)}^{\infty}\left\{\sum_{j=0}^{\infty}\varphi_j(t/n)\varphi_{(j+r)}((t+d)/n)\right\}^2$$

$$< \infty,$$

by the same argument used in Lemma P.2 [The second equality is attributed to (10)]. From $E(\varepsilon_t\varepsilon_t^\phi\varepsilon_s\varepsilon_s^\phi)=0,\ \forall t\neq s$, it follows that

$$E(M_{2n}^2) = \frac{\sigma_\varepsilon^2}{nh}\left\{\frac{1}{n}\sum_{t=p}^{n+p-1}\frac{1}{h}K^2\left(\frac{t+1}{n}-u\right)\left[\frac{1}{h}\left(\frac{t+1}{n}-u\right)\right]^{2l}\sigma_{\phi,d}^2(t/n)\right\},$$

the negligibility of which is evident from

$$\frac{1}{n}\sum_{t=p}^{n+p-1}\frac{1}{h}K^2\left(\frac{t+1}{n}-u\right)\left[\frac{1}{h}\left(\frac{t+1}{n}-u\right)\right]^{2l}\sigma_{\phi,d}^2(t/n)$$

$$\to \sigma_{\phi,d}^2(u)\int K^2(s)s^{2l}ds < \infty.$$

(iii) For the negligibility of M_{3n}, we only show $M_{31n}=o_p(1)$. The same argument is valid to show $M_{32n}=o_p(1)$. Observe that

$$
\begin{aligned}
M_{31n} &= \frac{1}{n}\sum_{t=p}^{n+p-1}\{K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l \tilde{\phi}_{d+r}(t/n,(t+d)/n;L)\varepsilon_t^2 \\
&\quad -K_h\left(\frac{t}{n}-u\right)\frac{1}{h^l}\left(\frac{t}{n}-u\right)^l \tilde{\phi}_{d+r}(t-1/n,(t+d-1)/n;L)\varepsilon_{t-1}^2\} \\
&\quad +\frac{1}{n}\sum_{t=p}^{n+p-1}\{K_h\left(\frac{t}{n}-u\right)\frac{1}{h^i}\left(\frac{t}{n}-u\right)^l - K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l\} \\
&\quad \times\tilde{\phi}_{d+r}(t-1/n,(t+d-1)/n;L)\varepsilon_{t-1}^2 \\
&\equiv M'_{31n}+M''_{31n},
\end{aligned}
$$

respectively. The telescoping sum M'_{31n} becomes

$$
\begin{aligned}
&\frac{1}{nh}\left[K\left(\frac{n+p-nu}{nh}\right)\left(\frac{n+p-nu}{nh}\right)^l \tilde{\phi}_{d+r}(n+p/n,(n+p+d)/n;L)\varepsilon_{n+p}^2\right] \\
&-\frac{1}{nh}\left[K\left(\frac{p-nu}{nh}\right)\left(\frac{p-nu}{nh}\right)^l \tilde{\phi}_{d+r}(p-1/n,(p-1+d)/n;L)\varepsilon_{p-1}^2\right].
\end{aligned}
$$

Both terms in the above equation are negligible, $o_p(1)$, because $\tilde{\phi}_{d+r}(t/n, (t+d)/n; L)\varepsilon_t^2 = O_p(1)$ by Lemma P.2, and $K(\cdot)$ is compactly supported and bounded by E.2.

Next, for the negligibility of M''_{31n}, we apply the Taylor expansion on $K^*(s) \equiv K(s)\, s^i$,

$$
K^*\left(s+\frac{1}{nh}\right) = K^*(s) + \frac{K^{*\prime}(s)}{nh} + O\left(\frac{1}{n^2h^2}\right),
$$

and obtain

$$
\begin{aligned}
&\left|K_h\left(\frac{t+1}{n}-u\right)\left[\frac{1}{h}\left(\frac{t+1}{n}-u\right)\right]^l - K_h\left(\frac{t}{n}-u\right)\left[\frac{1}{h}\left(\frac{t}{n}-u\right)\right]^l\right| \\
&= \frac{1}{h}\left|K^*\left(\frac{t-nu}{nh}\right) - K^*\left(\frac{t-nu}{nh}-\frac{1}{nh}\right)\right| \\
&= \frac{1}{nh^2}K^{*\prime}\left(\frac{t-nu}{nh}\right) + O\left(\frac{1}{n^2h^3}\right) = o(1), \text{ for all } t,
\end{aligned}
$$

under the assumption that $nh^2 \to \infty$. Now,

$$M''_{31n} \le \frac{1}{h} \sup_t \left| K\left[\frac{1}{h}\left(\frac{t+1}{n} - u\right)\right]\left[\frac{1}{h}\left(\frac{t+1}{n} - u\right)\right]^l - K\left[\frac{1}{h}\left(\frac{t}{n} - u\right)\right]\left[\frac{1}{h}\left(\frac{t}{n} - u\right)\right]^l \right| \times |\tilde{\phi}_{d+r}(t-1/n, (t+d-1)/n; L)]\varepsilon_{t-1}^2| = o_p(1),$$

given that $\tilde{\phi}_{d+r}(t/n, (t+d)/n; L)\varepsilon_t^2 = O_p(1)$.

