Econometric Theory, **30**, 2014, 1315–1347.
doi:10.1017/S0266466614000188

NONPARAMETRIC ESTIMATION OF DYNAMIC PANEL MODELS WITH FIXED EFFECTS

YOONSEOK LEE
Syracuse University

This paper considers nonparametric estimation of autoregressive panel data models with fixed effects. A within-group type series estimator is developed and its convergence rate and asymptotic normality are derived. It is found that the series estimator is asymptotically biased and the bias could reduce the mean-square convergence rate compared with the cross-section cases. A bias corrected nonparametric estimator is developed.

1. INTRODUCTION

Notwithstanding the large literature on nonparametric time series models, little attention has been given to nonparametric dynamic panel models. One explanation is the difficulty of treating individual effects and the nonlinear autoregressive structure simultaneously in the context of nonparametric estimation, especially when the unobserved individual effects are specified as fixed effects. This paper suggests an easy-to-use solution to this problem by developing series estimation of nonparametric autoregressive fixed-effects models, particularly when the fixed effects are additively separable to the nonparametric component. Note that, although there are several studies on nonparametric panel models, they are restricted to nondynamic models with strict exogeneity (e.g., Porter, 1996; Baltagi and Li, 2002; Henderson and Ullah, 2005; Su and Ullah, 2006; Henderson et al., 2008; Lee and Mukherjee, 2012; Su and Jin, 2012) or they assume the individual effects to be uncorrelated with the regressors, which basically precludes the fixed-effect assumption (e.g., Li and Stengos, 1996; Li and Kniesner, 2002; Park, Sickles, and Simar, 2007).

This paper uses the within transformation (i.e., deviations from the individual sample mean) to eliminate the fixed effects and develops nonparametric

I thank Peter Phillips, Donald Andrews, Yuichi Kitamura, a co-editor, and anonymous referees for their valuable suggestions that greatly improved the paper. I also received helpful comments from seminar participants at Cornell, Michigan, MSU, OSU, Penn, PSU, Rochester, Binghamton, TAMU, UBC, UC-Irvine, UVa, UWa-Seattle, VaTech, WMU, Yale, NY Camp Econometrics II, and the 2005 Greater New York Metropolitan Area Econometrics Colloquium at Columbia University. I gratefully acknowledge financial support from the Cowles Foundation under the Carl Arvid Anderson Prize. All errors are solely mine. Address correspondence to Yoonseok Lee, Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244-1020, USA; e-mail: ylee41@maxwell.syr.edu.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

estimation for the within-transformed dynamic panel models using series estimation. Among the large literature on series estimation, Andrews (1991a), Newey (1997), and Chen and Shen (1998) are closely related to this paper. Series estimation is convenient in this context because the within transformation of the unknown function can be approximated by a linear combination of the within-transformed series functions. The within-transformation-based method, once the asymptotic bias is corrected, yields more efficient estimators in finite samples in comparison with the first-difference-based instrumental variables (IV) estimators, especially when the autoregressive parameters are near unit root. Moreover, as the standard within-group (WG) estimator, the new estimation procedure is based on least squares estimation and thus it is much easier to implement in practice than the IV-based methods.

Technically, we obtain a class of nonparametric functions in which a panel homogeneous Markov process is shown to be stationary β-mixing conditioning on the fixed effects. Based on this mixing property, we derive the mean-square convergence rate and pointwise asymptotic normality of the series estimator when both the cross-section size N and the length of time T are large and they are of comparable sizes. Similarly as pooled estimation in linear dynamic panel data models (e.g., Hahn and Kuersteiner, 2002; Alvarez and Arellano, 2003; Lee, 2012), an asymptotic bias is present, which could reduce the mean-square convergence rate compared with the cross-section cases. The asymptotic bias form is obtained, based on which we develop a bias corrected series estimator.

This paper is organized as follows. Section 2 introduces the basic model and discusses a mixing condition for the nonlinear autoregressive panel system. Section 3 develops WG series estimation for dynamic panel models. Section 4 examines the asymptotic properties of the estimator under large N and T and also proposes a bias corrected estimator. Section 5 summarizes the Monte Carlo experiments examining the performance of the WG series estimator and bias correction in finite samples. Section 6 concludes the paper with some remarks on the generalization to the partially linear models. All the mathematical proofs are provided in the Appendix.

2. NONPARAMETRIC DYNAMIC PANEL MODELS

We consider a panel process $\{y_{i,t}\}$ generated from a nonlinear autoregressive model given by

$$y_{i,t} = m\left(y_{i,t-1}\right) + \mu_i + u_{i,t} \tag{1}$$

for $i = 1, 2, \cdots, N$ and $t = 1, 2, \cdots, T$, where $y_{i,t} \in \mathcal{Y} \subset \mathbb{R}$ and $m(\cdot) \in \mathcal{M}$ is unknown with $\mathcal{M}$ being a specified class of functions from $\mathcal{Y}$ to $\mathbb{R}$. The objective is to estimate $m(\cdot)$. For all i and t, it is assumed that $\mathbb{E}(u_{i,t}|y_{i,t-1}, \cdots, y_{i,0}, \mu_i) = 0$ and the initial values, $y_{i,0}$, are observed. The individual specific effect, μ_i, is assumed to have finite variance and to be independent of $u_{i,t}$ for all i and t.

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

But it can be correlated with the regressor $y_{i,t-1}$ (i.e., the fixed effect), so that it captures the unobserved and possibly omitted individual heterogeneity.

The key assumption in (1) is that the fixed effect is additively separable from the unknown function $m(\cdot)$ so that the partially linear specification $\mathbb{E}(y_{i,t}|y_{i,t-1}, \cdots, y_{i,0}, \mu_i) = m\left(y_{i,t-1}\right) + \mu_i$ holds. This assumption is restrictive but it simplifies identification and estimation of the semiparametric model. Such specification is particularly interesting when the individuals share the common autoregressive dynamics so that the marginal effects are homogeneous across i, whereas the drifts can be heterogeneous. The growth regression (e.g., Lee, 2006) is a good example, in which the nonlinearity captures multiple regimes of growth convergence patterns over different levels of GDP whereas the fixed effects capture the country-specific production functions. Other examples are the panel extension of the functional coefficient autoregressive model (Chen and Tsay, 1993: $y_{i,t} = \varsigma(y_{i,t-1})y_{i,t-1} + \mu_i + u_{i,t}$ for unknown $\varsigma(\cdot)$) and the smoothed threshold autoregressive model (Chan and Tong, 1986: $y_{i,t} = \varsigma(y_{i,t-1} - c)\alpha_1 y_{i,t-1} + [1 - \varsigma(y_{i,t-1} - c)]\alpha_2 y_{i,t-1} + \mu_i + u_{i,t}$ for some $c > 0$ and unknown $\varsigma(\cdot)$).

We first assume the following conditions.

Assumption 1 (Unobservables)**.** (i) $\{u_{i,t}\}$ is i.i.d. with mean zero, variance $0 < \sigma^2 < \infty$, and $\mathbb{E}|u_{i,t}|^\nu < \infty$ for some $\nu > 4$. (ii) $\{u_{i,t}\}$ has a positive density and an absolutely continuous distribution with respect to the Lebesgue measure on $\mathbb{R}$. (iii) $u_{i,t}$ is independent of μ_i for all i and t. (iv) μ_i is i.i.d. with a finite second moment and has compact support.

To make the notation simple, we let $\{u_{i,t}\}$ be independent and identically distributed (*i.i.d.*) although it can be relaxed to be a martingale difference sequence conditional on μ_i for each i. The homoskedasticity assumption can also be relaxed provided that $N^{-1}\sum_{i=1}^N \sigma_i^2 \to_p \sigma^2$ as $N \to \infty$ for some $0 < \sigma^2 < \infty$, where $\mathbb{E}u_{i,t}^2 = \sigma_i^2$ for each i (e.g., Phillips and Sul, 2003). Assumption 1 implies that $u_{i,t}$ is independent of $\{y_{i,s}\}_{s \le t-1}$ for each i and thus sequentially exogenous since $y_{i,t}$ is a function of μ_i and $\{u_{i,s}\}_{s \le t}$. In addition, $\{y_{i,t}\}$ is mutually independent across i. For each i, however, we assume that the stochastic process $\{y_{i,t}\}$ given in (1) is a homogeneous Markov chain satisfying the following conditions.

Assumption 2 (Markov chain)**.** For each i and conditional on μ_i, (i) $\{y_{i,t}\}$ is a homogeneous Markov chain satisfying (1), whose τ-th step transition probability is denoted as $P_i^\tau(y, \cdot)$ and (ii) there exists an invariant probability measure π_i, from which the initial value $y_{i,0}$ is drawn. (iii) $m(\cdot)$ is Borel measurable and there exist constants $\overline{y} > 0$, $0 < c_1 < 1$, and $c_0 \ge 0$ such that $|m(y)| \le c_1|y| + c_0$ if $|y| > \overline{y}$; and $\sup_{y:|y| \le \overline{y}} |m(y)| < \infty$, where $[-\overline{y}, \overline{y}] \subset \mathcal{Y}$.

Assumption 2-(iii) is a standard condition for a stationary Markov chain to be geometrically (Harris) ergodic (e.g., Tjøstheim, 1990; Doukhan, 1994). In the time series context, a wide class of nonlinear autoregressive functions satisfy this assumption (e.g., Tong, 1990; Meyn and Tweedie, 1993; Fan and Yao, 2003).

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

Chen and Shen (1998) also use the same condition in developing a sieve extremum estimator for a weakly dependent nonlinear time series.

When individual effect μ_i is present in the dynamics as (1), the panel process $\{y_{i,t}\}$ cannot be simply assumed to be (unconditionally) ergodic because the time-invariant random constant μ_i makes the temporal dependence of $y_{i,t}$ persistent. Since we understand the transition probability $P_i^\tau(y,\cdot)$ and the invariant probability measure π_i as conditional probabilities given μ_i, however, μ_i can be regarded as a common nonrandom shift of the distribution of $\{y_{i,t}\}$ for each i and it no longer affects the temporal dependence of $\{y_{i,t}\}$. In this case, we consider that a stationary homogeneous panel Markov process $\{y_{i,t}\}$ in (1) is β-mixing (i.e., absolutely regular) conditional on the fixed effect μ_i for each i if there exists a sequence of nonnegative random variables $\{\beta_i(\tau)\}$ converging to zero almost surely as $\tau \to \infty$ such that

$$\int_{\mathcal{Y}} \pi_i(dy)||P_i^\tau(y,\cdot) - \pi_i(\cdot)||_{TV} \leq \beta_i(\tau) \text{ a.s.} \tag{2}$$

for all $\tau > 0$ and for each given i. Note that the left-hand-side of (2) is the conditional β-mixing coefficient of the stationary homogeneous Markov process $\{y_{i,t}\}$ for each i with $||\cdot||_{TV}$ being the total variation of a signed measure,[1] similarly as the definition of the standard β-mixing coefficient (e.g., Davydov, 1973; Doukhan, 1994). Under the conditions in Assumptions 1 and 2, we can conclude the following result.

PROPOSITION 2.1. *Under Assumptions 1 and 2, conditional on μ_i for each i, the panel process $\{y_{i,t}\}$ satisfying (1) is geometrically ergodic, stationary, and β-mixing with exponential decay.*

Since the homogeneous Markov process $\{y_{i,t}\}$ is geometrically ergodic and stationary, the invariant probability measure π_i uniquely exists, from which the initial value $y_{i,0}$ was drawn. Therefore, the process is the unique stationary solution for a given initial distribution π_i for each i. As discussed in Prakasa Rao (2009), however, conditional mixing does not imply unconditional mixing for a sequence of random variables and vice versa. We need to assume the following condition that guarantees the average of the mixing coefficients $\beta_i(\tau)$ goes to zero fast enough.

Assumption 3 (Beta-mixing)**.** $N^{-1}\sum_{i=1}^N \sum_{\tau=1}^\infty \beta_i(\tau)^{1-4/\nu} < \infty$ a.s. for some $\nu > 4$ defined in Assumption 1.

