



MONASH University

Australia

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**Semiparametric Estimation in Multivariate
Nonstationary Time Series Models**

Jiti Gao and Peter C.B. Phillips

September 2011

Working Paper 17/11

Semiparametric Estimation in Multivariate Nonstationary Time Series Models

Jiti Gao

University of Adelaide and Monash University

Peter C. B. Phillips

Yale University, University of Auckland,

University of Southampton, & Singapore Management University

September 5, 2011

Abstract

A system of multivariate semiparametric nonlinear time series models is studied with possible dependence structures and nonstationarities in the parametric and nonparametric components. The parametric regressors may be endogenous while the nonparametric regressors are assumed to be strictly exogenous. The parametric regressors may be stationary or nonstationary and the nonparametric regressors are nonstationary integrated time series. Semiparametric least squares (SLS) estimation is considered and its asymptotic properties are derived. Due to endogeneity in the parametric regressors, SLS is not consistent for the parametric component and a semiparametric instrumental variable (SIV) method is proposed instead. Under certain regularity conditions, the SIV estimator of the parametric component is shown to have a limiting normal distribution. The rate of convergence in the parametric component depends on the properties of the regressors. The conventional \sqrt{n} rate may apply even when nonstationarity is involved in both sets of regressors.

Key words and phrases: Endogeneity; integrated process, nonstationarity; partial linear model; simultaneity; vector semiparametric regression.

JEL Classification: C23, C25.

1 Introduction

Existing studies show that both nonstationarity and nonlinearity are common features of much economic data. Modeling such data in a way that allows for possible nonstationarity helps to avoid dependence on stationarity assumptions and mixing conditions for all of the variables in the system. At present there is a large literature on parametric linear modeling of nonstationary time series and interest has primarily focused on time series with a unit root or near unit root structure (for an overview, see, for example, Phillips and Xiao, 1998, and the references therein). In practical work, much attention is given to multivariate systems and cointegration models. Inferential methods for these linear systems include both parametric (e.g., Johansen, 1995) and semiparametric (e.g., Phillips and Hansen, 1990; Phillips, 1991, 1995, 2012) approaches.

In comparison with work on linear parametric models, there have been only a few studies of parametric nonlinear models with integrated variables. Park and Phillips (1988, 1989, 1999, 2001) introduced techniques for developing asymptotics for certain classes of nonlinear nonstationary parametric systems and aspects of this work have been extended by Pötscher (2004), Jeganathan (2004, 2008), and Berkes and Horváth (2006). Interest has also developed in nonparametric modeling methods to deal with nonlinearity of unknown form involving nonstationary variables. Existing studies in the field of nonparametric autoregression and cointegration estimation include Phillips and Park (1998), Karlsen and Tjøstheim (2001), Wang and Phillips (2009a, 2009b), Karlsen *et al* (2007), Kasparis and Phillips (2009), Cai *et al* (2009), Schienle (2009), and Phillips (2009). The last paper examines in a nonparametric setting spurious time series models of the type for which the asymptotic theory was given in Phillips (1986, 1998).

Among nonparametric studies of nonstationarity, two different mathematical approaches have been developed. In one approach, a so-called “Markov splitting technique” has been used in Karlsen and Tjøstheim (2001), and Karlsen *et al* (2007) to model univariate time series with a null-recurrent structure; and Chen *et al* (2008) consider univariate semiparametric regression modeling of null-recurrent time series, in which there is neither endogeneity nor heteroskedasticity. In the other approach, Phillips and Park (1998), Phillips (2009), and Wang and Phillips (2009a, 2009b) have developed ‘local-time’ methods to derive an asymptotic theory for nonparametric estimation of univariate models involving integrated time series.

In the case of independent and stationary time series data, semiparametric regression models have been intensively studied for more than two decades and there is a wide literature (Robinson 1988; Härdle *et al* 2000; Gao 2007; Li and Racine, 2007, among many others). In applied work, semiparametric methods have been shown to be particularly useful in modeling economic data in a way that retains generality where it is most needed while reducing dimensionality problems.

The present paper seeks to pursue these advantages in a wider context that allows for nonstationarities and endogeneities within a vector semiparametric regression model. The null recurrent structure of integrated time series typically reduces the amount of time that such time series spend in the vicinity of any one point, thereby

exacerbating the sparse data problem or “curse of dimensionality” in nonparametric and semiparametric modeling of multivariate integrated time series. On the other hand, recurrence means that nonlinear shape characteristics of unknown form may be captured over unbounded domains and endogeneity may be often accommodated without specialized methods (Wang and Phillips 2009b).

A common motivation for the use of semiparametric formulations such as (1.1) below is that they reduce nonparametric dimensionality through the presence of a linear parametric component. In our setting, the time series $\{Y_t, X_t, V_t\} : 1 \leq t \leq n\}$ are assumed to be modeled in a system of multivariate nonstationary time series models the form

$$\begin{aligned} Y_t &= A X_t + g(V_t) + e_t, \\ X_t &= H(V_t) + U_t, \quad t = 1, 2, \dots, n, \\ E[e_t|V_t] &= E[e_t] = 0 \quad \text{and} \quad E[U_t|V_t] = 0, \end{aligned} \tag{1.1}$$

where n is the sample size, A is a $p \times d$ -matrix of unknown parameters, $Y_t = (y_{t1}, \dots, y_{tp})'$, $X_t = (x_{t1}, \dots, x_{td})'$, and V_t is a sequence of univariate integrated time series regressors, $g(\cdot) = (g_1(\cdot), \dots, g_p(\cdot))'$ and $H(\cdot) = (h_1(\cdot), \dots, h_d(\cdot))'$ ¹ are all unknown functions, and both e_t and U_t are vectors of stationary time series. Note that $\{X_t\}$ can be stationary when $\{X_t\}$ and $\{V_t\}$ are independent. An extended version of model (1.1) is given in (2.21) in Section 2.3 below to deal with a more general case.

Model (1.1) corresponds to similar structures that have been used in the independent case (see Newey *et al* 1999; Su and Ullah 2008). The condition $E[e_t|V_t] = E[e_t]$ is generally needed to ensure that the model is identified. For, if there were an unknown function $\lambda(\cdot)$ such that $e_t = \lambda(V_t) + \varepsilon_t$ with $E[\varepsilon_t|V_t] = 0$, then only $g(\cdot) + \lambda(\cdot)$ would normally be estimable. However, recent research has revealed that some cases where e_t is correlated with V_t may be included. In particular, in studying nonparametric regressions of the form $Y_t = g(V_t) + e_t$, Wang and Phillips (2009b) consider a nonstationary endogenous regressor case where V_t is correlated with e_t and show that conventional nonparametric regression is applicable in spite of the endogeneity. Phillips and Su (2011) show that the same phenomena holds in cross section cases where there are continuous location shifts in the regressor, which play the role of an instrumental variable in tracing out the nonparametric regression function.

The identification condition $E[e_t|V_t] = E[e_t] = 0$ eliminates endogeneity between e_t and V_t while retaining endogeneity between e_t and X_t and potential nonstationarity in both X_t and V_t . The condition $E[e_t|V_t] = E[e_t] = 0$ in our setting corresponds to the condition $E[e_t|V_t, U_t] = E[e_t|U_t]$ that is assumed in Newey *et al* (1999) and Su and Ullah (2008), the former being implied by $E[e_t|V_t] = E(E[e_t|U_t, V_t]|V_t) = E(E[e_t|U_t]|V_t) = E(E[e_t|U_t]) = E[e_t]$ when U_t is independent of V_t and $E[e_t] = 0$. The identification conditions in (1.1) allow for both conditional heteroskedasticity and endogeneity in e_t , permitting e_t to depend on U_t ². These conditions are also less

¹ $F'(\cdot)$ denotes transpose of the vector function $F(\cdot)$, and $F^{(i)}(\cdot)$ denotes the i -th derivative of $F(\cdot)$.

²The additive case where $e_t = \lambda(U_t) + \mu_t$ with $E[\mu_t|V_t] = 0$ is covered in the first part of

restrictive than the exogeneity condition between e_t and (X_t, V_t) that is common in the literature for the stationary case (see, for example, Gao 2007).

The present paper treats model (1.1) as a vector semiparametric structural model and considers the case where X_t and V_t may be vectors of nonstationary regressors and X_t may be endogenous. In the case where endogeneity is involved in semiparametric regression modeling of independent data, some related developments include Newey *et al* (1999), Ai and Chen (2003), Newey and Powell (2003), Florens *et al* (2007), and Su and Ullah (2008). While estimation of partially linear models with endogeneity is discussed in each of these papers, neither the proposed structures nor the estimation methods may be used to deal with our case.

The contributions of the paper are as follows. We first consider a semiparametric least squares (SLS) estimator of A . When there is endogeneity in X_t , the SLS estimator of A is inconsistent. This may be seen from model (2.9) below when $E[U_t e_t'] \neq 0$. Accordingly, the paper proposes a semiparametric instrumental variable least squares (SIV) estimate of A to deal with endogeneity in X_t and a nonparametric estimator for the function $g(\cdot)$. The SIV estimator of A is shown to be consistent with a conventional \sqrt{n} rate of convergence in some cases even when X_t is stochastically nonstationary. This rate arises because nonstationarity in the regression may be eliminated by means of stochastic detrending.

The semiparametric procedure given here may be used on a system of nonlinear simultaneous equations with the following features: (i) nonstationarity and endogeneity in the parametric regressors; (ii) nonlinearity and nonstationarity in the nonparametric regressors; and (iii) stationary residuals. As such, the paper complements existing results on parametric modeling with endogeneity, nonparametric and semiparametric estimation of nonlinear time series (such as Fan and Yao 2003; Gao 2007), instrumental variable estimation of nonparametric models (such as Robinson 1988; Ai and Chen 2003; Newey and Powell 2003; Su and Ullah 2008), and nonparametric and semiparametric estimation of nonstationary time series (such as Phillips and Park 1998; Karlsen and Tjøstheim 2001; Karlsen *et al* 2007; Wang and Phillips 2009a, 2009b). For more references, including econometric interpretations of nonlinear and nonstationary effects, we refer to Phillips (2001) and Teräsvirta, Tjøstheim and Granger (2010).

In related work Chen *et al* (2008) consider the case where $\{V_t\}$ is a null recurrent Markov chain and assume the existence of an unknown functional $H(v) = E[X_t|V_t = v]$ that is independent of t in a scalar semiparametric regression $Y_t = X_t' \alpha + g(V_t) + e_t$ with $E[e_t|X_t, V_t] = 0$. By contrast, this paper imposes a set of general conditions in Assumption 3.3 below on the integrated process V_t . Note that a general integrated process is not a Markov chain unless it is of the explicit form $V_t = V_{t-1} + v_t$ with v_t being independent and identically distributed. Other related studies include Cai *et al* (2009) for a nonstationary varying coefficient time series model, Gao *et al* (2009a, 2009b) for model specification testing involving nonstationarity, and Phillips (2009) for nonparametric kernel estimation of the relationship between two integrated time

(1.1) because $E[e_t|V_t] = E[\lambda(U_t)|V_t] + E[\mu_t|V_t] = E[\lambda(U_t)] = E[e_t]$ when U_t is independent of V_t . The multiplicative case where $e_t = \sigma(U_t)\nu_t$ is also covered in the first part of (1.1) because $E[e_t|V_t] = E[\sigma(U_t)\nu_t|V_t] = E[e_t]$ when (U_t, ν_t) is assumed to be independent of V_t .

series in a spurious regression context.

The paper is organized as follows. Section 2 proposes estimators of the parameter matrix A and the nonlinear functions $g(\cdot)$. Section 3.1 establishes that the proposed semiparametric least squares (SLS) estimator of A achieve the conventional \sqrt{n} rate of convergence for the case where both the functional forms of $g(v)$ and $H(v)$ belong to a general class of functions. Section 3.2 briefly discusses cases where a super n rate of convergence for ordinary least squares (OLS) estimation of A is achievable when $g(v)$ is some ‘small’ function. One case involves an autoregressive version of model (1.1). A bandwidth selection method is developed in Section 4.1. Section 4.2 provides two examples to illustrate implementation. Conclusions are given and some limitations of the framework are discussed in Section 5. Proofs of the main results are given in Appendix A and subsidiary lemmas in Appendix B.

2 Semiparametric Estimation

Before addressing estimation, we provide a more detailed discussion of the model and its implications. Write (1.1) in full as:

$$Y_t = A X_t + g(V_t) + e_t \quad (2.1)$$

$$X_t = H(V_t) + U_t, \quad (2.2)$$

$$E[e_t|V_t] = E[e_t] = 0, \quad (2.3)$$

$$E[U_t|V_t] = 0. \quad (2.4)$$

When the variables $\{(X_t, V_t, e_t)\}$ are jointly stationary with finite second moments, the conditional expectation $H(V_t) = E[X_t|V_t]$ is well-defined. It is common to assume weak exogeneity, so that $E[e_t|(U_t, V_t)] = 0$, and letting $U_t = X_t - E[X_t|V_t]$, the decomposition of $X_t = H(V_t) + U_t$ is immediate. In consequence, the model (2.1)–(2.4) reduces to a standard semiparametric form

$$Y_t = A X_t + g(V_t) + e_t, \quad \text{with } E[e_t|(U_t, V_t)] = 0 \quad (2.5)$$

as discussed, for example, in Robinson (1988), Härdle *et al* (2000) and Gao (2007).

In the case where both X_t and V_t are nonstationary, the notion of a constant conditional expectation functional $E[X_t|V_t]$ may not be well defined. In (2.2), the dependence of X_t on V_t takes the general form of a nonlinear cointegrating system relating nonstationary variables. It follows from (2.1)–(2.4) that

$$\begin{aligned} E[Y_t|V_t = v] &= A H(v) + A E[U_t|V_t = v] + g(v) + E[e_t|V_t = v] \\ &= A H(v) + g(v), \end{aligned} \quad (2.6)$$

which implies that $\Psi(v) = E[Y_t|V_t = v]$ is well defined. In addition, (2.6) implies

$$g(v) = \Psi(v) - AH(v). \quad (2.7)$$

Thus, in view of equation (2.7), we can rewrite (2.1) as

$$Y_t - \Psi(V_t) = A (X_t - H(V_t)) + e_t = A U_t + e_t,$$

where $U_t = X_t - H(V_t)$, as assumed in (1.1). Introducing the “stochastically detrended” variable

$$W_t = Y_t - \Psi(V_t), \quad (2.8)$$

we can write (2.1) and (2.2) in semiparametrically contracted form as

$$W_t = A U_t + e_t. \quad (2.9)$$

Regarding (2.6)–(2.9), we make the following observations:

- As discussed in Section 1.2 of Härdle, Liang and Gao (2000), the stationarity of W_t and U_t in model (2.9) ensures that A is identifiable and estimable.
- The contracted form model (2.9) is semiparametric because both W_t and U_t are not observable and need to be estimated nonparametrically.
- Since $E[H(V_t)e'_t] = E\{H(V_t)E[e'_t|V_t]\} = 0$, we have

$$E[X_t e'_t] = E[H(V_t)e'_t] + E[U_t e'_t] = E[U_t e'_t] = E[U_t E(e'_t|U_t)]. \quad (2.10)$$

It follows that the unknown matrix A can be consistently estimated based on (2.9) when $E[U_t e'_t] = 0$. The following two cases show that this condition can still be satisfied even when e_t may depend on U_t .

Case 2.1. Consider a multiplicative relationship of the form $e_t = \sigma(U_t)\pi_t$, where π_t is a sequence of independent random errors with $E[\pi_t|U_t] = 0$ and $\sigma(U_t)$ is a positive definite matrix. In this case, we have $E[e_t|U_t] = \sigma(U_t)E[\pi_t|U_t] = 0$.

Case 2.2. Let $p(\cdot)$ be the marginal density of U_t and $\gamma(u) = E[e'_t|U_t = u]$. Then, $E[U_t e'_t] = E[U_t E(e'_t|U_t)] = E[U_t \gamma(U_t)] = \int_{-\infty}^{\infty} u \gamma(u) p(u) du = 0$ when $\gamma(u)p(u) = \gamma(-u)p(-u)$ for all u .

In such cases as these, there is no need to introduce instrumental variables (IVs) in the estimation of (2.9). Otherwise, endogeneity must be addressed and an IV procedure may be used to achieve consistent estimation of A . Section 2.1 proposes a semiparametric least squares (SLS) estimation method for the case where $E(e'_t|U_t) = 0$. Section 2.2 develops a semiparametric instrumental variable procedure (SIV) that is applicable in the case of nonstationary U_t .

2.1 SLS estimation

When $E(e'_t|U_t) = 0$, consistent estimation is possible based on (2.9). But since both W_t and U_t are unobservable, the unknown functions $\Psi(\cdot)$ and $H(\cdot)$ must be estimated nonparametrically. Substituting nonparametric kernel estimates into (2.9) gives an approximate semiparametric nonlinear time series model of the form

$$\tilde{Y}_t = A \tilde{X}_t + e_t, \quad (2.11)$$

where $\tilde{Y}_t = \widehat{W}_t F_t$ and $\tilde{X}_t = \widehat{U}_t F_t$, in which $\widehat{W}_t = Y_t - \widehat{\Psi}(V_t)$ and $\widehat{U}_t = X_t - \widehat{H}(V_t)$. In these formulae, F_t is the indicator $F_t = I(\widehat{p}_n(V_t) > b_n)$ where b_n is a sequence

of positive numbers that tend to zero as $n \rightarrow \infty$, $\hat{p}_n(v) = \frac{1}{\sqrt{nh}} \sum_{s=1}^n K\left(\frac{V_s - v}{h}\right)$, $\hat{\Psi}(v) = \sum_{s=1}^n w_{ns}(v)Y_s$ and $\hat{H}(v) = \sum_{s=1}^n w_{ns}(v)X_s$ with $w_{ns}(\cdot)$ being a sequence of probability weight functions of the form

$$w_{nt}(v) = \frac{K_{v,h}(V_t)}{\sum_{k=1}^n K_{v,h}(V_k)} \quad \text{with} \quad K_{v,h}(V_t) = \frac{1}{h} K\left(\frac{V_t - v}{h}\right), \quad (2.12)$$

in which $K(\cdot)$ is a probability kernel function and h is a bandwidth parameter. Note that since V_t is scalar, we need only use a single bandwidth parameter h .

Note that $\hat{p}(v)$ could be thought of as a density estimate of the invariant measure of $\{V_t\}$, and it is introduced to solve the so-called ‘‘random denominator’’ problem. This type of truncation method has been widely used in the literature for the independent sample case (see, for example, Robinson 1988).

The semiparametric least squares (SLS) estimator of A is defined by the equation

$$\hat{A} = \tilde{Y}' \tilde{X} (\tilde{X}' \tilde{X})^{-1}, \quad (2.13)$$

where $\tilde{X}' = (\tilde{X}_1, \dots, \tilde{X}_n)$, $\tilde{Y}' = (\tilde{Y}_1, \dots, \tilde{Y}_n)$, and throughout the paper D^{-1} is the inverse of D or a generalized inverse if D^{-1} does not exist. The vector of unknown functions $g(\cdot)$ is then estimated by

$$\hat{g}(v) = g_n(v; \hat{A}) \equiv \sum_{s=1}^n w_{ns}(v)Y_s - \hat{A} \sum_{s=1}^n w_{ns}(v)X_s. \quad (2.14)$$

By elementary calculation

$$(\hat{A} - A) \tilde{X}' \tilde{X} = \tilde{e}' \tilde{X} + \tilde{G}' \tilde{X}, \quad (2.15)$$

with $\tilde{G}' = (\tilde{G}_1, \dots, \tilde{G}_n) = (\tilde{g}(V_1), \dots, \tilde{g}(V_n))$, $\tilde{g}(V_t) = g(V_t) - \sum_{s=1}^n w_{ns}(V_t)g(V_s)$, $\tilde{e}' = (\tilde{e}_1, \dots, \tilde{e}_n)$ and $\tilde{e}_t = e_t - \sum_{s=1}^n w_{ns}(V_t)e_s$. This estimator in (2.13) is implemented in Example 4.1 below.

