Nonparametric Kernel Estimation of Evolutionary Autoregressive Processes*

Woocheol Kim†

Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin

November 2001

Abstract

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 mdified local linear smoother. The asymptotic normality and variance of the new estimator are derived by extending 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 finite-dimensional distributions of the kernel estimates. We also examine the finite sample performance of the method via numerical simulations. As an empirical illustration, the method is applied to the real data of US stock returns.

Some key words: Autoregressive models; Evolutionary linear processes; Local linear fits; Locally-stationary processes; Phillips and Solo device; Time-varying coefficients.

---

*I would like to thank Rainer Dahlhaus, Oliver Linton, Michael Neuman, Peter Phillips, Donald Andrews, and Wolfgang Härdle for helpful discussions and comments.

†Institute for Statistics and Econometrics, Humboldt University of Berlin, Spandauer Str. 1, 10178, Berlin, Germany. E-mail address: `woocheol@wiwi.hu-berlin.de`

1 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 has some appeal if the data measure deviations from what is believed to be a steady-state equilibrium. However, the notion of stationarity is best considered to be a mathematical idealization which is often too simple to capture the complicated dynamic structure of economic time series. The availability of 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 experience 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 seek 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 AR models to a general nonstationary process by allowing the autoregressive coefficients to change smoothly over time. An evolutionary AR(p) process, $\{y_t\}_{t=1}^n$ is defined to have the following DGP

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

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

Unrestricted nonstationarity, however, may entail so much arbitrariness in the time-dependent behavior of a process that it is impossible to develop a meaningful asymptotic theory. 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 some restrictions on the process to control the extent of the deviations from stationarity. A natural way of doing so, is to embed a stationary structure on the process in the vicinity of each time point. This idea is similar to 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

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

definition of local stationarity was recently made by Dahlhaus (1996), 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 like a stationary process in the neighborhood of each instant in time, but has global nonstationary behavior. In the example (1) above, the evolutionary AR model is locally stationary, if the coefficients are smooth functions of time. It will be shown that, 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 3).

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 and his collaborators (Cramér(1961), Priestley(1965, 1966), Subba Rao and Tong (1972), 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 nineties, the field has undergone some breakthroughs following a series of recent developments by Dahlhaus (1996a, 1996b, 1996c), which provided a more rigorous definition and treatment of locally stationary processes. Under this framework, Neumann and Von Sachs (1997) applied wavelet methods for adaptive estimation of evolutionary spectra.

The main contribution of this paper is to present 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, it is realistic to assume that we have no prior information on the time dependency of the parameters. Often, to empirical economists, the finding of evolution in the coefficients is itself of direct interest. Thus, the approach chosen in the paper is to impose no functional restrictions on the coefficients and 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 local linear smoother makes the derivation of the limiting theory relatively easy. The intuition is that, in a limiting case, kernel methods allow 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 rest of the paper is as follows. Section 2 defines the local linear smoother for estimating the AR coefficients. In section 3, 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 4 reports some results from numerical simulations and empirical illustrations. Technical conditions and proofs are collected in Section 5.

2 Kernel Estimation

Throughout the paper, we will use the following notation to represent coefficients as functions of a rescaled time index, i.e., $\alpha_{k,t,n} = \alpha_k(t/n)$ with $\alpha(\cdot) : [0,1] \to \mathbb{R}$. To estimate $\alpha(\cdot) \equiv (\alpha_1(\cdot), .., \alpha_p(\cdot))^T$, 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 $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$ be a nonnegative weight function on a compact support. Given the observations $\{y_t\}_{t=1}^{n+p}$, define the kernel-weighted least squares estimator of $\alpha_k(u)$'s and their first derivatives, $\alpha'(u)$'s, as

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

Minimizing (2) w.r.t. the a_{k0}'s and a_{k1}'s gives $\widehat{\alpha}(u)$ of the form,

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}(u) & \equiv \left[\widehat{\alpha}_1(u), .., \widehat{\alpha}_p(u), \widehat{\alpha}_1'(u), .., \widehat{\alpha}_p'(u)\right]^T \\ & = \left(Z^T W Z\right)^{-1} \left(Z^T W y\right), \end{aligned} \tag{3}$$

where

$$y = (y_{p+1}, ..., y_n)^T,$$

$$
\begin{aligned}
Y_{t-1} &= (y_{t-1},..,y_{t-p})^T, Y = (Y_p,..,Y_{n-1})^T \\
Z &= [I_n, D_n]Y \text{ with } D_n = diag[(1/n-u),..,(n/n-u)], \\
W &= diag\left[K_h\left(1/n-u\right),..,K_h\left(n/n-u\right)\right].
\end{aligned}
$$

