Econometric Theory, **24**, 2008, 1321–1342. Printed in the United States of America.
doi:10.1017/S0266466608080523

NONPARAMETRIC ESTIMATION OF VARYING COEFFICIENT DYNAMIC PANEL DATA MODELS

ZONGWU CAI
University of North Carolina at Charlotte
and
Xiamen University

QI LI
Texas A&M University
and
Tsinghua University

We suggest using a class of semiparametric dynamic panel data models to capture individual variations in panel data. The model assumes linearity in some continuous/discrete variables that can be exogenous/endogenous and allows for nonlinearity in other weakly exogenous variables. We propose a nonparametric generalized method of moments (NPGMM) procedure to estimate the functional coefficients, and we establish the consistency and asymptotic normality of the resulting estimators.

1. INTRODUCTION

There exists a rich literature on linear and nonlinear parametric dynamic panel data models that assume that all regression coefficients are constant, both over time and across individuals. The readers are referred to Arellano (2003), Baltagi (2005), and Hsiao (2003) for an overview of statistical inference and econometric interpretation of this widely used class of parametric panel data models. It is well known, however, that parametric panel data models may be misspecified, and estimators obtained from misspecified models are often inconsistent. To deal with this issue, some nonparametric/semiparametric dynamic panel data models have been proposed. For example, Robertson and Symons (1992) con-

We thank a co-editor and two anonymous referees for insightful comments that greatly improved this paper. We also thank Jeff Racine and Dennis Jansen for their help in editorial suggestions. Cai's research was supported, in part, by the National Science Foundation grants DMS-0072400 and DMS-0404954 and funds provided by the University of North Carolina at Charlotte. Li's research was partially supported by the Private Enterprise Research Center, Texas A&M University. The authors thank participants at the 11th International Panel Data Conference (June 2004 at Texas A&M University) for their helpful comments. Address correspondence to Qi Li, Department of Economics, Texas A&M University, College Station, TX 77843-4228, USA; e-mail: qi@econmail.tamu.edu.

 0266-4666/08 $15.00

sidered a model that assumes the coefficients of the dynamic part to be constant whereas the coefficients for the static part are allowed to change over individuals. Pesaran and Smith (1995) treated the case where coefficients of both the dynamic and the static parts can vary across individuals. Horowitz and Markatou (1996), Li and Hsiao (1998), and Kniesner and Li (2002) considered partially linear panel data models with exogenous regressors, and Li and Stengos (1996) and Baltagi and Li (2002) considered instrumental variable (IV) estimation of partially linear models. One of the advantages of the nonparametric/semiparametric approach is that little prior restriction is imposed on the model's structure. Also, this approach may offer useful insights for the construction of parametric models. Obviously there are many possible nonlinear semiparametric functional forms to be explored.

In this paper we contribute to this literature by extending a varying coefficient method to the analysis of dynamic panel data models. We consider a panel with N individual units and over T time periods. We consider the case of large N and allow for both fixed T and large T. Moreover, we allow for endogenous variables to enter the parametric part of the model. We propose a nonparametric generalized method of moments (NPGMM) approach that is a combination of the local linear fitting of Fan and Gijbels (1996) and the generalized method of moments (GMM) approach of Hansen (1982). We establish both the consistency and asymptotic normality of the proposed estimators. A related work to this paper is the paper by Cai, Das, Xiong, and Wu (2006). Cai et al. (2006) considered estimating a varying coefficient model, and they also allowed for endogenous variables to enter the parametric part of the paper. However, Cai et al. (2006) only considered the independent data case, whereas we consider a panel data model allowing for both small T and large T cases. Moreover, our estimation procedure is fundamentally different from the two-stage estimation procedure proposed by Cai et al. (2006). Their two-stage estimation method requires one to first estimate a high-dimension nonparametric model and then to estimate a varying coefficient model using the first-stage nonparametric estimates as generated regressors. Our estimation method only requires a one-step estimation of a varying coefficient model (a low-dimension semiparametric model). We will further discuss the comparison of our estimator with that of Cai et al. (2006) in Section 3 after we introduce our estimation method. Recently, Ai and Chen (2003) considered an efficient estimation of the parametric components in a general class of semiparametric models where the endogenous variable is allowed to appear inside an unknown function, i.e., the endogenous variable appears at the nonparametric part of the model. Their model is more challenging to handle than ours technically. However, the difference between the present paper and their paper is that Ai and Chen mainly considered the efficient estimation of the $\sqrt{n}$ (n is the sample size) asymptotic normality result for the finite-dimensional parameters but they did not provide asymptotic distribution of the nonparametric components because the exact leading bias term in series estimation is generally unknown, whereas in this paper we use the kernel method and we derive the asymptotic normal distribution of our (the nonparametric component of the model) semiparametric estimator.

Varying coefficient models are well known in the statistics/econometrics literature, and there are a variety of applications; see, e.g., Cai, Fan, and Yao (2000), Chen and Tsay (1993), and Hastie and Tibshirani (1993) for details. The structures of these models are analogous to those of random coefficients models (e.g., Hsiao, 2003; Granger and Teräsvirta, 1993). Recently, these models have been used in various empirical applications. For example, Hong and Lee (2003) explored inference and forecasting of exchange rates, Juhl (2005) studied the possible unit root behavior of U.S. unemployment data, Li, Huang, Li, and Fu (2002) modeled the production frontier using Chinese manufacturing data, and Cai et al. (2006) considered nonparametric two-stage IV estimators for returns to education.

The rest of this paper is organized as follows. In Section 2, we formally introduce the varying coefficient dynamic panel data model and discuss model identification issues. In Section 3, we propose a nonparametric IV estimation procedure that combines the local linear fitting scheme and GMM to estimate the coefficient functions, and we establish the consistency and asymptotic normality of the resulting estimators. All technical proofs are relegated to the Appendix.

2. VARYING COEFFICIENT DYNAMIC PANEL MODELS

We consider a class of semiparametric panel data models, called "varying coefficient dynamic panel data models," that assume the following form:

$$Y_{it} = \mathbf{X}_{it}'\mathbf{g}(\mathbf{Z}_{it}) + \epsilon_{it}, \qquad 1 \le i \le N, \quad \text{and} \quad 1 \le t \le T, \tag{1}$$

where $\mathbf{X}_{it}$ is of dimension $d \times 1$ with its first element $X_{it,1} = 1$, the prime denotes the transpose of a matrix or vector, the coefficient functions $\{g_j(\cdot)\}$ $(j = 1, \ldots, d)$ are unspecified smooth functions in $\Re^p$ $(p \ge 1, \mathbf{Z}_{it} \in \Re^p)$, the errors $\{\epsilon_{it}\}$ can be serially correlated and are assumed to be stationary (also strong mixing if T is large), and $\mathrm{E}(\epsilon_{it}|\mathbf{Z}_{it}) = 0$. The main focus in this paper is on estimating model (1) under the assumption that some or all components of $\mathbf{X}_{it}$ may be correlated with the error ϵ_{it}. More specifically, we assume that $\mathrm{E}(\epsilon_{it}|\mathbf{Z}_{it}) = 0$ but allow for $\mathrm{E}(\epsilon_{it}|\mathbf{X}_{it}) \ne 0$. If both $\mathbf{X}_{it}$ and $\mathbf{Z}_{it}$ are exogenous, and in particular do not contain lagged values of Y_{it}, then model (1) becomes a varying coefficient *static* panel data model.

The general setting in model (1) includes many familiar models in the literature. For example, it covers the following partially linear dynamic panel data model:

$$Y_{it} = g_1(\mathbf{Z}_{it}) + \widetilde{\mathbf{X}}_{it}'\boldsymbol{\beta} + \epsilon_{it}, \qquad 1 \le i \le N, \quad \text{and} \quad 1 \le t \le T, \tag{2}$$

where $\widetilde{\mathbf{X}}_{it}$ is $\mathbf{X}_{it}$ without the first component $\mathbf{X}_{it,1}$. Indeed, model (2) has been studied by many authors in the literature. For example, Li and Hsiao (1998) and Kniesner and Li (2002) studied model (2) under the assumption that $\mathrm{E}(\epsilon_{it}|\widetilde{\mathbf{X}}_{it},\mathbf{Z}_{it}) = 0$ (i.e., there is no endogenous regressor), and Li and Stengos (1996) and Baltagi and Li (2002) tackled it by allowing some or all components of $\widetilde{\mathbf{X}}_{it}$ to be correlated with the error ϵ_{it} (i.e., there exist some endogenous regressors). If some or all components of $\mathbf{X}_{it}$ are endogenous, model (1) covers the nonparametric IV models considered by Das (2005) for discrete endogenous regressors and Cai et al. (2006) for both discrete and continuous endogenous regressors, and the semiparametric IV models by Newey (1990) and Cai and Xiong (2006) with cross-sectional data. Finally, if there is no endogenous variable, model (1) includes the static panel transition regression model of González, Teräsvirta, and van Dijk (2005) and the threshold nondynamic panel model of Hansen (1999).

