SEMIPARAMETRIC ESTIMATION IN TIME SERIES OF SIMULTANEOUS EQUATIONS

By

Jiti Gao and Peter C. B. Phillips

September 2010

COWLES FOUNDATION DISCUSSION PAPER NO. 1769

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

Semiparametric Estimation in Time Series of Simultaneous Equations

Jiti Gao

School of Economics, University of Adelaide

Peter C. B. Phillips

Yale University, University of Auckland,

University of Southampton, & Singapore Management University

June 2, 2010

Abstract

A system of vector 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 strictly exogenous and represent trends. The parametric regressors may be stationary or nonstationary and the nonparametric regressors are nonstationary time series. This framework allows for the nonparametric treatment of stochastic trends and subsumes many practical cases. Semiparametric least squares (SLS) estimation is considered and its asymptotic properties are derived. Due to endogeneity in the parametric regressors, SLS is generally inconsistent for the parametric component and a semiparametric instrumental variable least squares (SIVLS) method is proposed instead. Under certain regularity conditions, the SIVLS estimator of the parametric component is shown to be consistent with a limiting normal distribution that is amenable to inference. The rate of convergence in the parametric component is the usual $\sqrt{n}$ rate and is explained by the fact that the common (nonlinear) trend in the system is eliminated nonparametrically by stochastic detrending.

Key words and phrases: Endogeneity; exogeneity; nonstationarity; partially linear model; simultaneous equation; stochastic detrending; 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 e.g. 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 1991; 1995, 2000) and semiparametric (e.g., Phillips and Hansen 1990, Phillips 1991; 1995, Watson 1994) 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), de Jong (2004), Jeganathan (2004), 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 (1998, 2001), Wang and Phillips (2009a, 2009b), Karlsen *et al* (2007), Kasparis and Phillips (2009), Cai, Li and Park (2009), Schienle (2009), and Phillips (2009). The last paper examines in a nonparametric setting spurious time series models of the type considered by Granger and Newbold (1974, 1977) in a linear parametric setting, for which the asymptotic theory was given in Phillips (1986, 1998).

Among the 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 (1998, 2001), and Karlsen *et al* (2007) to model univariate time series with some kind of 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; Linton 1995; Pagan and Ullah 1999; Härdle *et al* 2000; Yatchew 2003; 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 simultaneous equations of the form

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

where n is the sample size, A is a $p \times d$–matrix of unknown parameters, $Y_t = (y_{t1}, \cdots, y_{tp})'$, $X_t = (x_{t1},\ \cdots,\ x_{td})'$, and V_t is a sequence of univariate integrated time series regressors, $g(\cdot) = (g_1(\cdot), \cdots, g_p(\cdot))'$ and $H(\cdot) = (h_1(\cdot), \cdots, h_d(\cdot))'$[1] are all unknown functions, and both ϵ_t and U_t are vectors of stationary time series. 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 Linton 1995; Newey *et al* 1999; Su and Ullah 2008). The condition $E[\epsilon_t|V_t] = E[\epsilon_t]$ is generally needed to ensure that the model is identified. For, if there were an unknown function $\lambda(\cdot)$ such that $\epsilon_t = \lambda(V_t) + \varepsilon_t$ 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 ϵ_t is correlated with V_t may be included. In particular, in studying nonparametric regressions of the form $Y_t = g(V_t) + \epsilon_t$, Wang and Phillips (2009b) consider a nonstationary endogenous regressor case where V_t is correlated with ϵ_t and show that conventional nonparametric regression is applicable in spite of the endogeneity. Phillips and Su (2010) 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 $E[\epsilon_t|V_t] = E[\epsilon_t] = 0$ eliminates endogeneity between ϵ_t and V_t while retaining endogeneity between ϵ_t and X_t and potential nonstationarity in both X_t and V_t. The condition $E[\epsilon_t|V_t] = E[\epsilon_t] = 0$ in our setting corresponds

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

to the condition $E[\epsilon_t|V_t, U_t] = E[\epsilon_t|U_t]$ that is assumed in Newey *et al* (1999) and Su and Ullah (2008), the former being implied by $E[\epsilon_t|V_t] = E\left(E\left[\epsilon_t|U_t, V_t\right]|V_t\right) = E\left(E\left[\epsilon_t|U_t\right]|V_t\right) = E\left(E\left[\epsilon_t|U_t\right]\right) = E\left[\epsilon_t\right]$ when U_t is independent of V_t and $E[\epsilon_t] = 0$. The identification conditions in (1.1) allow for both conditional heteroskedasticity and endogeneity in ϵ_t, permitting ϵ_t to depend on U_t[2]. These conditions are also less restrictive than the exogeneity condition between ϵ_t and (X_t, V_t) that is common in the literature for the stationary case (see, for example, Gao 2007).

In comparison with a related paper by Chen *et al* (2008), the present paper considers a general multivariate model structure in (1.1) with emphasis on possible endogeneity and nonstationarity. 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 $Y_t = X_t'\alpha + g(V_t) + \epsilon_t$ with $E[\epsilon_t|X_t, V_t] = 0$. In that model, $X_t = H(V_t) + U_t$ holds automatically since $U_t = X_t - E[X_t|V_t]$, thereby restricting the (mean) impact on X_t of any nonstationarity in V_t through the conditional mean function $E[X_t|V_t]$ and the (stationary) error U_t. 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 form $V_t = V_{t-1} + v_t$ with v_t being independent and identically distributed.

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 endogenous, nonstationary regressors. In the case where endogeneity is involved in semiparametric regression modeling of independent data, some related developments include Robinson (1988), Newey *et al* (1999), Ai and Chen (2003), Newey and Powell (2003), Blundell *et al* (2007), 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. Accordingly, the paper proposes a semiparametric instrumental variable least squares (SIVLS) estimate of A to deal with endogeneity in X_t and a nonparametric estimator for the function $g(\cdot)$. The SIVLS estimator of A is shown to be consistent with a conventional $\sqrt{n}$–rate of convergence even when X_t is stochastically nonstationary. This rate arises because nonstationarity in the regression is 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 comple-

[2] The additive case where $\epsilon_t = \lambda(U_t) + \mu_t$ with $E[\mu_t|V_t] = 0$ is covered in the first part of (1.1) because $E\left[\epsilon_t|V_t\right] = E\left[\lambda(U_t)|V_t\right] + E[\mu_t|V_t] = E\left[\lambda(U_t)\right] = E\left[\epsilon_t\right]$ when U_t is independent of V_t. The multiplicative case where $\epsilon_t = \sigma(U_t)\nu_t$ is also covered in the first part of (1.1) because $E\left[\epsilon_t|V_t\right] = E\left[\sigma(U_t)\nu_t|V_t\right] = E\left[\epsilon_t\right]$ when (U_t, ν_t) is assumed to be independent of V_t.

ments existing results on parametric modeling with endogeneity (such as Phillips 1983; Sargan 1988), parameter estimation in simultaneous equations models (such as Greene 2005), nonparametric and semiparametric estimation of nonlinear time series (such as Tong 1990; Fan and Yao 2003; Gao 2007), parameter estimation in vector autoregressions and cointegration, instrumental variable estimation of nonparametric models (such as Robinson 1988; Newey *et al* 1999; Ai and Chen 2003; Newey and Powell 2003; Blundell *et al* 2007; Florens *et al* 2007; 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; Chen *et al* 2008; Phillips 2009; Wang and Phillips 2009a, 2009b).

The paper is organized as follows. Section 2 proposes estimators of the parameter matrix A and the nonlinear functions $g(\cdot)$. Asymptotic results of the proposed semiparametric estimators are established in Section 3. A bandwidth selection method is developed in Section 4.1. Section 4.2 gives 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 more detailed discussion of the model and its implications. Write (1.1) in full as:

$$Y_t = A\ X_t + g(V_t) + \epsilon_t \tag{2.1}$$

$$X_t = H(V_t) + U_t, \tag{2.2}$$

$$E[\epsilon_t|V_t] = E[\epsilon_t] = 0, \tag{2.3}$$

$$E[U_t|V_t] = 0. \tag{2.4}$$

When the variables $\{(X_t, V_t, \epsilon_t)\}$ 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 $E[\epsilon_t|(U_t, V_t)] = 0$, 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

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

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

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

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

$$g(v) = \Psi(v) - AH(v). \tag{2.7}$$

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

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

where $e_t = \epsilon_t$ and $U_t = X_t - H(V_t)$, as assumed in (1.1). Introducing the "stochastically detrended" variable

$$W_t = Y_t - \Psi(V_t), \tag{2.8}$$

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

$$W_t = A\; U_t + e_t. \tag{2.9}$$

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

- 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 $E\left[\lambda_t U_t'\right] = E\left\{\lambda_t E\left[U_t'|V_t\right]\right\} = 0$, we have

$$E\left[U_t \epsilon_t'\right] = E\left[U_t \lambda_t{}'\right] + E\left[U_t e_t'\right] = E\left[U_t e_t'\right] = E\left[U_t E\left(e_t'|U_t\right)\right]. \tag{2.10}$$

It follows that the unknown matrix A can be consistently estimated based on (2.9) when $E\left[U_t e_t'\right] = 0$. 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 $\gamma(u) = E\left[e_t'|U_t = u\right]$. Then, $E\left[U_t e_t'\right] = E\left[U_t E\left(e_t'|U_t\right)\right] = E\left[U_t \gamma(U_t)\right] = \int_{-\infty}^{\infty} u\gamma(u)p(u)du = 0$ 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 $E\left(e_t'|U_t\right) = 0$. Section 2.2 develops a semiparametric instrumental variable procedure (SIVLS) that is applicable in the case of nonstationary U_t.