The first p-elements of $\widehat{\boldsymbol{\alpha}}(u)$ 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 its derivatives, we denote the estimates of $\alpha(u)$ by

$$
\widehat{\alpha}(u) = [\widehat{\alpha}_1(u),..,\widehat{\alpha}_p(u)]^T = E_0\left(Z^TWZ\right)^{-1}\left(Z^TWy\right), \tag{4}
$$

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

$$
\widehat{\alpha}(u) = E_0 D_h[(ZD_h)^T W Z D_h]^{-1}[(ZD_h)^T W y] = E_0 S_n^{-1} t_n, \tag{5}
$$

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

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

Here, the estimation errors, $\widehat{\alpha}(r)-\alpha(r)$ are not so simple as those associated with the usual least squares framework, since the coefficients, $\alpha_{t,n}$, depend on the time index, t. The kernel estimate are subject to some bias as in 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,*

$$
\widehat{\alpha}(u) - \alpha(u) = B_n + \widetilde{t}_n + o_p(h^2), \tag{6}
$$

where

$$
\begin{aligned}
B_n &= \frac{h^2}{2} E_0 S_n^{-1} \left[S_{n2},\ S_{n3}\right]^T \alpha''(u), \\
\widetilde{t}_n &= E_0 S_n^{-1} \widetilde{\tau}_n, \\
\widetilde{\tau}_n &= \left[\widetilde{\tau}_{0n},\ \widetilde{\tau}_{1n}\right]^T, \\
\widetilde{\tau}_{nl} &= \frac{1}{n} \sum_{t=p+1}^{n+p} K_h\left(\frac{t}{n} - u\right) \left[\frac{1}{h}\left(\frac{t}{n} - u\right)\right]^l Y_{t-1}\varepsilon_t.
\end{aligned}
$$

3 Statistical Results

The asymptotic properties of our estimator, $\widehat{\alpha}_0(\cdot)$ 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 decomposition for the sample moments of S_{ni} and $\widetilde{\tau}_{nk}$ in (6). Let a function $\phi_k : [0,1] \longmapsto R$ to be defined as $\phi_k(u) = \lim_{n\to\infty} \phi_k([nu]/n)$ with $\phi_k(t/n) \equiv \sum_{j=0}^{\infty} \varphi_{tj}\varphi_{(t+k)(k+j)}$. Also, let $\Gamma(u)$ be a symmetric $p \times p$ matrix with the h-th off-diagonal elements being $[\phi_h(u), .., \phi_h(u)]_{1\times(p-h)}$, for $h = 1, .., p-1$, and the diagonal, $[\phi_0(u), .., \phi_0(u)]_{1\times p}$. 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 $\widetilde{\tau}_n$.

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,*

$$
B_n = \frac{h^2}{2} E_0 S_n^{-1} \left[S_{n2},\ S_{n3}\right]^T \alpha''(u) \xrightarrow{p} \frac{h^2}{2}\mu_K^2 \alpha''(u).
$$

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

Lemma 4. *Assume that E.1 through E.3 and A.1 hold. If $h \to 0$ and $nh^2 \to \infty$, then,*

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

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

Since $B_n = O_p(h^2)$ and $\widetilde{\tau}_n = O_p(\sqrt{nh})$, the above results means that $\widehat{\alpha}(u)$ is a consistent estimator, when $h \to 0$ and $nh^2 \to \infty$. Note that the asymptotic bias in Lemma 3 has the same form as the standard local linear fit. Lemma 3 and 4 gives the following theorem.