Proof of Lemma 3. By Lemma P.2 and E.3, it holds that

$$S_n \xrightarrow{p} \sigma_\varepsilon^2 \begin{bmatrix} \Gamma(u) & O_{p\times p} \\ O_{p\times p} & \mu_K^2 \Gamma(u) \end{bmatrix},$$

and

$$S_n^{-1} \xrightarrow{p} \sigma_\varepsilon^{-2} \begin{bmatrix} \Gamma^{-1}(u) & O_{p\times p} \\ O_{p\times p} & \mu_K^{-2} \Gamma^{-1}(u) \end{bmatrix}.$$

By the continuous mapping theorem, the bias term,

$$h^{-2}B_n = \frac{1}{2} E_0 S_n^{-1} \begin{bmatrix} S_{n2}^T \\ S_{n3}^T \end{bmatrix} \alpha''(u)$$

$$\xrightarrow{p} \frac{\alpha''(u)}{2\sigma_\varepsilon^2} [I_p, O_{p\times p}] \begin{bmatrix} \Gamma^{-1}(u) & O_{p\times p} \\ O_{p\times p} & \mu_K^{-2} \Gamma^{-1}(u) \end{bmatrix} \begin{bmatrix} \sigma_\varepsilon^2 \mu_K^2 \Gamma(u) \\ O_{p\times p} \end{bmatrix}$$

$$= \frac{\alpha''(u)}{2} \mu_K^2 I_p = \frac{\mu_K^2}{2} \alpha''(u).$$

Let $\mathcal{F}_t$ be the natural filtration of $\{y_t\}_{t=1}^n$.

Lemma P.3. (Central limit theorem for martingale differences: Corollary 6) Let, for every $n>0$, the sequence $\eta^n = (\eta_{nk}, F_k)$ be a square integrable martingale difference, that is,

$$E(\eta_{nk} | \mathcal{F}_{k-1}) = 0,\ E(\eta_{nk}^2) < \infty,\ 1 \le k \le n \tag{14}$$

and let

$$\sum_{k=1}^{n} E(\eta_{nk}^2) = 1,\ \forall n \ge n_0 > 0. \tag{15}$$

The conditions

$$\sum_{k=1}^{n} E(\eta_{nk}^2 \mid \mathcal{F}_{k-1}) \xrightarrow{p} 1,\ \text{as}\ n \to \infty, \tag{16}$$

$$\sum_{k=1}^{n} E(\eta_{nk}^2 I[\,|\eta_{nk}| > \varepsilon) \mid \mathcal{F}_{k-1}]) \xrightarrow{p} 0,\ \text{as}\ n \to \infty, \forall \varepsilon > 0, \tag{17}$$

are sufficient for convergence

$$\sum_{k=1}^{n} \eta_{nk} \xrightarrow{D} N(0,1),\ \text{as}\ n \to \infty.$$

Proof of Lemma 4. Considering the use of the Cramer-Wold device, it suffices to show

$$\sqrt{nh}\, a^T \tilde{\tau}_{n0} \xrightarrow{D} N(0, a^T \Sigma a),$$

as $n \to \infty$, for any vector $a \in \mathbb{R}^p$ with unit Euclidean norm, $\| a \|^2 = 1$. Fix such a vector $a \in \mathbb{R}^p$. Now that $E(Y_{t-1} Y_{t-1}^T \varepsilon_t^2) = E(Y_{t-1} Y_{t-1}^T E(\varepsilon_t^2 | \mathcal{F}_{t-1})) = \sigma_\varepsilon^4 \Gamma\{(t-1)/n\} < \infty$, we define

$$V_n(u) \equiv Var(\sqrt{nh}\, a^T \tilde{\tau}_{n0}) = \frac{1}{nh} \sum_{t=p+1}^{n+p} K^2\left(\frac{1}{h}\left(\frac{t}{n} - u\right)\right) \sigma_\varepsilon^4 a^T \Gamma\left(\frac{t-1}{n}\right) a.$$

Denote the normalized errors by

$$\eta_t \equiv V_n^{-1/2}(u) \frac{1}{\sqrt{nh}} K\left(\frac{1}{h}\left(\frac{t}{n} - u\right)\right) a^T Y_{t-1} \varepsilon_t.$$

In the following, we will check with each condition of Lemma P.3 for the asymptotic normality of η_t. The first part of (14) is evident from $E(y_{t-1}\ \varepsilon_t |$

$\mathcal{F}_{t-1})=0$, by A.1. Also,

$$E(\eta_t^2) = V_n^{-1}(u)\,\frac{1}{nh}\,K^2\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)\sigma_\varepsilon^4 a^T\Gamma\left(\frac{t-1}{n}\right)a$$
$$<\infty, \quad \text{for } 1\le t\le n,$$

which implies (14). (15) follows immediately from the way we construct η_{nt} and $E(\eta_{nt}^2)<\infty$, for $1\le t\le n$.