2.1 SLS estimation

When $E\left(e_t'|U_t\right)=0$, 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

$$\widetilde{Y}_t = A\ \widetilde{X}_t + e_t, \tag{2.11}$$

where $\widetilde{Y}_t = \left(Y_t - \widehat{\Psi}(V_t)\right) F_t$ and $\widetilde{X}_t = \left(X_t - \widehat{H}(V_t)\right) F_t$. In these formulae, F_t is the indicator $F_t = I\left(\widehat{p}_n(V_t) > b_n\right)$ where b_n is a sequence of positive numbers that tend to zero as $n \to \infty$, $\widehat{p}_n(v) = \frac{1}{\sqrt{n}h}\sum_{s=1}^n K\left(\frac{V_s - v}{h}\right)$, $\widehat{\Psi}(v) = \sum_{s=1}^n w_{ns}(v)Y_s$ and $\widehat{H}(v) = \sum_{s=1}^n w_{ns}(v)X_s$ with $w_{ns}(\cdot)$ being a sequence of probability weight functions of the form

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

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.

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

$$\widehat{A} = \widetilde{Y}'\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}, \tag{2.13}$$

where $\widetilde{X}' = (\widetilde{X}_1, \cdots, \widetilde{X}_n)$, $\widetilde{Y}' = (\widetilde{Y}_1, \cdots, \widetilde{Y}_n)$, and throughout the paper D^{-1} is the inverse of D or a generalized inverse if D^{-1} does not exist. This type of truncated least squares estimation method has been widely used in the literature for the independent sample case (see, for example, Robinson 1988).

The vector of unknown functions $g(\cdot)$ is then estimated by

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

By elementary calculation

$$(\widehat{A} - A)\ \widetilde{X}'\widetilde{X} = \widetilde{e}'\widetilde{X} + \widetilde{G}'\widetilde{X}, \tag{2.15}$$

with $\widetilde{G}' = (\widetilde{G}_1, \cdots, \widetilde{G}_n) = (\widetilde{g}(V_1), \cdots, \widetilde{g}(V_n))$, $\widetilde{g}(V_t) = g(V_t) - \sum_{s=1}^n w_{ns}(V_t)g(V_s)$, $\widetilde{e}' = (\widetilde{e}_1, \cdots, \widetilde{e}_n)$ and $\widetilde{e}_t = e_t - \sum_{s=1}^n w_{ns}(V_t)e_s$. 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

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

where $e' = (e_1, \cdots, e_n)$ and $U = (U_1, \cdots, U_n)'$. 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 $\widehat{A}$ 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 least squares (SIVLS) estimation method for A is developed that is consistent and has desirable asymptotic distributional properties.

2.2 SIVLS 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

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

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

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

Equations (2.17) and (2.18) imply

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

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

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

where $\eta' = (\eta_1, \cdots, \eta_n)$. 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 system of vector semiparametric stationary IV time series models. Each η_t may be regarded as the stationary component of a suitable 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

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

where η_t is assumed to satisfy (2.18), $Q_t = (q_{t1}, \cdots, q_{td})'$ is a vector of possible instrumental variables for X_t generated by a reduced form equation involving V_t, and $J(\cdot) = (J_1(\cdot), \cdots, J_d(\cdot))'$ 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 (see, for example, Li and Stengos, 1996; Baltagi and Li, 2002) 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

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

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

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

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

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

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 $\widehat{B}$ and the matrix $\Gamma = \Delta\Delta'$ by

$$\widehat{\Gamma} = \frac{1}{n} \sum_{t=1}^{n} \left(\widetilde{Y}_t - \widehat{B}\widetilde{X}_t\right) \left(\widetilde{Y}_t - \widehat{B}\widetilde{X}_t\right)'. \tag{2.25}$$

As shown in Corollary 3.3 below, we have $\widehat{\Gamma} \to_P \Gamma$ as $n \to \infty$. The matrix Σ is then consistently estimated by $\widehat{\Sigma} = I - \widehat{\Delta}$.

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 (SIVLS) estimator

$$\widehat{A}^* = \widehat{A}^*(h) = \widetilde{Y}'\widetilde{Q}\left(\widetilde{X}'\widetilde{Q}\right)^{-1}, \tag{2.26}$$

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

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

It follows from (2.26) that

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

As shown in Appendix A, we have the following decomposition

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

where $\eta = (\eta_1, \cdots, \eta_n)'$.

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 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 $\{\mathcal{Z}_t, t = 0, \pm 1, \cdots\}$ is α–mixing if the mixing numbers $\alpha(n) \to 0$ as $n \to \infty$, where

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

in which $\mathcal{F}_k^j$ is the σ–field generated by $\{\mathcal{Z}_t, k \le t \le j\}$. For the original definition, see Rosenblatt (1956), and for a recent discussion and related limit theorems, see Lin and Lu (1996) and the references therein.

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 $E\left[\|\xi_1\|^{4+\gamma_1}\right] < \infty$ 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*

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

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

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

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) $\Gamma_1 = E\left[U_1\eta_1'\right]$ be nonsingular.

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

Assumption 3.3. (i) *$\{V_t : t \geq 0\}$ is independent of $\{(e_s, U_s, \eta_s) : s \geq 1\}$.*

(ii) *If $f_{i,k}(\cdot)$ is the density function of $V_{i,k} = \varphi_{i-k}\left(V_i - V_k\right)$ for $i > k$ with $\varphi_m = \frac{L_s(m)}{\sqrt{m}}$ for $m \geq 1$, where $L_s(\cdot)$ is a slowly varying function, then*

$$\inf_{\delta>0} \limsup_{m\to\infty} \sup_{i\geq 1} \sup_{|v|\leq\delta} f_{i+m,i}(v) < \infty. \tag{3.4}$$

There exists a filtration $\{F_t,\ t \geq 0\}$ such that V_t is adapted to F_t and for which, with probability one,

$$\inf_{\delta>0} \limsup_{m\to\infty} \sup_{i\geq 1} \sup_{|v|\leq\delta} f_{i+m,i}(v|\mathcal{F}_i) < \infty, \tag{3.5}$$

where $f_{i,k}(v|F_k)$ is the conditional density function of $V_{i,k}$ given $\mathcal{F}_k$.

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,*

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

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

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

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

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

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

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

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* $n \to \infty$, *the following rate conditions*

$$h_n \to 0, \ nh_n^2 \to \infty, \ \ nh_n^6 \to 0, \tag{3.11}$$

$$b_n \to 0, \ \frac{L_s(n)}{\sqrt{n} b_n^2} \to 0, \ \ \frac{L_s^2(n)\sqrt{h}}{b_n^2} \to 0, \ \ \frac{L_s^6(n)}{nh^2 b_n^8} \to 0, \tag{3.12}$$

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

(iii) b_n *is chosen such that* $\sum_{t=1}^{n} P\left(\widehat{p}_n(V_t) \le b_n\right) = o(n)$.

(iv) *There exists a real function* $\lambda(x, y)$ *such that* $||g(x + yh) - g(x)|| \le h\lambda(y, x)$ *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.

Some 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* $n \to \infty$, *we have*

$$\sqrt{n}(\widehat{A}^* - A) \to_D N\left(0, \Sigma_1^*\right), \tag{3.13}$$

where $\Sigma_1^* = \left(I \otimes \Gamma_1^{-1}\right) \ \Omega_1^* \ \left(I \otimes \left(\Gamma_1^{-1}\right)'\right)$ *with* $\Omega_1^* = \sum_{j=0}^{\infty} E\left[\left(e_1 e_{1+j}'\right) \otimes \left(\eta_1 \eta_{1+j}'\right)\right]$ *and* $\Gamma_1 = E\left[U_1 \eta_1'\right]$.

Theorem 3.1 shows that the semiparametric IV estimator $\widehat{A}^*$ can be asymptotically normal in the limit even when the parametric and nonparametric regressors are both nonstationary. In addition, $\widehat{A}^*$ 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 consist of a system of vector semiparametric stationary IV time seres models 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 achievable.

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 $n \to \infty$*

$$\sqrt{n}(\widehat{A} - A) \to_D N\left(0, \Sigma_1^*\right), \tag{3.14}$$

where $\Sigma_1^* = \left(I \otimes \Gamma_1^{-1}\right) \Omega_1^* \left(I \otimes \Gamma_1^{-1}\right)$ *with* $\Omega_1^* = \sum_{j=0}^{\infty} E\left[e_1 e_{1+j}'\right] \otimes E\left[U_1 U_{1+j}'\right]$ *and* $\Gamma_1 = E\left[U_1 U_1'\right]$.

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

$$\sqrt{n}(\widehat{A} - A) \to_D N\left(0, \Sigma_{11} \otimes \Sigma_{22}^{-1}\right), \tag{3.15}$$

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

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) obtain 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 $\widehat{A}$ or $\widehat{A}^*$ 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, Andrews 1991; Phillips 1995). Such extensions are straightforward and are not considered here.

Recall that the nonparametric component is estimated by $\widehat{g}^*(v)$ as defined in (2.27). The asymptotic distribution of $\widehat{g}^*(v)$ 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 $n \to \infty$*

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

where $\Omega_g = \int K^2(u)du \cdot E\left[e_1 e_1'\right]$ *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 $\sqrt{n}h$ (see Wang and Phillips, 2009a). It follows that in the nonstationary case, the rate of convergence of $\widehat{g}^*(v)$ is $(\sqrt{n}h)^{\frac{1}{2}}$.