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

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

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

For a stationary AR(1) case, $\Gamma(u)$ is simplified to be $\sum_{j=0}^{\infty} \varphi_j^2 = \sum_{j=0}^{\infty} \alpha^{2j} = 1/(1-\alpha^2)$, 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 $\widehat{\varepsilon}_t = y_t - \sum_{k=1}^{p} \widehat{\alpha}_k(t/n)\, y_{t-k}$ and $\widehat{\sigma}_\varepsilon^2 = \sum_{t=p+1}^{n} \widehat{\varepsilon}_t^2/(n-p)$. By Lemma 2, $\Gamma(u)$ is consistently estimated by

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

and $\Sigma_\alpha(u)$ by

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

Since $\widehat{\alpha}(u_1)$ and $\widehat{\alpha}(u_2)$ are asymptotically uncorrelated for $u_1 \neq u_2$, their joint distribution is also asymptotically normal with a covariance of $\operatorname{diag}\{\Sigma_\alpha(u_1), \Sigma_\alpha(u_2)\}$. 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,*

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

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

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 $\mathrm{H}_0 : \alpha(u_i) = \alpha^*(u_i)$ for all $i = 1, .., d$, against the general alternative. A feasible Wald test statistic is given by

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

and follows a $\chi^2(dp)$ asymptotically, under the null hypothesis. Since $|\alpha(u_i) - \alpha^*(u_i)| \neq 0$ under the alternative, $\widehat{H}_n$ goes to infinity as $n \to \infty$; i.e., the test is consistent. In a similar way, we can set up a test for stationarity against general nonstationarity, by assuming a null hypothesis, H_0 : $\alpha(u_i) = \alpha^*$ for all $i = 1, .., d$. Since 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 whose convergence rate is $\sqrt{n}$ under H_0. In this case, the test statistics is given by

$$\widetilde{H}_n = \sum_{i=1}^{d} nh\left[\widehat{\alpha}_0(u_i) - \overline{\widehat{\alpha}} - \widehat{B}_n\right]' \widehat{\Sigma}_\alpha^{-1}\left[\widehat{\alpha}_0(u_i) - \overline{\widehat{\alpha}} - \widehat{B}_n\right], \tag{8}$$

where $\overline{\widehat{\alpha}} = \frac{1}{n-p}\sum_{t=p+1}^{n}\widehat{\alpha}_0(t/n)$, or $\overline{\widehat{\alpha}} = \left(Y_t'Y_t\right)^{-1}Y_t'y_t$. $\widetilde{H}_n$ weakly converges to $\chi^2(dp)$ under H_0, and the test is consistent, since $\left|\widehat{\alpha}_0(u_i) - \overline{\widehat{\alpha}}\right| \xrightarrow{p} \left|\alpha(u_i) - \frac{1}{d}\sum_{j=1}^{d}\alpha(u_j)\right| \neq 0$, under H_A.

4 Numerical Studies

Simulations We carry out some numerical simulations to investigate the finite sample performance of the kernel estimator defined in Section 2. In the simulations, we used three different types of time-varying AR(1) models with $y_t = \alpha(t/n)\, y_{t-1} + 0.5\varepsilon_t, \quad t = 1, \ldots, n,$where ε_t are i.i.d $N(0,1)$ and

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

For each model, we applied the local linear smoother to estimate the AR(1) coefficients and report their basic statistical results. Two sets of simulated data with different sample sizes ($n = 150, 300$) are generated from each model. There were 2500 and 1000 replications with sample sizes of $n = 150$ and 300, respectively. For the kernel estimators, 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. Fig. 1 shows the estimates for a typical sample ($n = 150$) along with asymptotic confidence intervals.

Figure 1, here

Considering the nonparametric nature of our smoothers, the estimators seem to work relatively well even in a sample as small as $n = 150$. Fig.1(c) indicates that the estimation of a sinusoidal trend in the coefficient involves more biases than others. The constant coefficients in Fig. 1.(d) is efficiently estimated by the parametric least squares, but the nonparametric fits are close to the truth except at the boundaries. The asymptotic confidence intervals cover the true functions at almost all the points, but seem somewhat narrow, especially for the sinusoidal specification. This is partly due to the disregarded biases in constructing confidence intervals. To check with the asymptotic results of Theorem 5, we also compute the probability that the true coefficients are included in the 95% asymptotic confidence intervals 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.

Table 1.: Coverage of True values in the 95% CI (Model I)

	at the whole points	at the randomly-chosen points
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

Real Application To illustrate an empirical example, a time-varying AR(1) model is fitted to to the monthly returns of the S&P 500 stock index.(1926:1-1997:12). A Gaussian kernel was used with bandwidth $h = 0.1472$. Fig 2 depicts both the kernel and OLS estimates. The OLS estimate is less than 0.1 with standard error, 0.034 and statistically significant (p-value = 1.5%). If the stock market is efficient, the returns follows a martingale difference process. So, the OLS estimate does not seem to be compatible with the efficient market hypothesis. In contrast, our kernel estimates indicate that the hypothesis contradicts the real data only during the pre-world-war period ($t \leq 229$, in the graph). Interestingly, the coefficients converge to zero along with the time, and seem to be insignificant at least in the latter periods.