Assuming that $g(\cdot)$ and $H(\cdot)$ are both differentiable and their first derivatives are all continuous, as shown in Appendix A, an approximate version of (2.15) has the form

$$(\hat{A} - A) U'U (1 + o_P(1)) = e'U (1 + o_P(1)), \quad (2.16)$$

where $e' = (e_1, \dots, e_n)$ and $U = (U_1, \dots, U_n)'$. This reduction shows that \sqrt{n} convergence is achievable when $E[e|U] = 0$ and some smoothness conditions are imposed on $g(\cdot)$ and $H(\cdot)$.

Equation (2.16) also shows that \hat{A} will be inconsistent when U is a matrix of endogenous regressors for which $E[e|U] \neq 0$. This case is now considered and a semiparametric instrumental variable (SIV) estimation method for A is developed that is consistent and has desirable asymptotic properties.

2.2 SIV estimation

In the case where U is a matrix of integrated regressors, a semiparametric version of the fully modified (FM) estimation procedure of Phillips and Hansen (1990) and Phillips (1995) may be used to consistently estimate A . That approach may be considered for the case where both X_t and V_t are univariate integrated regressors and are independent of each other. But when U is a matrix of stationary regressors, the FM method fails. We therefore propose here a semiparametric instrumental variable (SIV) approach.

To develop the SIV method, in the semiparametric model

$$W_t = AU_t + e_t \quad \text{with} \quad E[e_t|V_t] = 0 \quad \text{and} \quad E[e_t|U_t] \neq 0, \quad (2.17)$$

we assume the existence of a vector of stationary variables η_t for which

$$E[U_t\eta_t'] \neq 0 \quad \text{and} \quad E[e_t|\eta_t] = 0. \quad (2.18)$$

Equations (2.17) and (2.18) imply

$$W_t\eta_t' = AU_t\eta_t' + e_t\eta_t' \quad \text{with} \quad E[U_t\eta_t'] \neq 0 \quad \text{and} \quad E[e_t\eta_t'] = 0. \quad (2.19)$$

We focus on the case where the number of instruments equals the number of regressors and

$$\text{rank of } E[\eta'\eta] \equiv r = d \equiv \text{rank of } E[\eta'U], \quad (2.20)$$

where $\eta' = (\eta_1, \dots, \eta_n)$. The case where the number of instrumental variables is greater than the number of regressors may be analyzed in a similar way.

If W_t , U_t and η_t were all observed time series, models (2.17) and (2.19) would consist of a vector semiparametric system with stationary time series regressors. Here, each η_t may be regarded as the stationary component of a suitable instrumental variable (IV). In this setting, it is straightforward to construct a consistent estimator for A .

Since η_t may not be directly observable, we assume that there is a vector of observed instruments, Q_t , that satisfy an expanded version of the system (1.1) of the form

$$\begin{aligned} Y_t &= A X_t + g(V_t) + e_t \quad \text{with} \quad E[e_t|V_t] = E[e_t], \\ X_t &= H(V_t) + U_t \quad \text{with} \quad E[U_t|V_t] = 0, \\ Q_t &= J(V_t) + \eta_t \quad \text{with} \quad E[\eta_t|V_t] = 0, \end{aligned} \quad (2.21)$$

where η_t is assumed to satisfy (2.18), $Q_t = (q_{t1}, \dots, q_{td})'$ is a vector of possible instrumental variables for X_t generated by a reduced form equation involving V_t , and $J(\cdot) = (J_1(\cdot), \dots, J_d(\cdot))'$ is a vector of unknown functions.

The residual η_t may be interpreted as a sequence of stochastically detrended versions of Q_t and we therefore assume that η_t is strictly stationary even though Q_t itself may be a vector of nonstationary instruments. In effect, the nonstationarity in Q_t arises from the component $J(V_t)$ which depends on the nonstationary process V_t . It is particularly natural to choose a stationary IV like η_t as a residual when U_t

itself is assumed to be a stationary residual given by the stochastically detrended quantity $X_t - H(V_t)$. The augmented system (2.21) simply adds in this instrument generating equation to the original system (1.1). The new system obviously reduces to (1.1) when there is no endogeneity in X_t .

As discussed in the literature for the stationary case, the existence and choice of Q_t is often a difficult and important practical matter. In the nonstationary case, similar considerations apply. To clarify the issues involved, we look at the following special case.

Remark 2.1. Consider a pair (e_t, η_t) of the form

$$e_t = \Sigma U_t + \Delta \Pi_t \quad \text{and} \quad \eta_t = \Delta U_t - \Sigma \Pi_t, \quad (2.22)$$

where both Σ and $\Delta = I - \Sigma$ are deterministic, symmetric and positive definite matrices, and Π_t is a vector of stationary errors satisfying $E[\Pi_t] = 0$, $\text{cov}(U_t, \Pi_t) = \text{cov}(V_t, \Pi_t) = 0$ and $\text{cov}(\Pi_t, \Pi_t) = \text{cov}(U_t, U_t) = I$. In this case, we have

$$\begin{aligned} E[e_t U_t'] &= \Sigma E[U_t U_t'], & E[\eta_t U_t'] &= \Delta E[U_t U_t'], \\ E[e_t \eta_t'] &= \Sigma E[U_t U_t'] \Delta' - \Delta E[\Pi_t \Pi_t'] \Sigma' = 0. \end{aligned} \quad (2.23)$$

We discuss how to estimate Σ . Using the linear reduced form (2.17) and substituting (2.22) into (2.17), we have

$$W_t = A U_t + e_t = (A + \Sigma) U_t + (I - \Sigma)\Pi_t = B U_t + \Delta \Pi_t, \quad (2.24)$$

where $B = A + \Sigma$ and $\Delta = I - \Sigma$. Since $\text{cov}(U_t, \Pi_t) = 0$, we can estimate B using the same method as in (2.13) by \hat{B} and the matrix $\Gamma = \Delta \Delta'$ by

$$\hat{\Gamma} = \frac{1}{n} \sum_{t=1}^n (\tilde{Y}_t - \hat{B} \tilde{X}_t) (\tilde{Y}_t - \hat{B} \tilde{X}_t)'. \quad (2.25)$$

As shown in Corollary 3.3 below, we have $\hat{\Gamma} \rightarrow_P \Gamma$ as $n \rightarrow \infty$. The matrix Σ is then consistently estimated by $\hat{\Sigma} = I - \hat{\Delta}$ under constraints such that both $\hat{\Sigma}$ and $\hat{\Delta}$ are still positive definite matrices.

Let $J(v) = H(v)$. Then, $Q_t = J(V_t) + \eta_t$ is a vector of valid instrumental variables. This case, along with the estimation method proposed in (2.25), is implemented in Example 4.2.

We now construct a consistent estimator for A . In view of equations (2.17)–(2.21), and similar to (2.13), we define the semiparametric instrumental variable least squares (SIV) estimator

$$\hat{A}^* = \hat{A}^*(h) = \tilde{Y}' \tilde{Q} (\tilde{X}' \tilde{Q})^{-1}, \quad (2.26)$$

where $\tilde{Q}' = (\tilde{Q}_1, \dots, \tilde{Q}_n)$ with $\tilde{Q}_t = (Q_t - \sum_{s=1}^n w_{ns}(V_t) Q_s) F_t$. Correspondingly, the vector of unknown functions $g(\cdot)$ is estimated by

$$\hat{g}^*(v) = g_n(v; \hat{A}^*) \equiv \sum_{s=1}^n w_{ns}(v) Y_s - \hat{A}^* \sum_{s=1}^n w_{ns}(v) X_s. \quad (2.27)$$

It follows from (2.26) that

$$(\widehat{A}^* - A) \widetilde{X}' \widetilde{Q} = \widetilde{e}' \widetilde{Q} + \widetilde{G}' \widetilde{Q},$$

where \widetilde{G} and \widetilde{e} are defined analogously to \widetilde{Q} . As shown in Appendix A, we have the following decomposition

$$(\widehat{A}^* - A) U' \eta (1 + o_P(1)) = e' \eta (1 + o_P(1)), \quad (2.28)$$

where $\eta = (\eta_1, \dots, \eta_n)'$ and $e = (e_1, \dots, e_n)'$.

To establish the validity of the approximations given in (2.16) and (2.28), we impose certain regularity conditions which enable us to establish consistency and a limit distribution theory.

3 Main Results and Extensions

3.1 Asymptotic Theory

As pointed out in the Introduction, the limit theory in this kind of nonstationary semiparametric model depends on the probabilistic structure of the regressors and errors e_t , U_t , η_t and V_t as well as the functional forms of $g(\cdot)$, $H(\cdot)$ and $J(\cdot)$. It is convenient for the development that follows to make general conditions on the nonstationary process V_t rather than specify a particular generating mechanism. These conditions are discussed in Appendix A and include the usual integrated and near integrated process mechanisms that commonly appear in applications. It is also convenient to use mixing conditions to establish some of the main results in the paper and we recall that a matrix stationary process $\{\mathcal{Z}_t, t = 0, \pm 1, \dots\}$ is α -mixing if the mixing numbers $\alpha(n) \rightarrow 0$ as $n \rightarrow \infty$, where

$$\alpha(n) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^\infty} |P(AB) - P(A)P(B)|, \quad (3.1)$$

in which \mathcal{F}_k^j is the σ -field generated by $\{\mathcal{Z}_t, k \leq t \leq j\}$.

The following assumptions are used to develop the asymptotic theory. A detailed discussion of these conditions is provided in Appendix A.

Assumption 3.1. (i) $\xi_t = (U_t', \eta_t)'$ is a vector of (strictly) stationary time series with $E[\xi_1] = 0$ and $E[\|\xi_1\|^{4+\gamma_1}] < \infty$ for some $\gamma_1 > 0$, where $\|\cdot\|$ denotes the Euclidean norm. The process ξ_t is α -mixing with mixing numbers $\alpha_\xi(j)$ that satisfy

$$\sum_{j=1}^{\infty} \alpha_\xi^{\frac{\gamma_1}{4+\gamma_1}}(j) < \infty. \quad (3.2)$$

(ii) $\zeta_t = e_t$ or $e_t \eta_t'$ is a matrix of stationary time series with $E[\|\zeta_1\|^{4+\gamma_2}] < \infty$ for some $\gamma_2 > 0$. The process ζ_t is α -mixing with mixing numbers $\alpha_\zeta(j)$ that satisfy

$$\sum_{j=1}^{\infty} \alpha_\zeta^{\frac{\gamma_2}{4+\gamma_2}}(j) < \infty. \quad (3.3)$$

Assumption 3.2. (i) Let model (1.1) hold and Q_t be a vector of instrumental variables such that conditions (2.18), (2.20) and (2.21) are all satisfied.

(ii) $E[e_{s+t} \otimes \eta_t] = 0$ for all $s \geq 0$ and $E[e_s \otimes e_t \otimes \eta_u \otimes \eta_v] = 0$ when at least three of the date indices are different.

(iii) $\Gamma_1 = E[U_1 \eta'_1]$ be nonsingular.

(iv) $\Sigma_1^* = (I \otimes \Gamma_1^{-1}) \Omega_1^* (I \otimes (\Gamma_1^{-1})')$ and $\Omega_1^* = \sum_{j=0}^{\infty} E[(e_1 e'_{1+j}) \otimes (\eta_1 \eta'_{1+j})]$ are positive definite.

Assumption 3.3. (i) $\{V_t : t \geq 0\}$ is independent of $\{(e_t, U_t, \eta_t) : t \geq 1\}$.

(ii) Let $f_{i,k}(\cdot)$ be the density function of $V_{i,k} = \varphi_{i-k}(V_i - V_k)$ for $i > k$ with $\varphi_m = \frac{1}{\sqrt{m}}$ for $m \geq 1$. Let $f_{i,k}(x)$ is uniformly bounded by some function $\lambda_1(x)$ such that $\int_{-\infty}^{\infty} \lambda_1(x) dx < \infty$ and

$$\lim_{\delta \rightarrow 0} \limsup_{m \rightarrow \infty} \sup_{i \geq 1} \sup_{|v| \leq \delta} |f_{i+m,i}(v) - f_{i+m,i}(0)| = 0. \quad (3.4)$$

There exists a filtration $\{\mathcal{F}_t, t \geq 0\}$ such that V_t is adapted to \mathcal{F}_t . Let $f_{i,k}(v|\mathcal{F}_k)$ be the conditional density function of $V_{i,k}$ given \mathcal{F}_k , $\max_{i \geq 1; k \geq 1} f_{i,k}(v|\mathcal{F}_k)$ be bounded by some function $\lambda_2(x)$ such that $\int_{-\infty}^{\infty} \lambda_2(x) dx < \infty$, and with probability one,

$$\lim_{\delta \rightarrow 0} \limsup_{m \rightarrow \infty} \sup_{i \geq 1} \sup_{|v| \leq \delta} |f_{i+m,i}(v|\mathcal{F}_i) - f_{i+m,i}(0|\mathcal{F}_i)| = 0. \quad (3.5)$$

Assumption 3.4. (i) The vector function $g(v)$ is continuously differentiable for $v \in R$ and the derivative $g^{(1)}(v)$ satisfies, for large enough n ,

$$\sum_{t=1}^n \int \left\| g^{(1)}(\varphi_t^{-1}v) \right\|^2 f_{t,0}(v) dv = O(nh^{-1}), \quad (3.6)$$

where $\{f_{t,0}(v)\}$ is as defined in Assumption 3.3 above.

(ii) The vector function $H(v)$ is continuously differentiable for $v \in R$ and the derivative $H^{(1)}(v)$ satisfies for large enough n

$$\sum_{t=1}^n \int \left\| H^{(1)}(\varphi_t^{-1}v) \right\|^2 f_{t,0}(v) dv = O(nh^{-1}) \quad \text{and} \quad (3.7)$$

$$\sum_{t=1}^n \int \left\| (g^{(1)}(\varphi_t^{-1}v))' H^{(1)}(\varphi_t^{-1}v) \right\| f_{t,0}(v) dv = O\left(n^{\frac{1}{2}-\varepsilon_1} b_n^2 h^{-2}\right), \quad (3.8)$$

where $0 < \varepsilon_1 < \frac{1}{2}$ is some constant.

(iii) The vector function $J(v)$ is continuously differentiable for $v \in R$ with derivative $J^{(1)}(v)$ that satisfies for large enough n

$$\sum_{t=1}^n \int \left\| J^{(1)}(\varphi_t^{-1}v) \right\|^2 f_{t,0}(v) dv = O(nh^{-1}) \quad \text{and} \quad (3.9)$$

$$\sum_{t=1}^n \int \left\| (g^{(1)}(\varphi_t^{-1}v))' J^{(1)}(\varphi_t^{-1}v) \right\| f_{t,0}(v) dv = O\left(n^{\frac{1}{2}-\varepsilon_2} b_n^2 h^{-2}\right), \quad (3.10)$$

where $0 < \varepsilon_2 < \frac{1}{2}$ is some constant.

Assumption 3.5. (i) $K(\cdot)$ is a symmetric and bounded probability density function with compact support C_K and $K(u)$ is continuous for all $u \in C_K$.

(ii) The sequences $\{h_n\}$ and $\{b_n\}$ both satisfy, as $n \rightarrow \infty$, the following rate conditions

$$h_n \rightarrow 0, \quad nh_n^2 \rightarrow \infty, \quad nh_n^6 \rightarrow 0, \quad (3.11)$$

$$b_n \rightarrow 0, \quad \frac{1}{\sqrt{nb_n^2}} \rightarrow 0, \quad \frac{\sqrt{h}}{b_n^2} \rightarrow 0, \quad \frac{1}{nh^2b_n^8} \rightarrow 0, \quad (3.12)$$

where $L_s(n)$ is as defined in Assumption 3.3(ii).

(iii) b_n is also chosen such that $\sum_{t=1}^n P(\hat{p}_n(V_t) \leq b_n) = o(n)$.

(iv) There exists a real function $\lambda(x, y)$ such that $\|g(x + yh) - g(x)\| \leq h\lambda(y, x)$ for small enough h , all $y \in R = (-\infty, \infty)$ and $\int_{-\infty}^{\infty} \lambda(x, y)K(x)dx < \infty$ for any given y .

Assumptions 3.1–3.5 appear to be reasonably mild conditions and include the important case where $g(v)$, $H(v)$ and $J(v)$ are all linear functions. Some detailed discussion and technical justifications for Assumptions 3.1–3.5 are provided in Appendix A. Under these conditions, we have the following results, whose proofs are also given in Appendix A.

Theorem 3.1 Under Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5(i)(ii)(iii), as $n \rightarrow \infty$, we have

$$\sqrt{n}(\hat{A}^* - A) \rightarrow_D N(0, \Sigma_1^*), \quad (3.13)$$

where $\Sigma_1^* = (I \otimes \Gamma_1^{-1}) \Omega_1^* (I \otimes (\Gamma_1^{-1})')$, $\Omega_1^* = \sum_{j=0}^{\infty} E[(e_1 e'_{1+j}) \otimes (\eta_1 \eta'_{1+j})]$ and $\Gamma_1 = E[U_1 \eta'_1]$.

Theorem 3.1 shows that the semiparametric IV estimator \hat{A}^* is asymptotically normal in the limit even when the parametric and nonparametric regressors are both nonstationary. In addition, \hat{A}^* is consistent when there is endogeneity in the parametric regressors. The explanation for the \sqrt{n} convergence rate and the limiting normality is that A is estimated based on (2.17) and (2.18), which consists of a vector semiparametric system in which η_t is a vector of stochastically detrended versions of the instruments Q_t . Stationarity of (U_t, e_t, η_t) then ensures that standard asymptotic normality with a conventional \sqrt{n} convergence rate is achieved.

When X_t is strictly exogenous and U_t is independent of e_t , Theorem 3.1 has the following corollary.

Corollary 3.1 (i) Let Assumptions 3.1, 3.2, 3.3, 3.4(i)(ii) and 3.5(i)(ii)(iii) hold. Then as $n \rightarrow \infty$

$$\sqrt{n}(\hat{A} - A) \rightarrow_D N(0, \Sigma_1^*), \quad (3.14)$$

where $\Sigma_1^* = (I \otimes \Gamma_1^{-1}) \Omega_1^* (I \otimes \Gamma_1^{-1})$ with $\Omega_1^* = \sum_{j=0}^{\infty} E[e_1 e'_{1+j}] \otimes E[U_1 U'_{1+j}]$ and $\Gamma_1 = E[U_1 U'_1]$.

(ii) If, in addition, both U_t and e_t are independent and identically distributed, then as $n \rightarrow \infty$

$$\sqrt{n}(\hat{A} - A) \rightarrow_D N\left(0, \Sigma_{11} \otimes \Sigma_{22}^{-1}\right), \quad (3.15)$$

where $\Sigma_{11} = E[e_1 e_1']$ and $\Sigma_{22} = E[U_1 U_1']$.