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,* $\Sigma_{11} = E\left[e_1 e_1'\right]$ *is positive definite, then as* $n \to \infty$

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

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 $\mathrm{cov}(U_t, \Pi_t) = 0$. 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* $n \to \infty$

$$\widehat{\Gamma} = \frac{1}{n}\sum_{t=1}^{n}\left(\widetilde{Y}_t - \widehat{B}\widetilde{X}_t\right)\left(\widetilde{Y}_t - \widehat{B}\widetilde{X}_t\right)' \to_P \Gamma \tag{3.18}$$

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

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

As in other nonparametric and semiparametric contexts, bandwidth choice is important in practical implementation. In the case where V_i is stationary, many existing studies (see, for example, §2.1.3 of Härdle *et al* 2000) offer solutions. But in nonstationary regressor cases, the literature on bandwidth selection is limited (see, however, the discussion in Wang and Phillips, 2009a) and many issues are still 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:

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

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

where $W_{ns}^{(-t)}(V_t) = \frac{K\left(\frac{V_s - V_t}{h}\right)}{\sum\limits_{k=1,\neq t}^{n} K\left(\frac{V_k - V_t}{h}\right)}$. We then define the leave–one–out semiparametric instrumental variable least squares (SIVLS) estimator of A by

$$\widetilde{A} = \widetilde{A}(h) = \overline{Y}'\overline{Q}(\overline{X}'\overline{Q})^{-1}, \tag{4.3}$$

where $\overline{X}' = (\overline{X}_1, \cdots, \overline{X}_n)$, $\overline{Q}' = (\overline{Q}_1, \cdots, \overline{Q}_n)$, $\overline{Y}' = (\overline{Y}_1, \cdots, \overline{Y}_n)$, and

$$\overline{X}_t = \left(X_t - \widetilde{H}_t(V_t)\right)\overline{F}_t = \left(X_t - \sum_{s=1,\neq t}^{n} W_{ns}^{(-t)}(V_t)X_s\right)\overline{F}_t,$$

$$\overline{Q}_t = \left(Q_t - \widetilde{J}_t(V_t)\right)\overline{F}_t = \left(Q_t - \sum_{s=1,\neq t}^{n} W_{ns}^{(-t)}(V_t)Q_s\right)\overline{F}_t,$$

$$\overline{Y}_t = \left(Y_t - \widetilde{\Psi}_t(V_t)\right)\overline{F}_t = \left(Y_t - \sum_{s=1,\neq t}^{n} W_{ns}^{(-t)}(V_t)Y_s\right)\overline{F}_t,$$

in which $\overline{F}_t = I\left(\overline{p}_{n,t}(V_t) > b_n\right)$ with $\overline{p}_{n,t}(V_t) = \frac{1}{\sqrt{n}h}\sum_{s=1,\neq t}^{n} K\left(\frac{V_s - V_t}{h}\right)$.

The corresponding leave–one–out estimator of $g(\cdot)$ is obtained as

$$\widetilde{g}(\cdot; h) = g_n(\cdot; \widetilde{A}(h)). \tag{4.4}$$

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

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

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

$$\text{CV}(\widetilde{h}) = \min_{h \in H_n} \text{CV}(h), \tag{4.6}$$

where H_n is a set of smoothing parameter values. The corresponding data-determined estimators of A and $g(\cdot)$ are then given by

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

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

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

4.2 Examples of implementation

Example 4.1 below demonstrates how the functional forms of $g(\cdot)$ and $H(\cdot)$ may affect the rate of convergence of $\widehat{A}$ in the exogenous case. 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 correlated with the detrended residuals. The estimation method proposed in Section 2.2 is implemented.

Example 4.1. Consider the semiparametric simultaneous equation model

$$Y_t = A\ X_t + G(V_t) + \epsilon_t, \tag{4.8}$$

where A is a matrix of 2×2 of unknown parameters of the form

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

$X_t = (X_{t1}, X_{t2})'$ 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 $v_t = \gamma\ v_{t-1} + \nu_t$, for $t = 1, 2, \cdots$, where $\gamma = 0,\ 0.5, 0.9$, $v_0 = 0$ and ν_t is a sequence of independent errors generated from $N(0, 1)$, $G(\cdot) = (g_1(\cdot), g_2(\cdot))'$ is a vector of unknown functions, and ϵ_t is a vector of stationary time series errors generated from

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

Independently from ϵ_t, generate U_t as

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

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

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

The following functions are used in the model specification:

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

The process X_t is generated by $X_t = H(V_t) + U_t$ and Y_t is generated by (4.8). The proposed estimation method in Section 2.1 is then applied to estimate A, and $G(\cdot)$ and $H(\cdot)$. We assess finite sample performance using the measures

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

where $\widehat{a}_{ij}$ is the (i,j)–th element of $\widehat{A}$.

Table 4.1. Simulation results 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

The simulation results for both the absolute errors and standard deviations given in Table 4.1 were performed 1000 times and the means are tabulated in Table 4.1. In the case of (4.12), the conditions of Theorem 3.1 all hold. Table 4.1 provides the finite sample evidence relating to the limit theory of Theorem 3.1 for both stationary nonparametric regressors and integrated nonparametric regressors. In addition, Table 4.1 shows that the dependence structure of v_t has some effect on the rate of convergence, particularly in the integrated case and when γ is as large as 0.9.

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

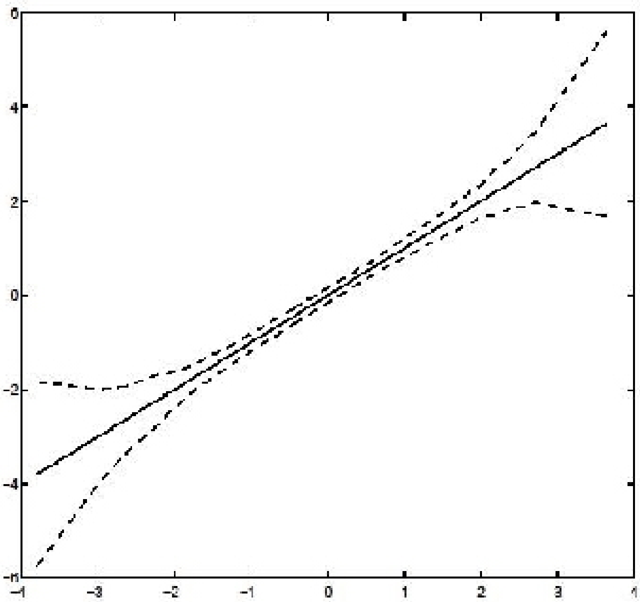

Fig 4.1(a) plots $\widehat{H}_1$ and its 95% CI against $(V_{(1)}, \cdots, V_{(n)})$ for $\gamma = 0$ and $n = 502$.

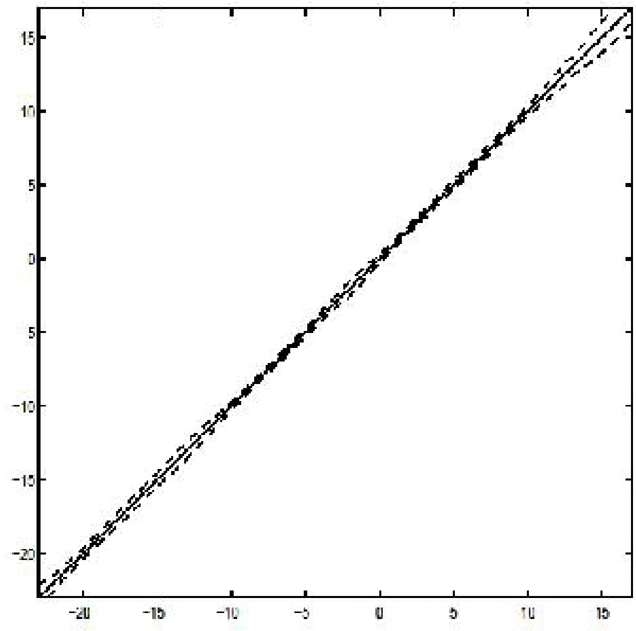

Fig 4.1(b) plots $\widehat{H}_2$ and its 95% CI against $(V_{(1)}, \cdots, V_{(n)})$ for $\gamma = 0.5$ and $n = 502$.

Example 4.2. We consider a vector simultaneous equations model of the form

$$Y_t = A\ X_t + G(V_t) + \epsilon_t, \tag{4.13}$$

where A is a matrix of 2×2 of unknown parameters of the form

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

$X_t = (X_{t1}, X_{t2})'$ is a vector of time series regressors, V_t is a sequence of integrated time series regressors of the form $V_t = V_{t-1} + v_t$ with $V_0 = 0$ and v_t a sequence of stationary disturbances generated by $v_t = \gamma\ v_{t-1} + \nu_t$, for $t = 1, 2, \cdots$, where $\gamma = 0.1, 0.5, 0.9$, $v_0 = 0$ and ν_t is a sequence of independent errors generated from $N(0, 1)$, $G(\cdot) = (g_1(\cdot), g_2(\cdot))'$ is a vector of unknown functions, and ϵ_t is generated by $\epsilon_t = \rho\ U_t + \mu_t$ with values of ρ taken from $\{0, 0.5, 0.9\}$ and where μ_t and U_t are two vectors of stationary time series 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:

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

The process X_t follows $X_t = H(V_t) + U_t$ and Y_t is generated by (4.13). We estimate A by $\widehat{A}^*$ of (2.26) with the choice of $Q_t = J(V_t) + \eta_t$ and $\eta_t = U_t - \rho\ \mu_t$, in which ρ is estimated by (2.25) in computing $\widehat{A}^*$ and (4.15) below.

Table 4.2. Simulation results based on model (4.13) with $\rho = 0$

	absolute error			standard deviation		
	$\gamma = 0.1$					
n	202	502	802	202	502	802
ASE_1^*	0.0719	0.0472	0.0371	0.0345	0.0227	0.0183
ASE_2^*	0.0112	0.0044	0.0029	0.0112	0.0046	0.0032
ASE_3^*	0.0118	0.0046	0.0029	0.0117	0.0050	0.0030
ASE_4^*	0.0714	0.0471	0.0371	0.0349	0.0223	0.0178
	$\gamma = 0.5$					
ASE_1^*	0.0423	0.0262	0.0213	0.0213	0.0121	0.0102
ASE_2^*	0.0068	0.0025	0.0016	0.0072	0.0025	0.0017
ASE_3^*	0.0064	0.0025	0.0016	0.0068	0.0023	0.0017
ASE_4^*	0.0428	0.0263	0.0213	0.0221	0.0121	0.0103
	$\gamma = 0.9$					
ASE_1^*	0.0106	0.0154	0.0045	0.0067	0.0356	0.0023
ASE_2^*	0.0030	0.0094	0.0003	0.0037	0.0293	0.0003
ASE_3^*	0.0029	0.0105	0.0003	0.0040	0.0352	0.0003
ASE_4^*	0.0105	0.0167	0.0045	0.0068	0.0484	0.0022

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 Tables 4.2–4.4:

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

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

The absolute errors and the standard deviations in Tables 4.2–4.4 together show that the proposed estimation method performs well for the linear endogenous case where

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

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 ρ.