Figure 2, here

5 Conditions and Proofs

5.1 Section 2

Conditions:

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

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

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 $E_1 \equiv [O_{p\times p}, I_p]$

$$\begin{aligned}
E_0\left(Z^TWZ\right)^{-1}\left(Z^TWZ\right)E_0^T &= I_p,\ ZE_0^T = Y,\\
E_0\left(Z^TWZ\right)^{-1}\left(Z^TWZ\right)E_1^T &= O_{p\times p},\ ZE_1^T = D_nY,
\end{aligned}$$

it follows that

$$\alpha(u) = E_0\left(Z^TWZ\right)^{-1}\left(Z^TWZ\right)E_0^T\alpha(u) = E_0\left(Z^TWZ\right)^{-1}Z^TWY\alpha(u),$$

and

$$0 = E_0\left(Z^TWZ\right)^{-1}\left(Z^TWZ\right)E_1^T\alpha'(u) = E_0\left(Z^TWZ\right)^{-1}Z^TWD_nY\alpha'(u).$$

The estimation error is then

$$\begin{aligned}
\widehat{\alpha}(u) - \alpha(u) &= E_0\left(Z^TWZ\right)^{-1}\left(Z^TWy\right) - E_0\left(Z^TWZ\right)^{-1}Z^TWY\alpha(u)\\
&= E_0\left(Z^TWZ\right)^{-1}Z^TW[y - Y\alpha(u)] = E_0\left(Z^TWZ\right)^{-1}Z^TW\left[y - Y\alpha(u) - D_nY\alpha'(u)\right]\\
&= E_0D_h[(ZD_h)^TWZD_h]^{-1}(ZD_h)^TW\left[y - Y\alpha(u) - D_nY\alpha'(u)\right]\\
&= E_0[(ZD_h)^TWZD_h]^{-1}(ZD_h)^TW\left[y - Y\alpha(u) - D_nY\alpha'(u)\right].
\end{aligned}$$

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

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

Due to the Taylor expansion of $\alpha\left(\frac{t}{n}\right)$ 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 thus the estimation error is decomposed into two parts:

$$
\begin{aligned}
& \widehat{\alpha}(u)-\alpha(u) \\
= \; & E_0[(ZD_h)^T WZD_h]^{-1}\frac{1}{n}\sum_{t=1}^{n} K_h\left(\frac{t}{n}-r\right)\left[\frac{1}{h^\lambda}\left(\frac{t}{n}-r\right)^\lambda\right]_{\lambda=2,3} Y_{t-1}Y_{t-1}^T\left[\frac{h^2}{2}\alpha''(u)\right] \\
& +E_0[(ZD_h)^T WZD_h]^{-1}\frac{1}{n}\sum_{t=1}^{n} K_h\left(\frac{t}{n}-r\right)\left[\frac{1}{h^\lambda}\left(\frac{t}{n}-u\right)^\lambda\right]_{\lambda=0,1} Y_{t-1}\varepsilon_t + o_p(h^2).
\end{aligned}
$$

5.2 Evolutionary Linear Processes and BN-Decompositions

When the roots of $\sum_{k=1}^{p}\alpha_k(u)z^j$ are uniformly bounded away from unit circle, it follows under the conditions on the bounded derivatives for $\alpha_k(\cdot)$ and $\sigma(\cdot)$ (see Miller, 1969; Hallin, 1978, 1984; 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

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

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

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

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

Proof of Lemma P.1 Let

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

. 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 MA coefficients given by the MA representation of the AR process. Then, from the stability condition, ii) is satisfied, and, by construction, it holds that $\frac{\sigma_\varepsilon}{\sqrt{2\pi}}\sum_{j=0}^{\infty}\varphi_j(u)\exp(-i\lambda j) = A(u,\lambda)$, for all u. The smoothness of $\varphi_j(\cdot)$ comes 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 $A_{t,n}^{0}(\lambda) \equiv \frac{\sigma_\varepsilon}{\sqrt{2\pi}}\sum_{j=0}^{\infty}\varphi_{j,t,n}(t/n)\exp(-i\lambda j)$. Since $\{y_{t,n}\}$ in (1) is locally stationary with a time-varying spectral density of $f(u,\lambda)$-by Dahlhaus (1996, Theorem 2.3), it follows that, for some constant K_1,

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

which implies

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

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 $\prod_{k=0}^{j}\alpha\left[(t-k)/n\right]$, but $\varphi_j(t/n) = \alpha(t/n)^j$. The above lemma means that $\sum_{j=0}^{\infty}[\varphi_{j,t,n} - \varphi_j(t/n)]\varepsilon_{t-j} = 0$ does not hold in a finite sample, but it does asymptotically.