Hahn and Kuersteiner (2011, Rem. 1) also introduce conditional mixing, who assume that the conditional α-mixing coefficient is uniformly bounded (i.e., $\sup_{1\leq i\leq N}|\alpha_i(\tau)| \leq Ca^\tau$ a.s. for some $0 < a < 1$ and $C > 0$, where $\alpha_i(\tau)$ is the conditional α-mixing coefficient defined in Lemma A1.1 in the Appendix). The conditional mixing idea in (2) is quite similar to theirs but it assumes a weaker

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

condition in that the mixing coefficient needs to be bounded on average (e.g., Assumption 3 is weaker than assuming $\sup_{1\leq i\leq N}\sum_{\tau=1}^{\infty}\beta_i(\tau)^{1-4/\nu}<\infty$ a.s.).

3. WITHIN-GROUP SERIES ESTIMATION

The key motivation of the within-group series estimation of $m(\cdot)$ is based on the interchangeability between two (global) linear operators: series approximation and within transformation. Roughly speaking, we let $\{g_{Kk}:\mathcal{Y}\to\mathbb{R}$ for $k=1,2,\cdots,K\}$ be a sequence of known measurable basis functions that approximate a smooth function $m(\cdot)$ as

$$m(y)\approx\sum_{k=1}^{K}\theta_{Kk}g_{Kk}(y) \tag{3}$$

for an arbitrary $y\in\mathcal{Y}$, where $\theta_{Kk}\in\mathbb{R}$ are unknown parameters. "$\approx$" indicates series approximation: namely, the choice of the sequence must be such that the approximation to $m(\cdot)$ improves as K gets larger in some metric, where $K=K(N,T)\to\infty$ as $N,T\to\infty$. We assume that K is not data dependent but we let it increase with N and T, where N and T maintain the following condition.

Assumption 4 (N and T)**.** $\lim_{N,T\to\infty}N/T=\kappa$, where $0<\kappa<\infty$.

To describe the basic idea of the estimator, we eliminate the individual effects μ_i in (1) by employing the standard within transformation as for the linear fixed-effects regression:[2]

$$y_{i,t}^{0}=m^{0}\left(y_{i,t-1}\right)+u_{i,t}^{0}, \tag{4}$$

where $m^{0}(y_{i,t-1})=m(y_{i,t-1})-(1/T)\sum_{s=1}^{T}m(y_{i,s-1})$, $y_{i,t}^{0}=y_{i,t}-(1/T)\sum_{s=1}^{T}y_{i,s}$, and similarly for $u_{i,t}^{0}$. From the series approximation (3), $m^{0}(y_{i,t-1})$ can be approximated as

$$m^{0}\left(y_{i,t-1}\right)\approx\sum_{k=1}^{K}\theta_{Kk}g_{Kk}^{0}\left(y_{i,t-1}\right) \tag{5}$$

using the transformed basis functions $g_{Kk}^{0}(y_{i,t-1})=g_{Kk}(y_{i,t-1})-(1/T)\sum_{s=1}^{T}g_{Kk}(y_{i,s-1})$. If one views $g_{Kk}(y_{i,t-1})$ as a regressor, $g_{Kk}^{0}(y_{i,t-1})$ corresponds to the demeaned regressor, as in the case of within-group (WG) estimation of the linear panel regression model. Using the linear approximation (5) for the regression (4), the unknown parameters $\theta_K=(\theta_{K1},\theta_{K2},\cdots,\theta_{KK})'$ can be estimated by least squares as

$$\widehat{\theta}_K=\left(\sum_{i=1}^{N}\sum_{t=1}^{T}g_K^{0}\left(y_{i,t-1}\right)g_K^{0}\left(y_{i,t-1}\right)'\right)^{-}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}g_K^{0}\left(y_{i,t-1}\right)y_{i,t}^{0}\right), \tag{6}$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

where $g_K^0(y) = (g_{K1}^0(y), g_{K2}^0(y), \cdots, g_{KK}^0(y))'$ and $(\cdot)^-$ denotes the generalized inverse. Under conditions given below, however, the denominator will be asymptotically nonsingular, and hence the generalized inverse will be the standard inverse for large N and T. The WG series estimator of $m(\cdot)$ is then obtained as

$$\widehat{m}(y) = \widehat{\theta}_K' g_K(y) = \sum_{k=1}^{K} \widehat{\theta}_{Kk} g_{Kk}(y) \tag{7}$$

for $y \in \mathcal{Y}$ from (3).

It is important to note that we exclude the constant term in the basis functions. Otherwise, the within transformation on $\{g_{Kk}(y)\}_{k=1}^K$ will make the denominator in (6) singular. This corresponds to the conventional linear panel regression case, where fixed effects and a constant term (in fact, any time-invariant terms) cannot be identified separately. Accordingly, we introduce the following normalization assumption.

Assumption 5 (Normalization)**.** $m(0) = 0$.

The normalization condition 5 allows μ_i to be unrestricted (e.g., it could have nonzero mean as long as it satisfies Assumption 1) but requires that $m(\cdot)$ passes through the origin. If $m(0) \neq 0$, we simply consider $(\mu_i + m(0))$ and $(m(y) - m(0))$ as fixed effects and the unknown function, respectively, to restore this condition.[3]

In order to conduct series estimation described above, we need to assume that $m(\cdot)$ is smooth enough in some sense so that the approximation (3) is uniform in some metric. In standard cross-sectional series estimation, the support of the unknown function is normally assumed to be compact (e.g., Andrews, 1991a; Newey, 1997) and the series approximation (3) is well-defined over the compact support in the standard sup-norm, provided that $m(\cdot)$ belongs to the Hölder class (e.g., Chen, 2007). However, such an assumption does not seem appropriate for the case of the autoregressive model (1) because it will not only restrict the support of the regressor $y_{i,t-1}$ but also the support of the dependent variable $y_{i,t}$. Since the error $u_{i,t}$ is defined over $\mathbb{R}$, restricting the support of $y_{i,t}$ to be bounded can be restrictive.

For characterizing the series approximation (3) over possibly unbounded support $\mathcal{Y}$, we introduce a weighted sup-norm as Chen, Hong, and Tamer (2005).[4] For a scalar function $h : \mathcal{Y} \to \mathbb{R}$, we let

$$\|h\|_{\infty,\omega} = \sup_{y \in \mathcal{Y}} |h(y)| (1 + \|y\|^2)^{-\omega/2} \text{ for some } \omega \geq 0,$$

where $\|\cdot\|$ is the Euclidean norm and $\omega = 0$ leads to the standard sup-norm for bounded $\mathcal{Y}$. We also define that h is $H(\gamma, \omega)$-smooth on $\mathcal{Y}$ if h belongs to $\Lambda_c^\gamma(\mathcal{Y}, \omega)$ for some $\gamma > 0$ and $\omega \geq 0$, where $\Lambda_c^\gamma(\mathcal{Y}, \omega) = \{h \in \Lambda^\gamma(\mathcal{Y}, \omega) : ||h(\cdot)[1 + \|\cdot\|^2]^{-\omega/2}||_{\Lambda^\gamma} \leq c < \infty\}$ is a weighted Hölder ball with radius c.

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

$\Lambda^{\gamma}(\mathcal{Y},\omega)$ denotes a weighted Hölder space such that $h(\cdot)[1+\|\cdot\|^2]^{-\omega/2}$ is in a Hölder space $\Lambda^{\gamma}(\mathcal{Y})$ of order γ with the Hölder norm $\|\cdot\|_{\Lambda^{\gamma}}$. See Chen, Hong, and Tamer (2005) for further details of the definitions. The following condition specifies a uniform rate of series approximation.

Assumption 6 (Series approximation)**.** For some $\gamma > 0$ and $\omega \geq 0$, (i) $m(\cdot)$ is $H(\gamma,\omega)$-smooth on $\mathcal{Y}$; *(ii)* $\mathbb{E}[\int(1+||y||^2)^{\omega}\pi_i(dy)] < \infty$ and (iii) there exist a parameter vector $\theta_K \in \mathbb{R}^K$ and a constant $\delta > 0$ such that $||m(y) - g_K(y)'\theta_K||_{\infty,\omega} = O\left(K^{-\delta}\right)$ as $K \to \infty$.

Note that Assumption 6-(ii) imposes conditions on the tail behavior of the marginal densities, which can also be found in Ai and Chen (2003), Chen et al. (2005), and Su and Jin (2012) to deal with unbounded support. δ in Assumption 6-(iii) is closely related to the smoothness of $m(y)$ and the dimensionality of y. For example, when $\omega = \gamma = 0$, this assumption is satisfied with $\delta = D_m/\dim(y)$ for regression splines and power series, where D_m is the number of continuous derivatives of $m(y)$ that exists and $\dim(y)$ is the dimension of y (e.g., Newey, 1997).

4. ASYMPTOTIC PROPERTIES

We start with a set of conditions on the series functions. We denote $\underline{g}_K\left(y_{i,t}\right) = g_K\left(y_{i,t}\right) - \mathbb{E}\left(g_K\left(y_{i,t}\right)|\mu_i\right)$ and thus $\mathbb{E}(\underline{g}_K\left(y_{i,t}\right)) = 0$ for any i and t.

Assumption 7 (Series functions)**.**

(i) For every K, there exist positive integers N^* and T^* such that for all $N \geq N^*$ and $T \geq T^*$, the $NT \times K$ vector $(g_K^0(y_{1,0}),\cdots,g_K^0(y_{N,T-1}))'$ is of full column rank.

(ii) For each K, $\{g_K(y)\}$ is fourth order stationary with $\int g_K(y)\,g_K(y)'\pi_i(dy) < \infty$ for each i; $\Gamma_K = \mathbb{E}\big(\underline{g}_K(y_{i,t})\underline{g}_K'(y_{i,t})\big)$ has the smallest eigenvalue bounded away from zero and a bounded largest eigenvalue.

(iii) $K = K(N,T)$ is nonrandom satisfying $K \to \infty$ and $K^2/NT \to 0$ as $N,T \to \infty$.

(iv) There is a positive constant ζ such that $\sup_{1\leq i\leq N}\sup_{1\leq k\leq K}\mathbb{E}\big(|\underline{g}_{Kk}(y_{i,t})|^{\nu}|\mu_i\big) < \zeta < \infty$ a.s. for some $\nu > 4$ defined in Assumption 1.

Similarly as in Andrews (1991a) and Newey (1997), the first two conditions are useful for controlling the inverse of the covariance matrix estimator, $\widehat{\Gamma}_K = (1/NT)\sum_{i=1}^N\sum_{t=1}^T g_K^0(y_{i,t})g_K^0(y_{i,t})'$, and its convergence in probability to its expectation in the Euclidean norm. Assumption 7-(iv) is, however, different from the aforementioned papers in that it does not assume bounded support of y.

We derive the main asymptotic results of the WG series estimator $\widehat{m}(y)$ defined in (7). The first theorem provides the mean-square convergence rate of $\widehat{m}(y)$.

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

THEOREM 3.1 (Mean square convergence rate). *Under Assumptions 1–7,*

$$\mathbb{E}\int (\widehat{m}(y)-m(y))^2\pi_i(dy) = O\left(\frac{K}{NT}+K^{-2\delta}+\frac{K}{T^2}\right) \quad \textbf{(8)}$$

as $N,T\to\infty$ jointly, where π_i denotes the marginal distribution of $y_{i,t}$ conditional on μ_i.

Theorem 3.1 implies that the probability limit of $\widehat{m}(y)-m(y)$ is zero for $K/NT\to 0$, $K/T^2\to 0$, and $K^{-2\delta}\to 0$ as $N,T\to\infty$. For the mean-square convergence rate in (8), the first two terms also appear in the standard series estimators for the cross-section case (e.g., Newey, 1997): the first term, K/NT, corresponds to the convergence rate of the variance, whereas the second term, $K^{-2\delta}$, corresponds to the squared approximation bias in (3). On the other hand, the last element, K/T^2, is new and it reflects the asymptotic order of the squared within-group bias (e.g., Nickell, 1981), which is introduced as we take the within transformation to eliminate the fixed effects. Since we assume that $N,T\to\infty$ at the same rate in Assumption 4, introduction of this last term does not necessarily make the rate of convergence slower. It thus follows that the rate of convergence of $\widehat{m}(y)$ remains the same as the standard series estimator case so that it attains Stone's (1982) optimal bound under Assumption 4.