Table 4.3. Simulation results 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

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

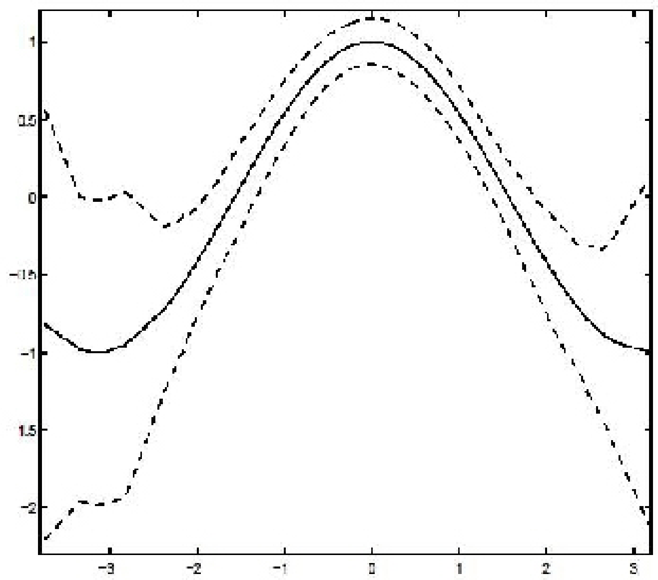

Fig 4.2(a) Plots of $\widehat{g}_1$ and its 95% CI against $(V_{(1)}, \cdots, V_{(n)})$ for $\rho = \gamma = 0$ and $n = 502$.

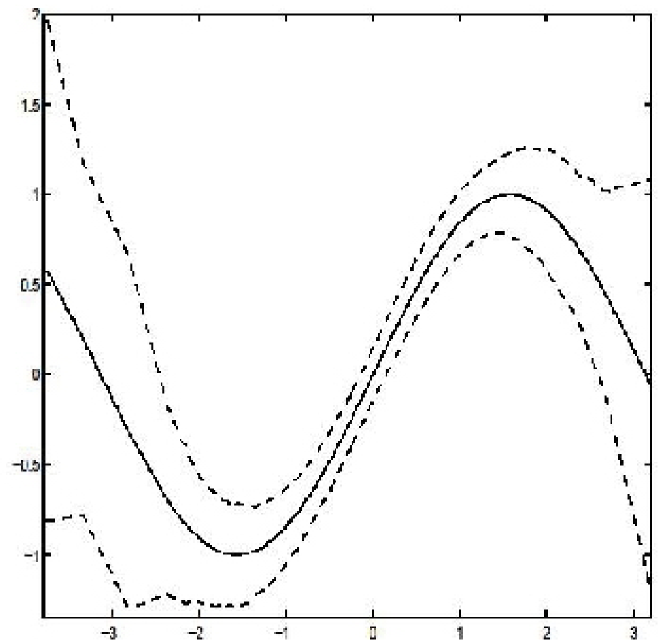

Fig 4.2(b) plots $\widehat{g}_2$ and its 95% CI against $(V_{(1)}, \cdots, V_{(n)})$ for $\rho = \gamma = 0.5$ and $n = 502$.

Table 4.4. Simulation results based on model (4.13) with $\rho = 0.9$

	absolute error			standard deviation		
	$\gamma = 0.1$					
n	202	502	802	202	502	802
ASE_1^*	0.0739	0.0479	0.0371	0.0373	0.0227	0.0175
ASE_2^*	0.0148	0.0062	0.0040	0.0150	0.0061	0.0044
ASE_3^*	0.0149	0.0059	0.0037	0.0152	0.0064	0.0036
ASE_4^*	0.0742	0.0478	0.0371	0.0365	0.0223	0.0174
	$\gamma = 0.5$					
ASE_1^*	0.0423	0.0269	0.0212	0.0223	0.0134	0.0105
ASE_2^*	0.0084	0.0034	0.0020	0.0088	0.0035	0.0020
ASE_3^*	0.0083	0.0033	0.0021	0.0085	0.0033	0.0021
ASE_4^*	0.0425	0.0268	0.0214	0.0232	0.0135	0.0105
	$\gamma = 0.9$					
ASE_1^*	0.0101	0.0060	0.0045	0.0058	0.0034	0.0024
ASE_2^*	0.0019	0.0018	0.0004	0.0021	0.0023	0.0004
ASE_3^*	0.0019	0.0018	0.0004	0.0020	0.0022	0.0004
ASE_4^*	0.0103	0.0058	0.0045	0.0058	0.0035	0.0024

5 Conclusions and Limitations

This paper explores the semiparametric 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 general multivariate case where both the parametric and nonparametric regressors may be nonstationary. A nonparametric kernel estimation method has been 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 method relevant for many potential applications in which 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. Consider, for example, the system

$$Y_t = aX_t + bg(V_t) + \epsilon_t, \quad g(V_t) = \frac{1}{1+V_t^4}, \tag{5.1}$$

$$X_t = H(V_t) + U_t, \quad H(V_t) = cV_t, \ V_t = \sum_{s=1}^{t} v_s, \tag{5.2}$$

where all variables are scalar and satisfy the conditions of Theorem 3.1. In this case, the simple IV estimator $a_{IV} = \left(\sum_{t=1}^{n} X_t V_t\right)^{-1} \left(\sum_{t=1}^{n} V_t Y_t\right)$ 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):

$$n^{-2}\sum_{t=1}^{n} X_t V_t \Rightarrow c\int_0^1 B_v^2,$$

$$n^{-1}\sum_{t=1}^{n} V_t \epsilon_t \Rightarrow \int_0^1 B_v dB_\epsilon,$$

$$\frac{1}{n^{1/4}}\sum_{t=1}^{n} \frac{V_t}{1+V_t^2} \Rightarrow \sqrt{\beta L_0^1} Z,$$

where $n^{-1/2}\sum_{t=1}^{[n\cdot]}(\epsilon_t, v_t) \Rightarrow (B_\epsilon, B_v)$, 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. The model (5.1) - (5.2) is 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 (5.1). 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 occured elsewhere in the sample space. The remaining components of the system are analogous to those in conventional cointegrated systems. Thus, (5.1) - (5.2) 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.

A further limitation is the assumption of exogeneity for the nonstationary regressor V_t. It will be useful to relax this condition in applications to allow the trending mechanism to be endogenous. A final limitation of the model is that each component of $g(\cdot)$ is a scalar function of V_t. For practial work, it will often be useful for g to be a function of several regressors involving both stationary and integrated components. These issues require different treatment of the asymptotic theory and some extension 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: DP0558602 and DP0879088. The work of the second author was supported by the Kelly Foundation and the NSF under Grant Nos. SES 06–47086 and SES 09-56687. The first author would also like to acknowledge useful discussions with Xiaohong Chen and Yuichi Kitamura while visiting Yale University. Thanks from the first author also go to Jiying Yin for excellent computing assistance.

Jiti Gao, School of Economics, The University of Adelaide, Adelaide SA 5005, Australia. Email: jiti.gao@adelaide.edu.au.
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 $E\left[e_t \otimes \eta_t\right] \neq 0$. 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). Since (e_s, U_s, η_s) is a vector of stochastically detrended stationary errors on the one hand and V_t is a sequence of nonstationary time series on the other hand, it is not unreasonable to impose the independence condition between the nonstationary V_t and the stationary $\{(e_s, U_s, \eta_s)\}$. Assumption 3.3(i) enables us to present a relatively clear and concise proof for each of the theorems.

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

$$V_t = V_{t-1} + v_t, \ t \geq 1, \tag{A.1}$$

where $V_0 = 0$ and $\{v_t\}$ is a stationary linear process with $E[v_1] = 0$ and $0 < E[v_1^2] < \infty$. Similar 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 $\widehat{\phi}_{i,k}(x)$ be the probability density function of $\frac{1}{\sqrt{k-i}\ \sigma_v}\sum_{t=i+1}^k v_t$ and $\widehat{\phi}_{i,k}(x|\mathcal{F}_i)$ 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 $k - i \to \infty$,

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

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

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 central limit 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 model of the form (A.1), Assumption 3.4(i) is easily verifiable. Let $g(v) = \boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 v + \boldsymbol{\theta}_2 v^{1+\lambda_0}$ for $0 < \lambda_0 < 1/2$, $n^{\lambda_0} h = O(1)$ and $f_{t,0}(v) = O(v^{-(1+2\lambda_0+\varepsilon_0)})$ for some $\varepsilon_0 > 0$ as $t \to \infty$ and $v \to \infty$. It then follows that

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

which implies Assumption 3.4(i).

Assumption 3.4(ii) is similarly verifiable. Consider the case where $g(v) = \boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 v$ and $H(v) = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 v + \boldsymbol{\phi}_2 v^{1+\lambda_1}$ for $1 < \lambda_1 < \frac{1}{2}$. Let $n^{\frac{1}{2}+\lambda_1-\varepsilon_1} h^2 = O(1)$ ($\varepsilon_1 < \frac{1}{2} - \lambda_1$) and $f_{t,0}(v) = O(v^{-(1+2\lambda_1+\varepsilon_1)})$ for some $\varepsilon_1 > 0$ as $t \to \infty$ and $v \to \infty$. It can now be seen that Assumption 3.4(ii) holds. 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 is commonly used in the stationary time series case. Assumption 3.5(ii) requires that the rate $b_n^{-2} \to \infty$ is slower than $\sqrt{h} \to 0$ and the rate $b_n^4 \to 0$ is slower than that of $\sqrt{n}h \to \infty$. 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 $P(\widehat{p}_n(v) > b_n) \geq P(\widehat{p}_n(v) > \lambda_0)$ for any positive constant $\lambda_0 > 0$ such that $\lambda_0 > b_n$. In view of this, in order to verify Assumption 3.5(iii), it suffices to show that

$$P(\widehat{p}_n(V_t) > \lambda_0) \to 1, \tag{A.5}$$

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

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

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

where $\widetilde{p}_{(n-t)}(0) = \frac{1}{\sqrt{n-t}h} \sum_{i=1}^{n-t} K\left(\frac{\widetilde{V}_{t+i}(t)}{h}\right)$, $p_s(0)$ is a positive local-time random variable, and we have used the point–wise convergence of $\widetilde{p}_m(0) \to p_s(0)$ as $m \to \infty$ by virtue of theorem 2.1 of Wang and Phillips (2009a). Equation (A.6) implies that uniformly in $1 \leq t \leq [\frac{n}{2}]$,

$$P(\widehat{p}_n(V_t) > \lambda_0) \to 1, \tag{A.7}$$

for some $\lambda_0 > 0$ as $n \to \infty$.