Corollary 3.1 extends existing results for the univariate case where both the parametric and nonparametric regressors are independent random variables (see, for example, Robinson 1988; Härdle *et al* 2000) to the vector case where both the parametric and nonparametric regressors may be nonstationary. Chen *et al* (2008) gave the univariate version of Corollary 3.1 under the assumption that V_t is a null recurrent Markov chain.

Note that when there is heteroskedasticity in e_t , either \hat{A} or \hat{A}^* may be replaced by a weighted semiparametric least squares estimator (see, for example Chapter 2 of Härdle *et al* 2000). In this case, it is necessary to estimate the covariance matrix Ω_1^* by suitable application of some existing methods (see, for example, Phillips 1995). Such extensions are not trivial, and therefore left for future research.

Recall that the nonparametric component is estimated by $\hat{g}^*(v)$ as defined in (2.27). The asymptotic distribution of $\hat{g}^*(v)$ is obtained along lines similar to those in Wang and Phillips (2009a) and Karlsen *et al* (2007) and is given in Theorem 3.2 below.

Theorem 3.2 *Let the conditions of Theorem 3.1 hold. If, in addition, Assumption 3.5(iv) holds, then as $n \rightarrow \infty$*

$$\sqrt{\sum_{t=1}^n K\left(\frac{v - V_t}{h}\right)} (\hat{g}^*(v) - g(v)) \rightarrow_D N(0, \Omega_g), \quad (3.16)$$

where $\Omega_g = \int K^2(u) du \cdot E[e_1 e_1']$ and $\lambda_s(\epsilon) = E[\epsilon_s]$.

Remark 3.2. The random normalization in (3.16) implies that the convergence rate depends on the order of the sample average $\sum_{t=1}^n K\left(\frac{v - V_t}{h}\right)$. In the stationary case, this quantity typically has order nh , whereas when V_t is a unit root or near integrated process it has order \sqrt{nh} (see Wang and Phillips, 2009a). It follows that in the nonstationary case, the rate of convergence of $\hat{g}^*(v)$ is $(\sqrt{nh})^{\frac{1}{2}}$.

Finally, we establish the following convergence results for the residual moment matrix.

Theorem 3.3 *Let Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5(i)(ii)(iii) hold. If, in addition, $\Sigma_{11} = E[e_1 e_1']$ is positive definite, then as $n \rightarrow \infty$*

$$\hat{\Sigma}_{11} = \frac{1}{n} \sum_{t=1}^n \left(Y_t - \hat{A}^* X_t - \hat{g}_n^*(V_t)\right) \left(Y_t - \hat{A}^* X_t - \hat{g}_n^*(V_t)\right)' \rightarrow_P \Sigma_{11}. \quad (3.17)$$

Since Π_t involved in (2.22) satisfies the same conditions as $\{(e_t, U_t)\}$, Theorem 3.3 can be used to deduce the following corollary when $\text{cov}(U_t, \Pi_t) = 0$. The Corollary below shows that the covariance matrix Σ involved in (2.22) representing the level of endogeneity in that model can be consistently estimated.

Corollary 3.2 *Let Assumptions 3.1, 3.2, 3.3, 3.4(i)(ii) and 3.5(i)(ii)(iii) hold. If, in addition, Σ is positive definite, then as $n \rightarrow \infty$*

$$\hat{\Gamma} = \frac{1}{n} \sum_{t=1}^n (\tilde{Y}_t - \hat{B}\tilde{X}_t) (\tilde{Y}_t - \hat{B}\tilde{X}_t)' \rightarrow_P \Gamma \quad (3.18)$$

when $\text{cov}(\Pi_t, \Pi_t) = \text{cov}(U_t, U_t) = I$, where \hat{B} is as defined in (2.25) and $\Gamma = \Delta\Delta'$.

3.2 Some Extensions

This Section establishes an asymptotically consistent estimator for \hat{A} with the conventional \sqrt{n} rate of convergence under the assumption that the nonparametric functional forms of $g(v)$, $H(v)$ and $J(v)$ are unknown and can include certain polynomial functions. In the case where $H(v)$ is a linear function of v and $g(v)$ behaves like some ‘small’ function from the linear component, we may provide an efficient estimator for A in the univariate case. Meanwhile, an autoregressive version of model (1.1) can also be considered when $g(v)$ behaves like some ‘small’ function.

Consider system (1.1) with $p = d = 1$ and $H(v) = \theta_0 + \theta_1 v$, where both θ_0 and θ_1 are unknown parameters. Before discussing estimation, we impose the following conditions.

Assumption 3.6. (i) *Let $V_t = V_{t-1} + v_t$, where $\varepsilon_t = (e_t, U_t, v_t)$ is a vector of stationary time series with $E[\varepsilon_1] = 0$ and $E[\|\varepsilon_1\|^{4+\gamma_1}] < \infty$ for some $\delta_\varepsilon > 0$. The process ε_t is α -mixing with mixing numbers $\alpha_\varepsilon(j)$ that satisfy $\sum_{j=1}^{\infty} \alpha_\varepsilon^{\frac{\gamma_1}{4+\delta_\varepsilon}}(j) < \infty$.*

(ii) *Let $E_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} e_t$ and $V_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} v_t$. There is a vector Brownian motion (B_e, B_v) such that $(E_n(r), V_n(r)) \Rightarrow_{\mathcal{D}} (B_e(r), B_v(r))$ on $D[0, 1]^2$ as $n \rightarrow \infty$, where the symbol “ \Rightarrow ” stands for weak convergence.*

(iii) *Let $vg(v)$ be an integrable function and satisfy $\int_{-\infty}^{\infty} vg(v) \neq 0$.*

Theorem 2.1 of Wang and Phillips (2009a) shows that as $n \rightarrow \infty$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n V_t g(V_t) \rightarrow_D L_{B_v}(1, 0) \int_{-\infty}^{\infty} vg(v) dv, \quad (3.19)$$

where $L_{B_v}(1, 0)$ is the local-time process of V_t .

Equation (3.19) then implies as $n \rightarrow \infty$

$$\frac{1}{n} \sum_{t=1}^n V_t g(V_t) = \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^n V_t g(V_t) \rightarrow_P 0. \quad (3.20)$$

Under Assumption 3.6, in view of (3.20), we have as $n \rightarrow \infty$

$$\begin{aligned}
n(\tilde{A} - A) &= n \left(\left(\sum_{t=1}^n X_t^2 \right)^{-1} \sum_{t=1}^n X_t Y_t - A \right) \\
&= n \left(\left(\sum_{t=1}^n X_t^2 \right)^{-1} \sum_{t=1}^n X_t e_t + \left(\sum_{t=1}^n X_t^2 \right)^{-1} \sum_{t=1}^n X_t g(V_t) \right) \\
&\Rightarrow_{\mathcal{D}} \left(\theta_1 \int_0^1 B_v^2(r) dr \right)^{-1} \cdot \left(\int_0^1 B_v(r) dB_e(r) \right), \tag{3.21}
\end{aligned}$$

which means that rate n convergence is achievable.

In the case where $\int vg(v) = 0$, we have a similar result. Consider, for example, the system

$$Y_t = a X_t + b g(V_t) + \epsilon_t \quad \text{with} \quad g(V_t) = \frac{1}{1 + V_t^4}, \tag{3.22}$$

$$X_t = H(V_t) + U_t \quad \text{and} \quad H(V_t) = c V_t, \tag{3.23}$$

with $V_t = \sum_{s=1}^t v_s$, where all variables are scalar and satisfy the conditions of Theorem 3.1. In this case, the simple IV estimator $a_{IV} = (\sum_{t=1}^n X_t V_t)^{-1} (\sum_{t=1}^n V_t Y_t)$ converges at the usual rate n for cointegrated systems and has a mixed normal limit distribution that is amenable to inference. To see this, we use the following three results (the first two are standard and the third follows from the limit theory for a zero energy functional of a partial sum process – see Jeganathan, 2008):

$$\begin{aligned}
\frac{1}{n^2} \sum_{t=1}^n X_t V_t &\Rightarrow c \int_0^1 B_v^2, \\
\frac{1}{n} \sum_{t=1}^n V_t \epsilon_t &\Rightarrow \int_0^1 B_v dB_\epsilon, \\
\frac{1}{\sqrt{\sqrt{n}}} \sum_{t=1}^n \frac{V_t}{1 + V_t^4} &\Rightarrow \sqrt{\beta} L_0^1 Z,
\end{aligned}$$

where $\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor n \rfloor} (\epsilon_t, v_t) \Rightarrow (B_\epsilon, B_v)$, bivariate Brownian motion, $L_0^1 = L_{B_v}^1(1, 0)$ is the local time of B_v at the origin over the unit time interval $[0, 1]$, Z is a standard normal variate, and the constant β depends on the distribution of the $\{v_t\}$. From these results, we have the limit theory

$$n(a_{IV} - a) = \left(c \int_0^1 B_v^2 \right)^{-1} \left(\int_0^1 B_v dB_\epsilon \right),$$

which has a mixed normal distribution under the exogeneity condition on V_t . In this case, direct IV estimation is (asymptotically infinitely) superior to semiparametric estimation involving nonparametric stochastic detrending.

Models (3.22) and (3.23) are of some practical interest. In particular, the function $g(V_t)$ is integrable and provides a ‘small’ nonlinear correction to the linear component of the cointegrating relation (3.22). This nonlinear component becomes

most relevant when the process V_t takes values near the origin. But the function could easily be reformulated so that the most relevant values occurred elsewhere in the sample space. The remaining components of the system are analogous to those in conventional cointegrated systems. Thus, (3.22) - (3.23) is a cointegrated system with small deviations from linearity that affect the relationship but do not disturb the properties of a simple IV estimator. In effect, estimation of the linear component aX_t may be conducted without concern for the nonlinear component. So nonlinear stochastic detrending is unnecessary here. Of course, when the functional form of the stochastic trending component is unknown then a parametric procedure like linear IV estimation may be unreliable and will normally result in inconsistency.

Meanwhile, autoregressive version of model (1.1) are also of general interest and do have various applications. In the stationary case where $\{V_t\}$ is stationary, the proposed SLS estimation methods still works well when some components of X_t can be the lagged variables of Y_t (see, for example, Gao 2007). In the case where $\{V_t\}$ is integrated, it may not be possible to assume $U_t = Y_{t-1} - E[Y_{t-1}|V_t]$ is stationary. In some simple cases, such as $g(v) = v$, U_t is not even integrated. This is mainly because the nonstationarity of Y_t induced from $g(V_t)$ can be of higher order, for example when the functional form of $g(v)$ is polynomial. In the case where $g(v)$ is some ‘small’ function, such as that satisfying Assumption 3.6(iii), the ordinary least squares estimator of A may be n -consistent or \sqrt{n} -consistent, depending on the functional form of $g(v)$.

We now discuss briefly the case where $p = d = 1$ and $X_t = Y_{t-1}$. In this case, model (1.1) becomes

$$Y_t = AY_{t-1} + g(V_t) + e_t = \sum_{j=0}^{\infty} A^j g(V_{t-j}) + \sum_{j=0}^{\infty} A^j e_{t-j}, \quad (3.24)$$

when $|A| < 1$, and

$$Y_t = AY_{t-1} + g(V_t) + e_t = Y_0 + \sum_{s=1}^t g(V_s) + \sum_{s=1}^t e_s, \quad (3.25)$$

when $A \equiv 1$. In model (3.25) for example, we have as $t \rightarrow \infty$

$$\frac{1}{\sqrt{t}} \sum_{s=1}^t g(V_s) \rightarrow_D \int_{-\infty}^{\infty} g(v) dv L_{B_V}(1, 0), \quad (3.26)$$

when Assumption 3.6(i) is satisfied and $\int_{-\infty}^{\infty} |g(v)| dv < \infty$. In this case, it can be shown that the OLS estimator of A is rate n consistent.

Thus, if $g(v)$ is a ‘small’ nonparametric departure function in the equation specification then rate n convergence is possible in the estimation of A . On the other hand, rate \sqrt{n} convergence for the SLS estimator of A is possible when $g(v)$ belongs to a general class of functions, including certain polynomial functions. In other words, (1.1) may be treated as either a semiparametric model with $g(V_t)$ being a stochastic trend component or as an approximate linear model with $g(v)$ being a

‘small’ departure function. In the latter case, a super n rate of convergence is achievable in the estimation of A . But in the former case, SLS estimation can only achieve the conventional \sqrt{n} rate of convergence.

Remark 3.3. As in other nonparametric and semiparametric estimation problems, bandwidth parameter choice is critical in the practical implementation of the proposed estimation procedure. In the case where V_t is stationary, existing studies (see, for example, §2.1.3 of Härdle *et al* 2000) may be used to provide solutions. Section 4.1 proposes a semiparametric cross-validation selection method and provides some examples of its implementation.

4 Examples of Implementation

4.1 Bandwidth parameter choice

In the case where V_i is stationary, many existing studies (see, for example, §2.1.3 of Härdle *et al* 2000) offer solutions to bandwidth choice. In nonstationary regressor cases, the literature on bandwidth selection is much more limited (see, however, the analysis in Wang and Phillips, 2009a, 2009b) and many issues remain to be investigated. The present section provides some discussion of the issue in the semiparametric setting considered here.

We start by introducing the leave-one-out estimators of $H(v)$, $\Psi(v)$ and $g(v)$ as follows:

$$\tilde{H}_t(V_t) = \sum_{s=1, \neq t}^n W_{ns}^{(-t)}(V_t) X_s \quad \text{and} \quad \tilde{\Psi}_t(V_t) = \sum_{s=1, \neq t}^n W_{ns}^{(-t)}(V_t) Y_s, \quad (4.1)$$

$$g_{tn}(V_t; A) = \sum_{s=1, \neq t}^n W_{ns}^{(-t)}(V_t) (Y_s - A X_s) = \tilde{\Psi}_t(V_t) - A \tilde{H}_t(V_t), \quad (4.2)$$

where $W_{ns}^{(-t)}(V_t) = K\left(\frac{V_s - V_t}{h}\right) / \sum_{k=1, \neq t}^n K\left(\frac{V_k - V_t}{h}\right)$. Define the leave-one-out semiparametric instrumental variable least squares (SIV) estimator of A by

$$\tilde{A} = \tilde{A}(h) = \bar{Y}' \bar{Q} (\bar{X}' \bar{Q})^{-1}, \quad (4.3)$$

where $\bar{X}' = (\bar{X}_1, \dots, \bar{X}_n)$, $\bar{X}_t = (X_t - \tilde{H}_t(V_t)) \bar{F}_t = \left(X_t - \sum_{s=1, \neq t}^n W_{ns}^{(-t)}(V_t) X_s\right) \bar{F}_t$, $\bar{Q}' = (\bar{Q}_1, \dots, \bar{Q}_n)$, $\bar{Q}_t = (Q_t - \tilde{J}_t(V_t)) \bar{F}_t = \left(Q_t - \sum_{s=1, \neq t}^n W_{ns}^{(-t)}(V_t) Q_s\right) \bar{F}_t$, $\bar{Y}' = (\bar{Y}_1, \dots, \bar{Y}_n)$ and $\bar{Y}_t = (Y_t - \tilde{\Psi}_t(V_t)) \bar{F}_t = \left(Y_t - \sum_{s=1, \neq t}^n W_{ns}^{(-t)}(V_t) Y_s\right) \bar{F}_t$, in which $\bar{F}_t = I(\bar{p}_{n,t}(V_t) > b_n)$ with $\bar{p}_{n,t}(V_t) = \frac{1}{\sqrt{nh}} \sum_{s=1, \neq t}^n K\left(\frac{V_s - V_t}{h}\right)$. The corresponding leave-one-out estimator of $g(\cdot)$ is

$$\tilde{g}(\cdot; h) = g_n(\cdot; \tilde{A}(h)). \quad (4.4)$$

The leave-one-out cross-validation (CV) function is defined by

$$\text{CV}(h) = \frac{1}{n} \sum_{t=1}^n \left(Y_t - \tilde{A}X_t - \tilde{g}_t(V_t) \right)' \left(Y_t - \tilde{A}X_t - \tilde{g}_t(V_t) \right), \quad (4.5)$$

where $\tilde{g}_t(V_t) = g_{tn}(V_t; \tilde{A})$. The optimal smoothing parameter \tilde{h} is chosen so that

$$\text{CV}(\tilde{h}) = \min_{h \in H_n} \text{CV}(h), \quad (4.6)$$

where $H_n = [c_1 n^{-1}, c_2 n^{-1+c_3}]$, in which $0 < c_i < \infty$ for $i = 1, 2$ and $0 < c_3 \leq 1$ are chosen such that \tilde{h} is achievable and locally unique in each individual case. The corresponding data-determined estimators of A and $g(\cdot)$ are then given by

$$\tilde{A}^* = \tilde{A}(\tilde{h}), \quad \text{and} \quad \tilde{g}^*(v) = g_n(v; \tilde{A}(\tilde{h})), \quad (4.7)$$

where $g_n(v; A)$ is defined in (2.14).

Preliminary study of the asymptotic behavior of \tilde{h} shows that \tilde{h} is proportional to $(\sqrt{n})^{-\frac{1}{5}}$ which, allowing for the nonstationarity of V_t , is comparable to the usual $n^{-\frac{1}{5}}$ bandwidth rate in the stationary case. The correspondence arises because in the integrated time series case, the amount of time spent by the process around any particular spatial point is of the order \sqrt{n} rather than n (see Phillips, 2001).

The following examples show how to implement the proposed procedure. Throughout these examples, the kernel is $K(x) = \frac{1}{2}I_{[-1,1]}(x)$, and the optimal bandwidth \tilde{h} is chosen as shown above.

4.2 Simulated examples

Example 4.1 below demonstrates how the functional forms of $g(\cdot)$ and $H(\cdot)$ may affect the rate of convergence of \hat{A} in the exogenous case using SLS estimation. In this case, $\eta_t = U_t$ and $J(\cdot) = H(\cdot)$. The following discussion looks at two pairs of $(G(\cdot), H(\cdot))$ such that the conditions in Assumption 3.4(i)(ii) are satisfied. Example 4.2 examines an endogenous case where the parametric variables are linearly related with the residuals. The SIV estimation method proposed in Section 2.2 is implemented.

Example 4.1. Consider the semiparametric simultaneous equation model

$$Y_t = A X_t + G(V_t) + e_t, \quad (4.8)$$

where A is the 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} -0.5 & 0.6 \\ 0.6 & -0.5 \end{pmatrix},$$

$X_t = (X_{t1}, X_{t2})'$ is a vector of time series regressors, V_t is a sequence of integrated time series regressors of the form $V_t = V_{t-1} + v_t$ with $V_0 = 0$ and v_t is a sequence of stationary disturbances generated by $v_t = \gamma v_{t-1} + \zeta_t$, for $t = 1, 2, \dots$, where

$\gamma \in \{0, 0.5, 0.9\}$, $v_0 = 0$ and ζ_t is a sequence of independent $N(0, 1)$ errors, $G(\cdot) = (g_1(\cdot), g_2(\cdot))'$ is a vector of functions (specified below), and e_t is an error vector generated from

$$e_t \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.6 \\ -0.6 & 1 \end{pmatrix} \right). \quad (4.9)$$

Independently from ϵ_t , the residuals U_t are generated as

$$U_t = \begin{pmatrix} 0.3 & 0 \\ 0 & -0.3 \end{pmatrix} U_{t-1} + \mu_t, \quad t = 1, 2, \dots, \quad (4.10)$$

where $U_0 = (0, 0)'$ and μ_t is a vector of i.i.d. normal errors of the form

$$\mu_t \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right). \quad (4.11)$$

Following functions are used in the model specification:

$$g_1(v) = \sin(v), \quad g_2(v) = \cos(v) \quad \text{and} \quad H_1(v) = H_2(v) = v. \quad (4.12)$$

The process X_t is generated by $X_t = H(V_t) + U_t$ and Y_t is generated by (4.8).