However, the rate of convergence could become slower when T is small (i.e., $T\ll N$), which could be a concern when we consider the finite sample properties with short panel data. In this case, the choice of the optimal K can be smaller than the standard case (or the case with Assumption 4). For example, when Assumption 4 holds, we can choose the optimal K such that $K^{-2\delta}=K/NT$ yielding $K^*=O((NT)^{1/(1+2\delta)})$ and thus the mean-square convergence rate becomes $O((NT)^{-2\delta/(1+2\delta)})$. On the other hand, when T is smaller than N, we instead choose the optimal K such that the two possibly dominating terms satisfy $K^{-2\delta}=K/T^2$ yielding $K^{**}=O(T^{2/(1+2\delta)})$. Note that K^{**} is smaller than K^* and the mean-square convergence rate in this case becomes $O(T^{-4\delta/(1+2\delta)})$, which can be slower than $O((NT)^{-2\delta/(1+2\delta)})$ when T is small.

Using a similar argument as Hahn and Kuersteiner (2002) or Lee (2012), although $\widehat{m}(y)$ is consistent from Theorem 3.1, it can be shown that $\sqrt{NT}(\widehat{m}(y)-m(y))$ is not centered at zero under Assumption 4 as in the standard linear dynamic fixed effect regression. The following theorem obtains asymptotic normality of the WG series estimator for the unknown function $m(\cdot)$, in which a proper bias correction is imposed. Note that "$\to_d$" means convergence in distribution. Γ_K is defined in Assumption 7-(ii). Recall that $\widehat{\Gamma}_K=(1/NT)\sum_{i=1}^N\sum_{t=1}^T g_K^0(y_{i,t})g_K^0(y_{i,t})'$.

THEOREM 3.2 (Asymptotic normality). *Let $\widehat{b}_K(y)=g_K(y)'\widehat{\Gamma}_K^{-1}\widehat{\Phi}_K$, where $\widehat{\Phi}_K$ is a $K\times 1$ vector defined as*

$$\widehat{\Phi}_K=\sum_{j=0}^{J}\frac{1}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\left(1-\frac{j}{J+1}\right)g_K^0\left(y_{i,t+j}\right)\widehat{u}_{i,t}^0$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

with $\widehat{u}^0_{i,t} = y^0_{i,t} - \widehat{\theta}'_K g^0_K\left(y_{i,t-1}\right)$ *and* $J = J(T)$ *such that* $J = O(T^{1/3})$. *If Assumptions 1–7 hold and* $\sqrt{NT}K^{-\delta} \to 0$, *then as* $N, T \to \infty$ *jointly, we have*

$$v_K(y)^{-1/2}\left(\widehat{m}(y) - m(y) + \frac{1}{T}\widehat{b}_K(y)\right) \to_d \mathcal{N}(0,1) \tag{9}$$

for $y \in \mathcal{Y}$, *where* $v_K(y) = \sigma^2 g_K(y)'\Gamma_K^{-1} g_K(y)/NT$. *The result still holds using a consistent estimator of* $v_K(y)$ *given by* $\widehat{v}_K(y) = \widehat{\sigma}^2 g_K(y)'\widehat{\Gamma}_K^{-1} g_K(y)/NT$, *where* $\widehat{\sigma}^2 = (1/NT)\sum_{i=1}^N \sum_{t=1}^T \left(\widehat{u}^0_{i,t}\right)^2$.

Unlike the standard asymptotic normality result of the series estimators in the *i.i.d.* cross-section case (e.g., Andrews, 1991a; Newey, 1997), $\widehat{m}(y)$ has nondegenerating asymptotic bias especially when $\lim_{N,T\to\infty} N/T \neq 0$, which thus requires a bias correction term $(1/T)\widehat{b}_K(y)$ for each $y \in \mathcal{Y}$. Such asymptotic bias can be found even in the linear dynamic fixed-effects models (e.g., Hahn and Kuersteiner, 2002; Alvarez and Arellano, 2003; Lee, 2012), and it reflects the endogeneity problem after the within transformation. Note that within-transformed functions of sequentially exogenous regressors $g^0_K\left(y_{i,t-1}\right)$ in (6) have nonzero correlation with the within-transformed regression errors $u^0_{i,t}$.

The rate of convergence in (9) is not $\sqrt{NT}$ but slower than $\sqrt{NT}$ as the smoothing parameter $1/K$ shrinks like the standard nonparametric regression estimators. Roughly speaking, it is $\sqrt{NT/K}$ because $g_K(y)'\Gamma_K^{-1} g_K(y) = O(K)$ that yields $v_K(y) = O(K/NT)$. Note that $\lambda_{\min}(\Gamma_K^{-1})||g_K(y)||^2 \leq g_K(y)'\Gamma_K^{-1} g_K(y) \leq \lambda_{\max}\left(\Gamma_K^{-1}\right)||g_K(y)||^2$ with $\lambda_{\min}\left(\Gamma_K^{-1}\right) = \lambda^{-1}_{\max}(\Gamma_K) < \infty$ and $\lambda_{\max}\left(\Gamma_K^{-1}\right) = \lambda^{-1}_{\min}(\Gamma_K) < \infty$ from Assumption 7-(ii). Moreover, it can also be verified that $||\widehat{b}_K(y)||^2 = O_p(K)$ and thus the asymptotic distribution (9) has constant noncentrality as in the standard parametric dynamic panel case since $v_K(y)^{-1/2}(1/T)\widehat{b}_K(y) = O\left((K/NT)^{-1/2}\right)O_p\left(K^{1/2}/T\right) = O_p\left(\sqrt{N/T}\right) = O_p(1)$ under Assumption 4.

More precisely, from Lemma A1.6 in the Appendix, it can be shown that $||\widehat{\Phi}_K - \Phi_K|| \to_p 0$ as $N, T \to \infty$, where $\Phi_K = \sum_{j=1}^{\infty} cov\left(g_K\left(y_{i,t+j}\right), u_{i,t}\right)$ is the one-sided long-run covariance between $g_K\left(y_{i,t-1}\right)$ and $u_{i,t}$. A simple Bartlett weight function $(1 - j/(J+1))$ is used here (e.g., Newey and West, 1987) although it could be generalized using kernel functions as Andrews (1991b) if the kernel and the truncation parameter J satisfy proper conditions. Note that the required order for J corresponds to the optimal rate for the Bartlett kernel (e.g., Andrews, 1991b), whereas Newey and West (1987) need $J = o(T^{1/4})$ for the consistency of the long-run autocovariance matrix estimator. In fact, we could impose a weaker condition that $J = o((NT/K)^{1/3})$ in this case.

Therefore, the form of the asymptotic bias can be found as $(1/T)g_K(y)'\Gamma_K^{-1}\Phi_K$ for each K since $||\widehat{\Gamma}_K^{-1} - \Gamma_K^{-1}||_1 \to_p 0$ as $N, T \to \infty$ from Lemma A1.2 in the Appendix, where $||B||_1 = \sup_{\ell:||\ell||\leq 1}||B\ell||$, provided that the truncation parameter satisfies $J = O(T^{1/3})$. If we consider the linear dynamic panel

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

regression case (i.e., $y_{i,t} = \mu_i + \theta y_{i,t-1} + u_{i,t}$), then this asymptotic bias expression simplified to $(1/T)[\sigma^2/(1-\theta^2)]^{-1}[\sigma^2/(1-\theta)] = (1/T)(1+\theta)$, which corresponds to the findings by Hahn and Kuersteiner (2002). In comparison, under the additive separability assumption on the fixed effect, such bias does not appear in the static models like $y_{i,t} = m\left(x_{i,t}\right) + \mu_i + u_{i,t}$ when $x_{i,t}$ is strictly exogenous (i.e., $\mathbb{E}(u_{i,t}|x_{i,1}, \cdots, x_{i,T}) = 0$).[5]

Finally, Theorem 3.2 suggests a bias-corrected estimator $\widetilde{m}(y)$ given by

$$\widetilde{m}(y) = \widetilde{m}(y) = \widehat{m}(y) + \frac{1}{T}\widehat{b}_K(y),$$

which satisfies $\widehat{v}_K(y)^{-1/2}(\widetilde{m}(y) - m(y)) \to_d \mathcal{N}(0,1)$ as $N, T \to \infty$. Using the delta method, this pointwise normality result can be generalized to the asymptotic normality of a functional of the bias corrected series estimator $\widetilde{m}(y)$ as Andrews (1991a) and Newey (1997) under proper smoothness conditions on the functional. Note that the bias corrected estimator $\widetilde{m}(\cdot)$ retains the standard rate of convergence given by $\mathbb{E}\int(\widetilde{m}(y) - m(y))^2\pi_i(dy) = O\left(K/NT + K^{-2\delta}\right)$. This gives the optimal rate of convergence $O_p((NT)^{-2\delta/(1+2\delta)})$ with the optimal choice $K^* = (NT)^{1/(1+2\delta)}$ for $\delta > 0$, although we would need larger K than the optimal K^* (i.e., undersmoothing) in order to satisfy asymptotic normality in Theorem 3.2 that requires $K^{-\delta}\sqrt{NT} \to 0$.

5. SIMULATIONS

To evaluate the finite sample performance of the WG series estimator, we conduct simulation studies for $y_{i,t} = m(y_{i,t-1}) + \mu_i + u_{i,t}$ in two different dynamic structures given by

M1: $m(y) = \exp(y)/(1+\exp(y)) - 0.5$
M2: $m(y) = 0.3y\exp(-0.1y^2)$

for $i = 1, \cdots, N$ and $t = 1, \cdots, T$. The nonlinear functions are smooth and all parameter values are chosen to satisfy stationarity as well as Assumption 2. Each nonlinear function is properly centered to satisfy $m(0) = 0$. Fixed effects μ_i are randomly drawn from $\mathcal{U}(0,1)$ and $u_{i,t}$ from $\mathcal{N}(0,1)$. The first model is the logistic function, which is also investigated in Ai and Chen (2003) in the cross-section case. The second model is the amplitude-dependent exponential autoregressive model discussed in Tong (1990), which is also the case of the functional coefficient autoregressive model by Chen and Tsay (1993).

Five sets of samples of $N = 100$ with $T = 5, 10, 25, 50, 100$ are generated. We estimate the unknown function $m(\cdot)$ by WG series estimation and replicate the entire procedure 1000 times. For series estimation, we use power series, particularly the Hermite polynomials. Following the discussions in Andrews (1991a) and Newey (1997), the first set of simulation chooses the optimal K^* as 4, which

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

TABLE 1. Integrated mean squared error and Integrated absolute bias

			Fixed K (K = 4)				Cross Validation K				
			IMSE		IAB		IMSE		IAB		
	N	T	n/c	b/c	n/c	b/c	n/c	b/c	n/c	b/c	$\overline{K}_{CV}$
M1	100	5	2.275	1.365	2.919	2.152	2.165	1.278	2.912	2.144	2.00
		10	0.660	0.292	1.477	0.647	0.573	0.243	1.410	0.581	2.06
		25	0.249	0.176	0.706	0.322	0.208	0.157	0.601	0.218	2.38
		50	0.143	0.121	0.367	0.132	0.151	0.138	0.302	0.106	3.11
		100	0.099	0.092	0.211	0.104	0.124	0.121	0.154	0.060	3.67
M2	100	5	2.284	1.373	2.945	2.174	2.143	1.268	2.908	2.135	2.17
		10	0.662	0.293	1.485	0.652	0.634	0.307	1.417	0.586	2.74
		25	0.257	0.179	0.727	0.344	0.264	0.192	0.698	0.316	3.51
		50	0.156	0.130	0.402	0.181	0.178	0.155	0.374	0.144	3.81
		100	0.116	0.105	0.283	0.101	0.144	0.138	0.190	0.071	4.21

Note: The results are based on within-group series estimation over 1000 iterations using power series. "n/c"(No Correction) displays integrated mean squared error (IMSE) and integrated absolute bias (IAB) before bias correction; "b/c"(Bias Correction) displays IMSE and IAB after bias correction. K is the number of series functions used for the approximation: K = 4 for the fixed choice is assigned to meet the asymptotic order and $\overline{K}_{CV}$ is the average value over K at each replication from the leave-one-out Cross Validation.

is roughly $(NT)^{1/6}$ in our case. We fix K over the five different sample sizes for a better comparison because using different number of basis functions could change the curve fitting performance.[6]

The simulation results are summarized in Table 1. The integrated mean squared errors (IMSE: $\int \mathbb{E}(m(x) - m(x))^2 dx$) are calculated using the discrete expression for the integration (e.g., Ai and Chen, 2003) given by $\sum_{j=1}^{NT-1}\{(1/R)\sum_{r=1}^{R}(y_{j+1}^r - y_j^r)(m(y_j^r) - \widehat{m}(y_j^r))^2\}$, where $y_1^r \leq \cdots \leq y_j^r \leq \cdots \leq y_{NT}^r$ are sorted $y_{i,t}$ at the rth replication $(r = 1, \cdots, R)$. In order to investigate the bias correction performance, we also consider the integrated absolute bias (IAB: $\int |\mathbb{E}m(x) - m(x)|dx$) given by $\sum_{j=1}^{NT-1}|(1/R)\sum_{r=1}^{R}(m(y_j^r) - \widehat{m}(y_j^r))|(y_{j+1}^r - y_j^r)$. The left panel in Table 1 exhibits several interesting findings. First, both the IMSE and the IAB get smaller after bias corrections, which shows that the bias correction improves overall fitting. Although there could be possible increase in the finite sample variance after the bias correction when we use bias estimator to correct the asymptotic bias, the bias reduction is quite significant to offset such effects. Second, the overall curve fit improves with T even without bias correction. This is because the endogeneity bias from the within transformation reduces with T, which is also the case in the linear dynamic fixed-effects models.