When $\mathrm{E}(\epsilon_{it}|\mathbf{X}_{it}) \neq 0$, it is clear from (1) that $\mathrm{E}(Y_{it}|\mathbf{X}_{it},\mathbf{Z}_{it}) \neq \mathbf{X}_{it}'\mathbf{g}(\mathbf{Z}_{it})$. Therefore, one cannot consistently estimate the coefficient functions $\{g_j(\cdot)\}$ by projecting Y_{it} on $\mathbf{X}_{it}'\mathbf{g}(\mathbf{Z}_{it})$ (in the $\mathcal{L}_2(\mathbf{X},\mathbf{Z})$ projection space). To obtain a consistent estimator of the coefficient functions $\{g_j(\cdot)\}$, we assume that there exists a $q \times 1$ vector of instrumental variables $\mathbf{W}_{it}$ with the first component $W_{it,1} \equiv 1$ such that $\mathrm{E}(\epsilon_{it}|\mathbf{W}_{it}) = 0$. Then, we have the following orthogonality condition:

$$\mathrm{E}(\epsilon_{it}|\mathbf{V}_{it}) = 0, \qquad \textbf{(3)}$$

where $\mathbf{V}_{it} = (\mathbf{W}_{it}',\mathbf{Z}_{it}')'$. Multiplying (1) by $\pi(\mathbf{V}_{it}) \equiv \mathrm{E}(\mathbf{X}_{it}|\mathbf{V}_{it})$ on both sides and taking expectations, conditional on $\mathbf{Z}_{it} = \mathbf{z}$, we obtain

$$\begin{aligned}\mathrm{E}(\pi(\mathbf{V}_{it})Y_{it}|\mathbf{Z}_{it} = \mathbf{z}) &= \mathrm{E}(\pi(\mathbf{V}_{it})\mathbf{X}_{it}'|\mathbf{Z}_{it} = \mathbf{z})\mathbf{g}(\mathbf{z})\\ &= \mathrm{E}(\pi(\mathbf{V}_{it})\pi(\mathbf{V}_{it})'|\mathbf{Z}_{it} = \mathbf{z})\mathbf{g}(\mathbf{z}),\end{aligned}$$

where we have made use of the law of iterated expectations. Under the assumption that $\mathrm{E}(\pi(\mathbf{V}_{it})\pi(\mathbf{V}_{it})'|\mathbf{Z}_{it} = \mathbf{z})$ is positive definite, we obtain

$$\mathbf{g}(\mathbf{z}) = [\mathrm{E}(\pi(\mathbf{V}_{it})\pi(\mathbf{V}_{it})'|\mathbf{Z}_{it} = \mathbf{z})]^{-1}\mathrm{E}(\pi(\mathbf{V}_{it})Y_{it}|\mathbf{Z}_{it} = \mathbf{z}). \qquad \textbf{(4)}$$

The condition that $\mathrm{E}(\pi(\mathbf{V}_{it})\pi(\mathbf{V}_{it})'|\mathbf{Z}_{it} = \cdot)$ is positive definite guarantees that $\mathbf{g}(\cdot)$ is identified. In principle one can also construct an estimator of $\mathbf{g}(z)$ based on (4). However, such an estimator will require a two-stage nonparametric estimation procedure: the first step is to estimate the conditional mean $\pi(\mathbf{V}_{it})$, and the second stage is to estimate another conditional mean function of $\hat{\pi}(\mathbf{V}_{it})Y_{it}$ conditional on $\mathbf{Z}_{it} = \mathbf{z}$ where $\hat{\pi}(\mathbf{V}_{it})$ is the nonparametric estimate obtained at the first step; see, e.g., Cai et al. (2006). Such a double nonparametric estimation procedure complicates the asymptotic analysis of such an estimator. To overcome this shortcoming, in the next section we propose a simple estimator for $\mathbf{g}(\cdot)$ that requires only one nonparametric estimation procedure.

3. STATISTICAL PROPERTIES

3.1. NPGMM Estimation

In the remaining part of the paper we assume that the model is identified. It follows from the orthogonality condition (3) that, for any vector function $\mathbf{Q}(\mathbf{V}_{it})$ with dimension m_1 specified later, we have

$$\mathbf{0} = \mathrm{E}(\mathbf{Q}(\mathbf{V}_{it})\epsilon_{it}|\mathbf{V}_{it}) = \mathrm{E}[\mathbf{Q}(\mathbf{V}_{it})\{Y_{it} - \mathbf{X}_{it}'\mathbf{g}(\mathbf{Z}_{it})\}|\mathbf{V}_{it}]. \tag{5}$$

If $\mathbf{Q}(\mathbf{V}_{it})$ is chosen to be $\pi(\mathbf{V}_{it})$, solving $\mathbf{g}(\cdot)$ from equation (5) leads to equation (4). However, for computational simplicity, we will not choose $\mathbf{Q}(\cdot)$ as $\pi(\cdot)$.

Clearly, (5) provides conditional moment restrictions and can lead to an estimation approach similar to the GMM of Hansen (1982) for parametric models. We propose an estimation procedure to combine the orthogonality conditions given in (5) and the local linear fitting scheme of Fan and Gijbels (1996) to estimate the coefficient functions. This nonparametric estimation procedure is termed nonparametric generalized method of moments (NPGMM).

We apply local linear fitting to estimate the coefficient functions $\{g_j(\cdot)\}$, although other smoothing methods such as the Nadaraya–Watson kernel method and spline method are applicable. The main reason for preferring local linear fitting is because it possesses some attractive properties, such as high statistical efficiency in an asymptotic minimax sense, design adaptation, and automatic boundary corrections (e.g., Fan and Gijbels, 1996). The detailed description of this approach can be found in Fan and Gijbels (1996), and its basic idea is illustrated next. Note that although a general local polynomial technique is applicable here, Fan and Gijbels (1996) argued that local linear fitting might be sufficient for most applications, whereas the theory developed for the local linear estimator holds for the local polynomial estimator with a slight modification. Therefore, in this paper we focus only on local linear estimation.

We assume throughout that $\{g_j(\cdot)\}$ are twice continuously differentiable. Then, for a given point $\mathbf{z} \in \Re^p$ and for $\{\mathbf{Z}_{it}\}$ in a neighborhood of $\mathbf{z}$, using Taylor expansions, $g_j(\mathbf{Z}_{it})$ can be approximated by a linear function $a_j + \mathbf{b}_j'(\mathbf{Z}_{it} - \mathbf{z})$ with $a_j = g_j(\mathbf{z})$ and $\mathbf{b}_j = \nabla g_j(\mathbf{z}) = \partial g_j(\mathbf{z})/\partial \mathbf{z}$, the derivative of $g_j(\mathbf{z})$. Hence, model (1) is approximated by the following working linear model:

$$Y_{it} \simeq \mathbf{U}_{it}'\boldsymbol{\alpha} + \epsilon_{it},$$

where $\mathbf{U}_{it} = \begin{pmatrix} \mathbf{X}_{it} \\ \mathbf{X}_{it} \otimes (\mathbf{Z}_{it} - \mathbf{z}) \end{pmatrix}$ is an $m_2 \times 1$ $(m_2 = d(p+1))$ vector, $\otimes$ denotes the Kronecker product, and $\boldsymbol{\alpha} = (a_1, \ldots, a_d, \mathbf{b}_1', \ldots, \mathbf{b}_d')'$ is an $m_2 \times 1$ vector of parameters. Therefore, for $\{\mathbf{Z}_{it}\}$ in a neighborhood of $\mathbf{z}$, the orthogonality conditions in (5) implies the following locally weighted orthogonality conditions:

$$\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbf{Q}(\mathbf{V}_{it})(Y_{it} - \mathbf{U}_{it}'\boldsymbol{\alpha})K_h(\mathbf{Z}_{it} - \mathbf{z}) = \mathbf{0}, \tag{6}$$

where $K_h(\cdot) = h^{-p}K(\cdot/h)$, $K(\cdot)$ is a kernel function in $\Re^p$, and $h = h_n > 0$ is a bandwidth that controls the amount of smoothing used in the estimation.

We will estimate $\mathbf{g}(\mathbf{z})$ based on (6). Although (6) is the sample analog of an unconditional zero population mean equation, it is equivalent to the conditional mean restriction of (3) if one requires that (6) hold true for all measurable functions $\mathbf{Q}(\cdot)$. For a specific choice of $\mathbf{Q}(\cdot)$, (6) is weaker than (3). It might be possible to relax the conditional mean restriction (3) to a weak unconditional population mean restriction based on (6). However, this will complicate the asymptotic analysis. Therefore, we will impose the orthogonal condition (3) throughout the paper to simplify the asymptotic analysis.

Equation (6) can be viewed as the IV version of the nonparametric estimation equations discussed in Cai (2003) and the locally weighted version of equation (9.2.29) in Hamilton (1994, p. 243) or equation (14.2.20) in Hamilton (1994, p. 419) for parametric IV models. To ensure that equation (6) has a unique solution, the dimension of $\mathbf{Q}(\cdot)$ must satisfy $m_1 \geq m_2$ because the number of parameters in (6) is m_2. However, when $m_1 > m_2$, the model is overidentified, and there may not exist a unique $\boldsymbol{\alpha}$ to satisfy (6). To obtain a unique $\boldsymbol{\alpha}$ satisfying (6), we premultiply (6) by an $m_2 \times m_1$ matrix $\mathbf{S}_n'$, where with $\mathbf{Q}_{it} = \mathbf{Q}(\mathbf{V}_{it})$ and $n = NT$,

$$\mathbf{S}_n = \frac{1}{n}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{Q}_{it}\mathbf{U}_{it}'K_h(\mathbf{Z}_{it} - \mathbf{z}).$$

Then solving for $\boldsymbol{\alpha}$ we obtain

$$\hat{\boldsymbol{\alpha}} = (\mathbf{S}_n'\mathbf{S}_n)^{-1}\mathbf{S}_n'\mathbf{T}_n, \tag{7}$$

where

$$\mathbf{T}_n = \frac{1}{n}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{Q}_{it}K_h(\mathbf{Z}_{it} - \mathbf{z})Y_{it}.$$

The estimator $\hat{\boldsymbol{\alpha}}$ defined in (7) is termed the NPGMM estimate of $\boldsymbol{\alpha}$, and it gives the NPGMM estimate of $\mathbf{g}(\mathbf{z})$ and its first-order derivatives $\nabla g_j(\mathbf{z})$ $(1 \leq j \leq d)$.

We now compare our estimation procedure with the two-stage estimation method proposed by Cai et al. (2006), described briefly as follows. At the first stage, one estimates $\pi(\mathbf{V}_{it})$ nonparametrically (say, by a kernel method). Let $\hat{\pi}(\mathbf{V}_{it})$ denote the resulting nonparametric estimator. Then at the second stage, one estimates $\mathbf{g}(\cdot)$ based on the varying coefficient model: $Y_{it} = \hat{\pi}(\mathbf{V}_{it})'\mathbf{g}(\mathbf{Z}_{it}) + u_{it}$. Recall that the dimensions of $\mathbf{W}_{it}$ and $\mathbf{Z}_{it}$ are q and p, respectively. Hence, the first stage of Cai et al. (2006) requires the estimation of a nonparametric regression model of dimension $q + p$. Also, their two-step method requires the use of two sets of smoothing parameters and that first-step estimation should be undersmoothed. In contrast, our proposed method only involves a one-step estimation procedure of a varying coefficient model with nonparametric components of dimension p. In empirical applications, it is likely that $\mathbf{W}_{it}$ is a high-dimension vector, whereas $\mathbf{Z}_{it}$ is a scalar (or a low-dimension vector). In such situations our proposed estimator is expected to have much better *finite-sample* performance than that for the two-stage estimator of Cai et al. (2006) because our estimator only involves low-dimensional nonparametric estimations.