Similarly, we have uniformly in $\left[\frac{n}{2}\right] + 1 \leq t \leq n$,

$$\begin{aligned}
\widehat{p}_n(V_t) &= \frac{1}{\sqrt{n}h} \sum_{k=1}^{t-1} K\left(\frac{\overline{V}_k(t)}{h}\right) + \frac{1}{\sqrt{n}h} \sum_{k=t+1}^{n} K\left(\frac{\widetilde{V}_k(t)}{h}\right) + \frac{1}{\sqrt{n}h} K(0) \\
&\geq \frac{1}{\sqrt{n}h} \sum_{k=1}^{t-1} K\left(\frac{\overline{V}_k(t)}{h}\right) + o_P(1) = \frac{\sqrt{t-1}}{\sqrt{n}} \frac{1}{\sqrt{t-1}h} \sum_{i=1}^{t-1} K\left(\frac{\overline{V}_i(t)}{h}\right) + o_P(1) \\
&\equiv \frac{\sqrt{t-1}}{\sqrt{n}} \overline{p}_{(t-1)}(0) + o_P(1) = \frac{\sqrt{t-1}}{\sqrt{n}} p_s(0) + o_P(1),
\end{aligned} \tag{A.8}$$

where $\overline{p}_{(t-1)}(0) = \frac{1}{\sqrt{t-1}h} \sum_{i=1}^{t-1} K\left(\frac{\overline{V}_i(t)}{h}\right)$, and we again use the pointwise convergence of $\overline{p}_m(0) \to p_s(0)$ as $m \to \infty$ as in (A.6). This implies that equation (A.7) also holds uniformly in $[\frac{n}{2}] + 1 \leq t \leq n$. Therefore, Assumption 3.5(iii) is verified.

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* $n \to \infty$

$$\frac{1}{n}\widetilde{X}'\widetilde{Q} = \frac{1}{n}U'\eta + o_P(1) \to_P E\left[U_1\eta_1'\right]. \tag{A.9}$$

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} e_t \otimes \eta_t \to_D N\left(0, \Omega_1^*\right), \tag{A.10}$$

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*

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

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

$$\left(\widehat{A}^* - A\right)\widetilde{X}^\tau\widetilde{Q} \quad = \quad \widetilde{e}'\widetilde{Q} + \widetilde{G}'\widetilde{Q} = \sum_{t=1}^{n} e_t\widetilde{Q}_t'F_t + \sum_{t=1}^{n}\widetilde{G}_t\widetilde{Q}_t' - \sum_{t=1}^{n}\overline{e}_t\widetilde{Q}_t'F_t,$$

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

$$\sum_{t=1}^{n}\widetilde{G}_t\widetilde{Q}_t'F_t \quad = \quad o_P(\sqrt{n}), \tag{A.11}$$

$$\sum_{t=1}^{n}\overline{e}_t\widetilde{Q}_t'F_t \quad = \quad o_P(\sqrt{n}), \tag{A.12}$$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} e_t\widetilde{Q}_t'F_t \quad \to_D \quad N\left(0, \Omega_1^*\right), \tag{A.13}$$

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

In order to prove (A.11)–(A.13), it suffices to show that for large enough n

$$\sum_{t=1}^{n} \widetilde{G}_t \eta_t' F_t = o_P(\sqrt{n}), \tag{A.14}$$

$$\sum_{t=1}^{n} \widetilde{G}_t \overline{\eta}_t' F_t = o_P(\sqrt{n}), \tag{A.15}$$

$$\sum_{t=1}^{n} \widetilde{G}_t \widetilde{J}_t' F_t = o_P(\sqrt{n}), \tag{A.16}$$

$$\sum_{t=1}^{n} \overline{e}_t \eta_t' F_t = o_P(\sqrt{n}), \tag{A.17}$$

$$\sum_{t=1}^{n} \overline{e}_t \overline{\eta}_t' F_t = o_P(\sqrt{n}), \tag{A.18}$$

$$\sum_{t=1}^{n} \overline{e}_t \widetilde{J}_t' F_t = o_P(\sqrt{n}), \tag{A.19}$$

$$\sum_{t=1}^{n} e_t \overline{\eta}_t' F_t = o_P(\sqrt{n}), \tag{A.20}$$

$$\sum_{t=1}^{n} e_t \widetilde{J}_t' F_t = o_P(\sqrt{n}), \tag{A.21}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t \eta_t' F_t \to_D N\left(0, \Omega_1^*\right), \tag{A.22}$$

where $\overline{\eta}_t = \sum\limits_{s=1}^{n} w_{ns}(V_t)\eta_s$. Since the finite dimensionality of p and d does not affect the validity of (A.14)–(A.22), 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

$$\frac{1}{\sqrt{n}h} \sum_{j=1}^{n} K\left(\frac{V_j - v}{h}\right)(g(V_j) - g(v)) \tag{A.23}$$
$$= \frac{g^{(1)}(v)}{\sqrt{n}h} \sum_{j=1}^{n} K\left(\frac{V_j - v}{h}\right)(V_j - v)(1 + o_P(1)).$$

In view of (A.23), in order to prove (A.14), it suffices to show that for n large enough

$$\sum_{t=1}^{n} \Delta_n(V_t)\eta_t F_t = o_P(\sqrt{n}), \tag{A.24}$$

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

$$\sum_{t=1}^{\infty} |E[\eta_1 \eta_t]| < \infty, \tag{A.25}$$

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

$$\begin{aligned}
E\left(\sum_{t=1}^{n}\eta_t\Delta_n(V_t)F_t\right)^2 &= \sum_{t=1}^{n}E\left[\eta_t^2\ \Delta_n(V_t)F_t\right]^2 \\
&+ \sum_{t_1=1}^{n}\sum_{t_2\neq t_1}E\left[\eta_{t_1}\eta_{t_2}\cdot\Delta_n(V_{t_1})F_{t_1}\Delta_n(V_{t_2})F_{t_2}\right] \\
&\leq Cb_n^{-2}\sum_{t=1}^{n}E\left[\eta_t^2\right]E\left[\Gamma_n(V_t)F_t\right]^2 \\
&+ Cb_n^{-2}\frac{1}{2}\sum_{t_1=1}^{n}\sum_{t_2\neq t_1}\left|E\left[\eta_{t_1}\eta_{t_2}\right]\right|E\left[\Gamma_n^2(V_{t_1})F_{t_1}+\Gamma_n^2(V_{t_2})F_{t_2}\right] \\
&\leq Cb_n^{-2}\sum_{t=1}^{n}E\left[\Gamma_n(V_t)F_t\right]^2,
\end{aligned} \tag{A.26}$$

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

By Assumption 3.3(i), (A.23)–(A.26) and the definition of $\Delta_n(V_t)$, we have

$$E\left(\sum_{t=1}^{n}\eta_t\Delta_n(V_t)F_t\right)^2\leq\Delta_{n,1}+\Delta_{n,2}, \tag{A.27}$$

where

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

and

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

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

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

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

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

Similarly,

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

We have therefore shown that

$$\Delta_{n,1} = o(n). \tag{A.28}$$

Next consider $\Delta_{n,2}$. Analogous to the calculation of $\Delta_{n,1}$, we need only to deal with the case of $k_2 > k_1 > t$ and the other cases can be dealt with in a similar way. By

Assumptions 3.3(ii), 3.4(i) and 3.5(i)(ii), we have

$$
\begin{aligned}
& b_n^{-2} n^{-1} h^{-2} \sum_{t=1}^{n-2} E\left(\left[g^{(1)}(V_t)\right]^2 \sum_{k_1=t+1}^{n} \sum_{k_2=k_1+1}^{n}\right. \\
& \times \left. E\left[(V_{k_2}-V_t)(V_{k_1}-V_t) K\left(\frac{V_{k_2}-V_t}{h}\right) K\left(\frac{V_{k_1}-V_t}{h}\right) |\mathcal{F}_t\right]\right) \\
\leq\ & C b_n^{-2} n^{-1} h^2 \sum_{t=1}^{n-2} E\left[g^{(1)}(V_t)\right]^2 \sum_{k_1=t+1}^{n} \sum_{k_2=k_1+1}^{n} \varphi_{k_2-k_1} \varphi_{k_1-t} \\
\leq\ & C b_n^{-2} L_s^2(n) h^2 \sum_{t=1}^{n} E\left[g^{(1)}(V_t)\right]^2 \leq O\left(b_n^{-2} n L_s^2(n) h\right) = o(n).
\end{aligned} \tag{A.29}
$$

The detailed calculation of (A.29) 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.28), implies that (A.14) holds.