Although we assume that K is not data dependent in the main theorems, we also consider flexible choice of K based on the data in order for the sensitivity check.

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

More precisely, we choose K by the leave-one-out Cross Validation (CV): $K_{CV} = \arg\min_{1\le K\le K_{\max}} \sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{i,t}^{0} - \widetilde{m}_{-i}^{0}(y_{i,t-1})\right)^2$, where $\widetilde{m}_{-i}^{0}(y) = g_K^0(y)'\widetilde{\theta}_{-i,K}$ is the bias-corrected estimate computed using all the observations but the ith one. We let $K_{\max} = 10$ for the power series.[7] The right panel in Table 1 summarizes the same simulation result with K_{CV}. At each replication, different K is chosen based on the CV; the average values of 1000 choices of K over the replications are reported as $\overline{K}_{CV}$ in Table 1. Note that $\overline{K}_{CV}$ are not much different from the fixed choice discussed above. But we can find that $\overline{K}_{CV}$ are rather close to $(T^2)^{1/6}$ instead of $(NT)^{1/6}$ when T is quite smaller than N, which corresponds to the optimal choice of K with small T as we discussed in Section 4. We also find that the bias reduction still works well in this case, which demonstrates that the estimation and the bias correction work even under the data-dependent choice of K.

6. CONCLUDING REMARKS

This paper proposes an alternative to the common but rather restricted specification in dynamic panel models—linear autoregressive panel models. We employ nonparametric estimation on the lagged terms. Even though we allow for a general functional form in the regression, we require enough smoothness of the unknown function, which excludes some interesting specifications such as dynamic discrete choice models. In addition, we still postulate an additively separable structure so that neither an individual effect nor the error term is included in the unknown function m. Nonseparability can be allowed for at the cost of more restrictions on m, which is required for identification.

In the current setup, however, we could allow for exogenous variables $x_{i,t} \in \mathcal{X} \subset \mathbb{R}^r$ by considering a partially linear model

$$y_{i,t} = m\left(y_{i,t-1}\right) + \gamma' x_{i,t} + \mu_i + u_{i,t}, \tag{10}$$

where γ is an $r \times 1$ vector of parameters and $\mathcal{X}$ is compact. In this case, using the conditions for the nonlinear $ARX(p,q)$ process in Doukhan (1994, Sect. 2.4.2, Thm. 7), we can derive the conditions under which the Markov process $\{y_{i,t}\}$ in (10) satisfies Proposition 2.1. For example, we let the observed $\{x_{i,t}\}$ be given by $x_{i,t} = \xi(x_{i,t}^*, \mu_i)$ for some smooth, bounded, and measurable function $\xi : \mathcal{X}^* \times \mathbb{R} \to \mathcal{X}$, where $x_{i,t}^* \in \mathcal{X}^* \subseteq \mathcal{X}$ is a latent *i.i.d.* stochastic process independent of μ_i and $u_{i,t}$. Note that under this specification, $\{x_{i,t}\}$ can be serially correlated as well as correlated with μ_i. If there exists a locally bounded and measurable function $f : \mathcal{X}^* \to \mathbb{R}_+$ satisfying $\mathbb{E}f(x_{i,t}^*) < \infty$ and $|\gamma'\xi(x_{i,t}^*, \mu)| \le f(x_{i,t}^*) + c|\mu|$ a.s. for a given $\mu_i = \mu$ and a constant $c < \infty$, then it can be shown that the Markov process $\{y_{i,t}\}$ in (10) is also stationary and β-mixing for each i. The extension of the asymptotic results to the partially linear case is discussed in Lee (2006).

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

NOTES

1. For a signed measure σ on a σ-field $\mathcal{B}$, we define $\|\sigma\|_{TV} \stackrel{\circ}{=} \sup_{B\in\mathcal{B}} \sigma(B) - \inf_{B\in\mathcal{B}} \sigma(B)$. If σ_1 and σ_2 are two probability measures and $\sigma = \sigma_1 - \sigma_2$, we have $\|\sigma\|_{TV} = 2\sup_{B\in\mathcal{B}} |\sigma_1(B) - \sigma_2(B)|$ in view of Scheffe's theorem (e.g., Liebscher, 2005, p. 671)

2. It is the standard within transformation in the sense that we subtract the global mean over $\mathcal{Y}$. In comparison, Lee and Mukherjee (2012) develop a local within transformation for nonparametric models, which uses the local information around some particular value $y \in \mathcal{Y}$.

3. In Porter (1996) and Henderson et al. (2008), it is instead assumed that $\mathbb{E}(\mu_i) = 0$ so that $m(\cdot)$ is identified from $\mathbb{E}(m(y_{i,t-1})) = \mathbb{E}(y_{i,t})$. But since $m(0)$ could be nonzero in this case, $m(\cdot)$ can be identified only up to a constant addition, which is because the constant term in $m(\cdot)$ is eliminated by the within transformation and we cannot restore it unless it is zero.

4. A similar approach can be found in Su and Jin (2012) in a static panel model. An alternative approach to handle the unbounded support is considering a sequence of increasing bounded supports $\mathcal{Y}_{N,T}$ such that $\cup_{N,T\to\infty}\mathcal{Y}_{N,T} = \mathcal{Y}$ using some trimming parameter or restricting the function class that m belongs to (e.g., Gallant and Nychka, 1987; Newey and Powell, 2003).

5. In the nonlinear parametric setup, if the additive separability assumption is violated, we are supposed to have bias with the same asymptotic order even with strictly exogenous regressors (e.g., Hahn and Newey, 2004). It is mainly because, in the maximum likelihood setup, the information orthogonality does not hold between the parameter of interest and the incidental parameters in general under the nonseparability. See Hahn and Kuersteiner (2011) for more general result.

6. As discussed in Section 3, we exclude the constant term (i.e., the first element) in the basis functions. Therefore, $K = 4$ for power series means 5th order polynomial without constant term. However, the Hermite polynomials have nonzero constants in higher order terms. We thus vertically shift $\widehat{m}$ so that $(\widehat{m}(y) - \widehat{m}(0))$ is the new estimate and $(\widehat{\mu}_i + \widehat{m}(0))$ is the new fixed effect estimate.

7. By looking at the leave-one-out over i, instead of both i and t, we could avoid any complication in the serial dependence of the Cross Validation since it is assumed that $y_{i,t}$ is independent over i. We also tried Mallow's C_p criterion in addition to the Cross Validation but the choice of K does not seem to work properly. Based on the (asymptotic) similarity between AIC and C_p, it would be related with the incidental parameters problem in the model selection as studied in Lee and Phillips (2014).

8. More precisely, we define $\alpha_i(\tau)$ such that $\sup_t \sup_{A\in\mathcal{G}^t_{i,-\infty}, B\in\mathcal{G}^\infty_{i,t+\tau}} |\mathbb{P}_i(A\cap B) - \mathbb{P}_i(A)\,\mathbb{P}_i(B)| \le \alpha_i(\tau)$ a.s. for $\tau > 0$ and for each i, then $\alpha_i(\tau)$ decays exponentially. Note that $\mathcal{G}^{t_2}_{i,t_1}$ is the σ-field generated by $\{y_{i,t} : t_1 \le t \le t_2\}$ and $\mathbb{P}_i(A) = \int_A \pi_i(dy)$ is a conditional probability given μ_i.

9. More precisely, we can define an indicator function $\mathbb{I}_{N,T}$ for the smallest eigenvalue of $\widehat{\Gamma}_K$ being away from zero such that $\mathbb{P}(\mathbb{I}_{N,T} = 1) \to 1$ as $N, T \to \infty$. Whenever $\widehat{\Gamma}_K$ appears in the proof, we then consider $\mathbb{I}_{N,T}\widehat{\Gamma}_K$ as in Newey (1997).

10. Note that we can write $\widehat{\theta}_K - \theta_K = \widehat{\Gamma}_K^{-1}(1/NT)\sum_{i=1}^N\sum_{t=1}^T g_K^0(y_{i,t-1})\{u^0_{i,t} + (m^0(y_{i,t-1}) - g_K^0(y_{i,t-1})'\theta_K)\}$, where $g_K^0(y_{i,t-1})u^0_{i,t} = \underline{g}_K(y_{i,t-1})u_{i,t} - (1/T)\sum_{s=1}^T \underline{g}_K(y_{i,t-1})u_{i,s}$ by stationarity.

REFERENCES

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

Alvarez, J. & M. Arellano (2003) The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121–1159.

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

Andrews, D.W.K. (1991a) Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica* 59, 307–345.
Andrews, D.W.K. (1991b) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
Baltagi, B.H. & D. Li (2002) Series estimation of partially linear panel data models with fixed effects. *Annals of Economics and Finance* 3, 103–116.
Bhattacharya, R. & C. Lee (1995) On geometric ergodicity of nonlinear autoregressive models. *Statistics & Probability Letters* 22, 311–315.
Chan, K.S. & H. Tong (1986) On estimating thresholds in autoregressive models. *Journal of Time Series Analysis* 7, 179–190.
Chen, X. (2007) Large sample sieve estimation of semi-nonparametric models. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 6B, pp. 2549–2632. Elsevier.
Chen, X., H. Hong, & E. Tamer (2005) Measurement error models with auxiliary data. *Review of Economic Studies* 72, 343–366.
Chen, X. & X. Shen (1998) Sieve extremum estimators for weakly dependent data. *Econometrica* 66, 289–314.
Chen. R. & R.S. Tsay (1993) Nonlinear additive ARX models. *Journal of American Statistical Association* 88, 955–967.
Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.
Davydov, Y. (1973) Mixing conditions for Markov chains. *Theory of Probability and Its Applications* 18, 312–328.
Doukhan, P. (1994) *Mixing: Properties and Examples*. Springer-Verlag.
Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag.
Gallant, A.R. & D.W. Nychka (1987) Semi-nonparametric maximum likelihood estimation. *Econometrica* 55, 363–390.
Hahn, J. & G. Kuersteiner (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects. *Econometrica* 70, 1639–1657.
Hahn, J. & G. Kuersteiner (2011) Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27, 1152–1191.
Hahn, J. & W. Newey (2004) Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72, 1295–1319.
Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. Academic Press.
Henderson, D.J., R.J. Carroll, & Q. Li (2008) Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics* 144, 257–275.
Henderson, D.J. & A. Ullah (2005) A nonparametric random effects estimator. *Economics Letters* 88, 403–407.
Lee, Y. (2006) General approaches to dynamic panel modelling and bias correction. Ph.D. dissertation, Yale University.
Lee, Y. (2012) Bias in dynamic panel models under time series misspecification. *Journal of Econometrics* 169, 54–60.
Lee, Y. & D. Mukherjee (2012) New Nonparametric Estimation of the Marginal Effects in Fixed-Effects Panel Models: An Application on the Environmental Kuznets Curve. Working paper.
Lee, Y. & P.C.B. Phillips (2014) Model Selection in the Presence of Incidental Parameters. *Journal of Econometrics, forthcoming*.
Lewis, R. & G.C. Reinsel (1985) Prediction of multivariate time-series by autoregressive model-fitting. *Journal of Multivariate Analysis* 16, 393–411.
Li, Q. & T.J. Kniesner (2002) Nonlinearity in dynamic adjustment: Semiparametric estimation of panel labor supply. *Empirical Economics* 27, 131–148.
Li, Q. & T. Stengos (1996) Semiparametric estimation of partially linear panel data models. *Journal of Econometrics* 71, 389–397.