Note that if there is no endogenous variable (all components of $\mathbf{X}_{it}$ are exogenous), then one can choose $\mathbf{W}_{it}$ to be $\mathbf{X}_{it}$ and choose $\mathbf{Q}(\mathbf{V}_{it})$ to be $\mathbf{U}_{it}$. In this case, equation (6) becomes

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{U}_{it}K_h(\mathbf{Z}_{it}-\mathbf{z})(Y_{it}-\mathbf{U}_{it}'\boldsymbol{\alpha})=\mathbf{0},$$

which is the normal equation of the following locally linear least squares problem for the varying coefficient panel data model:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}K_h(\mathbf{Z}_{it}-\mathbf{z})(Y_{it}-\mathbf{U}_{it}'\boldsymbol{\alpha})^2.$$

Therefore, in this case the NPGMM estimator given by (6) reduces to the ordinary local linear estimator.

We now turn to the question of how to best choose $\mathbf{Q}(\mathbf{V}_{it})$ in (6). Motivated by local linear fitting, a simple choice of $\mathbf{Q}(\mathbf{V}_{it})$ is

$$\mathbf{Q}_{it}=\mathbf{Q}(\mathbf{V}_{it})=\begin{pmatrix}\mathbf{W}_{it}\\ \mathbf{W}_{it}\otimes(\mathbf{Z}_{it}-\mathbf{z})/h\end{pmatrix}, \tag{8}$$

so that the dimension of $\mathbf{Q}_{it}$ is $m_1=q(p+1)$. Therefore, the identification condition $m_1\ge m_2$ becomes $q\ge d$. Note that the choice of $\mathbf{Q}(\mathbf{V}_{it})$ given in (8) is computationally simple but it may not be optimal in the sense of minimizing the estimation asymptotic variance. For fixed orthogonality conditions, optimal instruments can be constructed by following approaches similar to Newey (1990) and Ai and Chen (2003). In this paper we focus only on the simple case whereby $\mathbf{Q}(\mathbf{V}_{it})$ has the form given in (8).

Before we derive the asymptotic distribution of $\hat{\boldsymbol{\alpha}}$, we first introduce some notation. Let $\mathbf{H}=\operatorname{diag}\{\mathbf{I}_d,h\mathbf{I}_{dp}\}$, which is of dimension $m_2\times m_2$, where $\mathbf{I}_j$ denotes a $j\times j$ identity matrix. Substituting (8) into (6), multiplying $\mathbf{H}$ on both sides of (7), and also inserting $\mathbf{H}\mathbf{H}^{-1}$ in the middle on the right-hand side of (7), we obtain

$$\begin{aligned}\mathbf{H}\hat{\boldsymbol{\alpha}}&=\mathbf{H}(\mathbf{S}_n'\mathbf{S}_n)^{-1}\mathbf{H}\mathbf{H}^{-1}\mathbf{S}_n'\mathbf{T}_n=[(\mathbf{S}_n\mathbf{H}^{-1})'\mathbf{S}_n\mathbf{H}^{-1}]^{-1}(\mathbf{S}_n\mathbf{H}^{-1})'\mathbf{T}_n\\&=[\tilde{\mathbf{S}}_n'\tilde{\mathbf{S}}_n']^{-1}\tilde{\mathbf{S}}_n'\mathbf{T}_n,\end{aligned} \tag{9}$$

where

$$\tilde{\mathbf{S}}_n=\mathbf{S}_n\mathbf{H}^{-1}=\frac{1}{n}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{Q}_{it}\tilde{\mathbf{U}}_{it}'K_h(\mathbf{Z}_{it}-\mathbf{z})$$

with

$$\widetilde{\mathbf{U}}_{it} = \mathbf{H}^{-1}\mathbf{U}_{it} = \begin{pmatrix} \mathbf{X}_{it} \\ \mathbf{X}_{it} \otimes (\mathbf{Z}_{it} - \mathbf{z})/h \end{pmatrix}$$

(so that $\widetilde{\mathbf{U}}'_{it} = \mathbf{U}'_{it}\mathbf{H}^{-1}$). We are now ready to derive the asymptotic distribution of $\hat{\boldsymbol{\alpha}}$, which is the subject of the next section.

3.2. Asymptotic Theory

First, for ease of reference, we state the definition of a strongly mixing sequence. Let $\{\zeta_t\}$ be a strictly stationary stochastic process and $\mathcal{F}_s^t$ denote the sigma algebra generated by $(\zeta_s, \ldots, \zeta_t)$ for $s \le t$. A process $\{\zeta_t\}$ is said to be strongly mixing or α-mixing if

$$\alpha(\tau) = \sup_{s\in\mathcal{N}}\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^s, B \in \mathcal{F}_{s+\tau}^\infty\} \to 0$$

as $\tau \to \infty$.

Next, we introduce the following notation. Denote by $\boldsymbol{\mu}_2(K) = \int \mathbf{u}\mathbf{u}'K(\mathbf{u})d\mathbf{u}$ and $\nu_0 = \int K^2(\mathbf{u})d\mathbf{u}$. Define $\sigma^2(\mathbf{v}) = \mathrm{Var}(\epsilon_{it}|\mathbf{V}_{it} = \mathbf{v})$, $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\mathbf{z}) = \mathrm{E}(\mathbf{W}_{it}\mathbf{X}'_{it}|\mathbf{Z}_{it} = z)$, $\boldsymbol{\Omega}_1 = \boldsymbol{\Omega}_1(\mathbf{z}) = \mathrm{Var}\{\mathbf{W}_{it}\epsilon_{it}|\mathbf{Z}_{it} = \mathbf{z}\}$, $\sigma_{1t}(\mathbf{V}_{i1}, \mathbf{V}_{it}) = \mathrm{E}\{\epsilon_{i1}\epsilon_{it}|\mathbf{V}_{i1}, \mathbf{V}_{it}\}$, and

$$\mathbf{G}_{1t}(\mathbf{Z}_{i1}, \mathbf{Z}_{it}) = \mathrm{E}\{\mathbf{W}_{i1}\mathbf{W}'_{it}\sigma_{1t}(\mathbf{V}_{i1}, \mathbf{V}_{it})|\mathbf{Z}_{i1}, \mathbf{Z}_{it}\} = \mathrm{E}\{\mathbf{W}_{i1}\mathbf{W}'_{it}\epsilon_{i1}\epsilon_{it}|\mathbf{Z}_{i1}, \mathbf{Z}_{it}\}.$$

Then, it is obvious that $\boldsymbol{\Omega}_1 = \mathbf{G}_1(\mathbf{z}, \mathbf{z})$ and $\sigma^2(\mathbf{v}) = \sigma_{11}(\mathbf{v}, \mathbf{v})$. Set

$$\mathbf{S} = \mathbf{S}(\mathbf{z}) = \begin{pmatrix} \boldsymbol{\Omega} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} \otimes \boldsymbol{\mu}_2(K) \end{pmatrix}, \qquad \mathbf{S}^* = \mathbf{S}^*(\mathbf{z}) = \begin{pmatrix} \boldsymbol{\Omega}_1\nu_0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_1 \otimes \boldsymbol{\mu}_2(K^2) \end{pmatrix},$$

and

$$\mathbf{B}(\mathbf{z}) = \int \begin{pmatrix} \boldsymbol{\Omega}\mathbf{A}(\mathbf{u}, \mathbf{z}) \\ \{\boldsymbol{\Omega}\mathbf{A}(\mathbf{u}, \mathbf{z})\} \otimes \mathbf{u} \end{pmatrix} K(\mathbf{u})\, d\mathbf{u}, \quad \text{where } \mathbf{A}(\mathbf{u}, \mathbf{z}) = \begin{pmatrix} \mathbf{u}'\nabla^2 g_1(\mathbf{z})\mathbf{u} \\ \vdots \\ \mathbf{u}'\nabla^2 g_d(\mathbf{z})\mathbf{u} \end{pmatrix}$$

and $\nabla^2 g_j(\mathbf{z}) = \partial^2 g_j(\mathbf{z})/\partial\mathbf{z}\partial\mathbf{z}'$. We now impose some regularity conditions that are sufficient for deriving the consistency and asymptotic normality of the proposed estimators, although they might not be the weakest possible.

Assumption A.

A1. $\{(\mathbf{W}_{it}, \mathbf{X}_{it}, Y_{it}, \mathbf{Z}_{it}, \epsilon_{it})\}$ are independent and identically distributed across the i index for each fixed t and strictly stationary over t for each fixed i, $\mathrm{E}\|\mathbf{W}_{it}\mathbf{X}'_{it}\|^2 < \infty$, $\mathrm{E}\|\mathbf{W}_{it}\mathbf{W}'_{it}\|^2 < \infty$, and $\mathrm{E}|\epsilon_{it}|^2 < \infty$, where $\|\mathbf{A}\|^2 = \mathrm{tr}(\mathbf{A}\mathbf{A}')$ is the standard L_2-norm for a finite-dimensional matrix $\mathbf{A}$.

A2. For each $t \geq 1$, $\mathbf{G}_{1t}(\mathbf{z}_1, \mathbf{z}_2)$ and $f_{1t}(\mathbf{z}_1, \mathbf{z}_2)$, the joint density of $\mathbf{Z}_{i1}$ and $\mathbf{Z}_{it}$, are continuous at $(\mathbf{z}_1 = \mathbf{z}, \mathbf{z}_2 = \mathbf{z})$. Also, for each $\mathbf{z}$, $\boldsymbol{\Omega}(\mathbf{z}) > 0$ and $f(\mathbf{z}) > 0$, where $f(\mathbf{z})$ is the marginal density function of $\mathbf{Z}_{it}$. Further, $\sup_{t\geq 1} |\mathbf{G}_{1t}(\mathbf{z}, \mathbf{z}) f_{1t}(\mathbf{z}, \mathbf{z})| \leq \mathbf{M}(\mathbf{z}) < \infty$ for some function $\mathbf{M}(\mathbf{z})$. Finally, $\mathbf{g}(\mathbf{z})$and $f(\mathbf{z})$ are both twice continuously differentiable at $\mathbf{z} \in \mathcal{R}^p$.

A3. The kernel $K(\cdot)$ is a symmetric, nonnegative, and bounded second-order kernel function having compact support.