The estimation method proposed in Section 2.1 is applied to estimate A , and $G(\cdot)$ and $H(\cdot)$. We assess finite sample performance using the measures

$$\begin{aligned} \text{ASE}_1 &= |\hat{a}_{11} - a_{11}|, & \text{ASE}_2 &= |\hat{a}_{12} - a_{12}|, \\ \text{ASE}_3 &= |\hat{a}_{21} - a_{21}|, & \text{ASE}_4 &= |\hat{a}_{22} - a_{22}|, \end{aligned}$$

where \hat{a}_{ij} is the (i, j) -th element of \hat{A} averaged over the replications.

For $i = 1, 2$ and $1 \leq j \leq 1000$, let $\hat{H}_{i,j}(\cdot)$ be the estimate of $H_i(\cdot)$ at the j -th replication, $V_{(1)}(j) \leq V_{(2)}(j) \leq \dots \leq V_{(n)}(j)$ be the order statistics of V_t at the j -th replication, $\hat{H}_i(\cdot) = \frac{1}{1000} \sum_{j=1}^{1000} \hat{H}_{i,j}(\cdot)$ and $V_{(t)} = \frac{1}{1000} \sum_{j=1}^{1000} V_{(t)}(j)$. Figure 4.1(a) shows a plot for \hat{H}_1 and its 95% confidence interval (CI) against $(V_{(1)}, \dots, V_{(n)})$ for $\gamma = 0$ and $n = 502$, and Figure 4.1(b) shows a plot for \hat{H}_2 and its 95% confidence interval against $(V_{(1)}, \dots, V_{(n)})$ for $\gamma = 0.5$ and $n = 502$.

The simulation results for both the absolute errors and standard deviations given in Table 4.1 are based on averages over 1000 replications. In the case of (4.12), the conditions of Theorem 3.1 all hold. and the table provides finite sample evidence of the limit theory of Theorem 3.1 for integrated nonparametric regressors. In addition, Table 4.1 shows that the dependence structure of v_t can affect the magnitude of the errors - especially when γ is as large as 0.9, the signal in V_t is stronger and the error diagnostics are smaller.

Table 4.1. Finite Sample Performance of Semiparametric Least Squares Estimation based on model (4.8)

	absolute error			standard deviation		
	$\gamma = 0$					
n	202	502	802	202	502	802
ASE ₁	0.1279	0.1196	0.1186	0.0830	0.0606	0.0465
ASE ₂	0.1302	0.1181	0.1182	0.0816	0.0581	0.0476
ASE ₃	0.0812	0.0482	0.0374	0.0604	0.0362	0.0288
ASE ₄	0.0755	0.0467	0.0368	0.0568	0.0356	0.0277
	$\gamma = 0.5$					
ASE ₁	0.1060	0.0948	0.0894	0.0749	0.0547	0.0445
ASE ₂	0.1065	0.0901	0.0902	0.0756	0.0535	0.0444
ASE ₃	0.0744	0.0476	0.0379	0.0580	0.0359	0.0285
ASE ₄	0.0718	0.0459	0.0376	0.0560	0.0349	0.0276
	$\gamma = 0.9$					
ASE ₁	0.0693	0.0427	0.0333	0.0508	0.0333	0.0262
ASE ₂	0.0698	0.0419	0.0335	0.0511	0.0330	0.0254
ASE ₃	0.0699	0.0421	0.0329	0.0520	0.0316	0.0247
ASE ₄	0.0700	0.0422	0.0331	0.0521	0.0321	0.0249

Example 4.2. We consider a simultaneous system of the form

$$Y_t = A X_t + G(V_t) + e_t, \quad (4.13)$$

where A is the 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 1.0 & 0.6 \\ 0.6 & 1.0 \end{pmatrix},$$

$X_t = (X_{t1}, X_{t2})'$ is a vector of time series regressors, V_t is a sequence of integrated time series regressors following $V_t = V_{t-1} + v_t$ with $V_0 = 0$ and v_t a sequence of stationary disturbances generated by $v_t = \gamma v_{t-1} + \zeta_t$, for $t = 1, 2, \dots$, where $\gamma = 0.1, 0.5, 0.9$, $v_0 = 0$ and ζ_t is a sequence of independent $N(0, 1)$ errors, $G(\cdot) = (g_1(\cdot), g_2(\cdot))'$ is a vector of functions (specified below), and ϵ_t is generated by $e_t = \rho U_t + \mu_t$ with $\rho \in \{0, 0.5, 0.9\}$ and where μ_t and U_t are two errors independently generated as $\mu_t \sim N(0, I_2)$ and $U_t \sim N(0, I_2)$.

Choose $J(v) = H(v)$ and the following functions:

$$g_1(v) = \cos(v), \quad g_2(v) = \sin(v), \quad H_1(v) = v \cos(v), \quad H_2(v) = v \sin(v). \quad (4.14)$$

The process X_t follows $X_t = H(V_t) + U_t$ and Y_t is generated by (4.13). We estimate A by \hat{A}^* of (2.26) with the choice of $e_t = \rho I_2 U_t + \mu_t$, $Q_t = J(V_t) + \eta_t$ and $\eta_t = U_t - \rho I_2 \mu_t$,

in which I_2 denotes the two-dimensional identity matrix and ρ is estimated by (2.25) when computing \hat{A}^* and (4.15) below. Note that the estimation procedure is a restricted one such that $0 < \hat{\rho} < 1$.

Table 4.2. Finite Sample Performance of Semiparametric IV Estimation based on model (4.13) with $\rho = 0.5$

	absolute error			standard deviation		
	$\gamma = 0.1$					
n	202	502	802	202	502	802
ASE ₁ [*]	0.0741	0.0464	0.0378	0.0358	0.0222	0.0182
ASE ₂ [*]	0.0129	0.0051	0.0033	0.0130	0.0051	0.0035
ASE ₃ [*]	0.0128	0.0048	0.0032	0.0132	0.0045	0.0033
ASE ₄ [*]	0.0733	0.0466	0.0378	0.0358	0.0225	0.0182
	$\gamma = 0.5$					
ASE ₁ [*]	0.0420	0.0276	0.0211	0.0219	0.0138	0.0106
ASE ₂ [*]	0.0069	0.0029	0.0018	0.0071	0.0029	0.0018
ASE ₃ [*]	0.0072	0.0030	0.0018	0.0077	0.0030	0.0018
ASE ₄ [*]	0.0417	0.0278	0.0210	0.0220	0.0136	0.0103
	$\gamma = 0.9$					
ASE ₁ [*]	0.0103	0.0058	0.0044	0.0059	0.0033	0.0022
ASE ₂ [*]	0.0016	0.0017	0.0004	0.0017	0.0021	0.0004
ASE ₃ [*]	0.0016	0.0016	0.0004	0.0017	0.0022	0.0004
ASE ₄ [*]	0.0102	0.0059	0.0044	0.0059	0.0034	0.0022

Define the following quantities:

$$\begin{aligned} \text{ASE}_1^* &= |\hat{a}_{11}^* - a_{11}|, & \text{ASE}_2^* &= |\hat{a}_{12}^* - a_{12}|, \\ \text{ASE}_3^* &= |\hat{a}_{21}^* - a_{21}|, & \text{ASE}_4^* &= |\hat{a}_{22}^* - a_{22}|, \end{aligned} \quad (4.15)$$

where \hat{a}_{ij}^* is the (i, j) -th element of \hat{A}^* .

The simulation results for both the absolute errors and standard deviations are based on 1000 replications and the means of the following quantities are tabulated in Table 4.2 for the case of $\rho = 0.5$. Corresponding results for the cases of $\rho = 0$ and $\rho = 0.9$ are available upon request.

The absolute errors and the standard deviations in Table 4.2 together show that the proposed estimation method performs well for the linear endogenous case where

$$Y_t = AX_t + G(V_t) + e_t \quad \text{and} \quad e_t = \rho U_t + \mu_t, \quad (4.16)$$

where U_t and μ_t are vectors of mutually independent time series errors. In addition, the results show that the proposed estimation method is quite robust with respect to the values of γ and (although not reported here) ρ .

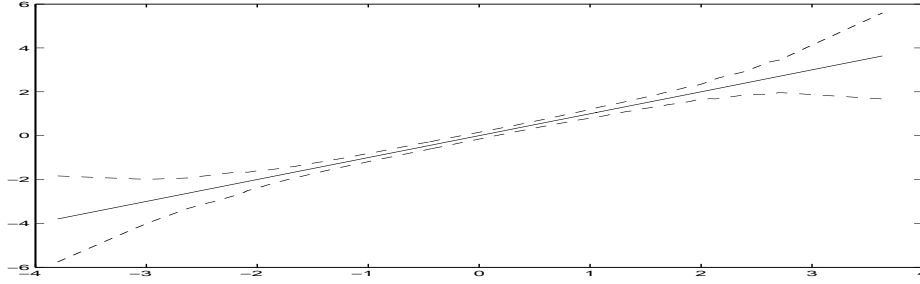


Figure 4.1(a) Nonparametric estimate and 95% confidence interval for $H_1(v) = v$ in the case of $\gamma = 0$.

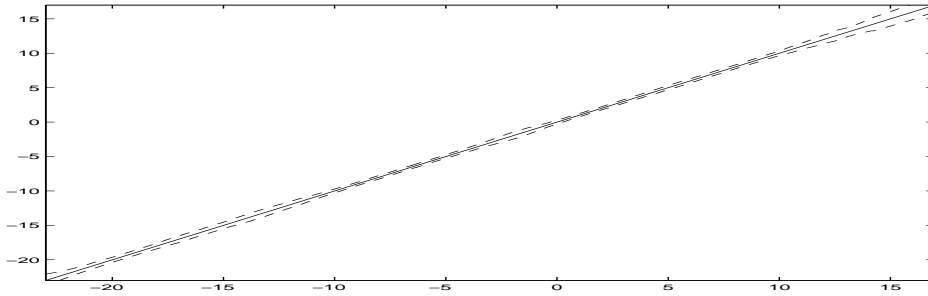


Figure 4.1(b) Nonparametric estimate and 95% confidence interval for $H_1(v) = v$ in the case of $\gamma = 0.5$.

For $i = 1, 2$ and $1 \leq j \leq 1000$, let $\hat{g}_{i,j}(\cdot)$ be the estimate of $g_i(\cdot)$ at the j -th replication, $V_{(1)}(j) \leq V_{(2)}(j) \leq \dots \leq V_{(n)}(j)$ be the order statistics of V_t at the j -th replication, $\hat{g}_i(\cdot) = \frac{1}{1000} \sum_{j=1}^{1000} \hat{g}_{i,j}(\cdot)$ and $V_{(t)} = \frac{1}{1000} \sum_{j=1}^{1000} V_{(t)}(j)$. Figures 4.2(a) shows a plot for \hat{g}_1 and its 95% confidence interval against $(V_{(1)}, \dots, V_{(n)})$ for $\rho = \gamma = 0$ and $n = 502$, and Figure 4.2(b) shows a plot for \hat{g}_2 and its 95% confidence interval against $(V_{(1)}, \dots, V_{(n)})$ for $\rho = \gamma = 0.5$ and $n = 502$.

5 Conclusions and Discussions

This paper explores estimation of a finite dimensional parameter matrix and nonparametric function estimation in the context of a multiple equation nonlinear simultaneous equations model of the form (1.1) in which stochastic trends of unknown form may be present. The proposed semiparametric instrumental variable (SIV) least squares procedure addresses endogeneity in the parametric regressors and enables asymptotically consistent estimation of the nonparametric functions.

The framework here extends univariate semiparametric regression with both independent and stationary regressors and errors to a multivariate case where both the parametric and nonparametric regressors may be nonstationary. A nonparametric kernel estimation method is used to eliminate the nonlinear components and construct an approximating parametric model which leads to the SIV estimator. The SIV estimator resolves endogeneity in the parametric regressors in a semiparametric setting that allows for possible stochastic trends in the generating mechanism for both the endogenous and exogenous regressors, thereby making the model and

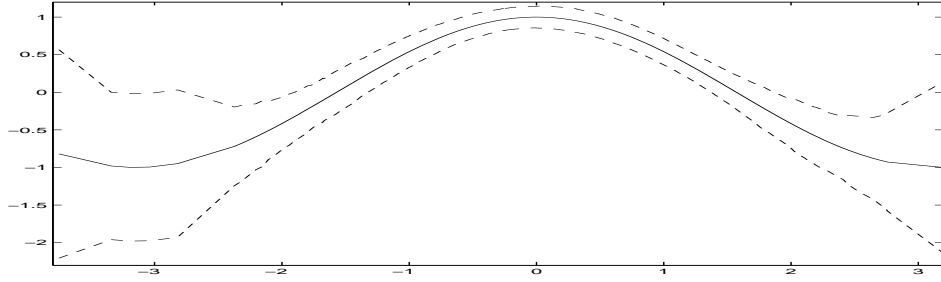


Figure 4.2(a) Nonparametric estimate and 95% confidence interval for $g_1(v) = \cos(v)$ in the case of $\rho = \gamma = 0$.

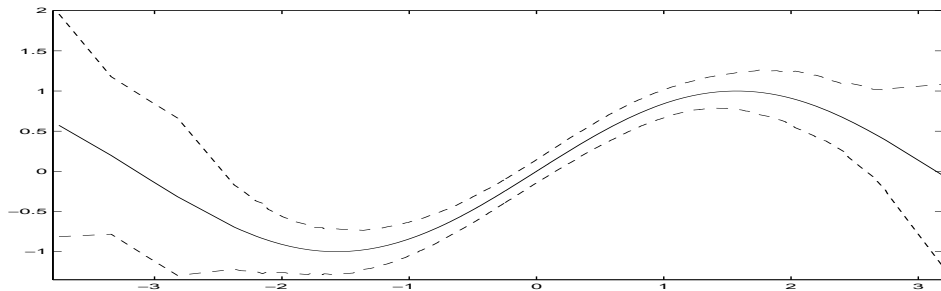


Figure 4.2(b) Nonparametric estimate and 95% confidence interval for $g_2(v) = \sin(v)$ in the case of $\rho = \gamma = 0$.

method relevant in many potential applications where the regressors may be endogenous, stochastic trends may be present in the data, and nonlinearities may occur in the generating mechanism. Simulations reveal that the proposed estimation method is easily implemented in practice and performs well in relation to the asymptotic theory for moderately sized samples.

While the nonparametric stochastic detrending approach explored here has the advantage of imposing only weak conditions on the trend functions, the \sqrt{n} convergence rate is below the usual n rate for cointegrated system estimation and may be improved in some cases. This has been briefly discussed in Section 3.2. A further limitation is the assumption of exogeneity for the nonstationary regressor V_t . It will certainly be useful for empirical applications to show that this condition may be relaxed to allow the trending mechanism to be endogenous. Another limitation is that each component of $g(\cdot)$ is a scalar function of V_t . For practical work, it will often be useful for $g(\cdot)$ to be a function of several regressors involving both stationary and integrated components. A further generalization of the present model is to a functional coefficient system

$$Y_t = A(U_t, V_t)X_t + e_t, \quad (5.1)$$

where $A(u, v)$ is a matrix of unknown coefficients, both V_t and X_t are integrated, $\{U_t\}$ is a vector of stationary regressors, and $\{e_t\}$ is the same as in (1.1). The system (5.1) extends the functional coefficient model of Cai *et al* (2009). These issues require different treatment of the asymptotics and some further development of the methods discussed here, so they are left for future research.

6 Acknowledgments

This work was started when the first author was visiting the Cowles Foundation for Research in Economics at Yale University between September and November 2007. The work of the first author was supported financially by the Cowles Foundation and two Australian Research Council Discovery Grants under Grant Numbers: DP0879088 and DP1096374. The work of the second author was supported by the Kelly Foundation and the NSF under Grant Nos. SES 06–47086 and 09–56687. The first author would also like to acknowledge useful discussions with Xiaohong Chen and Yuichi Kitamura. Thanks from the first author also go to Jiying Yin for excellent computing assistance.

Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East, VIC 3145, Australia. Email: jiti.gao@monash.edu.
Peter C B Phillips, Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520, USA. Email: peter.phillips@yale.edu.

7 Appendix A

7.1 Discussion of Assumptions 3.1–3.5

Assumption 3.1 is quite general allowing for a stationary dependence structure for ξ_t and ζ_t . Under some additional technical conditions, these time series might be stationary linear processes that are also α -mixing (see Corollary 4 of Withers 1981 for example).

Assumption 3.2(i) is needed to ensure that Q_t is a vector of valid instrumental variables when $E[e_t \otimes \eta_t] \neq 0$. Assumption 3.2(ii) is needed to deal with quadratic forms involving e_s and η_t . As pointed out in the beginning of Section 2.2, η_t is a vector of stationary detrended errors. Thus, it is not unreasonable to require η_t to be stationary, although Q_t can be nonstationary. Assumptions 3.2(ii)–(iv) are needed for the main theorems.

Assumption 3.3(i) imposes independence between V_t and (e_s, U_s, η_s) , which is restrictive in a cointegrating regression context. However, recent findings by Wang and Phillips (2009b) lead us to conjecture that some of our limit theory may extend to the case where V_t is endogenous.

Assumption 3.3(ii) allows for a general nonstationary structure by imposing conditions on both the marginal and conditional density functions of a normalized increment of V_t . To justify Assumption 3.3(ii), consider the case where V_t is generated by a random walk model of the form

$$V_t = V_{t-1} + v_t, \quad t \geq 1, \quad (\text{A.1})$$

where $V_0 = 0$ and $\{v_t\}$ is a stationary linear process with $E[v_1] = 0$ and $0 < E[v_1^2] < \infty$. Similarly to arguments used in the proofs of Corollaries 2.1 and 2.2 of Wang and Phillips (2009a), Assumption 3.3(ii) can be verified under (A.1). The rest of this verification considers the case where v_t is a sequence of i.i.d. errors. In this case, Assumption 3.3(ii) implies the following useful results: For $k > i$, let $\hat{\phi}_{i,k}(x)$ be the probability density function of $\frac{1}{\sqrt{k-i} \sigma_v} \sum_{t=i+1}^k v_t$ and $\hat{\phi}_{i,k}(x|\mathcal{F}_i)$ be the conditional probability density function of $\frac{1}{\sqrt{k-i} \sigma_v} \sum_{t=i+1}^k v_t$ given $\{\mathcal{F}_i\}$, which is a sequence of σ -fields generated by $\{v_j : 1 \leq j \leq$

$i\}$ such that V_i is adapted to \mathcal{F}_i , and $\sigma_v^2 = \text{var}(v_1)$. Then as $k - i \rightarrow \infty$,

$$\sup_{x \in \mathbb{R}^1} \left| \widehat{\phi}_{i,k}(x) - \phi(x) \right| \rightarrow_{a.s.} 0 \quad \text{and} \quad (\text{A.2})$$

$$\max_{i \geq 1} \sup_{x \in \mathbb{R}^1} \left| \widehat{\phi}_{i,k}(x | \mathcal{F}_i) - \phi(x) \right| \rightarrow_{a.s.} 0, \quad (\text{A.3})$$

where $\phi(\cdot)$ is the probability density function of the standard normal $N(0, 1)$. The derivation of (A.2) and (A.3) follows from standard theory (see, for example, the first part of the proof of Corollary 2.2 in Wang and Phillips 2009a).