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

Liebscher, E. (2005) Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis* 26, 669–689.
Meyn, S.P. & R.L. Tweedie (1993) *Markov Chains and Stochastic Stability*. Springer-Verlag.
Newey, W.K. (1997) Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147–168.
Newey, W.K. & J.L. Powell (2003) Instrumental variable estimation of nonparametric models. *Econometrica* 71, 1565–1578.
Newey, W.K. & K.D. West (1987) A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
Nickell, S. (1981) Biases in dynamic models with fixed effects. *Econometrica* 49, 1417–1425.
Nummelin, E. (1984) *General Irreducible Markov Chains and Nonnegative Operators*. Cambridge University Press.
Park, B., R.C. Sickles, & L. Simar (2007) Semiparametric efficient estimation of dynamic panel data models. *Journal of Econometrics* 136, 281–301.
Phillips, P.C.B. & H.R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.
Phillips, P.C.B. & D. Sul (2003) Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal* 6, 217–259.
Porter, J.R. (1996) Essays in econometrics. Ph.D. dissertation, MIT.
Prakasa Rao, B.L.S. (2009) Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics* 61, 441–460.
Stone, C.J. (1982) Optimal global rates of convergence for nonparametric regression. *Annals of Statistics* 10, 1040–1053.
Su, L & S. Jin (2012) Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics* 169, 34–47.
Su, L. & A. Ullah (2006) Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters* 92, 75–81.
Tjøstheim, D. (1990) Non-linear time series and Markov chains. *Advances in Applied Probability* 22, 587–611.
Tong, H. (1990) *Non-linear Time Series: A Dynamic System Approach*. Oxford University Press.
Tweedie, R.L. (1983) Criteria for rates of convergence of Markov chains, with application to queueing theory. In J.F.C. Kingman & G.E.H. Reuter (eds.), *Papers in Probability, Statistics and Analysis*, pp. 260–276. Cambridge University Press.
White, H. & I. Domowitz (1984) Nonlinear regression with dependent observations. *Econometrica* 52, 143–161.

APPENDIX: Mathematical Proofs

A.1 Useful Lemmas

Throughout the Appendix, we let C denote a generic positive constant that may be different in different uses. We let $\|\cdot\|$ be the Euclidean norm such that $\|\ell\| = \left(\ell'\ell\right)^{1/2}$ for a vector ℓ and $\|B\| = \left(tr\left(B'B\right)\right)^{1/2}$ for a matrix B, where $tr\left(\cdot\right)$ is the trace operator. We also introduce a matrix norm $\|B\|_1 = \sup_{\ell:\|\ell\|\leq 1}\|B\ell\|$. When B is symmetric and positive definite, $\|B\|_1$ is the largest eigenvalue of B. We denote $\underline{g}_K\left(y_{i,t}\right) = g_K\left(y_{i,t}\right) - \mathbb{E}\left(g_K\left(y_{i,t}\right)|\mu_i\right)$; the conditional covariance and the conditional expectation as $cov_i(\cdot,\cdot) = cov(\cdot,\cdot|\mu_i)$ and $\mathbb{E}_i(\cdot) = \mathbb{E}(\cdot|\mu_i)$, respectively. Lemma A1.1 provides the convergence rate of the denominator part of $\widehat{\theta}_K$ in (6).

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

LEMMA A1.1. *Under Assumptions 1–3 and 7,* $||(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\underline{g}_K(y_{i,t-1})\underline{g}_K(y_{i,t-1})' - \Gamma_K||^2 = O_p(K^2/NT)$ *and* $||(NT^2)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\underline{g}_K(y_{i,t-1})\underline{g}_K(y_{i,s-1})'||^2 = O_p(K^2/NT^2)$ *for large* N *and* T, *where* $\Gamma_K = \mathbb{E}\left(\underline{g}_K(y_{i,t})\underline{g}_K(y_{i,t})'\right)$.

Proof of Lemma A1.1. First, we have

$$
\begin{aligned}
&\mathbb{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\underline{g}_K(y_{i,t-1})\underline{g}_K(y_{i,t-1})' - \Gamma_K\right\|^2 \\
&\quad= \sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\underline{g}_{Kj}(y_{i,t-1})\underline{g}_{Kk}(y_{i,t-1}) - \Gamma_{K,jk}\right\}\right)^2 \\
&\quad= \frac{1}{NT}\sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\left(\frac{1}{T}\sum_{t=1}^{T}\left\{\underline{g}_{Kj}(y_{i,t-1})\underline{g}_{Kk}(y_{i,t-1}) - \Gamma_{K,jk}\right\}\right)^2\right] \\
&\quad= \frac{1}{NT}\sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}\left[\underline{g}_{Kj}(y_{i,0})\underline{g}_{Kk}(y_{i,0}) - \Gamma_{K,jk}\right]^2 + \frac{2}{NT}\sum_{j=1}^{K}\sum_{k=1}^{K} \\
&\qquad\times\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{\tau=1}^{T-1}\left(1-\frac{\tau}{T}\right)cov_i\left(\underline{g}_{Kj}(y_{i,0})\underline{g}_{Kk}(y_{i,0}), \underline{g}_{Kj}(y_{i,\tau})\underline{g}_{Kk}(y_{i,\tau})\right)\right] \\
&\quad\equiv A_{11}(N,T,K) + A_{12}(N,T,K)
\end{aligned}
$$

by the fourth order stationarity, where $\Gamma_{K,jk} = \mathbb{E}(\underline{g}_{Kj}(y_{i,t})\underline{g}_{Kk}(y_{i,t}))$ is the (j,k)th element of the $K \times K$ matrix Γ_K. It can be shown that

$$
\begin{aligned}
A_{11}(N,T,K) &\le \frac{1}{NT}\sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}\left(\underline{g}_{Kj}^2(y_{i,0})\underline{g}_{Kk}^2(y_{i,0})\right) \\
&\le \frac{K^2}{NT}\mathbb{E}\left(\sup_{1\le k\le K}\mathbb{E}_i\left(\underline{g}_{Kk}^4(y_{i,0})\right)\right) \le \frac{\zeta K^2}{NT}
\end{aligned}
$$

by Cauchy–Schwarz and Assumption 7. For $A_{12}(N,T,K)$, since β-mixing implies α-mixing (e.g., Doukhan, 1994, p. 4, Prop. 1), Proposition 2.1 implies that $\{y_{i,t}\}$ is an α-mixing process with exponential decay for each i given μ_i.[8] Furthermore, since $g_{Kk}(\cdot)$ are Borel measurable, $\{g_{Kk}(y_{i,t})\}$ is also α-mixing of the same size (i.e., whose mixing coefficient is of the same order of magnitude as that of $\{y_{i,t}\}$) for all $k = 1, \cdots, K$ (e.g., White and Domowitz, 1984, Lem. 2.1). Therefore, using the Davydov α-mixing inequality (e.g., Hall and Heyde, 1980, Cor. A.2) and under Assumptions 1–3 and 7, we obtain

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

$$\begin{aligned}
&\left|cov_i\left(\underline{g}_{Kj}\left(y_{i,0}\right)\underline{g}_{Kk}\left(y_{i,0}\right),\underline{g}_{Kj}\left(y_{i,\tau}\right)\underline{g}_{Kk}\left(y_{i,\tau}\right)\right)\right| \\
&\quad\le 8\alpha_i\left(\tau\right)^{1-2/p}\left(\mathbb{E}_i\left|\underline{g}_{Kj}\left(y_{i,0}\right)\underline{g}_{Kk}\left(y_{i,0}\right)\right|^p\right)^{2/p}\ \text{a.s.} \\
&\quad\le 8\alpha_i\left(\tau\right)^{1-2/p}\left(\mathbb{E}_i\left|\underline{g}_{Kk}\left(y_{i,0}\right)\right|^{2p}\right)^{2/p}\ \text{a.s.} \\
&\quad\le 8\alpha_i\left(\tau\right)^{1-2/p}\zeta^{2/p}\ \text{a.s.} \qquad \textbf{(A.1)}
\end{aligned}$$

for some $p>2$ since $\sup_{1\le i\le N}\sup_{1\le k\le K}\mathbb{E}_i|\underline{g}_{Kk}\left(y_{i,0}\right)|^{2p}\le\zeta$ a.s., where the second inequality is by Cauchy–Schwarz. It follows that

$$\begin{aligned}
&\left|\sum_{\tau=1}^{T-1}\left(1-\frac{\tau}{T}\right)cov_i\left(\underline{g}_{Kj}\left(y_{i,0}\right)\underline{g}_{Kk}\left(y_{i,0}\right),\underline{g}_{Kj}\left(y_{i,\tau}\right)\underline{g}_{Kk}\left(y_{i,\tau}\right)\right)\right| \\
&\quad\le 8\zeta^{2/p}\sum_{\tau=1}^{T-1}\left(1-\frac{\tau}{T}\right)\alpha_i\left(\tau\right)^{1-2/p}\ \text{a.s.},
\end{aligned}$$

where $\sum_{\tau=1}^{T-1}\left(1-\left(\tau/T\right)\right)\alpha_i\left(\tau\right)^{1-2/p}\to\sum_{\tau=1}^{\infty}\alpha_i\left(\tau\right)^{1-2/p}\le\left(1/2\right)\sum_{\tau=1}^{\infty}\beta_i\left(\tau\right)^{1-4/2p}<\infty$ a.s. as $T\to\infty$ using the Kronecker lemma, and from Proposition 1 of Doukhan (1994, p. 4) and Assumption 3. Therefore,

$$\left|A_{12}\left(N,T,K\right)\right|\le\frac{8\zeta^{2/p}K^2}{NT}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{\tau=1}^{\infty}\beta_i\left(\tau\right)^{1-2/p}\right]=O\left(\frac{K^2}{NT}\right),$$

which yields the first result.

For the second result, since $\|X\|^2\le 2\|X-Z\|^2+2\|Z\|^2$, we have

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\underline{g}_K\left(y_{i,t-1}\right)\sum_{s=1}^{T}\underline{g}_K\left(y_{i,s-1}\right)'\right\|^2 \\
&\quad=\frac{1}{NT^4}\sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\underline{g}_{Kj}\left(y_{i,t-1}\right)\underline{g}_{Kk}\left(y_{i,s-1}\right)\right\|^2\right]\le\frac{2}{NT^4}\sum_{j=1}^{K}\sum_{k=1}^{K} \\
&\qquad\times\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\underline{g}_{Kj}\left(y_{i,t-1}\right)\underline{g}_{Kk}\left(y_{i,s-1}\right)-\mathbb{E}_i\left(\underline{g}_{Kj}\left(y_{i,t-1}\right)\underline{g}_{Kk}\left(y_{i,s-1}\right)\right)\right]\right\|^2\right] \\
&\qquad+\frac{2}{NT^4}\sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_i\left(\underline{g}_{Kj}\left(y_{i,t-1}\right)\underline{g}_{Kk}\left(y_{i,s-1}\right)\right)\right\|^2\right] \\
&\quad\equiv A_{21}\left(N,T,K\right)+A_{22}\left(N,T,K\right).
\end{aligned}$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