A4. The IV $\mathbf{W}_{it}$ satisfies the conditions that $\mathrm{E}(\epsilon_{it} | \mathbf{W}_{it}, \mathbf{Z}_{it}) = 0$ and that $\mathrm{E}[\pi(\mathbf{V}_{it})\pi(\mathbf{V}_{it})' | \mathbf{Z}_{it} = \mathbf{z}]$ is of full rank for all $\mathbf{z}$, where $\mathbf{V}_{it} = (\mathbf{W}_{it}', \mathbf{Z}_{it}')$ and $\pi(\mathbf{V}_{it}) = \mathrm{E}(\mathbf{X}_{it} | \mathbf{V}_{it})$.

A5. $h \to 0$ and $Nh^p \to \infty$ as $N \to \infty$.

Assumption B.

B1. $T \to \infty$ and $nh^p \to \infty$ as $N \to \infty$.

B2. There exists some $\delta > 0$ such that $\mathrm{E}\{|\epsilon_{it}\mathbf{W}_{it}|^{2(1+\delta)} | \mathbf{Z} = \mathbf{u}\}$ is continuous at $\mathbf{u} = \mathbf{z}$.

B3. For each fixed i, the process $\{(\mathbf{W}_{it}, \mathbf{X}_{it}, Y_{it}, \mathbf{Z}_{it}, \epsilon_{it})\}$ is α-mixing with the mixing coefficient satisfying the condition $\alpha(k) = O(k^{-\tau})$, where $\tau = (2 + \delta)(1 + \delta)/\delta$.

B4. $NT^{(\tau+1)/\tau} h^{p(2+\delta)/(1+\delta)} \to \infty$.

Remark 1 (Discussion of conditions). Assumption A1 requires that observations are independent and identically distributed across i and stationary across t, which is a standard assumption in the panel data literature. Note that we do not assume that $\{\epsilon_{1t}\}$ is a martingale (random walk) difference process, which is imposed by Kniesner and Li (2002). Assumption A1 also gives some standard moment conditions. Assumption A2 includes some smoothness conditions on functionals involved. The requirement in A3 that $K(\cdot)$ be compactly supported is imposed for the sake of brevity of proofs and can be removed at the cost of lengthier arguments. In particular, the Gaussian kernel is allowed. Assumption A4 is a necessary and sufficient condition for model identification. Assumption A5 allows for T either fixed (bounded) or going to infinity. When T is fixed, the theoretical results are similar to the cross-sectional data case. But for large T ($T \to \infty$), the mathematical derivation is more involved. Therefore, for large T, we need some additional (stronger) conditions such as B2–B4. In particular, B2 requires the existence of some high-order moments. The α-mixing condition is one of the weakest mixing conditions for weakly dependent stochastic processes. Stationary time series or Markov chains fulfilling certain (mild) conditions are α-mixing with exponentially decaying coefficients; see Cai (2002) and Carrasco and Chen (2002) for additional examples. On the other hand, the assumption on the convergence rate of $\alpha(k)$ in B3 might not be the weakest possible and is imposed to simplify the proof. Conditions B2–B4 are similar to those needed for nonlinear time series models (e.g., Cai et al., 2000). Finally, we note that B4 is not restrictive; e.g., if we consider the optimal bandwidth such that $h_{opt} = O(n^{-1/(p+4)})$ (see Remark 3 in this section), then B4 is satisfied when $\delta \geq p/4 - 1$. Therefore, the conditions imposed here are quite mild and standard.

Before presenting some auxiliary results, we need to introduce some notation. Denote

$$R_j(\mathbf{Z}_{it}, \mathbf{z}) = g_j(\mathbf{Z}_{it}) - a_j - \mathbf{b}_j'(\mathbf{Z}_{it} - \mathbf{z}) - \tfrac{1}{2}(\mathbf{Z}_{it} - \mathbf{z})'\nabla^2 g_j(\mathbf{z})(\mathbf{Z}_{it} - \mathbf{z}),$$

$$\mathbf{B}_n = \frac{1}{n}\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(\mathbf{Z}_{it} - \mathbf{z})\mathbf{Q}_{it}\frac{1}{2}\sum_{j=1}^{d}(\mathbf{Z}_{it} - \mathbf{z})'\nabla^2 g_j(\mathbf{z})(\mathbf{Z}_{it} - \mathbf{z})X_{itj},$$

$$\mathbf{R}_n = \frac{1}{n}\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(\mathbf{Z}_{it} - \mathbf{z})\mathbf{Q}_{it}\sum_{j=1}^{d} R_j(\mathbf{Z}_{it}, \mathbf{z})X_{itj},$$

and

$$\mathbf{T}_n^* = \frac{1}{n}\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(\mathbf{Z}_{it} - \mathbf{z})\mathbf{Q}_{it}\epsilon_{it}.$$

Then $\mathbf{T}_n = \tilde{\mathbf{S}}_n\mathbf{H}\boldsymbol{\alpha} + \mathbf{T}_n^* + \mathbf{B}_n + \mathbf{R}_n$. Substituting this into (9), we obtain

$$\mathbf{H}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) - (\tilde{\mathbf{S}}_n'\tilde{\mathbf{S}}_n)^{-1}\tilde{\mathbf{S}}_n'\mathbf{B}_n - (\tilde{\mathbf{S}}_n'\tilde{\mathbf{S}}_n)^{-1}\tilde{\mathbf{S}}_n'\mathbf{R}_n = (\tilde{\mathbf{S}}_n'\tilde{\mathbf{S}}_n)^{-1}\tilde{\mathbf{S}}_n'\mathbf{T}_n^*. \quad \textbf{(10)}$$

To establish the asymptotic distribution of $\hat{\boldsymbol{\alpha}}$, we will show that the second term on the left-hand side of (10) contributes to the asymptotic bias, the third term on the left-hand side is negligible in probability, and the term on the right-hand side is asymptotically normal. To this end, we first provide some preliminary results stated here with their proofs relegated to the Appendix.

PROPOSITION 1. *Under Assumptions A1–A5, we have*

(i) $\tilde{\mathbf{S}}_n = f(\mathbf{z})\mathbf{S}\{1 + o_p(1)\}$,
(ii) $\mathbf{B}_n = (h^2/2) f(\mathbf{z})\mathbf{B}(\mathbf{z}) + o_p(h^2)$, *and*
(iii) $\mathbf{R}_n = o_p(h^2)$.

PROPOSITION 2.

(i) Under Assumptions A1–A4 and B1, and if $Th^p \to 0$, *then*

$$nh^p \operatorname{Var}(\mathbf{T}_n^*) \to f(\mathbf{z})\mathbf{S}^*. \quad \textbf{(11)}$$

(ii) If $Th^p \geq C > 0$, *and Assumptions A1–A4 and B1–B3 are satisfied, then (11) holds true.*

It follows from (10), Proposition 1, and Assumption A3 that

$$\mathbf{H}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})-\frac{h^2}{2}\begin{pmatrix}\mathbf{B}_g(\mathbf{z})\\ \mathbf{0}\end{pmatrix}+o_p(h^2)=f^{-1}(\mathbf{z})(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{T}_n^*\{1+o_p(1)\}, \tag{12}$$

where $\mathbf{B}_g(\mathbf{z})=\int\mathbf{A}(\mathbf{u},\mathbf{z})K(\mathbf{u})d\mathbf{u}=(\mathrm{tr}(\nabla^2 g_j(\mathbf{z})\boldsymbol{\mu}_2(K)))_{d\times 1}$ is a $d\times 1$ vector. The asymptotic sampling theory for the NPGMM estimators is established in Theorem 1 for consistency and in Theorems 2 and 3 for asymptotic normality with detailed proofs relegated to the Appendix.

THEOREM 1.

(i) If $Th^p\to 0$, under Assumptions A1–A5, we have

$$\mathbf{H}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})-\frac{h^2}{2}\begin{pmatrix}\mathbf{B}_g(\mathbf{z})\\ \mathbf{0}\end{pmatrix}=o_p(h^2)+O_p((nh^p)^{-1/2}). \tag{13}$$

(ii) If $Th^p\geq C$ for some $C>0$, and Assumptions A1–A5, B2, and B3 are satisfied, then (13) holds true.

The proof of Theorem 1 is straightforward from Proposition 2 and (12) and is therefore omitted.

Remark 2. Theorem 1 shows that $\hat{\boldsymbol{\alpha}}$ is consistent (with rate of convergence) for both large and small T cases. In particular, for the case where $Th^p\to 0$, it does not require any assumptions on the dependence structure such as Assumption B3. This is particularly useful in practice. For example, it covers models with serially correlated errors. The next two theorems give the asymptotic normal distribution for $\hat{\boldsymbol{\alpha}}$ for fixed and large T cases.

THEOREM 2. *If T is finite, under Assumptions A1–A5 and B2, we have*

$$\sqrt{nh^p}\left[\mathbf{H}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})-\frac{h^2}{2}\begin{pmatrix}\mathbf{B}_g(\mathbf{z})\\ \mathbf{0}\end{pmatrix}+o_p(h^2)\right]\xrightarrow{d}N(\mathbf{0},f^{-1}(\mathbf{z})\boldsymbol{\Delta}), \tag{14}$$

where $\boldsymbol{\Delta}=\mathrm{diag}\{\nu_0\boldsymbol{\Omega}_g,\boldsymbol{\Omega}_g\otimes[\boldsymbol{\mu}_2^{-1}(K)\boldsymbol{\mu}_2(K^2)\boldsymbol{\mu}_2^{-1}(K)]\}$ with $\boldsymbol{\Omega}_g=(\boldsymbol{\Omega}'\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}'\boldsymbol{\Omega}_1\boldsymbol{\Omega}(\boldsymbol{\Omega}'\boldsymbol{\Omega})^{-1}$.

THEOREM 3. *If $T\to\infty$ and Assumptions A1–A5 and B2–B4 are satisfied, then (14) holds true. In particular we have*

$$\sqrt{nh^p}\left[\hat{\mathbf{g}}(\mathbf{z})-\mathbf{g}(\mathbf{z})-\frac{h^2}{2}\mathbf{B}_g(\mathbf{z})+o_p(h^2)\right]\xrightarrow{d}N(\mathbf{0},\nu_0 f^{-1}(\mathbf{z})\boldsymbol{\Omega}_g).$$

To prove the asymptotic normality results stated in Theorems 2 and 3, given the results of Theorem 1, it suffices to show that $\sqrt{nh^p}\mathbf{T}_n^* \to N(\mathbf{0}, f(\mathbf{z})\mathbf{S}^*(\mathbf{z}))$, which is proved in the Appendix.