We next show that (A.15) holds. In view of (A.23), it suffices to show that

$$
\sum_{t=1}^{n} \widehat{\eta}_t \Delta_n(V_t) F_t = o_P(\sqrt{n}), \tag{A.30}
$$

where $\widehat{\eta}_t = \frac{1}{\sqrt{nh}\ \widehat{p}_n(V_t)} \left(\sum_{k=1}^{n} K\left(\frac{V_k - V_t}{h}\right) \eta_k\right)$. Similar to the arguments used in (A.26), we have

$$
\begin{aligned}
& E\left(\sum_{t=1}^{n} \widehat{\eta}_t \Delta_n(V_t) F_t\right)^2 \leq\ C b_n^{-4} h^{-4} n^{-2} \\
& \times E\left(\sum_{k=1}^{n}\left\{\sum_{t=1}^{n}\sum_{j=1}^{n}(V_j - V_t) K\left(\frac{V_k - V_t}{h}\right) K\left(\frac{V_j - V_t}{h}\right)\right\} g^{(1)}(V_k)\eta_k\right)^2 \\
=\ & C b_n^{-4} h^{-2} n^{-2} E\left(\sum_{k=1}^{n} M(V_k)\eta_k\right)^2,
\end{aligned} \tag{A.31}
$$

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.26), we have

$$
E\left(\sum_{k=1}^{n} M(V_k)\eta_k\right)^2 = E\left(E\left[\left(\sum_{k=1}^{n} M(V_k)\eta_k\right)^2 |\mathcal{F}_V\right]\right) \leq C \sum_{k=1}^{n} E\left(M(V_k)\right)^2, \tag{A.32}
$$

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

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

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

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

Equations (A.31) and (A.32) thus imply (A.30). Therefore, equation (A.15) is proved. By Assumption 3.3(ii) and (A.25), we have

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

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

$$\sum_{t=1}^{\infty}|E[e_1e_t]| < \infty \quad \text{and} \quad \sum_{t=1}^{\infty}|E[e_1\eta_1e_t\eta_t]| < \infty. \qquad \text{(A.34)}$$

By A4, (A.34) and using the same arguments as in the derivations for $\Delta_{n,1,1}$ and $\Delta_{n,1,2}$, we have

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

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

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

Thus, by (A.33), (A.35) and (A.35), we have

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

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

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

Similar to (A.24), equation (A.35) implies

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

by Assumption 3.5(i)(ii). Hence, (A.17) is proved.

We now show that

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

Note that

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

By Assumption 3.3(i), we have

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

By Assumptions 3.3(i) and applying the proof of (A.35), we can show that

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

Similarly, by (A.25) and (A.34), we have

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

It follows from (A.39)–(A.41) that

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

Observe that

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

By (A.25) and (A.34) as well as following the calculation of the order of $I_{n,1}^{(j)}$ above, we have

$$I_{n,2}^{(j)} = O\left(n^{\frac{5}{2}}L_s^3(n)h^{-1}\right), \quad j = 1, \cdots, 4. \tag{A.44}$$

By (A.43)–(A.44), we have

$$I_{n,2} = O\left(n^{\frac{5}{2}}L_s^3(n)h^{-1}\right).$$

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

$$E\left(\sum_{t=1}^{n}\left[\sum_{k=1}^{n} K_{V_t,h}(V_k)\eta_k\right]\left[\sum_{q=1}^{n} K_{V_t,h}(V_q)e_q\right]\right)^2 = O\left(n^{\frac{5}{2}}L_s^3(n)h^{-1}\right).$$

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

$$\begin{aligned} &\sum_{t=1}^{n}\left[\sum_{k=1}^{n} w_{nk}(V_t)\eta_k\right]\left[\sum_{q=1}^{n} w_{nq}(V_t)e_q\right] F_t \\ = \quad & O_P\left(n^{\frac{1}{4}}L_s^{\frac{3}{2}}(n)h^{-1/2}b_n^{-2}\right) = o_P(\sqrt{n}), \end{aligned}$$

which implies that (A.18) holds.

Finally, we prove (A.20) and (A.22). The proof of (A.20) 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

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

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

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

using the fact that

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

By (A.45) and (A.46), equation (A.22) is proved.

We finish the proof of Theorem 3.1 by completing the proofs of (A.16), (A.19) and (A.21). 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.27)–(A.29), we can show that

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

for some $0 < \varepsilon_1 < \frac{1}{2}$, which implies that (A.16) holds. The proofs of (A.19) and (A.21) are similar to that of (A.14) and so the details are omitted here.

7.4 Proof of Theorem 3.2

Observe that

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

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

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

$$\sqrt{\sum_{t=1}^{n} K\left(\frac{v - V_t}{h}\right)} \sum_{t=1}^{n} w_{nt}(v)\ U_t = \frac{1}{\sqrt{\sum_{t=1}^{n} K\left(\frac{v-V_t}{h}\right)}} \sum_{t=1}^{n} K\left(\frac{v - V_t}{h}\right) U_t$$
$$\to_D N\left(0, \Omega_u\right), \tag{A.48}$$

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

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

$$\sqrt{\sum_{t=1}^{n} K\left(\frac{v - V_t}{h}\right)} \sum_{t=1}^{n} w_{nt}(v)\ e_t = \frac{1}{\sqrt{\sum_{t=1}^{n} K\left(\frac{v-V_t}{h}\right)}} \sum_{t=1}^{n} K\left(\frac{v - V_t}{h}\right) e_t$$
$$\to_D N\left(0, \Omega_e\right), \tag{A.51}$$

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

The proof of (A.47) follows from existing results (see, for example, Theorem 5.1 of Karlsen and Tjøstheim 2001, 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

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

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 $\mathcal{V} = (V_1, \cdots, V_n)$ 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 $\widetilde{Z}_t = \left(Z_t - \sum_{s=1}^n w_{ns}(V_t) Z_s\right) F_t$, we have

$$\begin{aligned}
&\widetilde{Y}_t = A\widetilde{X}_t + \widetilde{g}(V_t) + \widetilde{\epsilon}_t = A\widetilde{X}_t + \widetilde{g}(V_t) + \widetilde{e}_t, \\
&Y_t - \widehat{A}^* X_t - \widehat{g}^*(V_t) = \widetilde{Y}_t - \widehat{A}^* \widehat{X}_t \\
= \quad &\left(A - \widehat{A}^*\right) \widetilde{X}_t + \widetilde{g}(V_t) + \widetilde{e}_t. \qquad \text{(A.53)}
\end{aligned}$$

Observe that

$$\begin{aligned}
&\sum_{t=1}^n \left(Y_t - \widehat{A}^* X_t - \widehat{g}_n^*(V_t)\right) \left(Y_t - \widehat{A}^* X_t - \widehat{g}_n^*(V_t)\right)' \\
= \quad &\sum_{t=1}^n \left(\left(A - \widehat{A}^*\right) \widetilde{X}_t + \widetilde{g}(V_t) + \widetilde{e}_t\right) \left(\left(A - \widehat{A}^*\right) \widetilde{X}_t + \widetilde{g}(V_t) + \widetilde{e}_t\right)' \\
= \quad &\sum_{t=1}^n \widetilde{e}_t \widetilde{e}_t' + \sum_{t=1}^n \left(A - \widehat{A}^*\right) \widetilde{X}_t \widetilde{X}_t' \left(A - \widehat{A}^*\right)' + \sum_{t=1}^n \widetilde{g}(V_t) \widetilde{g}(V_t)' \\
&+ 2\sum_{t=1}^n \left(A - \widehat{A}^*\right) \widetilde{X}_t \widetilde{e}_t' + 2\sum_{t=1}^n \left(A - \widehat{A}^*\right) \widetilde{X}_t \widetilde{g}(V_t)' + 2\sum_{t=1}^n \widetilde{g}(V_t) \widetilde{e}_t' \\
\equiv \quad &\sum_{j=1}^6 S_n(j). \qquad \text{(A.54)}
\end{aligned}$$

We show that as $n \to \infty$

$$\frac{1}{n} S_n(1) \to_P E\left[e_1 e_1'\right] \quad \text{and} \quad \frac{1}{n} S_n(j) \to_P 0 \qquad \text{(A.55)}$$

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

$$\sum_{t=1}^n \widetilde{e}_t \widetilde{e}_t' = \sum_{t=1}^n e_t e_t' F_t + \sum_{t=1}^n \overline{e}_t \overline{e}_t' F_t + 2\sum_{t=1}^n e_t \overline{e}_t' F_t, \qquad \text{(A.56)}$$

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

$$\frac{1}{n} \sum_{t=1}^n e_t e_t' F_t \to_P E\left[e_1 e_1'\right], \quad \frac{1}{n} \sum_{t=1}^n \overline{e}_t e_t' F_t \to_P 0 \quad \text{and} \quad \frac{1}{n} \sum_{t=1}^n \overline{e}_t \overline{e}_t' F_t \to_P 0. \qquad \text{(A.57)}$$

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 (B.2)–(**??**) imply (A.57) and the second part of (A.55) when U_s, η_t, $\widetilde{J}(V_t)$ and $\widetilde{H}(V_t)$ are replaced by e_s, e_t and $\widetilde{g}(V_t)$, 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

$$\sum_{t=1}^{n} \widetilde{X}_t \widetilde{Q}_t F_t = \sum_{t=1}^{n} \left(X_t - \sum_{k=1}^{n} w_{nk}(V_t) X_k \right) \left(Q_t - \sum_{q=1}^{n} w_{nq}(V_t) Q_q \right) F_t \quad \text{(B.1)}$$

$$= \sum_{t=1}^{n} \left(U_t + \widetilde{H}(V_t) - \sum_{k=1}^{n} w_{nk}(V_t) U_k \right) \left(\eta_t + \widetilde{J}(V_t) - \sum_{q=1}^{n} w_{nq}(V_t) \eta_q \right) F_t$$

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

Similar to (A.14)–(A.22), in order to prove Lemma A.1(i), it suffices to show that

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

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

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

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

$$\sum_{t=1}^{n} \left(\sum_{s=1}^{n} w_{ns}(V_t) \eta_s \right) \widetilde{H}(V_t) \ F_t = o_P(n), \tag{B.6}$$

$$\sum_{t=1}^{n} \widetilde{J}(V_t) U_t F_t = o_P(n), \tag{B.7}$$

$$\sum_{t=1}^{n} \widetilde{H}(V_t) \eta_t F_t = o_P(n), \tag{B.8}$$

$$\sum_{t=1}^{n} \widetilde{H}(V_t) \widetilde{J}(V_t) F_t = o_P(n), \tag{B.9}$$

$$\frac{1}{n} \sum_{t=1}^{n} U_t \eta_t F_t \to_P \Sigma_{u\eta}, \tag{B.10}$$

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

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

$$\sum_{t=1}^{n} \left(\sum_{s=1}^{n} \widehat{w}_{ns}(V_t) U_s \right) \left(\sum_{k=1}^{n} \widehat{w}_{nk}(V_t) \eta_k \right) F_t = o_P(n). \tag{B.11}$$