Next, to examine the condition (16), note that

$$\sum_{t=1}^{n} E(\eta_{nt}^2 \mid \mathcal{F}_{k-1}) = V_n^{-1}(u)\,\frac{1}{nh}\sum_{t=p+1}^{n+p} K^2\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)\sigma_\varepsilon^2 a^T Y_{t-1}Y_{t-1}^T a$$
$$= V_n^{-1}(u)a^T\tilde{V}_n(u)a,$$

where

$$\tilde{V}_n(u) = \sigma_\varepsilon^2\,\frac{1}{nh}\sum_{t=p+1}^{n+p} K^2\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)Y_{t-1}Y_{t-1}^T.$$

Applying the results from Lemma 2, we obtain the convergence of $\tilde{V}_n(u)$,

$$\tilde{V}_n(u) \xrightarrow{p} \sigma_\varepsilon^4\left(\int K^2(r)dr\right)\Gamma(u).$$

Also, by using integration by substitution and the Dominated Convergence Theorem,

$$V_n(u) = \frac{1}{nh}\sum_{t=p+1}^{n+p} K^2\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)\sigma_\varepsilon^4 a^T\Gamma\left(\frac{t-1}{n}\right)a^T$$
$$\to \sigma_\varepsilon^4\left(\int K^2(r)dr\right)a^T\Gamma(u)a^T,$$

which implies (16).

Finally, we turn to show (17). Given that $V_n(u)\to a^T\Sigma a>0$, there exists n_0, such that $V_n(u)>(1/2)a^T\Sigma a$, for all $n>n_0$. If we assume $n>n_0$, we obtain

$$\eta_t^2 = V_n^{-1}(u)\,\frac{1}{nh}\,K^2\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)a^T Y_{t-1}Y_{t-1}^T a^T\varepsilon_t^2$$

$$\leq \frac{2}{V_n(u)} \frac{\|K\|_\infty}{nh} K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right) a^T Y_{t-1} Y_{t-1}^T a \varepsilon_t^2$$

$$\leq \frac{2}{V_n(u)} \frac{\|K\|_\infty}{nh} K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right) \|a\|^2 \|Y_{t-1}\varepsilon_t\|^2$$

$$\equiv \kappa_1 \frac{1}{nh} K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right) \|Y_{t-1}\varepsilon_t\|^2$$

where we used the facts that $K(\cdot)$ is bounded and compactly supported and $\|\alpha\|^2=1$. The last inequality relies on the Cauchy-Schwartz inequality. Considering

$$E[\eta_{nt}^2 I(|\eta_{nt}| \geq \delta) \mid \mathcal{F}_{t-1}]$$

$$\leq \kappa_1 \frac{1}{nh} K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right) \|Y_{t-1}\|^2 E[\varepsilon_t^2 I(\|Y_{t-1}\varepsilon_t\| \geq \delta\kappa_1^{-1/2}\sqrt{nh}\,\|K\|_\infty^{-1/2}) \mid \mathcal{F}_{k-1}]$$

$$\leq \kappa_1 \frac{1}{nh} K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right) \|Y_{t-1}\|^2 E[\varepsilon_t^2 I(|\varepsilon_t| \geq \delta^{1/2}\kappa_1^{-1/4}\|K\|_\infty^{-1/4}\sqrt[4]{nh}) \mid \mathcal{F}_{k-1}]$$

$$+ \kappa_1 \frac{1}{nh} K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right) \|Y_{t-1}\|^2 E[\varepsilon_t^2 I(\|Y_{t-1}\| \geq \delta^{1/2}\kappa_1^{-1/4}\|K\|_\infty^{-1/4}\sqrt[4]{nh}) \mid \mathcal{F}_{k-1}],$$

and

$$\sum_t^n E[\eta_{nt}^2 I(|\eta_{nt}| \geq \delta) \mid \mathcal{F}_{t-1}] \leq I_{1n} + I_{2n},$$

$$I_{1n} = \kappa_1 \frac{1}{nh} \sum_t^n K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right) \|Y_{t-1}\|^2 E[\varepsilon_t^2 I(|\varepsilon_t| \geq \delta^{1/2}\kappa_1^{-1/4}\|K\|_\infty^{-1/4}\sqrt[4]{nh}) \mid \mathcal{F}_{k-1}]$$

$$I_{2n} = \kappa_1 \frac{1}{nh} \sum_t^n K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right) \sigma_\varepsilon^2 \|Y_{t-1}\|^2 I(\|Y_{t-1}\| \geq \delta^{1/2}\kappa_1^{-1/4}\|K\|_\infty^{-1/4}\sqrt[4]{nh}).$$