Remark 3. From Theorem 2 it is easy to see that the asymptotic variance of $\sqrt{nh^p}(\hat{\mathbf{g}}(\mathbf{z}) - \mathbf{g}(\mathbf{z}))$ is $\nu_0 f^{-1}(z)\boldsymbol{\Omega}_g$, which is the same as that given in Theorem 3. Hence, Theorems 2 and 3 show that $\hat{\mathbf{g}}(\mathbf{z})$ has the same leading bias and variance expressions for both finite T and the large T cases. This implies that if one first lets $N \to \infty$ (for a fixed value of T), and then let $T \to \infty$, this sequential limit is the same as the joint limit of $N \to \infty$ and $T \to \infty$. Therefore, we know that the asymptotic mean squares error (AMSE) of $\hat{\mathbf{g}}(\mathbf{z})$, whether T is fixed or large, is given by

$$\text{AMSE} = h^4 \|\mathbf{B}_g(\mathbf{z})\|^2/4 + \nu_0 f^{-1}(\mathbf{z})\text{tr}(\boldsymbol{\Omega}_g)(nh^p)^{-1}.$$

Then it is easy to show that the optimal bandwidth h_{opt} that minimizes the preceding AMSE is given by

$$h_{opt} = (p\nu_0 f^{-1}(\mathbf{z})\text{tr}(\boldsymbol{\Omega}_g)\|\mathbf{B}_g(\mathbf{z})\|^{-2})^{1/(p+4)}\, n^{-1/(p+4)}$$

and the resulting optimal AMSE is

$$\begin{aligned}\text{AMSE}_{opt} &= (p^{4/(p+4)}/+p^{-p/(p+4)})(\nu_0\, \text{tr}(\boldsymbol{\Omega}_g)f^{-1}(\mathbf{z}))^{4/(p+4)} \\ &\quad \times \|\mathbf{B}_g(\mathbf{z})\|^{2p/(p+4)} n^{-4/(p+4)},\end{aligned}$$

which is the optimal rate of convergence.

Also, it can be shown that, when T is sufficiently large and N is small, the results of Theorem 3 still hold although the theoretical justification needs some modifications.

Finally, let us consider the special case when model (1) does not have any endogenous variable (e.g., $\mathbf{W}_{it} = \mathbf{X}_{it}$). For this case, we have the following asymptotic normality result for the local linear estimator of the coefficient functions, which covers the results in Cai et al. (2000).

THEOREM 4.

(i) Under Assumptions A1–A5 and B2, if T is finite, then we have

$$\sqrt{nh^p}\left[\hat{\mathbf{g}}(\mathbf{z}) - \mathbf{g}(\mathbf{z}) - \frac{h^2}{2}\mathbf{B}_g(\mathbf{z}) + o_p(h^2)\right] \to N(\mathbf{0},\, \nu_0 f^{-1}(\mathbf{z})\boldsymbol{\Omega}_g^*(\mathbf{z})), \tag{15}$$

where

$$\mathbf{\Omega}_g^*(\mathbf{z}) = [\mathrm{E}\{\mathbf{X}_{it}\mathbf{X}_{it}'|\mathbf{Z}_{it} = \mathbf{z}\}]^{-1}\mathrm{E}\{\sigma^2(\mathbf{V}_{it})\,\mathbf{X}_{it}\mathbf{X}_{it}'|\mathbf{Z}_{it} = \mathbf{z}\} \times [\mathrm{E}\{\mathbf{X}_{it}\mathbf{X}_{it}'|\mathbf{Z}_{it} = \mathbf{z}\}]^{-1}.$$

(ii) If $T \to \infty$, and Assumptions A1–A5 and B2–B4 are satisfied, then (15) holds true.

Remark 4. Note first that (i) and (ii) of Theorem 4 are special cases of Theorems 2 and 3 by letting $\mathbf{W}_{it} = \mathbf{X}_{it}$. Also, Remarks 2 and 3 are applicable here for Theorem 4.

Remark 5. It is quite difficult to compare the relative efficiency of our estimator and the two-stage estimator proposed by Cai et al. (2006) for the general setup. For the simplest case that both $\mathbf{X}_{it}$ and $\mathbf{W}_{it}$ are scalars and that the error is conditional homogenous, one can show that the Cai et al. (2006) two-stage estimator is asymptotically more efficient than the estimator proposed in this paper (in the sense of having smaller asymptotic variance). However, because the Cai et al. (2006) estimator requires one first to estimate a high-dimensional nonparametric regression model, this will affect the finite-sample performance of the Cai et al. (2006) estimator. If one focuses on the first-order condition of (5), then Cai and Li (2005) showed that the optimal choice of $\mathbf{Q}(\mathbf{V}_{it})$ is $\mathbf{Q}(\mathbf{V}_{it}) = \mathrm{E}(\mathbf{X}_{it}|\mathbf{V}_{it})/\sigma^2(\mathbf{V}_{it})$, where $\sigma^2(\mathbf{V}_{it}) = \mathrm{Var}(\epsilon_{it}|\mathbf{V}_{it})$, and the resulting estimator will be asymptotically more efficient than both the estimator discussed in Section 3 of this paper and the two-stage estimator of Cai et al. (2006). However, a general treatment of efficient estimation is complex because (5) does not take care of the correlation of the T moment conditions.[1] We leave the general efficient estimation problem as a topic for future research.

NOTE

1. We owe this observation to a referee.

REFERENCES

Ai, C. & X. Chen (2003) Efficient estimation of models with conditional moment restrictions conditioning unknown functions. *Econometrica* 71, 1795–1843.

Arellano, M. (2003) *Panel Data Econometrics*. Oxford University Press.

Baltagi, B. (2005) *Econometric Analysis of Panel Data*, 2nd ed. Wiley.

Baltagi, B. & Q. Li (2002) On instrumental variable estimation of semiparametric dynamic panel data models. *Economics Letters* 76, 1–9.

Cai, Z. (2002) Regression quantiles for time series data. *Econometric Theory* 18, 169–192.

Cai, Z. (2003) Nonparametric estimation equations for time series data. *Statistics and Probability Letters* 62, 379–390.

Cai, Z., M. Das, H. Xiong, & X. Wu (2006) Functional-coefficient instrumental variables models. *Journal of Econometrics* 133, 207–241.

Cai, Z., J. Fan, & Q. Yao (2000) Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* 95, 941–956.

Cai, Z. & Q. Li (2005) Nonparametric Estimation of Varying Coefficient Dynamic Panel Data Models. Working paper, Department of Economics, Texas A&M University.

Cai, Z. & H. Xiong (2006) Efficient Estimation of Partially Varying-Coefficient Instrumental Variables Models. Working paper, Department of Mathematics and Statistics, University of North Carolina at Charlotte.

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Chen, R. & R.S. Tsay (1993) Functional-coefficient autoregressive models. *Journal of the American Statistical Association* 88, 298–308.

Das, M. (2005) Instrumental variables estimators for nonparametric models with discrete endogenous regressors. *Journal of Econometrics* 124, 335–361.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modeling and Its Applications*. Chapman and Hall.

González, A., T. Teräsvirta, & D. van Dijk (2005) Panel Smooth Transition Regression Models. Working paper, Department of Economic Statistics, Stockholm School of Economics.

Granger, C.L.J. & T. Teräsvirta (1993) *Modelling Nonlinear Economic Relationships*. Oxford University Press.

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

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press.

Hansen, B.E. (1999) Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics* 93, 345–368.

Hansen, P.L. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

Hastie, T.J. & R. Tibshirani (1993) Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society, Series B* 55, 757–796.

Hong, Y. & T.-H. Lee (2003) Inference and forecast of exchange rates via generalized spectrum and nonlinear time series models. *Review of Economics and Statistics* 85, 1048–1062.

Horowitz, J.L. & M. Markatou (1996) Semiparametric estimation of regression models for panel data. *Review of Economic Studies* 63, 145–168.

Hsiao, C. (2003) *Analysis of Panel Data*, 2nd ed. Cambridge University Press.

Ibragimov, I.A. & Y.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables*. Walters-Noordhoff.

Juhl, T. (2005) Functional-coefficient models under unit root behavior. *Econometrics Journal* 8, 197–213.

Kniesner, T.J. & Q. Li (2002) Nonlinearity in dynamic adjustment: Semiparametric estimation of panel labor supply. *Empirical Economics* 27, 131–148.

Li, Q. & C. Hsiao (1998) Testing serial correlation in semiparametric panel data models. *Journal of Econometrics* 87, 207–237.

Li, Q., C. Huang, D. Li, & F. Fu (2002) Semiparametric smooth coefficient model. *Journal of Business & Economic Statistics* 20, 412–422.

Li, Q. & T. Stengos (1996) Semiparametric estimation of partially linear panel data model. *Journal of Econometrics* 71, 389–397.

Newey, W.K. (1990) Efficient instrumental variables estimation of nonlinear models. *Econometrica* 58, 809–837.

Pesaran, M.H. & R. Smith (1995) Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68, 79–113.

Robertson, D. & J. Symons (1992) Some strange properties of panel data estimators. *Journal of Applied Econometrics* 7, 175–189.

Shao, Q. & H. Yu (1996) Weak convergence of weighted empirical processes of dependent sequences. *Annals of Probability* 24, 2098–2127.

Volkonskii, V.A. & Y.A. Rozanov (1959) Some limit theorems for random functions, part I. *Theory of Probability and Its Applications* 4, 178–197.

APPENDIX

Before we prove the main results of the paper, for reference convenience, we first present some lemmas which will be used in the proofs of Theorems 2 and 3, although they are just stated here without proof. Indeed, Lemma 1 is the so-called Davydov's inequality, which is Corollary A2 in Hall and Heyde (1980, p.278), Lemma 2 is Lemma 1.1 of Volkonskii and Rozanov (1959) and also appears in the books by Ibragimov and Linnik (1971, Remark 17.2.1) and Fan and Gijbels (1996, Lem. 6.1), and Lemma 3 is a part of Theorem 4.1 of Shao and Yu (1996). For the detailed proofs, see the aforementioned books and papers.