Assumption 3.4 imposes certain conditions on the smoothness of $g(\cdot)$, $H(\cdot)$ and $J(\cdot)$ as well as on the density function $f_{t,0}(v)$. Such conditions are needed in the nonstationary case to make sure that each of the bias terms involved is negligible. When V_t is a random walk of the form (A.1), Assumption 3.4(i) is easily verifiable. Let $g(v) = \theta_0 + \theta_1 v + \theta_2 v^{1+\lambda_0}$ for $0 < \lambda_0 < \frac{1}{2}$, $n^{\lambda_0} h = O(1)$ and $f_{t,0}(v) = O(v^{-(1+2\lambda_0+\varepsilon_0)})$ for some $\varepsilon_0 > 0$ as $t \rightarrow \infty$ and $v \rightarrow \infty$. It then follows that

$$\sum_{t=1}^n \int \left\| g^{(1)}(\varphi_t^{-1} v) \right\|^2 f_{t,0}(v) dv = O\left(\sum_{t=1}^n \varphi_t^{-2\delta_0} \right) = O(n^{1+\lambda_0}), \quad (\text{A.4})$$

which implies Assumption 3.4(i).

The first part of Assumption 3.4(ii) is similarly verifiable. Moreover, the second part of Assumption 3.4(ii) covers the case where both $g(v) = \theta_0 + \theta_1 v$ and $H(v) = \phi_0 + \phi_1 v$. Technically, this is because one may choose $h = O\left(n^{-\frac{1+2\varepsilon_1}{4}} \log^{-2}(n)\right)$ and $b_n = \log^{-1}(n)$ such that $n^{\frac{1}{2}+\varepsilon_1} h^2 b_n^{-2} = O(1)$ for some small $\varepsilon_1 > 0$. The verification of Assumption 3.4(iii) follows in a similar way.

Assumption 3.5(i) is a natural condition on the kernel function and has been used by many authors in the stationary time series case. Assumption 3.5(ii) requires that the rate $b_n^{-2} \rightarrow \infty$ is slower than $\sqrt{h} \rightarrow 0$ and the rate $b_n^4 \rightarrow 0$ is slower than that of $\sqrt{nh} \rightarrow \infty$. Such conditions are satisfied in various cases. For instance, if $b_n = c_b \log^{-1}(n)$ and $h_n = c_h n^{-\zeta_0}$ for some $c_b > 0$, $c_h > 0$ and $\varepsilon_0 < \zeta_0 < \beta - \varepsilon_0$, then Assumption 3.5(ii) holds automatically.

We now verify Assumption 3.5(iii). Note that in order to verify Assumption 3.5(iii), it suffices to show that

$$P(\widehat{p}_n(V_t) \leq b_n) \rightarrow 0, \quad \text{or} \quad P(\widehat{p}_n(V_t) > b_n) \rightarrow 1, \quad (\text{A.5})$$

uniformly in all $t \geq 1$ as $n \rightarrow \infty$.

Consider (A.1) in the case where v_t is a sequence of i.i.d. errors. Note that $\widehat{p}_n(V_t) = \frac{1}{\sqrt{nh}} \sum_{k=1}^n K\left(\frac{V_k - V_t}{h}\right)$. Define $\bar{V}_k(t) = \sum_{i=k+1}^t v_i$ for $t > k$ and $\tilde{V}_k(t) = \sum_{j=t+1}^k v_j$ for $k > t$. Since the kernel function $K(\cdot)$ is symmetric and V_k has independent increments, we have uniformly in $1 \leq t \leq \lfloor \frac{n}{2} \rfloor$,

$$\begin{aligned} \widehat{p}_n(V_t) &= \frac{1}{\sqrt{nh}} \sum_{k=1}^{t-1} K\left(\frac{\bar{V}_k(t)}{h}\right) + \frac{1}{\sqrt{nh}} \sum_{k=t+1}^n K\left(\frac{\tilde{V}_k(t)}{h}\right) + \frac{1}{\sqrt{nh}} K(0) \\ &\geq \frac{1}{\sqrt{nh}} \sum_{k=t+1}^n K\left(\frac{\tilde{V}_k(t)}{h}\right) + o_P(1) = \frac{\sqrt{n-t}}{\sqrt{n}} \frac{1}{\sqrt{n-th}} \sum_{i=1}^{n-t} K\left(\frac{\tilde{V}_{t+i}(t)}{h}\right) + o_P(1) \\ &\equiv \frac{\sqrt{n-t}}{\sqrt{n}} \tilde{p}_{(n-t)}(0) + o_P(1), \end{aligned} \quad (\text{A.6})$$

where $\tilde{p}_{(n-t)}(0) = \frac{1}{\sqrt{n-th}} \sum_{i=1}^{n-t} K\left(\frac{\tilde{V}_{t+i}(t)}{h}\right) \rightarrow_D p_s(0) = L_{B_v}(1,0)$ by Theorem 2.1 of Wang and Phillips (2009a), in which $L_{B_v}(1,0)$ is the local-time process associated with the Gaussian process $B_v(r)$ as the weak limiting distribution of $V_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} v_t$.

7.2 Technical lemmas

To prove the main theorems, we use the following lemmas.

LEMMA A.1. (i) *Under the conditions of Theorem 3.1, as $n \rightarrow \infty$*

$$\frac{1}{n} \tilde{X}' \tilde{Q} = \frac{1}{n} U' \eta + o_P(1) \rightarrow_P E[U_1 \eta'_1]. \quad (\text{A.7})$$

(ii) *Under the conditions of Theorem 3.1, as $n \rightarrow \infty$,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n e_t \otimes \eta_t \rightarrow_D N(0, \Omega_1^*), \quad (\text{A.8})$$

where Ω_1^* is as defined in Assumption 3.2(iv).

LEMMA A.2 *Suppose that $E|X|^p < \infty$ and $E|Y|^q < \infty$, where $p, q > 1$, $p^{-1} + q^{-1} < 1$. Then*

$$|E(XY) - (EX)(EY)| \leq 8(E|X|^p)^{1/p} (E|Y|^q)^{1/q} \alpha^{1-p^{-1}-q^{-1}},$$

where $\alpha = \sup_{A \in \sigma(X), B \in \sigma(Y)} |P(AB) - P(A)P(B)|$.

Since Corollaries 3.1–3.3 in Section 3 are special cases of Theorems 3.1–3.3 respectively, we only prove Theorems 3.1 and 3.2 in this appendix.

7.3 Proof of Theorem 3.1

$$(\hat{A}^* - A) \tilde{X}' \tilde{Q} = \tilde{e}' \tilde{Q} + \tilde{G}' \tilde{Q} = \sum_{t=1}^n e_t \tilde{Q}'_t F_t + \sum_{t=1}^n \tilde{G}_t \tilde{Q}'_t - \sum_{t=1}^n \bar{e}_t \tilde{Q}'_t F_t,$$

in order to prove Theorem 3.1, we need only show that for large enough n

$$\sum_{t=1}^n \tilde{G}_t \tilde{Q}'_t F_t = o_P(\sqrt{n}), \quad (\text{A.9})$$

$$\sum_{t=1}^n \bar{e}_t \tilde{Q}'_t F_t = o_P(\sqrt{n}), \quad (\text{A.10})$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n e_t \tilde{Q}'_t F_t \rightarrow_D N(0, \Omega_1^*), \quad (\text{A.11})$$

where Ω_1^* is as defined in Assumption 3.2(iv), $\tilde{G}_t = G(V_t) - \sum_{k=1}^n w_{nk}(V_t)G(V_k)$, $\tilde{Q}_t = Q_t - \sum_{s=1}^n w_{ns}(V_t)Q_s$ and $\bar{e}_t = \sum_{s=1}^n w_{ns}(V_t)e_s$.

In order to prove (A.9)–(A.11), it suffices to show that for large enough n

$$\sum_{t=1}^n \tilde{G}_t \eta_t' F_t = o_P(\sqrt{n}), \quad (\text{A.12})$$

$$\sum_{t=1}^n \tilde{G}_t \bar{\eta}_t' F_t = o_P(\sqrt{n}), \quad (\text{A.13})$$

$$\sum_{t=1}^n \tilde{G}_t \tilde{J}_t' F_t = o_P(\sqrt{n}), \quad (\text{A.14})$$

$$\sum_{t=1}^n \bar{e}_t \eta_t' F_t = o_P(\sqrt{n}), \quad (\text{A.15})$$

$$\sum_{t=1}^n \bar{e}_t \bar{\eta}_t' F_t = o_P(\sqrt{n}), \quad (\text{A.16})$$

$$\sum_{t=1}^n \bar{e}_t \tilde{J}_t' F_t = o_P(\sqrt{n}), \quad (\text{A.17})$$

$$\sum_{t=1}^n e_t \bar{\eta}_t' F_t = o_P(\sqrt{n}), \quad (\text{A.18})$$

$$\sum_{t=1}^n e_t \tilde{J}_t' F_t = o_P(\sqrt{n}), \quad (\text{A.19})$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n e_t \eta_t' F_t \rightarrow_D N(0, \Omega_1^*), \quad (\text{A.20})$$

where $\bar{\eta}_t = \sum_{s=1}^n w_{ns}(V_t) \eta_s$. Since the finite dimensionality of p and d does not affect the validity of (A.12)–(A.20), we assume without loss of generality that $p = d = 1$ in the rest of the proof of Theorem 3.1 below. As a result, all the vectors involved reduce to scalars.

By Assumption 3.5(i) and the continuity of $g(\cdot)$ and $g^{(1)}(\cdot)$, we have

$$\begin{aligned} & \frac{1}{\sqrt{nh}} \sum_{j=1}^n K\left(\frac{V_j - v}{h}\right) (g(V_j) - g(v)) \\ &= \frac{g^{(1)}(v)}{\sqrt{nh}} \sum_{j=1}^n K\left(\frac{V_j - v}{h}\right) (V_j - v)(1 + o_P(1)). \end{aligned} \quad (\text{A.21})$$

In view of (A.21), in order to prove (A.12), it suffices to show that for n large enough

$$\sum_{t=1}^n \Delta_n(V_t) \eta_t F_t = o_P(\sqrt{n}), \quad (\text{A.22})$$

where $\Delta_n(V_t) = \frac{g^{(1)}(V_t)}{\sqrt{nh\hat{p}_n(V_t)}} \sum_{j=1}^n (V_j - V_t) K\left(\frac{V_j - V_t}{h}\right)$. By Assumption 3.1(i) and Lemma A.2, we have

$$\sum_{t=1}^{\infty} |E[\eta_1 \eta_t]| < \infty, \quad (\text{A.23})$$

which, along with the stationarity of $\{\eta_t\}$, implies that

$$\begin{aligned}
E \left(\sum_{t=1}^n \eta_t \Delta_n(V_t) F_t \right)^2 &= \sum_{t=1}^n E [\eta_t^2 \Delta_n(V_t) F_t]^2 \\
&+ \sum_{t_1=1}^n \sum_{t_2 \neq t_1}^n E [\eta_{t_1} \eta_{t_2} \cdot \Delta_n(V_{t_1}) F_{t_1} \Delta_n(V_{t_2}) F_{t_2}] \\
&\leq C b_n^{-2} \sum_{t=1}^n E [\eta_t^2] E [\Gamma_n(V_t) F_t]^2 \\
&+ C b_n^{-2} \frac{1}{2} \sum_{t_1=1}^n \sum_{t_2 \neq t_1}^n |E [\eta_{t_1} \eta_{t_2}]| E [\Gamma_n^2(V_{t_1}) F_{t_1} + \Gamma_n^2(V_{t_2}) F_{t_2}] \\
&\leq C b_n^{-2} \sum_{t=1}^n E [\Gamma_n(V_t) F_t]^2,
\end{aligned} \tag{A.24}$$

where $\Gamma_n(V_t) = \frac{g^{(1)}(V_t)}{\sqrt{nh}} \sum_{j=1}^n (V_j - V_t) K \left(\frac{V_j - V_t}{h} \right)$.

By Assumption 3.3(i), (A.21)–(A.24) and the definition of $\Delta_n(V_t)$, we have

$$E \left(\sum_{t=1}^n \eta_t \Delta_n(V_t) F_t \right)^2 \leq \Delta_{n,1} + \Delta_{n,2}, \tag{A.25}$$

where

$$\Delta_{n,1} = C b_n^{-2} n^{-1} h^{-2} \sum_{t=1}^n E \left(\left[g^{(1)}(V_t) \right]^2 \sum_{k=1}^n (V_k - V_t)^2 K^2 \left(\frac{V_k - V_t}{h} \right) \right)$$

and

$$\begin{aligned}
\Delta_{n,2} &= C b_n^{-2} n^{-1} h^{-2} \sum_{t=1}^n \\
&\times E \left(\left[g^{(1)}(V_t) \right]^2 \sum_{k_1 \neq k_2} (V_{k_1} - V_t)(V_{k_2} - V_t) K \left(\frac{V_{k_1} - V_t}{h} \right) K \left(\frac{V_{k_2} - V_t}{h} \right) \right).
\end{aligned}$$

First consider $\Delta_{n,1}$. Note that

$$\begin{aligned}
\Delta_{n,1} &= C b_n^{-2} n^{-1} h^{-2} \sum_{t=1}^n E \left(\left[g^{(1)}(V_t) \right]^2 \sum_{k=1}^n (V_k - V_t)^2 K^2 \left(\frac{V_k - V_t}{h} \right) \right) \\
&= C b_n^{-2} n^{-1} h^{-2} \sum_{t=1}^n E \left(\left[g^{(1)}(V_t) \right]^2 \sum_{k=t+1}^n (V_k - V_t)^2 K^2 \left(\frac{V_k - V_t}{h} \right) \right) \\
&+ C b_n^{-2} n^{-1} h^{-2} \sum_{t=1}^n E \left(\left[g^{(1)}(V_t) \right]^2 \sum_{k=1}^t (V_k - V_t)^2 K^2 \left(\frac{V_k - V_t}{h} \right) \right) \\
&=: \Delta_{n,1,1} + \Delta_{n,1,2}.
\end{aligned}$$

For $\Delta_{n,1,1}$, by Assumptions 3.3(ii), 3.4(i) and 3.5(i)(ii), we have

$$\begin{aligned}
\Delta_{n,1,1} &= C b_n^{-2} n^{-1} h^{-2} \sum_{t=1}^n E \left(\left[g^{(1)}(V_t) \right]^2 \sum_{k=t+1}^n (V_k - V_t)^2 K^2 \left(\frac{V_k - V_t}{h} \right) \right) \\
&= C b_n^{-2} n^{-1} h^{-2} \sum_{t=1}^n E \left(\left[g^{(1)}(V_t) \right]^2 \sum_{k=t+1}^n E \left[(V_k - V_t)^2 K^2 \left(\frac{V_k - V_t}{h} \right) \mid \mathcal{F}_t \right] \right)
\end{aligned}$$

$$\begin{aligned}
&= Cb_n^{-2}n^{-1} \sum_{t=1}^n E \left(\left[g^{(1)}(V_t) \right]^2 \sum_{k=t+1}^n \int \left(\frac{v}{\varphi_{k-t}h} \right)^2 K^2 \left(\frac{v}{\varphi_{k-t}h} \right) f_{k,t}(v|\mathcal{F}_t) dv \right) \\
&= Cb_n^{-2}n^{-1}h \sum_{t=1}^n E \left(\left[g^{(1)}(V_t) \right]^2 \sum_{k=t+1}^n \varphi_{k-t} \int u^2 K^2(u) f_{k,t}(u\varphi_{k-t}h|\mathcal{F}_t) du \right) \\
&\leq Cb_n^{-2}n^{-1}h \sum_{t=1}^n E \left[g^{(1)}(V_t) \right]^2 \sum_{k=t+1}^n \varphi_{k-t} \\
&\leq Cb_n^{-2}n^{-\frac{1}{2}}h \sum_{t=1}^n E \left[g^{(1)}(V_t) \right]^2 = o(n).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\Delta_{n,1,2} &= Cb_n^{-2}n^{-1}h^{-2} \sum_{t=1}^n E \left(\left[g^{(1)}(V_t) \right]^2 \sum_{k=1}^t (V_k - V_t)^2 K^2 \left(\frac{V_k - V_t}{h} \right) \right) \\
&= Cb_n^{-2}n^{-1}h^{-2} \sum_{t=1}^n E \left(\sum_{k=1}^t \left[g^{(1)}(V_k + V_t - V_k) \right]^2 (V_k - V_t)^2 K^2 \left(\frac{V_k - V_t}{h} \right) \right) \\
&\leq Cb_n^{-2}n^{-1}h^{-2} \sum_{k=1}^n E \left(\left[g^{(1)}(V_k) \right]^2 \sum_{t=k}^n (V_t - V_k)^2 K^2 \left(\frac{V_t - V_k}{h} \right) \right) \\
&\leq Cb_n^{-2}n^{-\frac{1}{2}}h \sum_{k=1}^n E \left[g^{(1)}(V_k) \right]^2 = o(n).
\end{aligned}$$

We have therefore shown that

$$\Delta_{n,1} = o(n). \quad (\text{A.26})$$

Next consider $\Delta_{n,2}$. Analogous to the calculation of $\Delta_{n,1}$, we need only deal with the case of $k_2 > k_1 > t$ and the other cases can be handled similarly. By Assumptions 3.3(ii), 3.4(i) and 3.5(i)(ii), we have

$$\begin{aligned}
&b_n^{-2}n^{-1}h^{-2} \sum_{t=1}^{n-2} E \left(\left[g^{(1)}(V_t) \right]^2 \sum_{k_1=t+1}^n \sum_{k_2=k_1+1}^n \right. \\
&\quad \times \left. E \left[(V_{k_2} - V_t) (V_{k_1} - V_t) K \left(\frac{V_{k_2} - V_t}{h} \right) K \left(\frac{V_{k_1} - V_t}{h} \right) \middle| \mathcal{F}_t \right] \right) \\
&\leq Cb_n^{-2}n^{-1}h^2 \sum_{t=1}^{n-2} E \left[g^{(1)}(V_t) \right]^2 \sum_{k_1=t+1}^n \sum_{k_2=k_1+1}^n \varphi_{k_2-k_1} \varphi_{k_1-t} \\
&\leq Cb_n^{-2}h^2 \sum_{t=1}^n E \left[g^{(1)}(V_t) \right]^2 \leq O(b_n^{-2}nh) = o(n).
\end{aligned} \quad (\text{A.27})$$

The detailed calculation of (A.27) is similar to the derivations for $\Delta_{n,1,1}$ and $\Delta_{n,1,2}$. Hence, we have shown that $\Delta_{n,2} = o(n)$ holds, which, together with (A.26), implies that (A.12) holds.