Note that (i) because ε_t is i.i.d. with $E(\varepsilon_t^2)<\infty$,

$$E[\varepsilon_t^2 I(|\varepsilon_t| \geq \delta^{1/2}\kappa_1^{-1/4}\|K\|_\infty^{-1/4}\sqrt[4]{nh}) \mid \mathcal{F}_{t-1}] = o(1),$$

where $o(1)$ does not depend on t, and (ii) by Lemma 2,

$$\frac{1}{nh}\sum_{t}^{n} K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)\left\|Y_{t-1}\right\|^2 \xrightarrow{p} \sigma_\varepsilon^2\, tr(\Gamma(u)),$$

which yields

$$I_{1n}=o_p(1).$$

Considering that $I_{n2}\geq 0$ for all n, and given that $E(\| Y_t \|^2)<\infty$,

$$E(I_{n2})\simeq \kappa_1\sigma_\varepsilon^2\, E[\left\|Y_{[nu]}\right\|^2\ I(\left\|Y_{[nu]}\right\| \geq \delta^{1/2}\kappa_1^{-1/4}\ \left\|K\right\|_\infty^{-1/4}\ \sqrt[4]{nh}))$$
$$\to 0.$$

This condition implies $I_{n2}=o_p(1)$, which completes the proof for

$$\sum_{t=p+1}^{n+p} \eta_{nt} \xrightarrow{D} N(0,1) \text{ as } n\to\infty,$$

that is,

$$\sqrt{nh}\,\tilde{\tau}_n \xrightarrow{D} N(0,\Sigma).$$

Proof of Theorem 5. Lemma 4, along with the result of Lemma 2, yields

$$\sqrt{nh}[\hat{\alpha}(u)-\alpha(u)-B_n] = \sqrt{nh}E_0S_n^{-1}\tilde{\tau}_n \xrightarrow{D} N(0,\Sigma_\alpha),$$

and

$$\Sigma_\alpha = [\sigma_\varepsilon^{-2}\Gamma^{-1}(u)\ O_{p\times p}]\sigma_\varepsilon^4\left\|K\right\|_2^2\,\Gamma(u)\,[\sigma_\varepsilon^{-2}\Gamma^{-1}(u)\ O_{p\times p}]^T = \left\|K\right\|_2^2\,\Gamma^{-1}(u).$$

(Received 13 October 2011; Revised 31 October 2012; Accepted 7 November 2012)

References

Brillinger, D., and Hatanaka, M. “An Harmonic Analysis of Nonstationary

Multibariate Economic Processes." *Econometrica* 37 (No. 1 1969): 131-41.

Cramér, H. On Some Classes of Non-stationary Stochastic Processes. Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 2, Los Angeles, CA: University of California Press, pp. 57-78, 1961.

Dahlhaus, R. "On the Kullback-Leibler Information Divergence of Locally Stationary Processes." *Stochastic Processes and Their Application* 62 (No. 1 1996a): 139-68.

________. "Maximum Likelihood Estimation and Model Selection for Locally Stationary Processes." *Journal of Nonparametric Statistics* 6 (Nos. 2-3 1996b): 171-91.

________. "Fitting Time Series Models to Nonstationary Processes." *Annals of Statistics* 25 (No. 1 1997): 1-37.

Granger, C., and Hatanaka, M. *Spectral Analyses of Economic Time Series.* Princeton, NJ: Princeton University Press, 1964.

Mélard, G. "An Example of the Evolutionary Spectrum Theory." *Journal of Time Series Analysis* 6 (No. 2 1985): 81-90.

Neumann, M. H., and Von Sachs, R. "Wavelet Thresholding in Anisotropic Function Classes and Application to Adaptive Estimation of Evolutionary Spectra." *Annals of Statistics* 25 (No. 1 1997): 38-76.

Phillips, P. C. B., and Solo, V. "Asymptotics for Linear Processes." *Annals of Statistics* 20 (No. 2 1992): 971-1001.

Priestley, M. B. "Evolutionary Spectra and Non-stationary Processes." *Journal of Royal Statistical Society. Series B* 27 (No. 2 1965): 204-37.

________. *Spectral Analysis and Time Series 2.* London: Academic Press, 1981.

Priestley, M. B., and Tong, H. "On the Analysis of Bivariate Non-stationary Processes (with Discussion)." *Journal of Royal Statistical Society. Series B* 35 (No. 2 1973): 153-66.