The approximate MA representation in Lemma P.1 now allows us to apply the Phillips-Solo device of the second order Beveridge-Nelson decompositions to the sample moments of S_{nl} and $\widetilde{\tau}_{nl}$ in (6). Recall that a function, $\phi_h : [0,1] \longmapsto R$ 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 finite fourth cumulant.*

A.2. *(a)* $\sup_{t \leq n} \sum_{j=0}^{\infty} j^{1/2} \varphi_j^2(t/n) < \infty$, *(b)* $\sup_{t \leq n} \sum_{j=0}^{\infty} j^{1/2} [\varphi_j'(t/n)]^2 = o(n^2)$.

Since $\phi(\cdot)$ 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.4.1 (a) is, except for some generalizing modifications, of the same kind used in Phillips and Solo (1992) for the validity of the Beveridge-Nelson decomposition. A.4.1 (b) is an additional condition required to restrict the changes in the time-varying coefficients. Note that $\phi_h(\cdot)$ is continuously differentiable, i.e., $\phi(\cdot) \in C^2$. We now show the validity of BN decomposition when applied to an evolutionary AR process. From Lemma P.1., it follows that

$$\begin{aligned}
y_t y_{t+h} &\simeq \sum_{j=0}^{\infty} \varphi_j(t/n)\varepsilon_{t-j} \sum_{k=0}^{\infty} \varphi_k((t+h)/n)\varepsilon_{t+h-k} \\
&= \sum_{j=0}^{\infty} \varphi_j(t/n)\varphi_{j+h}((t+h)/n)\varepsilon_{t-j}^2 \\
&\quad + \sum_{j=0}^{\infty} \sum_{k=0, k\neq h+j}^{\infty} \varphi_j(t/n)\varphi_k((t+h)/n)\varepsilon_{t-j}\varepsilon_{t+h-k} \\
&= \sum_{j=0}^{\infty} \varphi_j(t/n)\varphi_{j+h}((t+h)/n)\varepsilon_{t-j}^2 \\
&\quad + \sum_{j=0}^{\infty} \sum_{r=-\infty, r\neq 0}^{\infty} \varphi_j(t/n)\varphi_{j+h+r}((t+h)/n)\varepsilon_{t-j}\varepsilon_{t-j-r}, \qquad (10)
\end{aligned}$$

where it is assumed 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 get

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

Observe that

$$\phi_{h+r}(t/n, (t+h)/n; L) \;\; = \;\; \phi_{h+r}(t/n, (t+h)/n; 1) - \widetilde{\phi}_{h+r}(t/n, (t+h)/n; L)(1-L)$$

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

where

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

This implies the two level BN decomposition:

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

whose validity depends on the condition:

$$
\text{(i)} \quad \widetilde{\phi}_{h+r}(t/n,(t+h)/n;L)\varepsilon_t\varepsilon_{t-r} \in L^2,
$$

$$
\text{(ii)} \quad [\widetilde{\phi}_{h+r}(t/n,(t+h)/n;L) - \widetilde{\phi}_{h+r}(t-1/n,(t+h-1)/n;L)]\varepsilon_{t-1}\varepsilon_{t-r-1} = o_p(1).
$$