We next show that (A.13) holds. In view of (A.21), it suffices to show that

$$\sum_{t=1}^n \hat{\eta}_t \Delta_n(V_t) F_t = o_P(\sqrt{n}), \quad (\text{A.28})$$

where $\hat{\eta}_t = \frac{1}{\sqrt{nh} \hat{p}_n(V_t)} \left(\sum_{k=1}^n K\left(\frac{V_k - V_t}{h}\right) \eta_k \right)$. Similar to the arguments used in (A.24), we have

$$\begin{aligned} & E \left(\sum_{t=1}^n \hat{\eta}_t \Delta_n(V_t) F_t \right)^2 \leq C b_n^{-4} h^{-4} n^{-2} \\ & \times E \left(\sum_{k=1}^n \left\{ \sum_{t=1}^n \sum_{j=1}^n (V_j - V_t) K\left(\frac{V_k - V_t}{h}\right) K\left(\frac{V_j - V_t}{h}\right) \right\} g^{(1)}(V_k) \eta_k \right)^2 \\ & = C b_n^{-4} h^{-2} n^{-2} E \left(\sum_{k=1}^n M(V_k) \eta_k \right)^2, \end{aligned} \quad (\text{A.29})$$

where $M(V_k) = g^{(1)}(V_k) \sum_{t=1}^n \sum_{j=1}^n \left(\frac{V_j - V_t}{h}\right) K\left(\frac{V_k - V_t}{h}\right) K\left(\frac{V_j - V_t}{h}\right)$. Let $\mathcal{F}_V = \sigma(V_t, 1 \leq t \leq n)$. By (A.24), we have

$$E \left(\sum_{k=1}^n M(V_k) \eta_k \right)^2 = E \left(E \left[\left(\sum_{k=1}^n M(V_k) \eta_k \right)^2 \middle| \mathcal{F}_V \right] \right) \leq C \sum_{k=1}^n E (M(V_k))^2, \quad (\text{A.30})$$

which implies that $E \left(\sum_{t=1}^n \hat{\eta}_t \Delta_n(V_t) F_t \right)^2$ is smaller than

$$C b_n^{-4} h^{-2} n^{-2} \sum_{k=1}^n E \left[g^{(1)}(V_k) \sum_{t=1}^n \sum_{j=1}^n \left(\frac{V_j - V_t}{h}\right) K\left(\frac{V_k - V_t}{h}\right) K\left(\frac{V_j - V_t}{h}\right) \right]^2.$$

Note that

$$\begin{aligned} & \sum_{k=1}^n E \left[g^{(1)}(V_k) \sum_{t=1}^n \sum_{j=1}^n \left(\frac{V_j - V_t}{h}\right) K\left(\frac{V_k - V_t}{h}\right) K\left(\frac{V_j - V_t}{h}\right) \right]^2 \\ & = \sum_{k=1}^n \sum_{t_1, t_2=1}^n \sum_{j_1, j_2=1}^n E \left([g^{(1)}(V_k)]^2 \left(\frac{V_{j_1} - V_{t_1}}{h}\right) \left(\frac{V_{j_2} - V_{t_2}}{h}\right) \right. \\ & \quad \left. K\left(\frac{V_k - V_{t_1}}{h}\right) K\left(\frac{V_k - V_{t_2}}{h}\right) K\left(\frac{V_{j_1} - V_{t_1}}{h}\right) K\left(\frac{V_{j_2} - V_{t_2}}{h}\right) \right). \end{aligned}$$

We consider the case where $t_1 > t_2 > j_1 > j_2 > k$ and the other cases can be dealt with analogously. By Assumptions 3.3(ii), 3.4(i) and 3.5, we have

$$\begin{aligned} & \sum_{k=1}^{n-4} \sum_{j_2=k+1}^{n-3} \sum_{j_1=j_2+1}^{n-2} \sum_{t_2=j_1+1}^{n-1} \sum_{t_1=t_2+1}^n E \left([g^{(1)}(V_k)]^2 \left(\frac{V_{j_1} - V_{t_1}}{h}\right) \left(\frac{V_{j_2} - V_{t_2}}{h}\right) \right. \\ & \quad \left. K\left(\frac{V_k - V_{t_1}}{h}\right) K\left(\frac{V_k - V_{t_2}}{h}\right) K\left(\frac{V_{j_1} - V_{t_1}}{h}\right) K\left(\frac{V_{j_2} - V_{t_2}}{h}\right) \right) \\ & \leq C h^4 \sum_{k=1}^n E [g^{(1)}(V_k)]^2 \sum_{j_2=k+1}^{n-3} \sum_{j_1=j_2+1}^{n-2} \sum_{t_2=t_1+1}^{n-1} \sum_{t_1=t_2+1}^n \varphi_{t_1-t_2} \varphi_{t_2-j_1} \varphi_{j_1-j_2} \varphi_{j_2-k} \\ & = O(n^3 h^3). \end{aligned}$$

Equations (A.29) and (A.30) thus imply (A.28) and equation (A.13) is proved. By Assumption 3.3(ii) and (A.23), we have

$$\begin{aligned} E \left(\sum_{t=1}^n \left\{ \sum_{k=1}^n K_{V_t, h}(V_k) e_k \right\} \eta_t \right)^2 &= \sum_{t=1}^n E \left(\left\{ \sum_{k=1}^n K_{V_t, h}(V_k) e_k \right\}^2 \eta_t^2 \right) \\ &+ \sum_{t=1}^n \sum_{s=1, \neq t}^n \sum_{k=1}^n \sum_{l=1}^n E [K_{V_t, h}(V_k) e_k \eta_t K_{V_s, h}(V_l) e_l \eta_s] \\ &= : \Xi_{n,1} + \Xi_{n,2}. \end{aligned} \quad (\text{A.31})$$

By Assumption 3.1(ii) and Lemma A.2, we can show that

$$\sum_{t=1}^{\infty} |E[e_1 e_t]| < \infty \quad \text{and} \quad \sum_{t=1}^{\infty} |E[e_1 \eta_1 e_t \eta_t]| < \infty. \quad (\text{A.32})$$

By A4, (A.32) and using the same arguments as in the derivations for $\Delta_{n,1,1}$ and $\Delta_{n,1,2}$, we have

$$\begin{aligned} \Xi_{n,2} &= \sum_{t=1}^n \sum_{s=1, \neq t}^n \sum_{k=1}^n \sum_{l=1}^n E [K_{V_t, h}(V_k) K_{V_s, h}(V_l)] E [e_k \eta_t e_l \eta_s] \\ &= \frac{1}{h^2} \sum_{t=1}^n \sum_{s=1, \neq t}^n \sum_{k=1}^n \sum_{l=1}^n E \left[K \left(\frac{V_k - V_t}{h} \right) K \left(\frac{V_l - V_s}{h} \right) \right] E [e_k \eta_t e_l \eta_s] \\ &= O \left(nh^{-2} + n^{\frac{3}{2}} \right) = O \left(n^{\frac{3}{2}} h^{-1} \right). \end{aligned} \quad (\text{A.33})$$

Similarly, by Assumptions 3.1(ii), 3.2(ii), 3.3(i) and 3.5(i)(ii), we have

$$\begin{aligned} \Xi_{n,1} &= \frac{1}{h^2} \sum_{t=1}^n \sum_{k=1}^n E \left[K^2 \left(\frac{V_k - V_t}{h} \right) \right] E [e_k^2 \eta_t^2] \\ &+ \frac{1}{h^2} \sum_{t=1}^n \sum_{k=1}^n \sum_{l=1, \neq k}^n E \left[K \left(\frac{V_k - V_t}{h} \right) K \left(\frac{V_l - V_t}{h} \right) \right] E [e_k e_l \eta_t^2] \\ &= O \left(n^{\frac{3}{2}} h^{-1} \right). \end{aligned} \quad (\text{A.34})$$

Thus, by (A.31), (A.33) and (A.34), we have

$$E \left(\sum_{t=1}^n \left\{ \sum_{k=1}^n K_{V_t, h}(V_k) e_k \right\} \eta_t \right)^2 = O(n^{\frac{3}{2}} h^{-1}). \quad (\text{A.35})$$

Recall that $\hat{p}_n(v) = \frac{1}{\sqrt{nh}} \sum_{t=1}^n K \left(\frac{V_t - v}{h} \right)$ and

$$w_{nk}(v) = \frac{K \left(\frac{V_k - v}{h} \right)}{\sum_{t=1}^n K \left(\frac{V_t - v}{h} \right)} = \frac{\frac{1}{\sqrt{nh}} K \left(\frac{V_k - v}{h} \right)}{\frac{1}{\sqrt{nh}} \sum_{t=1}^n K \left(\frac{V_t - v}{h} \right)} = \frac{\frac{1}{\sqrt{nh}} K \left(\frac{V_k - v}{h} \right)}{\hat{p}_n(v)}.$$

Analogous to (A.22), equation (A.35) implies

$$\begin{aligned} &\sum_{t=1}^n \left\{ \sum_{k=1}^n \frac{\frac{1}{\sqrt{nh}} K \left(\frac{V_k - V_t}{h} \right)}{\hat{p}_n(V_t)} e_k \right\} \eta_t F_t \\ &= O_P \left(\frac{1}{\sqrt{n} b_n} \right) \cdot \sum_{t=1}^n \left\{ \sum_{k=1}^n K_{V_t, h}(V_k) e_k \right\} \eta_t \\ &= O_P \left(n^{\frac{1}{4}} h^{-1/2} b_n^{-1} \right) = o_P(\sqrt{n}), \end{aligned} \quad (\text{A.36})$$

by Assumption 3.5(i)(ii). Hence, (A.15) is proved.

We now show that

$$\sum_{t=1}^n \left[\sum_{k=1}^n w_{nk}(V_t) \eta_k \right] \left[\sum_{q=1}^n w_{nq}(V_t) e_q \right] F_t = o_P(\sqrt{n}). \quad (\text{A.37})$$

Note that

$$\begin{aligned} & E \left(\sum_{t=1}^n \left[\sum_{k=1}^n K_{V_t, h}(V_k) \eta_k \right] \left[\sum_{q=1}^n K_{V_t, h}(V_q) e_q \right] \right)^2 \\ &= \sum_{t=1}^n E \left[\left(\sum_{k=1}^n K_{V_t, h}(V_k) \eta_k \right)^2 \left(\sum_{q=1}^n K_{V_t, h}(V_q) e_q \right)^2 \right] \\ &+ \sum_{t_1=1}^n \sum_{t_2 \neq t_1} E \left[\left(\sum_{k_1=1}^n K_{V_{t_1}, h}(V_{k_1}) \eta_{k_1} \right) \left(\sum_{q_1=1}^n K_{V_{t_1}, h}(V_{q_1}) e_{q_1} \right) \right. \\ &\quad \left. \times \left(\sum_{k_2=1}^n K_{V_{t_2}, h}(V_{k_2}) \eta_{k_2} \right) \left(\sum_{q_2=1}^n K_{V_{t_2}, h}(V_{q_2}) e_{q_2} \right) \right] =: I_{n,1} + I_{n,2}. \end{aligned} \quad (\text{A.38})$$

By Assumption 3.3(i), we have

$$\begin{aligned} I_{n,1} &= \sum_{t=1}^n \sum_{k=1}^n \sum_{q=1}^n E \left[K_{V_t, h}^2(V_k) K_{V_t, h}^2(V_q) \right] E \left[\eta_k^2 e_q^2 \right] \\ &+ \sum_{t=1}^n \sum_{k_1=1}^n \sum_{k_2 \neq k_1}^n \sum_{q=1}^n E \left[K_{V_t, h}(V_{k_1}) K_{V_t, h}(V_{k_2}) K_{V_t, h}^2(V_q) \right] E \left[\eta_{k_1} \eta_{k_2} e_q^2 \right] \\ &+ \sum_{t=1}^n \sum_{q_1=1}^n \sum_{q_2 \neq q_1}^n \sum_{k=1}^n E \left[K_{V_t, h}(V_{q_1}) K_{V_t, h}(V_{q_2}) K_{V_t, h}^2(V_k) \right] E \left[\eta_k^2 e_{q_1} e_{q_2} \right] \\ &+ \sum_{t=1}^n \sum_{k_1=1}^n \sum_{k_2 \neq k_1}^n \sum_{q_1=1}^n \sum_{q_2=1, \neq q_1}^n E \left[K_{V_t, h}(V_{k_1}) K_{V_t, h}(V_{k_2}) K_{V_t, h}(V_{q_1}) K_{V_t, h}(V_{q_2}) \right] \\ &\quad \times E \left[\eta_{k_1} \eta_{k_2} e_{q_1} e_{q_2} \right] =: I_{n,1}^{(1)} + I_{n,1}^{(2)} + I_{n,1}^{(3)} + I_{n,1}^{(4)}. \end{aligned} \quad (\text{A.39})$$

By Assumptions 3.3(i) and applying the proof of (A.33), we can show that

$$\begin{aligned} I_{n,1}^{(1)} &= \sum_{t=1}^n \sum_{k=1}^n E \left[K_{V_t, h}^4(V_k) \right] E \left[\eta_k^2 e_k^2 \right] \\ &+ \sum_{t=1}^n \sum_{k=1}^n \sum_{q \neq k}^n E \left[K_{V_t, h}^2(V_k) K_{V_t, h}^2(V_q) \right] \times E \left[\eta_k^2 e_q^2 \right] \\ &= O \left(n^{\frac{3}{2}} h^{-3} + n^2 h^{-2} \right) = O \left(n^2 h^{-2} \right). \end{aligned} \quad (\text{A.40})$$

Similarly, by (A.23) and (A.32), we have

$$I_{n,1}^{(j)} = O(n^2 h^{-2}), \quad j = 2, 3, 4. \quad (\text{A.41})$$

It follows from (A.39)–(A.41) that

$$I_{n,1} = O(n^2 h^{-2}). \quad (\text{A.42})$$

Observe that

$$\begin{aligned}
I_{n,2} &= \sum_{t_1=1}^n \sum_{t_1 \neq t_2}^n \sum_{k=1}^n \sum_{q=1}^n E \left[K_{V_{t_1},h}(V_k) K_{V_{t_2},h}(V_k) K_{V_{t_1},h}(V_q) K_{V_{t_2},h}(V_q) \right] \\
&\times E \left[\eta_k^2 e_q^2 \right] \\
&+ \sum_{t_1=1}^n \sum_{t_1 \neq t_2}^n \sum_{k_1=1}^n \sum_{k_2 \neq k_1}^n \sum_{q=1}^n E \left[K_{V_{t_1},h}(V_{k_1}) K_{V_{t_2},h}(V_{k_2}) K_{V_{t_1},h}(V_q) K_{V_{t_2},h}(V_q) \right] \\
&\times E \left[\eta_{k_1} \eta_{k_2} e_q^2 \right] \\
&+ \sum_{t_1=1}^n \sum_{t_1 \neq t_2}^n \sum_{q_1=1}^n \sum_{q_2 \neq q_1}^n \sum_{k=1}^n E \left[K_{V_{t_1},h}(V_k) K_{V_{t_2},h}(V_k) K_{V_{t_1},h}(V_{q_1}) K_{V_{t_2},h}(V_{q_2}) \right] \\
&\times E \left[\eta_k^2 e_{q_1} e_{q_2} \right] \\
&+ \sum_{t_1=1}^n \sum_{t_1 \neq t_2}^n \sum_{k_1=1}^n \sum_{k_2 \neq k_1}^n \sum_{q_1=1}^n \sum_{q_2 \neq q_1}^n E \left[K_{V_{t_1},h}(V_{k_1}) K_{V_{t_2},h}(V_{k_2}) K_{V_{t_1},h}(V_{q_1}) K_{V_{t_2},h}(V_{q_2}) \right] \\
&\times E \left[\eta_{k_1} \eta_{k_2} e_{q_1} e_{q_2} \right] =: I_{n,2}^{(1)} + I_{n,2}^{(2)} + I_{n,2}^{(3)} + I_{n,2}^{(4)}.
\end{aligned} \tag{A.43}$$

By (A.23) and (A.32) as well as following the calculation of the order of $I_{n,1}^{(j)}$ above, we have

$$I_{n,2}^{(j)} = O\left(n^{\frac{5}{2}} h^{-1}\right), \quad j = 1, \dots, 4. \tag{A.44}$$

By (A.43)–(A.44), we have

$$I_{n,2} = O\left(n^{\frac{5}{2}} h^{-1}\right).$$

This, combined with (A.38) and (A.42), leads to

$$E \left(\sum_{t=1}^n \left[\sum_{k=1}^n K_{V_t,h}(V_k) \eta_k \right] \left[\sum_{q=1}^n K_{V_t,h}(V_q) e_q \right] \right)^2 = O\left(n^{\frac{5}{2}} h^{-1}\right).$$

As a result, by Assumption 3.5(ii) we have

$$\sum_{t=1}^n \left[\sum_{k=1}^n w_{nk}(V_t) \eta_k \right] \left[\sum_{q=1}^n w_{nq}(V_t) e_q \right] F_t = O_P\left(n^{\frac{1}{4}} h^{-1/2} b_n^{-2}\right) = o_P(\sqrt{n}),$$

which implies that (A.16) holds.

Finally, we prove (A.18) and (A.20). The proof of (A.18) is similar to (A.36). By the central limit theorem for stationary α -mixing random variables (see Corollary 5.1 of Hall and Heyde 1980) and Assumption 3.1, we have

$$P \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^n \eta_t e_t < z \right\} \rightarrow \Phi \left(\frac{z}{\sigma_1} \right), \tag{A.45}$$

where $\sigma_1^2 = \Sigma_{e,\eta} > 0$ when the dimension of $\{\eta_t\}$ is assumed to be $d = 1$.

Meanwhile, by Assumptions 3.1(ii) and 3.5(iii) as well as Lemma A.2, we have

$$\begin{aligned}
& E \left(\sum_{t=1}^n \eta_t e_t (1 - F_t) \right)^2 = \sum_{t=1}^n E (\eta_t e_t (1 - F_t))^2 \\
& + 2 \sum_{t=2}^n \sum_{s=1}^{t-1} E (\eta_t \eta_s e_t e_s (1 - F_t)(1 - F_s)) \leq C \sum_{t=1}^n E(1 - F_t) \\
& + 2 \sum_{t=2}^n \sum_{s=1}^{t-1} E (\eta_t e_t \eta_s e_s) E [(1 - F_t)(1 - F_s)] \leq C \sum_{t=1}^n E(1 - F_t) \\
& + \sum_{t=2}^n \sum_{s=1}^{t-1} (\alpha_\zeta (|t - s|))^{\gamma_1/(2+\gamma_2)} E [(1 - F_t)(1 - F_s)] \\
& \leq C \sum_{t=1}^n E(1 - F_t) + C \sum_{t=2}^n \sum_{s=1}^{t-1} (\alpha_U (|t - s|))^{\gamma_1/(2+\gamma_1)} (\alpha_\epsilon (|t - s|))^{\gamma_2/(2+\gamma_2)} E [(1 - F_t)] \\
& \leq C \sum_{t=1}^n E [(1 - F_t)] = C \sum_{t=1}^n P(\hat{p}_n(V_t) \leq b_n) = o(n),
\end{aligned} \tag{A.46}$$

using the fact that

$$E[(1 - F_t)(1 - F_s)] \leq \frac{1}{2} \left(E[(1 - F_t)^2] + E[(1 - F_s)^2] \right) = \frac{1}{2} (E[(1 - F_t)] + E[(1 - F_s)]).$$

By (A.45) and (A.46), equation (A.20) is proved.

We finish the proof of Theorem 3.1 by completing the proofs of (A.14), (A.17) and (A.19). Let $\Lambda_n(V_t)$ be defined as $\Delta_n(V_t)$ with $g^{(1)}(\cdot)$ replaced by $H^{(1)}(\cdot)$. Similarly to the derivations in (A.25)–(A.27), we can show that

$$\begin{aligned}
E \left(\sum_{t=1}^n |\Lambda_n(V_t) \Delta_n(V_t) F_t| \right) &= O \left(b_n^{-2} h^2 \sum_{t=1}^n E \left| H^{(1)}(V_t) g^{(1)}(V_t) \right| \right) \\
&= O \left(n^{\frac{1}{2} - \varepsilon_1} \right) = o(\sqrt{n}),
\end{aligned}$$

for some $0 < \varepsilon_1 < \frac{1}{2}$, which implies that (A.14) holds. The proofs of (A.17) and (A.19) are similar to that of (A.12) and so the details are omitted here.