LEMMA 1. *Suppose that* U *and* V *are random variables that are* $\mathcal{F}_{-\infty}^{t}$ *and* $\mathcal{F}_{t+\tau}^{\infty}$*-measurable, respectively, and that* $\|U\|_p < \infty$, $\|V\|_q < \infty$, *where* $\|U\|_p = \{\mathrm{E}[|U|^p]\}^{1/p}$ *and* $p, q > 1$, $p^{-1} + q^{-1} < 1$. *Then*

$$|\mathrm{E}(UV) - \mathrm{E}(U)\mathrm{E}(V)| \le 8[\alpha(\tau)]^r \|U\|_p \|V\|_q,$$

where $r = 1 - p^{-1} - q^{-1}$.

LEMMA 2. *Let* $V_1, \ldots, V_L$ *be* α*-mixing stationary random variables that are* $\mathcal{F}_{i_1}^{j_1}, \ldots, \mathcal{F}_{i_L}^{j_L}$*-measurable, respectively, with* $1 \le i_1 < j_1 < i_2 \cdots < j_L$, $i_{l+1} - j_l \ge \tau$, *and* $|V_l| \le 1$ *for* $l = 1, \ldots, L$. *Then*

$$\left| \mathrm{E}\left(\prod_{l=1}^{L} V_l \right) - \prod_{l=1}^{L} \mathrm{E}(V_l) \right| \le 16(L-1)\alpha(\tau).$$

LEMMA 3. *Let* $2 < p < r \le \infty$ *and* V_t *be an* α*-mixing process with* $\mathrm{E}(V_t) = 0$ *and* $\|V_t\|_r < \infty$. *Define* $S_n = \sum_{t=n}^{n} V_t$ *and assume that* $\alpha(\tau) = O(\tau^{-\theta})$ *for some* $\theta > pr/(2(r-p))$. *Then*

$$\mathrm{E}|S_n|^p \le K n^{p/2} \max_{t \le n} \|V_t\|_r^p,$$

where K *is a finite positive constant.*

We use the same notation introduced in Sections 2 and 3. Throughout this Appendix, we denote by C a generic positive constant, which may take different values at different places.

Proof of Proposition 1. By the stationarity given in Assumption A1, we have

$$\begin{aligned}
\mathrm{E}(\widetilde{\mathbf{S}}_n) &= \mathrm{E}\{\mathbf{Q}_{it}\widetilde{\mathbf{U}}_{it}' K_h(\mathbf{Z}_{it} - \mathbf{z})\} \\
&= \mathrm{E}\begin{pmatrix} \mathbf{WX}' & \mathbf{WX}' \otimes (\mathbf{Z}-\mathbf{z})'/h \\ \mathbf{XW}' \otimes (\mathbf{Z}-\mathbf{z})/h & \mathbf{WX}' \otimes (\mathbf{Z}-\mathbf{z})(\mathbf{Z}-\mathbf{z})'/h^2 \end{pmatrix} K_h(\mathbf{Z}-\mathbf{z}) \\
&= \mathrm{E}\begin{pmatrix} \boldsymbol{\Omega}(\mathbf{Z}) & \boldsymbol{\Omega}(\mathbf{Z}) \otimes (\mathbf{Z}-\mathbf{z})'/h \\ \boldsymbol{\Omega}'(\mathbf{Z}) \otimes (\mathbf{Z}-\mathbf{z})/h & \boldsymbol{\Omega}(\mathbf{Z}) \otimes (\mathbf{Z}-\mathbf{z})(\mathbf{Z}-\mathbf{z})'/h^2 \end{pmatrix} K_h(\mathbf{Z}-\mathbf{z}) \\
&= \int \begin{pmatrix} \boldsymbol{\Omega}(\mathbf{u}) & \boldsymbol{\Omega}(\mathbf{u}) \otimes (\mathbf{u}-\mathbf{z})'/h \\ \boldsymbol{\Omega}'(\mathbf{u}) \otimes (\mathbf{u}-\mathbf{z})/h & \boldsymbol{\Omega}(\mathbf{u}) \otimes (\mathbf{u}-\mathbf{z})(\mathbf{u}-\mathbf{z})'/h^2 \end{pmatrix} K_h(\mathbf{u}-\mathbf{z}) f(\mathbf{u})\, d\mathbf{u} \\
&= \int \begin{pmatrix} \boldsymbol{\Omega}(\mathbf{z}+h\mathbf{u}) & \boldsymbol{\Omega}(\mathbf{z}+h\mathbf{u}) \otimes \mathbf{u}' \\ \boldsymbol{\Omega}'(\mathbf{z}+h\mathbf{u}) \otimes \mathbf{u} & \boldsymbol{\Omega}(\mathbf{z}+h\mathbf{u}) \otimes \mathbf{uu}' \end{pmatrix} K(\mathbf{u}) f(\mathbf{z}+h\mathbf{u})\, d\mathbf{u} \to f(\mathbf{z})\mathbf{S}(\mathbf{z})
\end{aligned}$$

by Assumptions A2 and A3. To establish the assertion in (i), we now define, for $1 \le l \le q$ and $1 \le j \le d$,

$$s_{n,lj} = \frac{1}{n} \sum_{i=1}^{N} \sum_{t=1}^{T} W_{itl} X_{itj} K_h(\mathbf{Z}_{it} - \mathbf{z}),$$

where W_{itl} is the lth element of $\mathbf{W}_{it}$ and X_{itj} is the jth element of $\mathbf{X}_{it}$. Then, by the stationarity given in Assumption A1, we have

$$\begin{aligned} nh^p \operatorname{Var}(s_{n,lj}) &= \frac{h^p}{T} \operatorname{Var}\left(\sum_{t=1}^{T} W_{itl} X_{itj} K_h(\mathbf{Z}_{it} - \mathbf{z}) \right) = h^p \operatorname{Var}(W_{i1l} X_{i1j} K_h(\mathbf{Z}_{i1} - \mathbf{z})) \\ &\quad + \frac{2h^p}{T} \sum_{t=1}^{T-1} (T-t) \operatorname{Cov}(W_{i1l} X_{i1j} K_h(\mathbf{Z}_{i1} - \mathbf{z}), W_{i(t+1)l} X_{i(t+1)j} K_h(\mathbf{Z}_{i(t+1)} - \mathbf{z})) \\ &\equiv I_1 + I_2. \end{aligned}$$

By Assumptions A1 and A2, it is easy to see that $I_1 \le C$. Next, we consider I_2. By Cauchy–Schwarz inequality, $|I_2| \le CT$. Therefore, $\operatorname{Var}(s_{n,lm}) \le C/(Nh^p) \to 0$ by Assumption A5, so that

$$\frac{1}{n} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{W}_{it} \mathbf{X}_{it}' K_h(\mathbf{Z}_{it} - \mathbf{z}) = f(\mathbf{z}) \boldsymbol{\Omega}(\mathbf{z}) + o_p(1). \qquad \textbf{(A.1)}$$

Similarly, one can show that

$$\frac{1}{n} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{W}_{it} \mathbf{X}_{it}' \otimes \left(\frac{\mathbf{Z}_{it} - \mathbf{z}}{h} \right) K_h(\mathbf{Z}_{it} - \mathbf{z}) = o_p(1)$$

and

$$\frac{1}{n} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{W}_{it} \mathbf{X}_{it}' \otimes \left(\frac{\mathbf{Z}_{it} - \mathbf{z}}{h} \right) \left(\frac{\mathbf{Z}_{it} - \mathbf{z}}{h} \right)' K_h(\mathbf{Z}_{it} - \mathbf{z}) = f(\mathbf{z}) \boldsymbol{\Omega}(\mathbf{z}) \otimes \boldsymbol{\mu}_2(K) + o_p(1).$$

Hence, we have proved (i).

Next, we prove (ii). Indeed, it is easy to see that

$$\begin{aligned} h^{-2} \mathrm{E}(\mathbf{B}_n) &= \frac{1}{2} \mathrm{E} \left\{ \begin{pmatrix} \mathbf{W}_{it} \mathbf{X}_{it}' \mathbf{A}((\mathbf{Z}_{it} - \mathbf{z})/h) \\ \mathbf{W}_{it} \mathbf{X}_{it}' \mathbf{A}((\mathbf{Z}_{it} - \mathbf{z})/h) \otimes (\mathbf{Z}_{it} - \mathbf{z})/h \end{pmatrix} K_h(\mathbf{Z}_{it} - \mathbf{z}) \right\} \\ &= \frac{1}{2} \int \begin{pmatrix} \boldsymbol{\Omega}(\mathbf{u}) \mathbf{A}((\mathbf{u} - \mathbf{z})/h) \\ \boldsymbol{\Omega}(\mathbf{u}) \mathbf{A}((\mathbf{u} - \mathbf{z})/h) \otimes (\mathbf{u} - \mathbf{z})/h \end{pmatrix} K_h(\mathbf{u} - \mathbf{z}) f(\mathbf{u})\, d\mathbf{u} \\ &= \frac{1}{2} \int \begin{pmatrix} \boldsymbol{\Omega}(\mathbf{z} + h\mathbf{u}) \mathbf{A}(\mathbf{u}) \\ \boldsymbol{\Omega}(\mathbf{z} + h\mathbf{u}) \mathbf{A}(\mathbf{u}) \otimes \mathbf{u} \end{pmatrix} K(\mathbf{u}) f(\mathbf{z} + h\mathbf{u})\, d\mathbf{u} \to \frac{1}{2} f(\mathbf{z}) \mathbf{B}(\mathbf{z}). \end{aligned}$$

Similar to (A.1), it is easy to show that any component of the variance of $h^{-2}\mathbf{B}_n$ converges to zero so that (ii) holds true.