To prove (i), we first consider

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

Then, it suffices to show that

$$
\begin{aligned}
&\sum_{j=0}^{\infty}[\sum_{s=j+1}^{\infty} \varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)]^2 \\
= \; &\sum_{j=0}^{\infty}[\sum_{s=j+1}^{\infty} s^{1/4}\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)/s^{1/4}]^2 \\
\leq \; &\sum_{j=0}^{\infty}\left(\sum_{s=j+1}^{\infty} s^{1/2}\varphi_s^2(t/n)\right)\left(\sum_{s=j+1}^{\infty} \varphi_{s+h+r}^2((t+h)/n)/s^{1/2}\right) \\
\leq \; &\sum_{s=1}^{\infty} s^{1/2}\varphi_s^2(t/n)\left(\sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty} \varphi_{s+h+r}^2((t+h)/n)/s^{1/2}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{s=1}^{\infty} s^{1/2}\varphi_s^2(t/n)\left(\sum_{s=1}^{\infty}\varphi_{s+h+r}^2((t+h)/n)/s^{1/2}\sum_{j=0}^{s-1}1\right)\\
&= \sum_{s=1}^{\infty} s^{1/2}\varphi_s^2(t/n)\left(\sum_{s=1}^{\infty}\varphi_{s+h+r}^2((t+h)/n)s^{1/2}\right)\\
&\le \left(\sup_t \sum_{s=1}^{\infty} s^{1/2}\varphi_s^2(t/n)\right)^2 < \infty
\end{aligned}
$$

To prove ii), noting that

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

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

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

and then it holds that

$$
\begin{aligned}
&\sum_{j=0}^{\infty}\left(\sum_{s=j+1}^{\infty}\Delta_t[\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)]\right)^2\\
&\le 2\left\{\sum_{j=0}^{\infty}\left(\sum_{s=j+1}^{\infty}\Delta_t\varphi_s(t/n)\varphi_{s+h+r}((t+h)/n)\right)^2 + \sum_{j=0}^{\infty}\left(\sum_{s=j+1}^{\infty}\varphi_s(t-1/n)\Delta_t\varphi_{s+h+r}((t+h)/n)\right)^2\right\}\\
&\le 2\{\sum_{s=1}^{\infty}s^{1/2}[\Delta_t\varphi_s(t/n)]^2\left(\sum_{s=1}^{\infty}\varphi_{s+h+r}^2((t+h)/n)s^{1/2}\right)\\
&\quad + \sum_{s=1}^{\infty}s^{1/2}\varphi_s^2(t-1/n)\left(\sum_{s=1}^{\infty}[\Delta_t\varphi_{s+h+r}((t+h)/n)]^2 s^{1/2}\right)\}\\
&\le 2\sup_t\sum_{s=1}^{\infty}s^{1/2}\varphi_s^2(t/n)\left\{\sum_{s=1}^{\infty}s^{1/2}[\Delta_t\varphi_s(t/n)]^2 + \sum_{s=1}^{\infty}[\Delta_t\varphi_{s+h+r}((t+h)/n)]^2 s^{1/2}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \quad 4\left(\sup_t \sum_{s=1}^{\infty} s^{1/2}\varphi_s^2(t/n)\right)\left(\sup_t \sum_{s=1}^{\infty} s^{1/2}[\Delta_t\varphi_s(t/n)]^2\right) \\
&\to \quad 4\left(\sup_t \sum_{s=1}^{\infty} s^{1/2}\varphi_s^2(r)\right)\frac{1}{n^2}\left(\sup_t \sum_{s=1}^{\infty} s^{1/2}[\varphi_s'(r)]^2\right) \\
&\to \quad 0,
\end{aligned}
$$

as n goes to ∞ (with $t = [nr]$), since $[\frac{\Delta_t\varphi_s(t/n)}{n}]^2 \to [\varphi_s'(r)]^2$ and $\sup_t \sum_{s=1}^{\infty} s^{1/2}[\varphi_s'(r)]^2 = o(n^2)$.

Now, (11) and (12) imply

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

Lemma P.2 (The validity of second order BN decomposition) *Under E.1, A.1 and A.2, the BN decomposition in (13) is valid, i.e.,*

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

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

5.3 Section 3

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

$$
\begin{aligned}
S_{nl,d} \quad &= \quad \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(\frac{t}{n}-u\right)\frac{1}{h^l}\left(\frac{t}{n}-u\right)^l y_{t-1}y_{t-1+d} \\
&= \quad \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l y_t y_{t+d}.
\end{aligned}
$$

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

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

where

$$\begin{aligned}
M_{1n} &= \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l \phi_d(t/n,(t+d)/n;1)\varepsilon_t^2,\\
M_{2n} &= \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l \varepsilon_t\varepsilon_t^{\phi},\\
\varepsilon_t^{\phi} &= \sum_{r=-\infty,r\neq 0}^{\infty} \phi_{d+r}(t/n,(t+d)/n;1)\varepsilon_{t-r}\\
M_{3n} &= -M_{31n} - M_{32n},\\
M_{31n} &= \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l (1-L)\widetilde{\phi}_{d+r}(t/n,(t+d)/n;L)\varepsilon_t^2,\\
M_{32n} &= \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l (1-L)\sum_{r=-\infty,r\neq 0}^{\infty}\widetilde{\phi}_{d+r}(t/n,(t+d)/n;L)\varepsilon_t\varepsilon_{t-r}.
\end{aligned}$$