7.4 Proof of Theorem 3.2

Observe that

$$\begin{aligned}
\hat{g}^*(v) - g(v) &= \sum_{t=1}^n w_{nt}(v) \left(Y_t - \hat{A}^* X_t \right) - g(v) \\
&= \sum_{t=1}^n w_{nt}(v) \epsilon_t + (A - \hat{A}^*) \sum_{t=1}^n w_{nt}(v) X_t + \sum_{t=1}^n w_{nt}(v) g(V_t) - g(v) \\
&= \sum_{t=1}^n w_{nt}(v) \epsilon_t + (A - \hat{A}^*) \sum_{t=1}^n w_{nt}(v) U_t \\
&+ (A - \hat{A}^*) \sum_{t=1}^n w_{nt}(v) H(V_t) + \sum_{t=1}^n w_{nt}(v) [g(V_t) - g(v)].
\end{aligned}$$

Note from Theorem 3.1 that $\hat{A}^* - A = O_P \left(n^{-\frac{1}{2}} \right)$,

$$\sum_{t=1}^n K \left(\frac{v - V_t}{h} \right) = O_P(\sqrt{nh}) \quad \text{and} \quad \sum_{t=1}^n K^2 \left(\frac{v - V_t}{h} \right) = O_P(\sqrt{nh}), \tag{A.47}$$

$$\begin{aligned} \sqrt{\sum_{t=1}^n K\left(\frac{v-V_t}{h}\right)} \sum_{t=1}^n w_{nt}(v) U_t &= \frac{1}{\sqrt{\sum_{t=1}^n K\left(\frac{v-V_t}{h}\right)}} \sum_{t=1}^n K\left(\frac{v-V_t}{h}\right) U_t \\ &\rightarrow_D N(0, \Omega_u), \end{aligned} \quad (\text{A.48})$$

$$\sqrt{\sum_{t=1}^n K\left(\frac{v-V_t}{h}\right)} \sum_{t=1}^n w_{nt}(v) H(V_t) = \sum_{t=1}^n \frac{K\left(\frac{v-V_t}{h}\right) H(V_t)}{\sqrt{\sum_{t=1}^n K\left(\frac{v-V_t}{h}\right)}}, \quad (\text{A.49})$$

$$\sqrt{\sum_{t=1}^n K\left(\frac{v-V_t}{h}\right)} \sum_{t=1}^n w_{nt}(v) [g(V_t) - g(v)] = o(1), \quad (\text{A.50})$$

$$\begin{aligned} \sqrt{\sum_{t=1}^n K\left(\frac{v-V_t}{h}\right)} \sum_{t=1}^n w_{nt}(v) e_t &= \frac{1}{\sqrt{\sum_{t=1}^n K\left(\frac{v-V_t}{h}\right)}} \sum_{t=1}^n K\left(\frac{v-V_t}{h}\right) e_t, \\ &\rightarrow_D N(0, \Omega_e), \end{aligned} \quad (\text{A.51})$$

where $\Omega_u = \int K^2(u) du \cdot E[U_1 U_1']$ and $\Omega_e = \int K^2(u) du \cdot E[e_1 e_1']$.

The proof of (A.47) follows from existing results (see, for example, Theorem 5.1 of Karlsen and Tjøstheim 2001, and Theorem 2.1 of Wang and Phillips 2009a). Similar to the proof of (5.16) and (5.18) of Wang and Phillips (2009a), the proof of (A.50) follows from Assumption 3.5(i)(ii)(iv). The proof of (A.48) is the same as that of (A.51), whose proof is given below. Using Taylor expansions and Assumption 3.4(ii), it can be shown that for n large enough

$$\sum_{t=1}^n w_{nt}(v) H(V_t) = H(v) \sum_{t=1}^n w_{nt}(v) (1 + o_P(1)) = O_P(1). \quad (\text{A.52})$$

In view of (A.47)–(A.52), in order to complete the proof of Theorem 3.2, it suffices to prove (A.51). Let us define $a_{nt}(v) = K\left(\frac{v-V_t}{h}\right)$ and $L_n \equiv \sum_{t=1}^n a_{nt}(v) e_t$. Note that the conditional variance matrix of L_n given $\mathcal{V} = (V_1, \dots, V_n)$ is $\Omega_{11} \cdot \sum_{t=1}^n K^2\left(\frac{v-V_t}{h}\right)$.

Note also that $\{e_t\}$ is assumed to be stationary and α -mixing. Thus, applying existing results (for example, Corollary 5.1 of Hall and Heyde 1980) completes the proof. Alternatively, by the standard small-block and large-block arguments as in the proof of Theorem 2.22 of Fan and Yao (2003), in order to prove (A.51), it suffices to verify the Feller and Lindberg conditions.

7.5 Proof of Theorem 3.3

In view of the definition $\tilde{Z}_t = (Z_t - \sum_{s=1}^n w_{ns}(V_t) Z_s) F_t$, we have

$$\begin{aligned} \tilde{Y}_t &= A\tilde{X}_t + \tilde{g}(V_t) + \tilde{\epsilon}_t = A\tilde{X}_t + \tilde{g}(V_t) + \tilde{\epsilon}_t, \\ Y_t - \hat{A}^* X_t - \hat{g}^*(V_t) &= \tilde{Y}_t - \hat{A}^* \tilde{X}_t \\ &= (A - \hat{A}^*) \tilde{X}_t + \tilde{g}(V_t) + \tilde{\epsilon}_t. \end{aligned} \quad (\text{A.53})$$

Observe that

$$\sum_{t=1}^n (Y_t - \hat{A}^* X_t - \hat{g}_n^*(V_t)) (Y_t - \hat{A}^* X_t - \hat{g}_n^*(V_t))'$$

$$\begin{aligned}
&= \sum_{t=1}^n \left((A - \hat{A}^*) \tilde{X}_t + \tilde{g}(V_t) + \tilde{e}_t \right) \left((A - \hat{A}^*) \tilde{X}_t + \tilde{g}(V_t) + \tilde{e}_t \right)' \\
&= \sum_{t=1}^n \tilde{e}_t \tilde{e}_t' + \sum_{t=1}^n (A - \hat{A}^*) \tilde{X}_t \tilde{X}_t' (A - \hat{A}^*)' + \sum_{t=1}^n \tilde{g}(V_t) \tilde{g}(V_t)' \\
&\quad + 2 \sum_{t=1}^n (A - \hat{A}^*) \tilde{X}_t \tilde{e}_t' + 2 \sum_{t=1}^n (A - \hat{A}^*) \tilde{X}_t \tilde{g}(V_t)' + 2 \sum_{t=1}^n \tilde{g}(V_t) \tilde{e}_t' \\
&\equiv \sum_{j=1}^6 S_n(j). \tag{A.54}
\end{aligned}$$

We show that as $n \rightarrow \infty$

$$\frac{1}{n} S_n(1) \rightarrow_P E[e_1 e_1'] \quad \text{and} \quad \frac{1}{n} S_n(j) \rightarrow_P 0 \tag{A.55}$$

for all $2 \leq j \leq 6$. Note that

$$\sum_{t=1}^n \tilde{e}_t \tilde{e}_t' = \sum_{t=1}^n e_t e_t' F_t + \sum_{t=1}^n \bar{e}_t \bar{e}_t' F_t + 2 \sum_{t=1}^n e_t \bar{e}_t' F_t, \tag{A.56}$$

where $\bar{e}_t = \sum_{s=1}^n w_{ns}(V_t) e_s$. In view of (A.56), in order to prove the first part of (A.55), it suffices to show that as $n \rightarrow \infty$

$$\frac{1}{n} \sum_{t=1}^n e_t e_t' F_t \rightarrow_P E[e_1 e_1'], \quad \frac{1}{n} \sum_{t=1}^n \bar{e}_t \bar{e}_t' F_t \rightarrow_P 0 \quad \text{and} \quad \frac{1}{n} \sum_{t=1}^n e_t \bar{e}_t' F_t \rightarrow_P 0. \tag{A.57}$$

Since the remainder of the proof of (A.57) and the second part of (A.55) is a special case of the proof of Lemma A.1(i) below, we do not repeat it here. In fact, equations (A.2)–(A.10) imply (A.57) and the second part of (A.55) when U_s , η_t , $\tilde{J}(V_t)$ and $\tilde{H}(V_t)$ are replaced by e_s , e_t and $\tilde{g}(V_t)$, respectively.

8 Appendix B

8.1 Proof of Lemma A.1(i)

As in previous proofs, we continue to consider the case $d = 1$ for convenience since the basic ideas hold for $d \geq 2$. Hence, all the vectors, including U_t and η_t , in the rest of the proof reduce to scalars.

Observe that

$$\begin{aligned}
\sum_{t=1}^n \tilde{X}_t \tilde{Q}_t F_t &= \sum_{t=1}^n \left(X_t - \sum_{k=1}^n w_{nk}(V_t) X_k \right) \left(Q_t - \sum_{q=1}^n w_{nq}(V_t) Q_q \right) F_t \tag{A.1} \\
&= \sum_{t=1}^n \left(U_t + \tilde{H}(V_t) - \sum_{k=1}^n w_{nk}(V_t) U_k \right) \left(\eta_t + \tilde{J}(V_t) - \sum_{q=1}^n w_{nq}(V_t) \eta_q \right) F_t \\
&= \sum_{t=1}^n U_t \eta_t F_t - \sum_{t=1}^n \left(\sum_{k=1}^n w_{nk}(V_t) U_k \right) \eta_t F_t - \sum_{t=1}^n \left(\sum_{k=1}^n w_{nk}(V_t) \eta_k \right) U_t F_t \\
&\quad + \sum_{t=1}^n U_t \tilde{J}(V_t) F_t + \sum_{t=1}^n \eta_t \tilde{H}(V_t) F_t
\end{aligned}$$

$$\begin{aligned}
& - \sum_{t=1}^n \left(\sum_{k=1}^n w_{nk}(V_t) U_k \right) \tilde{J}(V_t) F_t - \sum_{t=1}^n \left(\sum_{k=1}^n w_{nk}(V_t) \eta_k \right) \tilde{H}(V_t) F_t \\
& + \sum_{t=1}^n \left(\sum_{k=1}^n w_{nk}(V_t) U_k \right) \left(\sum_{q=1}^n w_{nq}(V_t) \eta_q \right) F_t + \sum_{t=1}^n \tilde{H}(V_t) \tilde{J}(V_t) F_t.
\end{aligned}$$

Similar to (A.12)–(A.20), in order to prove Lemma A.1(i), it suffices to show that

$$\sum_{t=1}^n \left(\sum_{s=1}^n w_{ns}(V_t) U_s \right) \left(\sum_{k=1}^n w_{nk}(V_t) \eta_k \right) F_t = o_P(n), \quad (\text{A.2})$$

$$\sum_{t=1}^n \left(\sum_{s=1}^n w_{ns}(V_t) U_s \right) \eta_t F_t = o_P(n), \quad (\text{A.3})$$

$$\sum_{t=1}^n \left(\sum_{s=1}^n w_{ns}(V_t) \eta_s \right) U_t F_t = o_P(n), \quad (\text{A.4})$$

$$\sum_{t=1}^n \left(\sum_{s=1}^n w_{ns}(V_t) U_s \right) \tilde{J}(V_t) F_t = o_P(n), \quad (\text{A.5})$$

$$\sum_{t=1}^n \left(\sum_{s=1}^n w_{ns}(V_t) \eta_s \right) \tilde{H}(V_t) F_t = o_P(n), \quad (\text{A.6})$$

$$\sum_{t=1}^n \tilde{J}(V_t) U_t F_t = o_P(n), \quad (\text{A.7})$$

$$\sum_{t=1}^n \tilde{H}(V_t) \eta_t F_t = o_P(n), \quad (\text{A.8})$$

$$\sum_{t=1}^n \tilde{H}(V_t) \tilde{J}(V_t) F_t = o_P(n), \quad (\text{A.9})$$

$$\frac{1}{n} \sum_{t=1}^n U_t \eta_t F_t \rightarrow_P \Sigma_{u\eta}, \quad (\text{A.10})$$

where $\Sigma_{u\eta} = E[U_1 \eta_1']$.

In the rest of the proof of Lemma A.1(i), we verify each of the equations (A.2)–(A.9). Since some of the proofs are very similar, we only provide some representative proofs here. Define $\hat{w}_{nk}(V_t) = \frac{1}{\hat{p}_n(V_t) \sqrt{nh}} K\left(\frac{V_t - V_k}{h}\right)$. In order to verify (A.2), it suffices to show that for n large enough

$$\sum_{t=1}^n \left(\sum_{s=1}^n \hat{w}_{ns}(V_t) U_s \right) \left(\sum_{k=1}^n \hat{w}_{nk}(V_t) \eta_k \right) F_t = o_P(n). \quad (\text{A.11})$$

Observe that

$$\begin{aligned}
& E \left[\sum_{t=1}^n \left(\sum_{k=1}^n \hat{w}_{nk}(V_t) U_k \right) \left(\sum_{q=1}^n \hat{w}_{nq}(V_t) \eta_q \right) F_t \right]^2 \\
& = \sum_{t=1}^n E \left[\left(\sum_{k=1}^n \hat{w}_{nk}(V_t) U_k \right)^2 \left(\sum_{q=1}^n \hat{w}_{nq}(V_t) \eta_q \right)^2 F_t \right]
\end{aligned}$$

$$\begin{aligned}
& + \sum_{t_1=1}^n \sum_{t_2 \neq t_1} E \left[\left(\sum_{k_1=1}^n \hat{w}_{nk_1}(V_{t_1}) U_{k_1} \right) \left(\sum_{q_1=1}^n \hat{w}_{nq_1}(V_{t_1}) \eta_{q_1} \right) \right. \\
& \times \left. \left[\left(\sum_{k_2=1}^n \hat{w}_{nk_2}(V_{t_2}) U_{k_2} \right) \left(\sum_{q_2=1}^n \hat{w}_{nq_2}(V_{t_2}) \eta_{q_2} \right) F_{t_1} F_{t_2} \right] \right] \\
=: & \quad \Pi_{n,1} + \Pi_{n,2}. \tag{A.12}
\end{aligned}$$

By Assumption 3.3(i), we have

$$\begin{aligned}
\Pi_{n,1} & \leq C \sum_{t=1}^n \sum_{k=1}^n \sum_{q=1}^n E \left[\left(\hat{w}_{nk}^2(V_t) \hat{w}_{nq}^2(V_t) \right) F_t \right] E \left[U_k^2 \eta_q^2 \right] \\
& + C \sum_{t=1}^n \sum_{k=1}^n \sum_{q=1}^n E \left[\left(\hat{w}_{nk}^3(V_t) \hat{w}_{nq}(V_t) \right) F_t \right] E \left[U_k^3 \eta_q \right] \\
& + C \sum_{t=1}^n \sum_{k=1}^n \sum_{q=1}^n E \left[\left(\hat{w}_{nk}^3(V_t) \hat{w}_{nq}(V_t) \right) F_t \right] E \left[U_q \eta_k^3 \right] \\
& + C \sum_{t=1}^n \sum_{k_1=1}^n \sum_{k_2 < k_1} \sum_{k_3 < k_2} E \left[\hat{w}_{nk_1}^2(V_t) \hat{w}_{nk_2}(V_t) \hat{w}_{nk_3}(V_t) F_t \right] E \left[\eta_{k_1}^2 U_{k_2} U_{k_3} \right] \\
& + C \sum_{t=1}^n \sum_{k_1=1}^n \sum_{k_2 < k_1} \sum_{k_3 < k_2} E \left[\hat{w}_{nk_1}^2(V_t) \hat{w}_{nk_2}(V_t) \hat{w}_{nk_3}(V_t) F_t \right] E \left[U_{k_1}^2 \eta_{k_2} \eta_{k_3} \right] \\
& + C \sum_{t=1}^n \sum_{k_1=1}^n \sum_{k_2 < k_1} \sum_{k_3 < k_2} \sum_{k_4 < k_3} E \left[\hat{w}_{nk_1}(V_t) \hat{w}_{nk_2}(V_t) \hat{w}_{nk_3}(V_t) \hat{w}_{nk_4}(V_t) F_t \right] \\
& \times E \left[U_{k_1} U_{k_2} \eta_{k_3} \eta_{k_4} \right] \\
=: & \quad \Pi_{n,1}(1) + \Pi_{n,1}(2) + \Pi_{n,1}(3) + \Pi_{n,1}(4) + \Pi_{n,1}(5) + \Pi_{n,1}(6). \tag{A.13}
\end{aligned}$$

For $\Pi_{n,1}(1)$, note that

$$\begin{aligned}
\Pi_{n,1}(1) & = \sum_{t=1}^n \sum_{k=1}^n E \left[\hat{w}_{nk}^4(V_t) F_t \right] E \left[U_k^2 \eta_k^2 \right] \\
& + \sum_{t=1}^n \sum_{k=1}^n \sum_{q \neq k} E \left[\hat{w}_{nk}^2(V_t) \hat{w}_{nq}^2(V_t) F_t \right] E \left[U_k^2 \eta_q^2 \right] \\
=: & \quad \Pi_{n,1}(1,1) + \Pi_{n,1}(1,2). \tag{A.14}
\end{aligned}$$