Observe that

$$\begin{aligned}
& E\left[\sum_{t=1}^{n} \left(\sum_{k=1}^{n} \widehat{w}_{nk}(V_t) U_k \right) \left(\sum_{q=1}^{n} \widehat{w}_{nq}(V_t) \eta_q \right) F_t \right]^2 \qquad \text{(B.12)} \\
= \ & \sum_{t=1}^{n} E\left[\left(\sum_{k=1}^{n} \widehat{w}_{nk}(V_t) U_k \right)^2 \left(\sum_{q=1}^{n} \widehat{w}_{nq}(V_t) \eta_q \right)^2 F_t \right] \\
+ \ & \sum_{t_1=1}^{n} \sum_{t_2 \neq t_1} E\left[\left(\sum_{k_1=1}^{n} \widehat{w}_{nk_1}(V_{t_1}) U_{k_1} \right) \left(\sum_{q_1=1}^{n} \widehat{w}_{nq_1}(V_{t_1}) \eta_{q_1} \right) \right. \\
\times \ & \left. \left[\left(\sum_{k_2=1}^{n} \widehat{w}_{nk_2}(V_{t_2}) U_{k_2} \right) \left(\sum_{q_2=1}^{n} \widehat{w}_{nq_2}(V_{t_2}) \eta_{q_2} \right) F_{t_1} F_{t_2} \right] \right. \\
=: \ & \Pi_{n,1} + \Pi_{n,2}.
\end{aligned}$$

By Assumption 3.3(i), we have

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

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

$$
\begin{aligned}
\Pi_{n,1}(1) = & \sum_{t=1}^{n}\sum_{k=1}^{n} E\left[\widehat{w}_{nk}^{4}(V_t)F_t\right] E\left[U_k^2\eta_k^2\right] \\
+ & \sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{q\neq k} E\left[\widehat{w}_{nk}^{2}(V_t)\widehat{w}_{nq}^{2}(V_t)F_t\right] E\left[U_k^2\eta_q^2\right] \\
= & :\Pi_{n,1}(1,1)+\Pi_{n,1}(1,2).
\end{aligned} \tag{B.13}
$$

By Assumptions 3.3 and 3.5(ii), we have

$$
\begin{aligned}
& \Pi_{n,1}(1,1) = \sum_{t=1}^{n}\sum_{k=1}^{n} E\left[\widehat{w}_{nk}^{4}(V_t)F_t\right] E\left[U_k^2\eta_k^2\right] \le C\sum_{t=1}^{n}\sum_{k=1}^{n} E\left[\widehat{w}_{nk}^{4}(V_t)F_t\right] \\
= \; & O\left(n^{-2}h^{-4}b_n^{-4}\left(nK^4(0)+\sum_{t=2}^{n}\sum_{k=1}^{t-1} E\left[K^4(\tfrac{V_t-V_k}{h})\right]\right)\right) \\
= \; & O\left(n^{-2}h^{-4}b_n^{-4}\left(nK^4(0)+\sum_{t=2}^{n}\sum_{k=1}^{t-1}\int K^4\left(\tfrac{u}{\varphi_{t-k}h}\right)f_{t,k}(u)du\right)\right) \\
= \; & O\left(n^{-2}h^{-4}b_n^{-4}\left(nK^4(0)+\sum_{t=2}^{n}\sum_{k=1}^{t-1} h\varphi_{t-k}\int K^4(u)f_{t,k}\left(h\varphi_{t-k}u\right)du\right)\right) \\
= \; & O\left(n^{-2}h^{-4}b_n^{-4}\left(nK^4(0)+h\sum_{t=2}^{n}\sum_{k=1}^{t-1}\tfrac{1}{\sqrt{(t-k)}}L_s(t-k)\right)\right) \\
= \; & O\left(n^{-1}h^{-4}b_n^{-4}+n^{-\frac{1}{2}}L_s(n)h^{-3}b_n^{-4}\right) = o(n^2)
\end{aligned} \tag{B.14}
$$

and by Assumptions 3.3 and 3.5(ii) again

$$
\begin{aligned}
\Pi_{n,1}(1,2) &\le C\sum_{t=1}^{n}\sum_{k_1=1}^{n}\sum_{k_2\neq k_1} E\left[\widehat{w}_{nk_1}^2(V_t)\widehat{w}_{nk_2}^2(V_t)F_t\right] \\
&\le Cn^{-2}h^{-4}b_n^{-4}\sum_{t=1}^{n}\sum_{k_1=1}^{n}\sum_{k_2\neq k_1} E\left[K^2\left(\tfrac{V_t-V_{k_1}}{h}\right)K^2\left(\tfrac{V_t-V_{k_2}}{h}\right)\right] \\
&\le Cn^{-2}h^{-4}b_n^{-4}\sum_{t=2}^{n}\sum_{k_1=1}^{t-1} K^2(0)E\left[K^2\left(\tfrac{V_t-V_{k_1}}{h}\right)\right] \\
&+ Cn^{-2}h^{-4}b_n^{-4}\sum_{t=3}^{n}\sum_{k_1=2}^{t-1}\sum_{k_2=1}^{k_1-1} E\left[K^2\left(\tfrac{V_t-V_{k_1}}{h}\right)K^2\left(\tfrac{V_t-V_{k_2}}{h}\right)\right] \\
&\le Cn^{-\frac{1}{2}}L_s(n)h^{-3}b_n^{-4} + Cn^{-2}h^{-4}b_n^{-4}\sum_{t=3}^{n}\sum_{k_1=2}^{t-1}\sum_{k_2=1}^{k_1-1}\varphi_{t-k_1} \\
&\times E\left(\int K^2\left(\tfrac{w_1}{h}\right)K^2\left(\tfrac{w_1+V_{k_1}-V_{k_2}}{h}\right)f_{t,k_1}(\varphi_{t-k_1}w_1|\mathcal{F}_{k_1})\right) \\
&\le Cn^{-\frac{1}{2}}L_s(n)h^{-3}b_n^{-4} + Cn^{-2}h^{-4}b_n^{-4}\sum_{t=3}^{n}\sum_{k_1=2}^{t-1}\sum_{k_2=1}^{k_1-1} h^2\varphi_{t-k_1}\varphi_{k_1-k_2} \\
&\times \int K^2(u_1+u_2)K^2(u_2)du_1du_2 \\
&= O(n^{-\frac{1}{2}}L_s(n)h^{-3}b_n^{-4} + n^{-2}h^{-2}b_n^{-4}\sum_{t=3}^{n}\sum_{k_1=2}^{t-1}\sum_{k_2=1}^{k_1-1}(t-k_1)^{-\frac{1}{2}}L_s(t-k_1) \\
&\times (k_1-k_2)^{-\frac{1}{2}}L_s(k_1-k_2)) \\
&= O(n^{-\frac{1}{2}}L_s(n)h^{-3}b_n^{-4} + L_s^2(n)h^{-2}b_n^{-4}) = o(n^2).
\end{aligned} \tag{B.15}
$$

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

$$
\Pi_{n,1}(2) = O\left(L_s^2(n)h^{-2}b_n^{-4}\right) = o(n^2). \tag{B.16}
$$

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

$$
\begin{aligned}
E[U_{k_1}^2\eta_{k_2}\eta_{k_3}] &= E[U_{k_1}^2\eta_{k_2}\eta_{k_3}] - E[U_{k_1}^2]E[\eta_{k_2}\eta_{k_3}] \\
&+ E[U_{k_1}^2]E[\eta_{k_2}\eta_{k_3}] - E[U_{k_1}^2]E[\eta_{k_2}]E[\eta_{k_3}] \\
&\le C\left(\alpha_\zeta^{\gamma_2/(4+\gamma_2)}(k_1-k_2) + \alpha_\zeta^{\gamma_2/(4+\gamma_2)}(k_2-k_3)\right),
\end{aligned} \tag{B.17}
$$

which implies

$$
\begin{aligned}
\Pi_{n,1}(3) &\le C\sum_{t=1}^{n}\sum_{k_1=1}^{n}\sum_{k_2<k_1}\sum_{k_3<k_2} E[\widehat{w}_{nk_1}^2(V_t)\widehat{w}_{nk_2}(V_t)\widehat{w}_{nk_3}(V_t)F_t] \\
&\times \left(\alpha_\zeta^{\gamma_2/(4+\gamma_2)}(k_1-k_2) + \alpha_\zeta^{\gamma_2/(4+\gamma_2)}(k_2-k_3)\right) \\
&\le CL_s^2(n)h^{-1}b_n^{-4} = o(n^2).
\end{aligned} \tag{B.18}
$$