(i) Since ε_t is i.i.d., the standard argument of LLN implies

$$\begin{aligned}
M_{1n} &\xrightarrow{p} \frac{1}{n}\sum_{t=p}^{n+p-1} K_h\left(\frac{t+1}{n}-u\right)\frac{1}{h^l}\left(\frac{t+1}{n}-u\right)^l \phi_d(t/n,(t+d)/n;1)\sigma_\varepsilon^2\\
&\to \sigma_\varepsilon^2\int_0^1 K_h(r-u)\left[\frac{1}{h}(r-u)\right]^l \phi_d(r,r;1)dr\\
&= \sigma_\varepsilon^2\int_{-\infty}^{\infty} K(s)\, s^l\phi_h(u-hs,u-hs;1)ds\\
&\to \sigma_\varepsilon^2\phi_h(u,u;1)\int_{-\infty}^{\infty} K(s)\, s^l ds,
\end{aligned}$$

where the last equation is given by Dominated Convergence Theorem.

(ii) Since $E(M_{2n}) = 0$ (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

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

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

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

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

whose negligibility is obvious 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

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

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

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

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

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

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

and obtain

$$\begin{aligned}
&\left| K_h\left(\frac{t+1}{n} - u\right)\left[\frac{1}{h}\left(\frac{t+1}{n} - u\right)\right]^l - K_h\left(\frac{t}{n} - u\right)\left[\frac{1}{h}\left(\frac{t}{n} - u\right)\right]^l \right| \\
= \; & \frac{1}{h}\left| K^*\left(\frac{t - nu}{nh}\right) - K^*\left(\frac{t-nu}{nh} - \frac{1}{nh}\right)\right| \\
= \; & \frac{1}{nh^2} K^{*'}\left(\frac{t-nu}{nh}\right) + O(\frac{1}{n^2h^3}) = o(1), \text{ for all } t,
\end{aligned}$$

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

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

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

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,

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

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

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

Lemma P.3 (CLT for martingale differences: Lipster and Shirjaev, 1980, Corollary 6) *Let for every* $n > 0$, *the sequence* $\eta^n = (\eta_{nk,} F_k)$ *be a square integrable martingale difference, i.e.,*

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

and let

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

The conditions

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

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

are sufficient for convergence

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

Proof of Lemma 4 Due to the Cramer-Wold device, it suffices to show

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

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 $E(Y_{t-1}Y_{t-1}^T\varepsilon_t^2) = E(Y_{t-1}Y_{t-1}^T E(\varepsilon_t^2|\mathcal{F}_{t-1})) = \sigma_\varepsilon^4 \Gamma(\frac{t-1}{n}) < \infty$,, we define

$$V_n(u) \equiv Var\left(\sqrt{nh} a^T \widetilde{\tau}_{0n}\right) = \frac{1}{nh} \sum_{t=p+1}^{n+p} K^2\left(\frac{1}{h}\left(\frac{t}{n} - u\right)\right) \sigma_\varepsilon^4 a^T \Gamma(\frac{t-1}{n}) a.$$

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 asymptotic normality of η_t. The first part of (14) is obvious from $E(y_{t-1}\varepsilon_t|\mathcal{F}_{t-1}) = 0$, by A.1. Also,

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

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

$$\begin{aligned} \sum_{t=1}^{n} E(\eta_{nt}^2|\mathcal{F}_{k-1}) &= V_n^{-1}(u)\frac{1}{nh}\sum_{t=p+1}^{n+p} K^2\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)\sigma_\varepsilon^2 a^T Y_{t-1}Y_{t-1}^T a \\ &= V_n^{-1}(u)a^T\widetilde{V_n}(u)a, \end{aligned}$$

where $\widetilde{V_n}(u) = \sigma_\varepsilon^2\frac{1}{nh}\sum_{t=p+1}^{n+p} K^2\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)Y_{t-1}Y_{t-1}^T$. Applying the results from Lemma 2, we obtain the convergence of $\widetilde{V_n}(u)$,

$$\widetilde{V_n}(u) \xrightarrow{p} \sigma_\varepsilon^4(\int K^2(r)\,dr)\Gamma(u).$$

Also, using integration by substitution and Dominated Convergence Theorem,

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

which implies (16).