By Assumptions 3.3 and 3.5(ii), we have

$$\begin{aligned}
\Pi_{n,1}(1,1) & = \sum_{t=1}^n \sum_{k=1}^n E \left[\hat{w}_{nk}^4(V_t) F_t \right] E \left[U_k^2 \eta_k^2 \right] \leq C \sum_{t=1}^n \sum_{k=1}^n E \left[\hat{w}_{nk}^4(V_t) F_t \right] \\
& = O \left(n^{-2} h^{-4} b_n^{-4} \left(n K^4(0) + \sum_{t=2}^n \sum_{k=1}^{t-1} E \left[K^4 \left(\frac{V_t - V_k}{h} \right) \right] \right) \right) \\
& = O \left(n^{-2} h^{-4} b_n^{-4} \left(n K^4(0) + \sum_{t=2}^n \sum_{k=1}^{t-1} \int K^4 \left(\frac{u}{\varphi_{t-k} h} \right) f_{t,k}(u) du \right) \right) \\
& = O \left(n^{-2} h^{-4} b_n^{-4} \left(n K^4(0) + \sum_{t=2}^n \sum_{k=1}^{t-1} h \varphi_{t-k} \int K^4(u) f_{t,k}(h \varphi_{t-k} u) du \right) \right) \\
& = O \left(n^{-2} h^{-4} b_n^{-4} \left(n K^4(0) + h \sum_{t=2}^n \sum_{k=1}^{t-1} \frac{1}{\sqrt{(t-k)}} \right) \right) \\
& = O \left(n^{-1} h^{-4} b_n^{-4} + n^{-\frac{1}{2}} h^{-3} b_n^{-4} \right) = o(n^2) \tag{A.15}
\end{aligned}$$

and by Assumptions 3.3 and 3.5(ii) again

$$\begin{aligned}
\Pi_{n,1}(1, 2) &\leq C \sum_{t=1}^n \sum_{k_1=1}^n \sum_{k_2 \neq k_1} E \left[\widehat{w}_{nk_1}^2(V_t) \widehat{w}_{nk_2}^2(V_t) F_t \right] \\
&\leq C n^{-2} h^{-4} b_n^{-4} \sum_{t=1}^n \sum_{k_1=1}^n \sum_{k_2 \neq k_1} E \left[K^2 \left(\frac{V_t - V_{k_1}}{h} \right) K^2 \left(\frac{V_t - V_{k_2}}{h} \right) \right] \\
&\leq C n^{-2} h^{-4} b_n^{-4} \sum_{t=2}^n \sum_{k_1=1}^{t-1} K^2(0) E \left[K^2 \left(\frac{V_t - V_{k_1}}{h} \right) \right] \\
&+ C n^{-2} h^{-4} b_n^{-4} \sum_{t=3}^n \sum_{k_1=2}^{t-1} \sum_{k_2=1}^{k_1-1} E \left[K^2 \left(\frac{V_t - V_{k_1}}{h} \right) K^2 \left(\frac{V_t - V_{k_2}}{h} \right) \right] \\
&\leq C n^{-\frac{1}{2}} h^{-3} b_n^{-4} + C n^{-2} h^{-4} b_n^{-4} \sum_{t=3}^n \sum_{k_1=2}^{t-1} \sum_{k_2=1}^{k_1-1} \varphi_{t-k_1} \\
&\times E \left(\int K^2 \left(\frac{w_1}{h} \right) K^2 \left(\frac{w_1 + V_{k_1} - V_{k_2}}{h} \right) f_{t,k_1}(\varphi_{t-k_1} w_1 | \mathcal{F}_{k_1}) \right) \\
&\leq C n^{-\frac{1}{2}} h^{-3} b_n^{-4} + C n^{-2} h^{-4} b_n^{-4} \sum_{t=3}^n \sum_{k_1=2}^{t-1} \sum_{k_2=1}^{k_1-1} h^2 \varphi_{t-k_1} \varphi_{k_1-k_2} \\
&\times \int K^2(u_1 + u_2) K^2(u_2) du_1 du_2 \\
&= O(n^{-\frac{1}{2}} h^{-3} b_n^{-4} + n^{-2} h^{-2} b_n^{-4} \sum_{t=3}^n \sum_{k_1=2}^{t-1} \sum_{k_2=1}^{k_1-1} (t - k_1)^{-\frac{1}{2}} (k_1 - k_2)^{-\frac{1}{2}}) \\
&= O(n^{-\frac{1}{2}} h^{-3} b_n^{-4} + h^{-2} b_n^{-4}) = o(n^2).
\end{aligned} \tag{A.16}$$

By the Hölder inequality and similar to the calculation of $\Pi_{n,1}(1)$, we have

$$\Pi_{n,1}(2) = O(h^{-2} b_n^{-4}) = o(n^2). \tag{A.17}$$

By Assumption 3.1(ii) and the covariance inequality for α -mixing sequence in Lemma A.2, we have for $k_3 < k_2 < k_1$,

$$\begin{aligned}
E[U_{k_1}^2 \eta_{k_2} \eta_{k_3}] &= E[U_{k_1}^2 \eta_{k_2} \eta_{k_3}] - E[U_{k_1}^2] E[\eta_{k_2} \eta_{k_3}] \\
&+ E[U_{k_1}^2] E[\eta_{k_2} \eta_{k_3}] - E[U_{k_1}^2] E[\eta_{k_2}] E[\eta_{k_3}] \\
&\leq C \left(\alpha_\zeta^{\gamma_2/(4+\gamma_2)} (k_1 - k_2) + \alpha_\zeta^{\gamma_2/(4+\gamma_2)} (k_2 - k_3) \right),
\end{aligned} \tag{A.18}$$

which implies

$$\begin{aligned}
\Pi_{n,1}(3) &\leq C \sum_{t=1}^n \sum_{k_1=1}^n \sum_{k_2 < k_1} \sum_{k_3 < k_2} E[\widehat{w}_{nk_1}^2(V_t) \widehat{w}_{nk_2}(V_t) \widehat{w}_{nk_3}(V_t) F_t] \\
&\times \left(\alpha_\zeta^{\gamma_2/(4+\gamma_2)} (k_1 - k_2) + \alpha_\zeta^{\gamma_2/(4+\gamma_2)} (k_2 - k_3) \right) \\
&\leq C h^{-1} b_n^{-4} = o(n^2).
\end{aligned} \tag{A.19}$$

Noting that for $k_4 < k_3 < k_2 < k_1$,

$$\begin{aligned}
E[U_{k_1} U_{k_2} \eta_{k_3} \eta_{k_4}] &\leq \\
&C \left(\alpha_\zeta^{\gamma_1/(4+\gamma_2)} (k_1 - k_2) + \alpha_\zeta^{\gamma_2/(4+\gamma_2)} (k_2 - k_3) + \alpha_\zeta^{\gamma_1/(4+\gamma_2)} (k_3 - k_4) \right),
\end{aligned}$$

we have for $j = 4, 5, 6$

$$\Pi_{n,1}(j) = O(\sqrt{n} b_n^{-4}) = o(n^2). \tag{A.20}$$

By (A.13)–(A.20), we also have

$$\Pi_{n,1} = o(n^2). \quad (\text{A.21})$$

For $\Pi_{n,2}$, consider the following decomposition:

$$\begin{aligned}
\Pi_{n,2} &= \sum_{t_1=1}^n \sum_{t_2 \neq t_1} E \left[\left(\sum_{k_1=1}^n \widehat{w}_{nk_1}(V_{t_1}) U_{k_1} \right) \left(\sum_{q_1=1}^n \widehat{w}_{nq_1}(V_{t_1}) \eta_{q_1} \right) \right. \\
&\quad \times \left. \left(\sum_{k_2=1}^n \widehat{w}_{nk_2}(V_{t_2}) U_{k_2} \right) \left(\sum_{q_2=1}^n \widehat{w}_{nq_2}(V_{t_2}) \eta_{q_2} \right) F_{t_1} F_{t_2} \right] \\
&\leq C \sum_{t_1=1}^n \sum_{t_2 \neq t_1} \sum_{k=1}^n E \left[\widehat{w}_{nk}^2(V_{t_1}) \widehat{w}_{nk}^2(V_{t_2}) F_{t_1} F_{t_2} \right] E[U_k^2 \eta_k^2] \\
&\quad + C \sum_{t_1=1}^n \sum_{t_2 \neq t_1} \sum_{k_1=1}^n \sum_{k_2 \neq k_1} E \left[\widehat{w}_{nk_1}^2(V_{t_1}) \widehat{w}_{nk_1}(V_{t_2}) \widehat{w}_{nk_2}(V_{t_2}) F_{t_1} F_{t_2} \right] E[U_{k_1}^3 \eta_{k_2}] \\
&\quad + C \sum_{t_1=1}^n \sum_{t_2 \neq t_1} \sum_{k_1=1}^n \sum_{k_2 \neq k_1} E \left[\widehat{w}_{nk_1}^2(V_{t_1}) \widehat{w}_{nk_1}(V_{t_2}) \widehat{w}_{nk_2}(V_{t_2}) F_{t_1} F_{t_2} \right] E[\eta_{k_1}^3 U_{k_2}] \\
&\quad + C \sum_{t_1=1}^n \sum_{t_2 \neq t_1} \sum_{k_1=1}^n \sum_{k_2 \neq k_1} E \left[\widehat{w}_{nk_1}^2(V_{t_1}) \widehat{w}_{nk_2}^2(V_{t_2}) F_{t_1} F_{t_2} \right] E[U_{k_1}^2 \eta_{k_2}^2] \\
&\quad + C \sum_{t_1=1}^n \sum_{t_2 \neq t_1} \sum_{k_1=1}^n \sum_{k_2 \neq k_1} E \left[\widehat{w}_{nk_1}(V_{t_1}) \widehat{w}_{nk_1}(V_{t_2}) \widehat{w}_{nk_2}(V_{t_1}) \widehat{w}_{nk_2}(V_{t_2}) F_{t_1} F_{t_2} \right] \\
&\quad \times E \left[U_{k_1}^2 \eta_{k_2}^2 \right] \\
&\quad + C \sum_{t_1=1}^n \sum_{t_2 \neq t_1} \sum_{k_1=1}^n \sum_{k_2 < k_1} \sum_{q_1 < k_2} E \left[\widehat{w}_{nk_1}(V_{t_1}) \widehat{w}_{nk_1}(V_{t_2}) \widehat{w}_{nk_2}(V_{t_1}) \widehat{w}_{nq_1}(V_{t_2}) F_{t_1} F_{t_2} \right] \\
&\quad \times E[U_{k_1}^2 \eta_{k_2} \eta_{q_1}] \\
&\quad + C \sum_{t_1=1}^n \sum_{t_2 \neq t_1} \sum_{k_1=1}^n \sum_{k_2 < k_1} \sum_{q_1 < k_2} E \left[\widehat{w}_{nk_1}(V_{t_1}) \widehat{w}_{nk_1}(V_{t_2}) \widehat{w}_{nk_2}(V_{t_1}) \widehat{w}_{nq_1}(V_{t_2}) F_{t_1} F_{t_2} \right] \\
&\quad \times E[\eta_{k_1}^2 U_{k_2} U_{q_1}] \\
&\quad + C \sum_{t_1=1}^n \sum_{t_2 \neq t_1} \sum_{k_1=1}^n \sum_{k_2 < k_1} \sum_{q_1 < k_2} E \left[\widehat{w}_{nk_1}^2(V_{t_1}) \widehat{w}_{nk_2}(V_{t_2}) \widehat{w}_{nq_1}(V_{t_2}) F_{t_1} F_{t_2} \right] \\
&\quad \times E[U_{k_1}^2 \eta_{k_2} \eta_{q_1}] \\
&\quad + C \sum_{t_1=1}^n \sum_{t_2 \neq t_1} \sum_{k_1=1}^n \sum_{k_2 < k_1} \sum_{q_1 < k_2} E \left[\widehat{w}_{nk_1}^2(V_{t_1}) \widehat{w}_{nk_2}(V_{t_2}) \widehat{w}_{nq_1}(V_{t_2}) F_{t_1} F_{t_2} \right] \\
&\quad \times E[\eta_{k_1}^2 U_{k_2} U_{q_1}] \\
&\quad + C \sum_{t_1=1}^n \sum_{t_2 \neq t_1} \sum_{k_1=1}^n \sum_{k_2 < k_1} \sum_{q_1 < k_2} \sum_{q_2 < q_1} E \left[\widehat{w}_{nk_1}(V_{t_1}) \widehat{w}_{nk_2}(V_{t_2}) \widehat{w}_{nq_1}(V_{t_1}) \widehat{w}_{nq_2}(V_{t_2}) F_{t_1} F_{t_2} \right] \\
&\quad \times E[U_{k_1} U_{k_2} \eta_{q_1} \eta_{q_2}] \\
&\equiv \sum_{j=1}^9 \Pi_{n,2}(j).
\end{aligned}$$

By Assumption 3.3, we have

$$\begin{aligned}\Pi_{n,2}(1) &\leq Cn^{-2}h^{-4}b_n^{-4} \left(\sum_{t_1=1}^n \sum_{t_2 \neq t_1} K^2(0) E \left[K^2 \left(\frac{V_{t_1} - V_{t_2}}{h} \right) \right] \right) \\ &+ Cn^{-2}h^{-4}b_n^{-4} \left(\sum_{t_1=1}^n \sum_{t_2 \neq t_1} \sum_{k \neq t_1, t_2} E \left[K^2 \left(\frac{V_{t_1} - V_k}{h} \right) K^2 \left(\frac{V_{t_2} - V_k}{h} \right) \right] \right) \\ &= O \left(n^{-\frac{1}{2}}h^{-3}b_n^{-4} + h^{-2}b_n^{-4} \right) = o(n^2).\end{aligned}$$

Similarly, by the Hölder inequality we have

$$\Pi_{n,2}(2) = \Pi_{n,2}(3) = O \left(n^{\frac{1}{2}}h^{-1}b_n^{-4} \right) = o(n^2).$$

Analogously, we have

$$\Pi_{n,2}(4) = \Pi_{n,2}(5) = O \left(n^{\frac{1}{2}}h^{-1}b_n^{-4} \right) = o(n^2)$$

and

$$\Pi_{n,2}(6) = O \left(n^{\frac{1}{2}}h^{-1}b_n^{-4} \right) = o(n^2).$$

Applying the proofs of (A.19) and (A.20), we have

$$\Pi_{n,2}(7) = O \left(n^{\frac{1}{2}}b_n^{-4} \right) = o(n^2),$$

$$\Pi_{n,2}(8) = O \left(n^{\frac{1}{2}}b_n^{-4} \right) = o(n^2)$$

and

$$\Pi_{n,2}(9) = O \left(nhb_n^{-4} \right) = o(n^2).$$

The above arguments then imply

$$\Pi_{n,2} = o(n^2). \tag{A.22}$$

By (A.12), (A.21), (A.22) and the Markov inequality, we have shown (A.11), which implies that (A.2) holds.

Using the same arguments as in the proof of Theorem 3.1, we can prove (A.9). By the law of large numbers for stationary α -mixing process (for example, Hall and Heyde 1980) and Assumption 3.1(ii), we obtain

$$\frac{1}{n} \sum_{t=1}^n U_t \eta_t \rightarrow_P \Sigma_{u\eta}, \tag{A.23}$$

where $\Sigma_{u\eta} = E[U_1 \eta_1]$. By Assumption 3.5(iii), we can prove that

$$\frac{1}{n} \sum_{t=1}^n U_t \eta_t F_t^c = o_P(1). \tag{A.24}$$

By (A.23) and (A.24) and noting that

$$\frac{1}{n} \sum_{t=1}^n U_t \eta_t F_t + \frac{1}{n} \sum_{t=1}^n U_t \eta_t F_t^c = \frac{1}{n} \sum_{t=1}^n U_t \eta_t,$$

we have shown that (A.10) holds.

By the Cauchy–Schwarz inequality, (A.2), (A.9) and (A.10), we can show that (A.3)–(A.7) hold. This completes the proof of Lemma A.1(i).

8.2 Proof of Lemma A.1(ii)

The result is a multivariate version of Corollary 5.1 of Hall and Hedye (1980).

8.3 Proof of Lemma A.2

The lemma is a special case of Lemma A.1 of Gao (2007).

REFERENCES

- Ai, C., and Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71, 1795–1843.
- Berkes, I., Horváth, L., 2006. Convergence of integral functionals of stochastic processes. *Econometric Theory* 22, 304–322.
- Cai, Z., Li, Q., Park, J., 2009. Functional–coefficient models for nonstationary time series data. *Journal of Econometrics* 148, 101–113.
- Chen, J., Gao, J., Li, D. 2008. Semiparametric regression estimation in null–recurrent time series. Working paper available at <http://www.adelaide.edu.au/directory/jiti.gao>.
- Fan, J., Yao, Q., 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.
- Florens, J. P., Johannes, J., van Bellegem, S., 2007. Instrumental regression in partially linear models. Working paper available from the authors websites.
- Gao, J., 2007. *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman & Hall/CRC, London.
- Gao, J., King, M. L., Lu, D., Tjøstheim, D., 2009a. Specification testing in nonstationary time series autoregression. *Annals of Statistics* 37, 3893–3928.
- Gao, J., King, M. L., Lu, D., Tjøstheim, D., 2009b. Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* 25, 1869–1892.
- Hall, P., Heyde, C. C., 1980. *Martingale Limit Theory and Its Applications*. Academic Press, New York.

- Hansen, B. E., 1995. Regression with nonstationary volatility. *Econometrica* 63, 1113–1132.
- Hansen, B. E., 2008. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726–748.
- Härdle, W., Liang, H., Gao, J., 2000. *Partially Linear Models*. New York: Physica-Verlag.
- Jeganathan, P., 2004. Convergence of functionals of sums of r.v.s to local times of fractional stable motions. *Annals of Probability* 32, 1771–1795.
- Jeganathan, P., 2008. Limit theorems for functionals of sums that converge to fractional Brownian and stable motions. CFDP #1649, Yale University.
- Karlsen, H. A., Tjøstheim, D., 2001. Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372–416.
- Karlsen, H. A., Mykelbust, T., Tjøstheim, D., 2007. Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* 35, 252–299.
- Kasparis, I., Phillips, P. C. B., 2009. Dynamic misspecification in nonparametric cointegrating regression. Cowles Foundation Discussion Paper No. 1700.
- Li, Q., Racine, J., 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, New Jersey.
- Newey, W. K., Powell, J., 2003. Instrumental variable estimation of nonparametric models. *Econometrica* 71, 1565–1578.
- Newey, W. K., Powell, J., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equations models. *Econometrica* 67, 565–603.
- Park, J., Phillips, P. C. B., 1988. Statistical inference in regressions with integrated processes I. *Econometric Theory* 4, 468–497.
- Park, J., Phillips, P. C. B., 1989. Statistical inference in regressions with integrated processes II. *Econometric Theory* 5, 95–131.
- Park, J., Phillips, P. C. B., 1999. Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 15, 269–298.
- Park, J., Phillips, P. C. B., 2001. Nonlinear regressions with integrated time series. *Econometrica* 69, 117–162.
- Phillips, P. C. B., 1986. Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–340.
- Phillips, P. C. B., 1991. Optimal inference in cointegrated systems. *Econometrica* 59, 238–306.
- Phillips, P. C. B., 1995. Fully modified least squares and vector autoregression. *Econometrica* 63, 1023–1078.

- Phillips, P. C. B., 1998. New tools for understanding spurious regressions. *Econometrica* 66, 1299–1325.
- Phillips, P. C. B., 2001. Descriptive econometrics for non-stationary time series with empirical illustrations. *Journal of Applied Econometrics* 16, 389 - 413.
- Phillips, P. C. B., 2009. Local limit theory and spurious regressions. *Econometric Theory* 25, 1466–1497.
- Phillips, P. C. B., 2012. Optimal estimation of cointegrated systems with irrelevant instruments. Forthcoming in *Journal of Econometrics*
- Phillips, P. C. B., Hansen, B. E., 1990. Statistical inference in instrumental variables regression with $I(1)$ processes. *Review of Economic Studies* 57, 99–125.
- Phillips, P. C. B., Park, J., 1998. Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper, No. 1181, Yale University.
- Phillips, P. C. B., Su, L., 2011. Non-parametric regression under location shifts. Forthcoming in *Econometrics Journal*.
- Phillips, P. C. B., Xiao, Z., 1998. A primer on unit root testing. *Journal of Economic Surveys* 12, 423–469.
- Pötscher, B. M., 2004. Nonlinear functions and convergence to Brownian motion: beyond the continuous mapping theorem. *Econometric Theory* 20, 1–22.
- Robinson, P. M., 1988. Root-N-consistent semiparametric regression. *Econometrica* 56, 931–954.
- Schienle, M., 2009. Reaching for econometric generality: nonparametric nonstationary regression. Working paper available from the Department of Economics, University of Mannheim, Germany.
- Su, L., Ullah, A., 2008. Local polynomial estimation of nonparametric simultaneous equations models. *Journal of Econometrics* 144, 193–218.
- Teräsvirta, T., Tjøstheim, D., Granger, C. W. J., 2010. *Modelling Nonlinear Economic Time Series*. Oxford University Press, Oxford.
- Wang, Q. Y., Phillips, P. C. B., 2009a. Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710–738.
- Wang, Q. Y., Phillips, P. C. B., 2009b. Structural nonparametric cointegrating regression. *Econometrica* 77, 1901–1948.
- Withers, C. S., 1981. Conditions for linear processes to be strong mixing. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 57, 509–534.