Note that

$$A_{22}(N,T,K)$$
$$= \frac{2}{NT^4}\sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{E}_i\left[\sum_{t=1}^{T}\underline{g}_{Kj}(y_{i,t-1})\sum_{s=1}^{T}\underline{g}_{Kk}(y_{i,s-1})\right]\right)^2\right]$$
$$= \frac{2}{NT^4}\sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left(T\mathbb{E}_i\left(\underline{g}_{Kj}^2(y_{i,0})\right)\right.\right.$$
$$\left.\left.+T\sum_{\tau=1}^{T-1}\left(1-\frac{\tau}{T}\right)cov_i\left(\underline{g}_{Kj}(y_{i,0}),\underline{g}_{Kj}(y_{i,\tau})\right)\right)^2\right],$$

but since $\mathbb{E}_i(\underline{g}_{Kj}^2(y_{i,0})) \le [\sup_i \sup_k \mathbb{E}_i(\underline{g}_{Kj}^4(y_{i,0}))]^{1/2} \le \zeta^{1/2}$ by Jensen's and $|cov_i(\underline{g}_{Kj}(y_{i,0}),\underline{g}_{Kj}(y_{i,\tau}))| \le 8\alpha_i(\tau)^{1-2/p}(\sup_i \sup_k \mathbb{E}_i|\underline{g}_{Kj}(y_{i,0})|^{2p})^{1/p} \le 8\alpha_i(\tau)^{1-2/p}\zeta^{1/2}$, it follows that $|A_{22}(N,T,K)| = O(K^2/NT^2)$ similarly as $A_{12}(N,T,K)$ above. For $A_{21}(N,T,K)$, by letting $V_{j,t} = \underline{g}_{Kj}(y_{i,t})$, we note that

$$|A_{21}(N,T,K)| \le \frac{2}{NT^4}\sum_{j=1}^{K}\sum_{k=1}^{K}$$
$$\times\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\left|\mathbb{E}_i V_{j,0}V_{k,t-1}V_{j,s-1}V_{k,r-1} - \mathbb{E}_i V_{j,0}V_{k,t-1}\mathbb{E}_i V_{j,s-1}V_{k,r-1}\right|\right],$$

where since $V_{j,t}$ is assumed to be fourth order stationary and $\mathbb{E}_i V_{j,t} = 0$ by construction,

$$\Pi_i \equiv \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\left|\mathbb{E}_i V_{j,0}V_{k,t-1}V_{j,s-1}V_{k,r-1} - \mathbb{E}_i V_{j,0}V_{k,t-1}\mathbb{E}_i V_{j,s-1}V_{k,r-1}\right|$$
$$\le 3!\sum_{t=1}^{T}\sum_{s=t}^{T}\sum_{r=s}^{T}\left|\mathbb{E}_i V_{j,0}V_{k,t-1}V_{j,s-1}V_{k,r-1} - \mathbb{E}_i V_{j,0}V_{k,t-1}\mathbb{E}_i V_{j,s-1}V_{k,r-1}\right|$$
$$\le 6\sum_{t=1}^{T}\sum_{0\le s,r\le t}\left\{\left|\mathbb{E}_i V_{j,0}(V_{k,t}V_{j,t+s}V_{k,t+s+r})\right| + \left|\mathbb{E}_i\left(V_{j,0}V_{k,t}\right)\mathbb{E}_i\left(V_{j,t+s}V_{k,t+s+r}\right)\right|\right\}$$
$$+6\sum_{s=1}^{T}\sum_{0\le t,r\le s}\left|\mathbb{E}_i\left(V_{j,0}V_{k,t}\right)\left(V_{j,t+s}V_{k,t+s+r}\right) - \mathbb{E}_i(V_{j,0}V_{k,t})\mathbb{E}_i(V_{j,t+s}V_{k,t+s+r})\right|$$
$$+6\sum_{r=1}^{T}\sum_{0\le t,s\le r}\left\{\left|\mathbb{E}_i\left(V_{j,0}V_{k,t}V_{j,t+s}\right)V_{k,t+s+r}\right|\right.$$
$$\left.+\left|\mathbb{E}_i(V_{j,0}V_{k,t})\mathbb{E}_i(V_{j,t+s}V_{k,t+s+r})\right|\right\}$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

$$
\begin{aligned}
&= 6\sum_{t=1}^{T}\sum_{0\le s,r\le t}\{\left|cov_i\left(V_{j,0},V_{k,t}V_{j,t+s}V_{k,t+s+r}\right)\right| \\
&\qquad + \left|cov_i\left(V_{j,0},V_{k,t}\right)\right|\left|\mathbb{E}_i\left(V_{j,t+s}V_{k,t+s+r}\right)\right|\} \\
&\quad + 6\sum_{s=1}^{T}\sum_{0\le t,r\le s}\left|cov_i\left(V_{j,0}V_{k,t},V_{j,t+s}V_{k,t+s+r}\right)\right| \\
&\quad + 6\sum_{r=1}^{T}\sum_{0\le t,s\le r}\{\left|cov_i\left(V_{j,0}V_{k,t}V_{j,t+s},V_{k,t+s+r}\right)\right| \\
&\qquad + \left|\mathbb{E}_i\left(V_{j,0}V_{k,t}\right)\right|\left|cov_i\left(V_{j,t+s},V_{k,t+s+r}\right)\right|\}
\end{aligned}
$$

similarly as Andrews (1991b, Lem. 1). Using the mixing inequality referred above, however, we have $\left|cov_i\left(V_{j,0},V_{k,t}\right)\right| \le 8\alpha_i(t)^{1-2/p}(\mathbb{E}_i|V_{j,0}|^p\mathbb{E}_i|V_{k,0}|^p)^{1/p} \le 8\alpha_i(t)^{1-2/p}\zeta^{1/2}$ a.s. and $\left|cov_i\left(V_{j,t+s}V_{k,t+s+r}\right)\right| \le 8\alpha_i(r)^{1-2/p}\zeta^{1/2}$ a.s. for some $p > 2$. Similarly, we can also derive that $\left|cov_i\left(V_{j,0},V_{k,t}V_{j,t+s}V_{k,t+s+r}\right)\right| \le 8\alpha_i(t)^{1-4/\nu}(\mathbb{E}_i|V_{j,0}|^\nu\mathbb{E}_i|V_{k,0}|^\nu)^{2/\nu} \le 8\alpha_i(t)^{1-4/\nu}\zeta$ a.s., $\left|cov_i\left(V_{j,0}V_{k,t},V_{j,t+s}V_{k,t+s+r}\right)\right| \le 8\alpha_i(s)^{1-4/\nu}\zeta$ a.s. and $\left|cov_i\left(V_{j,0}V_{k,t}V_{j,t+s},V_{k,t+s+r}\right)\right| \le 8\alpha_i(r)^{1-4/\nu}\zeta$ a.s. for some $\nu > 4$. In addition, by Cauchy–Schwarz, $\left|\mathbb{E}_i\left(V_{j,t+s}V_{k,t+s+r}\right)\right| \le [\mathbb{E}_i V_{j,t+s}^2\mathbb{E}_i V_{k,t+s+r}^2]^{1/2} \le \zeta^{1/2}$ a.s. Therefore,

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\Pi_i &\le 96\zeta(K)\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{0\le s,r\le t}\alpha_i(t)^{1-4/\nu} \\
&\quad + 48\zeta(K)\frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{0\le t,r\le s}\alpha_i(s)^{1-4/\nu} \\
&\quad + 96\zeta(K)\frac{1}{N}\sum_{i=1}^{N}\sum_{r=1}^{T}\sum_{0\le t,s\le r}\alpha_i(r)^{1-4/\nu} < \infty \qquad \textbf{(A.2)}
\end{aligned}
$$

since $(1/N)\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{0\le s,r\le t}\alpha_i(t)^{1-4/\nu} \le (T^2/N)\sum_{i=1}^{N}\sum_{t=1}^{T}\alpha_i(t)^{1-4/\nu} < \infty$ from Assumption 3 similarly as $A_{12}(N,T,K)$ above. It follows that $|A_{21}(N,T,K)| = O(K^2/NT^2)$, which yields the desired result. ■

Andrews (1991a) and Newey (1997) show that the variance estimation for linear functions of the series estimator is essentially the same as it is in least squares estimation for nonrandom K. We also estimate Γ_K by $\widehat{\Gamma}_K = (1/NT)\sum_{i=1}^{N}\sum_{t=1}^{T} g_K^0\left(y_{i,t}\right) g_K^0\left(y_{i,t}\right)'$ for every K. Note that Assumption 7-(i) ensures that $\widehat{\Gamma}_K$ is nonsingular for large N and T. In what follows, therefore, we simply assume that $\widehat{\Gamma}_K$ is invertible.[9]

LEMMA A1.2. *Under Assumptions 1–3 and 7, for large N and T,*

(1) $||\Gamma_K||_1||\Gamma_K^{-1}||_1, ||\Gamma_K^{1/2}||_1, ||\Gamma_K^{-1/2}||_1$ *are all uniformly bounded over K;*
(2) $||\widehat{\Gamma}_K - \Gamma_K|| = O_p(K/\sqrt{NT})$ *and* $||\widehat{\Gamma}_K - \Gamma_K||_1$, $||\widehat{\Gamma}_K^{1/2} - \Gamma_K^{1/2}||$, $||\widehat{\Gamma}_K^{1/2} - \Gamma_K^{1/2}||_1$ *are all* $o_p(1)$*;*

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

(3) $||\widehat{\Gamma}_K^{-1} - \Gamma_K^{-1}||_1 = O_p(K/\sqrt{NT})$;
(4) $||\widehat{\Gamma}_K||_1$, $||\widehat{\Gamma}_K^{-1}||_1$, $||\widehat{\Gamma}_K^{1/2}||_1$, $||\widehat{\Gamma}_K^{-1/2}||_1$ *are all* $O_p(1)$.

Proof of Lemma A1.2.

(1) For any symmetric matrix A, we let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ be the minimum and the maximum eigenvalues of A, respectively. Then, when A is positive definite, $\|A\|_1 = \lambda_{\max}(A)$. Since Γ_K is positive definite and $0 < \lambda_{\min}(\Gamma_K) \leq \lambda_{\max}(\Gamma_K) < \infty$ uniformly over K from Assumption 7-(ii), we get $||\Gamma_K||_1 = \lambda_{\max}(\Gamma_K) < \infty$, $||\Gamma_K^{1/2}||_1 = \lambda_{\max}(\Gamma_K^{1/2}) = [\lambda_{\max}(\Gamma_K)]^{1/2} < \infty$, $||\Gamma_K^{-1}||_1 = \lambda_{\max}\left(\Gamma_K^{-1}\right) = [\lambda_{\min}(\Gamma_K)]^{-1} < \infty$, and $||\Gamma_K^{-1/2}||_1 = \lambda_{\max}(\Gamma_K^{-1/2}) = [\lambda_{\min}(\Gamma_K)]^{-1/2} < \infty$ for all K.

(2) We can write

$$\sum_{i=1}^{N}\sum_{t=1}^{T} g_K^0\left(y_{i,t-1}\right) g_K^0\left(y_{i,t-1}\right)' = \sum_{i=1}^{N}\sum_{t=1}^{T} \underline{g}_K\left(y_{i,t-1}\right) \underline{g}_K\left(y_{i,t-1}\right)' - \frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} \underline{g}_K\left(y_{i,t-1}\right)\sum_{s=1}^{T} \underline{g}_K\left(y_{i,s-1}\right)',$$

then the first result is easily derived from Lemma A1.1. The second result is from $K/\sqrt{NT} \to 0$ and that $\|A\|_1 = \lambda_{\max}(A) \leq (\sum_{k=1}^{K} \lambda_k^2)^{1/2} = \left(tr\left(A'A\right)\right)^{1/2} = \|A\|$, where λ_ks are the eigenvalues of a $K \times K$ positive definite matrix A. Finally, the third and the fourth results follow straightforward since $f(x) = x^{1/2}$ is a continuous mapping everywhere (see Davidson, 1994, Thm. 18.9).

(3) Similarly as Lewis and Reinsel (1985, Thm. 1), we write $\widehat{\Gamma}_K^{-1} - \Gamma_K^{-1} = \left\{\Gamma_K^{-1} + \left(\widehat{\Gamma}_K^{-1} - \Gamma_K^{-1}\right)\right\}\left(\widehat{\Gamma}_K - \Gamma_K\right)\Gamma_K^{-1}$ and thus $||\widehat{\Gamma}_K^{-1} - \Gamma_K^{-1}||_1 \leq \left(||\Gamma_K^{-1}||_1 + ||\widehat{\Gamma}_K^{-1} - \Gamma_K^{-1}||_1\right)||\widehat{\Gamma}_K - \Gamma_K||_1||\Gamma_K^{-1}||_1$ or

$$||\widehat{\Gamma}_K^{-1} - \Gamma_K^{-1}||_1 \left\{1 - ||\widehat{\Gamma}_K - \Gamma_K||_1||\Gamma_K^{-1}||_1\right\} \leq ||\Gamma_K^{-1}||_1^2||\widehat{\Gamma}_K - \Gamma_K||_1$$

with probability approaching one (w.p.a.1). But note that $|||\widehat{\Gamma}_K - \Gamma_K||_1||\Gamma_K^{-1}||_1 = ||\widehat{\Gamma}_K\Gamma_K^{-1} - I_K||_1 = \lambda_{\min}\left(\widehat{\Gamma}_K\Gamma_K^{-1} - I_K\right) < 1$ w.p.a.1 since $||\widehat{\Gamma}_K\Gamma_K^{-1} - I_K||_1 \to_p 0$. Therefore, $1/\{1 - ||\widehat{\Gamma}_K - \Gamma_K||_1||\Gamma_K^{-1}||_1\}$ is bounded and thus

$$||\widehat{\Gamma}_K^{-1} - \Gamma_K^{-1}||_1 \leq \frac{||\Gamma_K^{-1}||_1^2||\widehat{\Gamma}_K - \Gamma_K||_1}{1 - ||\widehat{\Gamma}_K - \Gamma_K||_1||\Gamma_K^{-1}||_1} = O_p\left(K/\sqrt{NT}\right)$$

from the results in parts (1) and (2) above.