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

$$\begin{aligned} \eta_t^2 &= V_n^{-1}(u)\frac{1}{nh}K^2\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)a^T Y_{t-1}Y_{t-1}^T a^T\varepsilon_t^2 \\ &\le \frac{2}{V_n(u)}\frac{\|K\|_\infty}{nh}K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)a^T Y_{t-1}Y_{t-1}^T a\varepsilon_t^2 \\ &\le \frac{2}{V_n(u)}\frac{\|K\|_\infty}{nh}K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)\|a\|^2\|Y_{t-1}\varepsilon_t\|^2 \\ &\equiv \kappa_1\frac{1}{nh}K\left(\frac{1}{h}\left(\frac{t}{n}-u\right)\right)\|Y_{t-1}\varepsilon_t\|^2 \end{aligned}$$

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

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

and

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

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

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

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

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

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

which leads to

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

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

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

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

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

i.e.,

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

Proof of 5. Lemma 4 with (**??**) gives

$$\sqrt{nh}\,[\widehat{\alpha}(u) - \alpha(u) - B_n] = \sqrt{nh}E_0 S_n^{-1}\widetilde{\tau}_n \xrightarrow{D} N(0, \Sigma_\alpha),$$

and

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

References

[1] Akaike, H. (1974). A new look at the statistical model identification. IEEE Tras. Automat. Control AC-19, 716-722

[2] Beveridge, Stephen and C.R. Nelson (1981), "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'Business Cycle'," Journal of Monetary Economics, 7, 151-74.

[3] Cramer, H. (1961) On some classes of non-stationary stochastic processes. Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability, Vol. 2 Los Angeles, CA: University of California Press, pp. 57-78.

[4] Corbae, D., S. Ouliaris and P.C.B. Phillips (1997). Band spectral regression with trending data, Cowles Foundation Discussion Paper No. 1163.

[5] Dahlhaus, R. (1996a). On the Kullback-Leibler information divergence of locally stationary processes. Stochastic Process. Appl. 62, 139-168.

[6] ——— (1996b). Maximum likelihood estimation and model selection for locally stationary processes. J. Nonparametric Statist. 6, 171-191

[7] ——— (1996c). Asymptomatic statistical inference for nonstationary processes with evolutionary spectra. In Athens Conference on Applied Probability and Time Series Analysis (P. M. Robinson and M. Rosenblatt, eds.) 2. Springer, New York.

[8] ——— (1997), "Fitting time series models to nonstationary processes," Annals of Statistics, 25, 1-37.

[9] Fan, Y., and Q. Li (1996): "Consistent model specification tests: Omitted variables and semiparametric functional forms," *Econometrica* **64**, 865-890.

[10] Fan, Y., and Q. Li (1997): "Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing ," mimeographed.

[11] Kitagawa, G. and Gersch, W. (1985). A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series. IEEE Trans. Automat. Control. AC-30, 48-56.

[12] Kunsch, H. R. (1995). A note on causal solutions for locally stationary AP-processes. ETH Zurich. Preprint

[13] Melard, G. (1985) An example of the evolutionary spectral theory. J. Time. Ser. Anal. 6, 81-90.

[14] Melard, G. and Herteleer de Schutter, A. (1989). Contributions to evolutionary spectral theory. J. Time Ser. Anal. 10, 41-63

[15] Meyer, R. A., Jr. (1972) Estimating coefficients that change over time. Int. Econ. Rev. 13, 705-10.

[16] Neumann, M. H. and Von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. Ann. Statist. 25, 38-76.

[17] Phillips, P.C.B. and Solo, V. (1992). Asymptotics for linear processes. Annals of Statistics 20, 971-1001.

[18] Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. J. Roy. Statist. Soc. Ser. B 27, 204-237

[19] Priestley, M. B. (1981). Spectral Analysis and Time Series 2. Academic Press, London.

[20] Priestley, M. B. (1988). Nonlinear and Nonstationary Time Series Analysis. Academic Press, London.

[21] Young, P. C. and Beven, K. J. (1994). Data-based mechanistic modelling and the rainfall-flow nonlinearity. Environmetrics 5, 335-363.