Finally, we establish (iii). To this end, it is not difficult to check that

$$h^{-2}\mathrm{E}(\mathbf{R}_n) = h^{-2}\mathrm{E}\left\{\begin{pmatrix}\mathbf{W}_{it}\mathbf{X}_{it}'\mathbf{R}(\mathbf{Z}_{it},\mathbf{z}) \\ \mathbf{W}_{it}\mathbf{X}_{it}'\mathbf{R}(\mathbf{Z}_{it},\mathbf{z}) \otimes (\mathbf{Z}_{it}-\mathbf{z})/h\end{pmatrix} K_h(\mathbf{Z}_{it}-\mathbf{z})\right\}$$

$$= h^{-2}\int \begin{pmatrix}\boldsymbol{\Omega}(\mathbf{u})\mathbf{R}(\mathbf{u},\mathbf{z}) \\ \boldsymbol{\Omega}(\mathbf{u})\mathbf{R}(\mathbf{u},\mathbf{z}) \otimes (\mathbf{u}-\mathbf{z})/h\end{pmatrix} K_h(\mathbf{u}-\mathbf{z}) f(\mathbf{u})\, d\mathbf{u}$$

$$= h^{-2}\int \begin{pmatrix}\boldsymbol{\Omega}(\mathbf{z}+h\mathbf{u})\mathbf{R}(\mathbf{z}+h\mathbf{u},\mathbf{z}) \\ \boldsymbol{\Omega}(\mathbf{z}+h\mathbf{u})\mathbf{R}(\mathbf{z}+h\mathbf{u},\mathbf{z}) \otimes \mathbf{u}\end{pmatrix} K(\mathbf{u}) f(\mathbf{z}+h\mathbf{u})\, d\mathbf{u} \to 0,$$

because

$$h^{-2}\mathbf{R}_j(\mathbf{z}+h\mathbf{u},\mathbf{z}) = h^{-2}\left[\mathbf{g}_j(\mathbf{z}+h\mathbf{u},\mathbf{z}) - \mathbf{g}_j(\mathbf{z}) - \nabla\mathbf{g}_j(\mathbf{z})'(h\mathbf{u}) - \left(\frac{1}{2}\right)h^2\mathbf{u}'\nabla^2 g_j(\mathbf{z})\mathbf{u}\right]$$

$$= o(1)$$

for any $\mathbf{u}$ and that any component of the variance of $h^{-2}\mathbf{R}_n$ converges to zero. (Recall that $R_j(\mathbf{Z}_{it},\mathbf{z}) = g_j(\mathbf{Z}_{it}) - a_j(\mathbf{z}) - \nabla g(\mathbf{z})'(\mathbf{Z}_{it}-\mathbf{z}) - (\frac{1}{2})(\mathbf{Z}_{it}-\mathbf{z})'\nabla^2 g_j(\mathbf{z})(\mathbf{Z}_{it}-\mathbf{z})$.) Therefore, (iii) is verified. This completes the proof of Proposition 1. ■

Proof of Proposition 2. Clearly, $\mathrm{E}(\mathbf{T}_n^*) = 0$ and

$$\begin{aligned} nh^p \operatorname{Var}(\mathbf{T}_n^*) &= \frac{h^p}{T}\operatorname{Var}\left(\sum_{t=1}^{T}\mathbf{Q}_{it}\epsilon_{it}K_h(\mathbf{Z}_{it}-\mathbf{z})\right) \\ &= h^p \operatorname{Var}(\mathbf{Q}_{i1}\epsilon_{i1}K_h(\mathbf{Z}_{i1}-\mathbf{z})) \\ &\quad + \frac{2h^p}{T}\sum_{t=1}^{T-1}(T-t)\operatorname{Cov}(\mathbf{Q}_{i1}\epsilon_{i1}K_h(\mathbf{Z}_{i1}-\mathbf{z}), \mathbf{Q}_{i(t+1)}\epsilon_{i(t+1)}K_h(\mathbf{Z}_{i(t+1)}-\mathbf{z})) \\ &\equiv \mathbf{I}_3 + \mathbf{I}_4. \end{aligned}$$

By Assumptions A1 and A2, $\mathbf{I}_3 \to f(\mathbf{z})\mathbf{S}_1^*$. Clearly,

$$|\mathbf{I}_4| \le Ch^p \sum_{t=1}^{T-1} |\operatorname{Cov}(\mathbf{Q}_{i1}\epsilon_{i1}K_h(\mathbf{Z}_{i1}-\mathbf{z}), \mathbf{Q}_{i(t+1)}\epsilon_{i(t+1)}K_h(\mathbf{Z}_{i(t+1)}-\mathbf{z}))|. \tag{A.2}$$

We now show that the right-hand side of the preceding inequality goes to zero. We consider it in two cases: (I) $Th^p \to 0$ and (II) $Th^p \ge C > 0$.

Case I. For any $t \geq 1$, by Assumption A2,

$$\mathrm{Cov}(\mathbf{Q}_{i1}\epsilon_{i1}K_h(\mathbf{Z}_{i1}-\mathbf{z}),\mathbf{Q}_{i(t+1)}\epsilon_{i(t+1)}K_h(\mathbf{Z}_{i(t+1)}-\mathbf{z}))$$
$$= \mathrm{E}\{\mathbf{Q}_{i1}\mathbf{Q}'_{i(t+1)}\sigma^2(\mathbf{W}_{i1},\mathbf{W}_{i(t+1)},\mathbf{Z}_{i1},\mathbf{Z}_{i(t+1)})K_h(\mathbf{Z}_{i1}-\mathbf{z})K_h(\mathbf{Z}_{i(t+1)}-\mathbf{z})\}$$
$$\rightarrow f_{1,t+1}(\mathbf{z},\mathbf{z})\begin{pmatrix}\mathbf{G}_{1,t+1}(\mathbf{z},\mathbf{z}) & \mathbf{0}\\ \mathbf{0} & \mathbf{G}_{1,t+1}(\mathbf{z},\mathbf{z})\otimes\boldsymbol{\mu}_2(K)\end{pmatrix}.$$

By Assumption A2, then

$$|\mathbf{I}_4| \leq Ch^pT \rightarrow \mathbf{0}. \tag{A.3}$$

Case II. First, we split the sum in (A.2) into two parts as $\mathbf{I}_5 = \sum_{t=1}^{d_n}(\cdots)$ and $\mathbf{I}_6 = \sum_{t>d_n}(\cdots)$, where d_n is a sequence of positive integers such that $d_nh^p \rightarrow 0$. First, we show that $\mathbf{I}_5 \rightarrow 0$, which can be done by an analog of (A.3). Next we consider the upper bound of $\mathbf{I}_6$. For this purpose, we denote $K_{it}^{\nu} = K_h(\mathbf{Z}_{it} - z)[(\mathbf{Z}_{it} - \mathbf{z})/h]^{\nu}$, where $\nu = 0$ or 1, and use Lemma 1 to obtain

$$|\mathrm{Cov}(\epsilon_{i1}W_{i1l}K_{i1}^{\nu_1},\epsilon_{i(t+1)}W_{i(t+1)m}K_{i(t+1)}^{\nu_2})|$$
$$\leq C[\alpha(t)]^{\delta/(2+\delta)}\|\epsilon_{i1}W_{i1l}K_{i1}^{\nu_1}\|_{2+\delta}\cdot\|\epsilon_{i1}W_{i1l}K_{i1}^{\nu_2}\|_{2+\delta}$$

for ν_1, $\nu_2 = 0$ or 1. Conditioning on $\mathbf{Z}_{i1}$ and using Assumption B2 yields

$$\mathrm{E}|\epsilon_{i1}W_{i1l}K_{i1}^{\nu}|^{2+\delta} = \mathrm{E}[\mathrm{E}\{|\epsilon_{i1}W_{i1l}|^{2+\delta}|\mathbf{Z}_{i1}\}K_h^{2+\delta}(\mathbf{Z}_{i1}-\mathbf{z})\{|\mathbf{Z}_{i1}-\mathbf{z}|/h\}^{\nu(2+\delta)}]$$
$$= h^{-p(1+\delta)}f(\mathbf{z})\mathrm{E}\{|\epsilon_{i1}W_{i1l}|^{2+\delta}|\mathbf{Z}_{i1}=\mathbf{z}\}\int|\mathbf{u}|^{\nu(2+\delta)}K^{2+\delta}(\mathbf{u})\,d\mathbf{u}$$
$$+\,o(h^{-p(1+\delta)})$$
$$\leq Ch^{-p(1+\delta)} = O(h^{-p(1+\delta)}). \tag{A.4}$$

Then,

$$|\mathrm{Cov}(\epsilon_{i1}W_{i1l}K_{i1}^{\nu_1},\epsilon_{i(t+1)}W_{i(t+1)m}K_{i(t+1)}^{\nu_2})| = O(\alpha^{\delta/(2+\delta)}(t)\,h^{-2p(1+\delta)/(2+\delta)}).$$

Therefore, the (l,j)th element of $\mathbf{I}_6$ becomes

$$|\mathbf{I}_{6(l,j)}| \leq Ch^{-p\delta/(2+\delta)}\sum_{t>d_n}[\alpha(t)]^{\delta/(2+\delta)}$$
$$= O(h^{-p\delta/(2+\delta)}d_n^{-\delta}) \rightarrow 0$$

by Assumption B3 and choosing d_n such that $h^pd_n^2 = O(1)$, so the requirement that $d_nh_n^p \rightarrow 0$ is satisfied. Therefore, $\mathbf{I}_4 \rightarrow \mathbf{0}$. This proves Proposition 2. ■

To prove Theorems 2 and 3, from (12), clearly it suffices to establish the asymptotic normality of $\sqrt{nh^p}\mathbf{T}_n^*$. To this end, we now employ the Cramér–Wold device because

$\mathbf{T}_n^*$ is multivariate. For any unit vector $\mathbf{d} \in \Re^{m_1}$, let $\omega_{it} = h^{p/2}\mathbf{d}'\mathbf{Q}_{it}K_h(\mathbf{Z}_{it} - \mathbf{z})\epsilon_{it}$, $1 \leq i \leq N$ and $1 \leq t \leq T$. Then,

$$\sqrt{nh^p}\mathbf{d}'\mathbf{T}_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{N} \sum_{t=1}^{T} \omega_{it},$$

and by Proposition 2 and (A.2), for any $1 \leq i \leq N$ and $1 \leq t \leq T$,

$$\text{Var}(\omega_{it}) = f(\mathbf{z})\mathbf{d}'\mathbf{S}^*(\mathbf{z})\mathbf{d}(1 + o(1)) \equiv \theta^2(\mathbf{z})(1 + o(1)) \quad \text{and}$$

$$\sum_{t=2}^{T} |\text{Cov}(\omega_{i1}, \omega_{it})| = o(1).$$

Therefore, $\text{Var}(\sqrt{nh^p}\mathbf{d}'\mathbf{T}_n^*) = \theta^2(\mathbf{z})(1 + o(1))$.