(4) We first write that $||\widehat{\Gamma}_K||_1 \leq ||\widehat{\Gamma}_K - \Gamma_K||_1 + ||\Gamma_K||_1$ and the result follows using parts (1) and (2). The similar argument can be applied for other results. ■

The following lemmas are useful to obtain the convergence rate of the numerator part of $\widehat{\theta}_K$.

LEMMA A1.3. *Under Assumptions 1–4 and 7,* $||(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} \underline{g}_K\left(y_{i,t-1}\right) u_{i,t}||^2 = O_p(K/NT)$, $||(NT^2)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \underline{g}_K\left(y_{i,t-1}\right) u_{i,s}||^2 = O_p(K/T^2)$, *and* $||(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{i,t}^2 \underline{g}_K\left(y_{i,t-1}\right) \underline{g}_K\left(y_{i,t-1}\right)' - \sigma^2\Gamma_K|| = o_p(1)$ *for large* N *and* T.

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

Proof of Lemma A1.3. Since $\mathbb{E}_i[\underline{g}_{Kk}(y_{i,t-1})u_{i,t}] = 0$, we note that $T^{-1}\mathbb{E}_i(\sum_{t=1}^{T}\underline{g}_{Kk}(y_{i,t-1})u_{i,t})^2 \leq \mathbb{E}_i[\underline{g}_{Kk}(y_{i,0})u_{i,1}]^2 + 2\sum_{\tau=1}^{T-1}(1-\tau/T)|cov_i(\underline{g}_{Kk}(y_{i,0})u_{i,1}, \underline{g}_{Kk}(y_{i,\tau})u_{i,1+\tau})|$ by stationarity. The first term is simply $O(1)$ since $\mathbb{E}_i[\underline{g}_{Kk}(y_{i,0})u_{i,1}]^2 = \mathbb{E}_i(\underline{g}^2_{Kk}(y_{i,0}))\mathbb{E}_i(u^2_{i,1}) \leq \sigma^2\zeta^{1/2}$ from Assumptions 1, 2, and 7, where $\mathbb{E}_i(u^2_{i,1}) = \mathbb{E}(u^2_{i,1}) = \sigma^2$ for $u_{i,t}$ being independent of μ_i. For the second term, since $\{g_{Kk}(y_{i,t})\}$ is α-mixing for any k and $\{u_{i,t}\}$ is *i.i.d.*, $\{g_{Kk}(y_{i,t-1})u_{i,t}\}$ is at most α-mixing with the same mixing coefficient since $g_k(y_{i,t-1})$ and $u_{i,t}$ are mutually independent for all i and t. In this case, we simply need to consider richer σ-fields $\mathcal{G}^t_{i,-\infty}$ and $\mathcal{G}^{\infty}_{i,t+\tau}$ for $\tau > 0$, which are generated by $\{y_{i,s-1}\}_{s\leq t} \cup \{u_{i,s}\}_{s\leq t}$ and $\{y_{i,s-1}\}_{s\geq t+\tau+1} \cup \{u_{i,s}\}_{s\geq t+\tau+1}$, respectively, for given i. Moreover, for some $p > 2$, $\mathbb{E}_i[|\underline{g}_{Kk}(y_{i,t-1})u_{i,t}|^p] \leq \zeta^{1/2}\mathbb{E}_i[|u_{i,t}|^p] = \zeta^{1/2}\mathbb{E}|u_{i,t}|^p$ since $u_{i,t}$ is independent of μ_i. Similarly as the proof of Lemma A1.1, we thus have $|cov_i(\underline{g}_{Kk}(y_{i,0})u_{i,1}, \underline{g}_{Kk}(y_{i,\tau})u_{i,1+\tau})| \leq 8\alpha_i(\tau)^{1-2/p}\left(\zeta^{1/2}\mathbb{E}|u_{i,t}|^p\right)^{2/p}$ a.s. Therefore,

$$\mathbb{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\underline{g}_K(y_{i,t-1})u_{i,t}\right\|^2$$

$$= \frac{1}{NT^2}\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\left(\sum_{t=1}^{T}\underline{g}_{Kk}(y_{i,t-1})u_{i,t}\right)^2\right]$$

$$\leq \frac{K}{NT}\left\{\sigma^2\zeta^{1/2} + \frac{8}{N}\sum_{i=1}^{N}\sum_{\tau=1}^{\infty}\alpha_i(\tau)^{1-2/p}\left(\zeta^{1/2}\mathbb{E}|u_{i,t}|^p\right)^{2/p}\right\} = O\left(\frac{K}{NT}\right)$$

since $(\mathbb{E}|u_{i,t}|^p)^2 \leq \mathbb{E}|u_{i,t}|^{2p} < \infty$ for $p > 2$ by Assumption 1-(i).

For the second result, we note that since $u_{i,t}$ is *i.i.d.* over i and t and independent of μ_i, $\mathbb{E}_i(\sum_{t=1}^{T}u_{i,s})^4 = \sum_{t=1}^{T}\mathbb{E}(u^4_{1,t}) + \sum_{t=1}^{T}\sum_{s\neq t}\mathbb{E}(u^2_{1,t}u^2_{1,s}) = O(T^2)$. Moreover, using the second results of Lemma A1.1 above, we can obtain that

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\left(\sum_{t=1}^{T}\underline{g}_{Kk}(y_{i,t-1})\right)^4$$

$$\leq \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\left|cov_i\left(\underline{g}_{Kk}(y_{i,t-1})\underline{g}_{Kk}(y_{i,s-1}), \underline{g}_{Kk}(y_{i,r-1})\underline{g}_{Kk}(y_{i,q-1})\right)\right|$$

$$+ \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_i\left(\underline{g}_{Kk}(y_{i,t-1})\underline{g}_{Kk}(y_{i,s-1})\right)\right)^2 = O\left(T^2\right).$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

Therefore, it immediately follows that

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\underline{g}_K\left(y_{i,t-1}\right)u_{i,s}\right\|^2 \\
&\quad\le \frac{1}{NT^4}\mathbb{E}\left[\sum_{i=1}^{N}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\underline{g}_K\left(y_{i,t-1}\right)u_{i,s}\right\|^2\right] \\
&\quad= \frac{1}{T^4}\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\left(\sum_{t=1}^{T}\underline{g}_{Kk}\left(y_{i,t-1}\right)\sum_{s=1}^{T}u_{i,s}\right)^2\right] \\
&\quad\le \frac{1}{T^4}\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left\{\mathbb{E}_i\left(\sum_{t=1}^{T}\underline{g}_{Kk}\left(y_{i,t-1}\right)\right)^4\right\}^{1/2}\left\{\mathbb{E}_i\left(\sum_{s=1}^{T}u_{1,s}\right)^4\right\}^{1/2}\right] \\
&\quad\le \frac{1}{NT^4}\sum_{k=1}^{K}\mathbb{E}\left[\left\{\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\left(\sum_{t=1}^{T}\underline{g}_{Kk}\left(y_{i,t-1}\right)\right)^4\right\}^{1/2}\left\{\mathbb{E}\left(\sum_{s=1}^{T}u_{1,s}\right)^4\right\}^{1/2}\right] \\
&\quad= O\left(\frac{K}{T^2}\right),
\end{aligned}$$

where the inequalities are by Lyapunov's, Jensen's, and Cauchy–Schwarz inequalities in turn.

The last result can be readily derived using Lemma A1.1 by writing

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{i,t}^2\underline{g}_K\left(y_{i,t-1}\right)\underline{g}_K\left(y_{i,t-1}\right)'-\sigma^2\Gamma_K\right\|^2 \\
&\quad= \frac{1}{NT}\sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\left(u_{i,t}^2\underline{g}_{Kj}\left(y_{i,0}\right)\underline{g}_{Kk}\left(y_{i,0}\right)-\sigma^2\Gamma_{K,jk}\right)^2\right] \\
&\qquad+\frac{2}{NT}\sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{\tau=1}^{T-1}\left(1-\frac{\tau}{T}\right)\right. \\
&\qquad\qquad\left.\times\ cov_i\left(u_{i,1}^2\underline{g}_{Kj}\left(y_{i,0}\right)\underline{g}_{Kk}\left(y_{i,0}\right),u_{i,1+\tau}^2\underline{g}_{Kj}\left(y_{i,\tau}\right)\underline{g}_{Kk}\left(y_{i,\tau}\right)\right)\right].
\end{aligned}$$

Note that, from Assumption A1-(i), we let a constant $C>0$ such that $\mathbb{E}(u_{i,t}^{\delta})<C<\infty$ for $\delta>4$. Then $\mathbb{E}_i[u_{i,t}^2\underline{g}_{Kj}\left(y_{i,0}\right)\underline{g}_{Kk}\left(y_{i,0}\right)-\sigma^2\Gamma_{K,jk}]^2\le\mathbb{E}_i[u_{i,t}^4\underline{g}_{Kj}^2\left(y_{i,0}\right)\underline{g}_{Kk}^2\left(y_{i,0}\right)]\le C\zeta<\infty$, and similarly as (A.1)

$$\begin{aligned}
&\left|cov_i\left(u_{i,1}^2\underline{g}_{Kj}\left(y_{i,0}\right)\underline{g}_{Kk}\left(y_{i,0}\right),u_{i,1+\tau}^2\underline{g}_{Kj}\left(y_{i,\tau}\right)\underline{g}_{Kk}\left(y_{i,\tau}\right)\right)\right| \\
&\quad\le 8\alpha_i\left(\tau\right)^{1-2/p}C^{2/p}\zeta^{2/p}\ \text{ a.s.}
\end{aligned}$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

for $p > 2$. It thus follows that

$$\mathbb{E}\left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{i,t}^2 \underline{g}_K\left(y_{i,t-1}\right)\underline{g}_K\left(y_{i,t-1}\right)' - \sigma^2\Gamma_K \right\|^2 = O\left(\frac{K^2}{NT}\right) \to 0.$$

■

LEMMA A1.4. *Under Assumptions 1–7, for large N and T,* $\|m\left(y_{i,t}\right) - g_K\left(y_{i,t}\right)'\theta_K\|^2 = O_p(K^{-2\delta})$ *and* $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\{m^0\left(y_{i,t-1}\right) - g_K^0\left(y_{i,t-1}\right)'\theta_K\}^2 = O_p(K^{-2\delta})$.

Proof of Lemma A1.4. First note that

$$\begin{aligned}
\mathbb{E}\left[m\left(y_{i,t}\right) - g_K\left(y_{i,t}\right)'\theta_K\right]^2 &= \mathbb{E}\left[\int\left[m\left(y\right) - g_K\left(y\right)'\theta_K\right]^2\pi_i\left(dy\right)\right] \\
&= \mathbb{E}\left[\int\left[\left(m\left(y\right) - g_K\left(y\right)'\theta_K\right)\left(1+\|y\|^2\right)^{-\omega/2}\right]^2\right. \\
&\qquad \left.\times\left(1+\|y\|^2\right)^{\omega}\pi_i\left(dy\right)\right] \\
&\leq \left\|m\left(y\right) - g_K\left(y\right)'\theta_K\right\|_{\infty,\omega}^2\mathbb{E}\left[\int\left(1+\|y\|^2\right)^{\omega}\pi_i\left(dy\right)\right] \\
&= O(K^{-2\delta}) \qquad \textbf{(A.3)}
\end{aligned}$$

from Assumption 6. Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{m^0\left(y_{i,t-1}\right) - g_K^0\left(y_{i,t-1}\right)'\theta_K\}^2\right] \\
&\quad= \mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{m\left(y_{i,t-1}\right) - g_K\left(y_{i,t-1}\right)'\theta_K\right\}^2\right] \\
&\qquad-\mathbb{E}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{m\left(y_{i,t-1}\right) - g_K\left(y_{i,t-1}\right)'\theta_K\right\}\right. \\
&\qquad\qquad\left.\times\left\{m\left(y_{i,s-1}\right) - g_K\left(y_{i,s-1}\right)'\theta_K\right\}\right] \\
&\quad\leq \mathbb{E}\left[m\left(y_{i,t}\right) - g_K\left(y_{i,t}\right)'\theta_K\right]^2 + \left[\mathbb{E}\left(m\left(y_{i,t}\right) - g_K\left(y_{i,t}\right)'\theta_K\right)^2\right. \\
&\qquad\qquad\left.\times\mathbb{E}\left(m\left(y_{i,t}\right) - g_K\left(y_{i,t}\right)'\theta_K\right)^2\right]^{1/2} \\
&\quad= O\left(K^{-2\delta}\right),
\end{aligned}$$

where the inequality is from the Cauchy–Schwarz. ■

The following two lemmas establish the building blocks for deriving asymptotic distribution of $\widehat{m}(y)$.