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

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

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

$$\Pi_{n,1}(j) = O\left(\sqrt{n} L_s^3(n) b_n^{-4}\right) = o(n^2). \tag{B.19}$$

By (B.13)–(B.19), we also have

$$\Pi_{n,1} = o(n^2). \tag{B.20}$$

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

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

$$+ C\sum_{t_1=1}^{n}\sum_{t_2\neq t_1}\sum_{k_1=1}^{n}\sum_{k_2<k_1}\sum_{q_1<k_2} E[\widehat{w}_{nk_1}(V_{t_1})\widehat{w}_{nk_1}(V_{t_2})\widehat{w}_{nk_2}(V_{t_1})\widehat{w}_{nq_1}(V_{t_2})F_{t_1}F_{t_2}]$$
$$\times\ E[U_{k_1}^2\eta_{k_2}\eta_{q_1}]$$
$$+ C\sum_{t_1=1}^{n}\sum_{t_2\neq t_1}\sum_{k_1=1}^{n}\sum_{k_2<k_1}\sum_{q_1<k_2} E[\widehat{w}_{nk_1}(V_{t_1})\widehat{w}_{nk_1}(V_{t_2})\widehat{w}_{nk_2}(V_{t_1})\widehat{w}_{nq_1}(V_{t_2})F_{t_1}F_{t_2}]$$
$$\times\ E[\eta_{k_1}^2 U_{k_2}U_{q_1}]$$
$$+ C\sum_{t_1=1}^{n}\sum_{t_2\neq t_1}\sum_{k_1=1}^{n}\sum_{k_2<k_1}\sum_{q_1<k_2} E[\widehat{w}_{nk_1}^2(V_{t_1})\widehat{w}_{nk_2}(V_{t_2})\widehat{w}_{nq_1}(V_{t_2})F_{t_1}F_{t_2}]$$
$$\times\ E[U_{k_1}^2\eta_{k_2}\eta_{q_1}]$$
$$+ C\sum_{t_1=1}^{n}\sum_{t_2\neq t_1}\sum_{k_1=1}^{n}\sum_{k_2<k_1}\sum_{q_1<k_2} E[\widehat{w}_{nk_1}^2(V_{t_1})\widehat{w}_{nk_2}(V_{t_2})\widehat{w}_{nq_1}(V_{t_2})F_{t_1}F_{t_2}]$$
$$\times\ E[\eta_{k_1}^2 U_{k_2}U_{q_1}]$$
$$+ C\sum_{t_1=1}^{n}\sum_{t_2\neq t_1}\sum_{k_1=1}^{n}\sum_{k_2<k_1}\sum_{q_1<k_2}\sum_{q_2<q_1} E[\widehat{w}_{nk_1}(V_{t_1})\widehat{w}_{nk_2}(V_{t_2})\widehat{w}_{nq_1}(V_{t_1})\widehat{w}_{nq_2}(V_{t_2})F_{t_1}F_{t_2}]$$
$$\times\ E[U_{k_1}U_{k_2}\eta_{q_1}\eta_{q_2}]$$
$$\equiv \sum_{j=1}^{9}\Pi_{n,2}(j).$$

By Assumption 3.3, we have

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

Similarly, by the Hölder inequality we have

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

Analogously, we have

$$\Pi_{n,2}(4) = \Pi_{n,2}(5) = O\left(n^{\frac{1}{2}}L_s^3(n)h^{-1}b_n^{-4}\right) = o(n^2)$$

and

$$\Pi_{n,2}(6) = O\left(n^{\frac{1}{2}}L_s^3(n)h^{-1}b_n^{-4}\right) = o(n^2).$$

Applying the proofs of (B.18) and (B.19), we have

$$\Pi_{n,2}(7) = O\left(n^{\frac{1}{2}} L_s^3(n) b_n^{-4}\right) = o(n^2),$$

$$\Pi_{n,2}(8) = O\left(n^{\frac{1}{2}} L_s^3(n) b_n^{-4}\right) = o(n^2)$$

and

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

From the above arguments, we obtain

$$\Pi_{n,2} = o(n^2). \tag{B.21}$$

By (B.12), (B.20), (B.21) and the Markov inequality, we have shown (B.11), which implies that (B.2) holds.

Using the same arguments as in the proof of Theorem 3.1, we can prove (B.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

$$\frac{1}{n}\sum_{t=1}^{n} U_t \eta_t \to_P \Sigma_{u\eta}, \tag{B.22}$$

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

$$\frac{1}{n}\sum_{t=1}^{n} U_t \eta_t F_t^c = o_P(1). \tag{B.23}$$

By (B.22) and (B.23) and noting that

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

(B.10) follows directly.

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

8.2 Proof of Lemma A.1(ii)

The result is simply 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 X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Baltagi, B. H., and Q. Li (2002): "On Instrumental Variable Estimation of Semiparametric Dynamic Panel Data Models," *Economics Letters*, 76, 1–9.

Berkes, I., and L. Horváth (2006): "Convergence of Integral Functionals of Stochastic Processes," *Econometric Theory*, 22, 304–322.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi–nonparametric IV Estimation of Shape–Invariant Engel Curves," *Econometrica*, 75, 1613–1671.

Chen, J., J. Gao, and D. Li (2008): "Semiparametric Regression Estimation in Null–Recurrent Time Series," Working paper available at http://www.adelaide.edu.au/directory/jiti.gao.

Fan, J., and Q. Yao (2003): *Nonlinear Time Series: Nonparametric and Parametric Methods.* New York: Springer.

Florens, J. P., J. Johannes, and S. van Bellegem (2007): "Instrumental Regression in Partially Linear Models," Working paper available from the authors websites.

Gao, J. (2007): *Nonlinear Time Series: Semiparametric and Nonparametric Methods.* London: Chapman & Hall/CRC.

Gao, J., Z. Lu, and D. Tjøstheim (2006): "Estimation in Semiparametric Spatial Regression," *Annals of Statistics*, 34, 1395–1435.

Granger, C. W. J., and P. Newbold (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120.

Granger, C. W. J., and P. Newbold (1977): *Forecasting Economic Time Series.* New York: Academic Press.

Greene, W. H. (2005): *Econometric Analysis*, Fifth Edition. Prentice Hall International, Inc.

Hall, P., and C. C. Heyde (1980): *Martingale Limit Theory and Its Applications.* New York: Academic Press.

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., H. Liang, and J. Gao (2000): *Partially Linear Models.* New York: Physica-Verlag.

Heckman, J., H. Ichimura, H., J. Smith, and P. Todd (1998): "Characterizing Selection Bias using Experimental Data," *Econometrica*, 66, 1017–1098.

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.

Johansen, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

Johansen, S. (1995): *Likelihood–based Inference in Cointegrated Vector Autoregressive Models.* New York: Oxford University Press.

Johansen, S. (2000): "A Small Sample Correction for the Test of Cointegrating Rank in the Vector Autoregressive Model," *Econometrica*, 70, 1929–1961.

Juhl, T., and Z. Xiao (2005): "Partially Linear Models with Unit Roots," *Econometric Theory*, 21, 877–906.

Karlsen, H. A., and D. Tjøstheim (1998): "Nonparametric Estimation in Null Recurrent Time Series," Sonderforschungsbereich 373, 50. Humboldt Univ., Berlin.

Karlsen, H. A., and D. Tjøstheim (2001): "Nonparametric Estimation in Null Recurrent Time Series," *Annals of Statistics*, 29, 372-416.

Karlsen, H. A., T. Mykelbust, and D. Tjøstheim (2007): "Nonparametric Estimation in a Nonlinear Cointegration Type Model," *Annals of Statistics*, 35, 252–299.

Kasparis, I., and P. C. B. Phillips (2009): "Dynamic Misspecification in Nonparametric Cointegrating Regression," Cowles Foundation Discussion Paper No. 1700.

Li, Q., and J. Racine (2007): *Nonparametric Econometrics: Theory and Practice.* Princeton: Princeton University Press.

Li, Q., and T. Stengos (1996): "Semiparametric Estimation in Partially Linear Panel Data Models," *Journal of Econometrics*, 71, 389–397.

Lin, Z., and C. Lu (1996): *Limit Theory for Mixing Dependent Random Variables.* New York: Dordrecht: Kluwer.

Linton, O. (1995): "Second Order Approximation in the Partially Linear Regression Model," *Econometrica*, 63, 1079–1112.

Newey, W. K., and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Newey, W. K., J. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565–603.

Pagan, A., and A. Ullah (1999): *Nonparametric Econometrics.* Cambridge: Cambridge University Press.

Park, J., and P. C. B. Phillips (1988): "Statistical Inference in Regressions with Integrated Processes," I. *Econometric Theory*, 4, 468–497.

Park, J., and P. C. B. Phillips (1989): "Statistical Inference in Regressions with Integrated Processes," II. *Econometric Theory*, 5, 95–131.

Park, J., and P. C. B. Phillips (1999): "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory*, 15, 269–298.

Park, J., and P. C. B. Phillips (2001): "Nonlinear Regressions with Integrated Time Series," *Econometrica*, 69, 117-162.

Phillips, P. C. B. (1983): "Exact small Sample Theory in the Simultaneous Equations Model," In *Handbook of Econometrics* Vol I, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: Elsevier, pp. 449–518.

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. (2009): "Local Limit Theory and Spurious Regressions," *Econometric Theory*, 25, 1466–1497.

Phillips, P. C. B., and B. E. Hasen (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, 99–125.

PHILLIPS, P. C. B., AND J. PARK (1998): "Nonstationary Density Estimation and Kernel Autoregression," Cowles Foundation Discussion Paper, No. 1181, Yale University.

PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

PHILLIPS, P. C. B., AND L. SU (2010): "Nonparametric Regression under Location Shifts," unpublished paper, Yale University.

PHILLIPS, P. C. B., AND Z. XIAO (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.

ROSENBLATT, M. (1956): "A Central Limit Theorem and a Strong Mixing Condition," *Proc. Nat. Acad. Sci., U.S.A.*, 42, 43-47.

SARGAN, D. (1988): *Lectures on Advanced Econometric Theory* (edited by Meghnad Desai). New York: Basil Blackwell Inc..

SCHIENLE, M. (2009): "Reaching for Econometric Generality: Nonparametric Nonstationary Regression," Working paper available from the Department of Economics, University of Mannheim, Germany.

SU, L., AND A, ULLAH (2008): "Local Polynomial Estimation of Nonparametric Simultaneous Equations Models," *Journal of Econometrics*, 144, 193–218.

TONG, H. (1990): *Nonlinear Time Series*. Oxford: Oxford University Press.

WANG, Q. Y., AND P. C. B. PHILLIPS (2009a): "Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression," *Econometric Theory*, 25, 710–738.

WANG, Q. Y., AND P. C. B. PHILLIPS (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.

YATCHEW, A. (2003): *Semiparametric Regression for the Applied Econometrician*. Oxford: Oxford University.