Proof of Theorem 2. Define $\omega_{n,i}^* = T^{-1/2} \sum_{t=1}^{T} \omega_{it}$. Then, $\{\omega_{n,i}^*\}$ are independent double array random variables because T is finite and $\sqrt{nh^p}\mathbf{d}'\mathbf{T}_n^* = N^{-1/2} \sum_{i=1}^{N} \omega_{n,i}^*$. To show the asymptotic normality, it suffices to check Lyapounov's condition. By Minkowski's inequality and using derivations similar to those used in the proof of (A.4), we have

$$\text{E}|\omega_{n,i}^*|^{2+\delta} \leq CT^{(2+\delta)/2}\text{E}|\omega_{i1}|^{2+\delta} \leq CT^{(2+\delta)/2}h^{-p\delta/2}$$

by Assumption B2. Therefore, $n^{-(2+\delta)/2} \sum_{i=1}^{N} \text{E}|\omega_{n,i}^*|^{2+\delta} \leq C\ (Nh^p)^{-\delta/2} \to 0$ by Assumption A5. Thus, we have shown that Lyapounov's condition holds and Theorem 2 follows. ■

Proof of Theorem 3. When $T \to \infty$, for each i, $\{\omega_{i,t}\}_{t=1}^{T}$ is a stationary α-mixing sequence. Therefore, the proof is more complicated; see Hall and Heyde (1980) and Ibragimov and Linnik (1971). The common approach to prove asymptotic normality for a stationary α-mixing sequence is to employ Doob's small-block and large-block technique; see Ibragimov and Linnik (1971, Ch. 18), Cai (2002, 2003), and Cai et al. (2000) for details. For this setting, we partition $\{1, \ldots, T\}$ into $2q_T + 1$ subsets with large block of size r_T and small block of size $s_T < T$ with $r_T + s_T < T$ and r_T and s_T specified later. Set $q_T = \lfloor T/(r_T + s_T) \rfloor$ and define the random variables, for $0 \leq j \leq q_T - 1$,

$$\eta_{ij} = \sum_{t=j(r_T+s_T)+1}^{j(r_T+s_T)+r_T} \omega_{it}, \qquad \xi_{ij} = \sum_{t=j(r_T+s_T)+r_T+1}^{(j+1)(r_T+s_T)} \omega_{it}, \quad \text{and} \quad \zeta_{iq_T} = \sum_{t=q_T(r_T+s_T)+1}^{T} \omega_{it}.$$

Then,

$$\sqrt{nh^p}\mathbf{d}'\mathbf{T}_n^* = \frac{1}{\sqrt{n}} \left\{ \sum_{i=1}^{N} \sum_{j=0}^{q_T-1} \eta_{ij} + \sum_{i=1}^{N} \sum_{j=0}^{q_T-1} \xi_{ij} + \sum_{i=1}^{N} \zeta_{iq_T} \right\} \equiv \frac{1}{\sqrt{n}} \{Q_{n,1} + Q_{n,2} + Q_{n,3}\}.$$

To establish the asymptotic normality of $\sqrt{nh^p}\mathbf{d}'\mathbf{T}_n^*$, it suffices to show the following: as $N \to \infty$ and $T \to \infty$,

$$\frac{1}{n}\mathrm{E}[Q_{n,2}]^2 \to 0, \qquad \frac{1}{n}\mathrm{E}[Q_{n,3}]^2 \to 0, \tag{A.5}$$

$$\left|\mathrm{E}\left[\exp\left(it\sum_{j=0}^{q_T-1}\eta_{kj}\right)\right] - \prod_{j=0}^{q_T-1}\mathrm{E}[\exp(it\eta_{kj})]\right| \to 0, \quad \text{for any } 1 \le k \le N, \tag{A.6}$$

$$\frac{1}{n}\sum_{i=1}^{N}\sum_{j=0}^{q_T-1}\mathrm{E}(\eta_{ij}^2) \to \theta^2(\mathbf{z}), \tag{A.7}$$

and

$$\frac{1}{n}\sum_{i=1}^{N}\sum_{j=0}^{q_T-1}\mathrm{E}[\eta_{ij}^2 I\{|\eta_{ij}| \ge \epsilon\theta(\mathbf{z})\sqrt{N}\}] \to 0 \tag{A.8}$$

for every $\epsilon > 0$. The explanations of the equations (A.5)–(A.8) are as follows: (A.5) implies that $Q_{n,2}/\sqrt{n}$ and $Q_{n,3}/\sqrt{n}$ are asymptotically negligible in probability; (A.6) shows that $\{\eta_{kj}\}$ in $Q_{n,1}/\sqrt{n}$ are asymptotically independent; and (A.7) and (A.8) are the standard Lindeberg–Feller conditions for the asymptotic normality of $Q_{n,1}/\sqrt{n}$ for the independent setup. It follows from the proof of Theorem 18.4.1 in Ibragimov and Linnik (1971) that a combination of (A.6)–(A.8) concludes $Q_{n,2}/\sqrt{n} \to N(0, \theta^2(\mathbf{z}))$. Therefore, because both $Q_{n,1}/\sqrt{n}$ and $Q_{n,3}/\sqrt{n}$ converge to zero in probability, by applying Slusky's theorem, we prove the asymptotic normality of $\sqrt{nh^p}\mathbf{d}'\mathbf{T}_n^*$.

The remaining parts of the proof are to verify equations (A.5)–(A.8). First, let us establish (A.5). For this purpose, we choose the large-block size r_T by $r_T = \lfloor T^{1/\tau} \rfloor$ and the small-block size by $s_T = \lfloor T^{1/(\tau+1)} \rfloor$, where τ is given in Assumption B3 and $\lfloor x \rfloor$ denotes the integer part of x. Then, it can easily be shown from Assumption B3 that

$$s_T/r_T \to 0, \qquad r_T/T \to 0, \quad \text{and} \quad q_T\alpha(s_T) \le CT^{-1/(\tau+1)\tau} \to 0. \tag{A.9}$$

Observe that

$$N^{-1}\mathrm{E}[Q_{n,2}]^2 = \sum_{j=0}^{q_T-1}\mathrm{Var}(\xi_{ij}) + 2\sum_{0\le k<j\le q_T-1}\mathrm{Cov}(\xi_{ik},\xi_{ij}) \equiv J_1 + J_2. \tag{A.10}$$

It follows from stationarity and Proposition 2 that

$$J_1 = q_T\,\mathrm{Var}(\xi_{i1}) = q_T\,\mathrm{Var}\left(\sum_{t=1}^{s_T}\omega_{it}\right) = q_T s_T[\theta^2(\mathbf{z}) + o(1)]. \tag{A.11}$$

Next consider the second term J_2 on the right-hand side of (A.10). Let $r_j^* = j(r_T + s_T)$; then $r_j^* - r_k^* \ge r_T$ for all $j > k$. Therefore, we have

$$|J_2| \le 2 \sum_{0\le k<j\le q_T-1} \sum_{j_1=1}^{s_T} \sum_{j_2=1}^{s_T} |\text{Cov}(\omega_{i,r_k^*+r_T+j_1}, \omega_{i,r_j^*+r_T+j_2})|$$

$$\le 2 \sum_{j_1=1}^{T-r_T} \sum_{j_2=j_1+r_T}^{T} |\text{Cov}(\omega_{ij_1}, \omega_{i,j_2})|.$$

By stationarity and (A.2), one obtains

$$|J_2| \le 2T \sum_{j=r_T+1}^{T} |\text{Cov}(\omega_{i1}, \omega_{ij})| = o(T). \tag{A.12}$$

Hence, by (A.9)–(A.12), we have

$$\frac{1}{n}\,\text{E}(Q_{n,2})^2 = O(q_T s_T/T) + o(1) = o(1). \tag{A.13}$$

It follows from stationarity, (A.9), and Proposition 2 that

$$\text{Var}(Q_{n,3}) = N\,\text{Var}\left(\sum_{t=1}^{T-q_T(r_T+s_T)} \omega_{it}\right) = O(N(T-q_T(r_T+s_T))) = o(n). \tag{A.14}$$

Combining (A.9), (A.13), and (A.14), we have established (A.5).

To establish (A.6), we use Lemma 2 to obtain

$$\left|\text{E}\left[\exp\left(it\sum_{j=0}^{q_T-1}\eta_{kj}\right)\right] - \prod_{j=0}^{q_T-1}\text{E}[\exp(it\eta_{kj})]\right| \le 16 q_T \alpha(s_T),$$

which goes to zero as $T \to \infty$ by (A.9). Therefore, (A.6) is proved.

As for (A.7), by stationarity, (A.2), (A.9), and Proposition 2, it is easily seen that

$$\frac{1}{n}\sum_{i=1}^{N}\sum_{j=0}^{q_T-1}\text{E}(\eta_{ij}^2) = \frac{q_T}{T}\,\text{E}(\eta_{i1}^2) = \frac{q_T r_T}{T}\cdot\frac{1}{r_T}\,\text{Var}\left(\sum_{t=1}^{r_T}\omega_{it}\right) \to \theta^2(\mathbf{z}),$$

so that (A.7) is proved.

It remains to establish (A.8). For this purpose, we employ Lemma 3 and Assumption B3 to obtain

$$\text{E}[\eta_{i1}^2 I\{|\eta_{i1}| \ge \epsilon\theta(\mathbf{z})\sqrt{N}\}] \le CN^{-\delta/2}\text{E}(|\eta_{i1}|^{2+\delta}) \le CN^{-\delta/2} r_T^{1+\delta/2} \|\omega_{i1}\|_{2+2\delta}^{2+\delta}.$$

Similar to (A.4), one can show easily that

$$\text{E}(|\omega_{i1}|^{2+2\delta}) \le Ch^{-p\delta},$$

which in conjunction with the preceding result implies that

$$\text{E}[\eta_{i1}^2 I\{|\eta_{i1}| \ge \epsilon\theta(\mathbf{z})\sqrt{n}\}] \le Cn^{-\delta/2} r_T^{1+\delta/2} h^{-p\delta(2+\delta)/2(1+\delta)}.$$

Thus, by the definition of r_T, we obtain

$$\frac{1}{n}\sum_{i=1}^{N}\sum_{j=0}^{q_T-1} \mathrm{E}[\eta_{ij}^2 I\{|\eta_{ij}| \geq \epsilon\theta(\mathbf{z})\sqrt{n}\}] \leq C\,(NT^{(\tau+1)/\tau}h^{p(2+\delta)/(1+\delta)})^{-\delta/2}$$

tending to zero by Assumption B4. Thus this completes the proof of Theorem 3. ■