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

LEMMA A1.5. *Under Assumptions 1–4 and 7,* $(NT)^{-1/2}\rho'\Gamma_K^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\underline{g}_K\left(y_{i,t-1}\right)u_{i,t} \to_d \mathcal{N}\left(0,\sigma^2\right)$ *as* $N,T\to\infty$, *for some* $K\times 1$ *nonrandom vector* ρ *satisfying* $\|\rho\|=1$.

Proof of Lemma A1.5. We first define a random variable $Z_{i,t}=\rho'\Gamma_K^{-1/2}\underline{g}_K\left(y_{i,t-1}\right)u_{i,t}/\sigma$, which is independent across i. For each i, moreover, $\{Z_{i,t},\mathcal{F}_{i,t}\}$ is a martingale difference sequence with variance one, where $\mathcal{F}_{i,t}=\sigma(\{Z_{i,s}\}_{s\le t})$. Note that $y_{i,t}$ is a nonlinear function of $\{u_{i,s}\}_{s\le t}$ and $u_{i,t}$ is serially uncorrelated. Therefore, it holds that $\mathbb{E}(Z_{i,t}|\mathcal{F}_{i,t-1})=\rho'\Gamma_K^{-1/2}\underline{g}_K\left(y_{i,t-1}\right)\mathbb{E}(u_{i,t}|\mathcal{F}_{i,t-1})/\sigma=0$ and $\mathbb{E}(Z_{i,t}^2|\mathcal{F}_{i,t-1})=\rho'\Gamma_K^{-1/2}\underline{g}_K\left(y_{i,t-1}\right)\mathbb{E}(u_{i,t}^2|\mathcal{F}_{i,t-1})\underline{g}_K\left(y_{i,t-1}\right)\Gamma_K^{-1/2}\rho/\sigma^2=1$.

Directly applying the conventional Lindeberg condition to the double indexed process $Z_{i,t}$ is not straightforward. We instead apply Phillips and Moon (1999, Thm. 2) for limit theories with large N and T by checking the generalized Lindeberg condition for joint asymptotic normality. In this case, the generalized Lindeberg condition (Phillips and Moon, 1999, eqn. 3.20) is given as

$$\frac{1}{s_{NT}^2}\sum_{i=1}^{N}\mathbb{E}\left[\eta_{iT}^2\mathbf{1}\{|\eta_{iT}|>s_{NT}\epsilon\}\right]\to 0 \tag{A.4}$$

as $N,T\to\infty$ for every $\epsilon>0$, where $\eta_{iT}=T^{-1/2}\sum_{t=1}^{T}Z_{i,t}$ and $s_{NT}^2=var(\sum_{i=1}^{N}\eta_{iT})$. By Davidson (1994, p. 372, Thm. 23.10), however, a sufficient condition for the Lindeberg condition (A.4) is that η_{iT} is uniformly integrable and s_{NT}^2/N is bounded above zero. For the second condition, note that

$$\begin{aligned}s_{NT}^2=var\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{i,t}\right)&=\frac{N}{T}var\left(\sum_{t=1}^{T}Z_{i,t}\right)\\&=N\left\{\mathbb{E}\left(Z_{i,t}^2\right)+2\sum_{\ell=1}^{T-1}\left(1-\frac{\ell}{T}\right)cov\left(Z_{i,1},Z_{i,1+\ell}\right)\right\}=N\end{aligned}$$

since $Z_{i,t}$ is martingale difference with variance one, and thus $s_{NT}^2/N=1>0$. For the first condition, similarly, we have $\mathbb{E}\left(\eta_{iT}^2\right)=\mathbb{E}\left(T^{-1/2}\sum_{t=1}^{T}Z_{i,t}\right)^2=1$, which gives $\sup_{1\le i\le N}\mathbb{E}\left(\eta_{iT}^2\right)<\infty$. Therefore, (A.4) holds and we can conclude that $\left(1/\sqrt{NT}\right)\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{i,t}\to_d\mathcal{N}(0,1)$ as $N,T\to\infty$ by Phillips and Moon (1999). ■

The following lemmas provide consistency of the estimators of σ^2 and $\Phi_K=\sum_{j=0}^{\infty}cov\left(g_K\left(y_{i,t+j}\right),u_{i,t}\right)$, which justify the bias correction formula in Theorem 3.2.

LEMMA A1.6. *For each* K, *we let* $\widehat{\Phi}_K=\sum_{j=0}^{J}(N(T-j))^{-1}\sum_{i=1}^{N}w_j\sum_{t=1}^{T-j}g_K^0\left(y_{i,t+j}\right)\widehat{u}_{i,t}^0$, *where* $w_j=1-(j/(J+1))$ *and* $\widehat{u}_{i,t}^0=y_{i,t}^0-\widehat{\theta}_K'g_K^0\left(y_{i,t-1}\right)$. *We have* $||(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\underline{g}_K\left(y_{i,t-1}\right)u_{i,s}-\Phi_K||\to_p 0$ *as* $N,T\to\infty$ *jointly under Assumptions 1–4 and 7. Moreover, if* $J^3/T\to a<\infty$, *then* $||\widehat{\Phi}_K-\Phi_K||\to_p 0$ *as* $N,T\to\infty$ *jointly under Assumptions 1, 2, 7, 6, and 4.*

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

Proof of Lemma A1.6. We let a $K \times 1$ vector

$$\begin{aligned}\Psi_K(N,T) &= \mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\underline{g}_K(y_{i,t-1})u_{i,s}\right] \\ &= \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=2}^{T}\sum_{\tau=1}^{t-1}\mathbb{E}_i\left(\underline{g}_K\left(y_{i,t-\tau}\right)u_{i,1}\right)\right\}\right] \\ &= \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{\tau=1}^{T-1}\left(1-\frac{\tau}{T}\right)cov_i\left(g_K\left(y_{i,\tau}\right),u_{i,1}\right)\right].\end{aligned}$$

Then, for the first result, we have

$$\begin{aligned}&\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\underline{g}_K\left(y_{i,t-1}\right)u_{i,s}-\Phi_K\right\| \\ &\quad\le\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\underline{g}_K\left(y_{i,t-1}\right)u_{i,s}-\Psi_K(N,T)\right\|+\left\|\Psi_K(N,T)-\Phi_K\right\|\to_p 0\end{aligned}$$

as $N,T\to\infty$, where the first term converges in probability to zero by the definition of $\Psi_K(N,T)$ and the second term is negligible for large N and T by the Kronecker lemma since $\Phi_K=\sum_{\tau=1}^{\infty}cov\left(g_K\left(y_{i,\tau}\right),u_{i,1}\right)$.

For the second result, by triangular inequality, we write

$$\begin{aligned}&\left\|\widehat{\Phi}_K-\Phi_K\right\| \\ &\le\left\|\sum_{j=0}^{J}w_j\mathbb{E}[\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}]-\sum_{j=0}^{\infty}cov\left(g_K\left(y_{i,t+j}\right),u_{i,t}\right)\right\| \\ &\quad+\left\|\sum_{j=0}^{J}\frac{w_j}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\left(\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}-\mathbb{E}[\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}]\right)\right\| \qquad \textbf{(A.5)} \\ &\quad+\left\|\sum_{j=0}^{J}\frac{w_j}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\left(\underline{g}_K\left(y_{i,t+j}\right)-g_K^0\left(y_{i,t+j}\right)\right)u_{i,t}\right\| \qquad \textbf{(A.6)} \\ &\quad+\left\|\sum_{j=0}^{J}\frac{w_j}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}g_K^0\left(y_{i,t+j}\right)\left(\widehat{u}_{i,t}^0-u_{i,t}\right)\right\|, \qquad \textbf{(A.7)}\end{aligned}$$

where the first term converges to zero as $J\to\infty$ using Kronecker lemma. Note that $cov\left(g_K\left(y_{i,t+j}\right),u_{i,t}\right)=cov\left(\underline{g}_K\left(y_{i,t+j}\right),u_{i,t}\right)$. Therefore, we need to show that (A.5), (A.6), and (A.7) are all $o_p(1)$ for large N,T and J.

For (A.5), we first note that $\left|w_j\right|\le C_w$ for all $j=1,2,\cdots,J$ and for some constant $0<C_w<\infty$. Thus similarly as the proof of Theorem 2 in Newey and West (1987), for any $\varepsilon>0$, we have

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

$$\mathbb{P}\left(\left\|\sum_{j=0}^{J}\frac{w_j}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\left(\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}-\mathbb{E}\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}\right)\right\|>\varepsilon\right)$$
$$\leq \mathbb{P}\left(\sum_{j=0}^{J}|w_j|\left\|\frac{1}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\left(\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}-\mathbb{E}\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}\right)\right\|>\varepsilon\right)$$
$$\leq \sum_{j=0}^{J}\mathbb{P}\left(\left\|\frac{1}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\left(\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}-\mathbb{E}\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}\right)\right\|>\frac{\varepsilon}{C_w J}\right)$$
$$\leq \sum_{j=0}^{J}\left(\frac{C_w J}{\varepsilon}\right)^2\mathbb{E}\left\|\frac{1}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\left(\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}-\mathbb{E}\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}\right)\right\|^2$$
$$\leq \sum_{j=0}^{J}\left(\frac{C_w J}{\varepsilon}\right)^2\left(\frac{T}{T-j}\right)^2\mathbb{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\underline{g}_K\left(y_{i,t+j}\right)u_{i,t}\right\|^2=\frac{C_w^2J^3}{\varepsilon^2}\times O_p\left(\frac{K}{NT}\right),$$

where the third inequality is by Chebyshev's inequality and the last equality is from Lemma A1.3. The desired result follows since $J^3K/NT=(J^3/T)(K/\sqrt{NT})\sqrt{T/N}\to 0$ as $N,T,J\to\infty$. Similarly, for (A.6), we have $\underline{g}_K\left(y_{i,t+j}\right)-g_K^0\left(y_{i,t+j}\right)=(1/T)\sum_{s=1}^{T}\underline{g}_K\left(y_{i,s+j}\right)$ and thus for any $\varepsilon>0$

$$\mathbb{P}\left(\left\|\sum_{j=0}^{J}\frac{w_j}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\left(\underline{g}_K\left(y_{i,t+j}\right)-g_K^0\left(y_{i,t+j}\right)\right)u_{i,t}\right\|>\varepsilon\right)$$
$$=\mathbb{P}\left(\left\|\sum_{j=0}^{J}\frac{w_j}{NT(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\sum_{s=1}^{T}\underline{g}_K\left(y_{i,s+j}\right)u_{i,t}\right\|>\varepsilon\right)$$
$$\leq\sum_{j=0}^{J}\left(\frac{C_w J}{\varepsilon}\right)^2\left(\frac{T}{T-j}\right)^2\mathbb{E}\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\sum_{s=1}^{T}\underline{g}_K\left(y_{i,s+j}\right)u_{i,t}\right\|^2$$
$$=\frac{C_w^2J^3}{\varepsilon^2}\times O_p\left(\frac{K}{T^2}\right)\to 0$$

from Lemma A1.3 since $J^3K/T^2=(J^3/T)(K/\sqrt{NT})\sqrt{N/T}\to 0$ as $N,T,J\to\infty$. Finally, for (A.7), since $\widehat{u}_{i,t}^0-u_{i,t}=(m^0\left(y_{i,t-1}\right)-\widehat{m}^0\left(y_{i,t-1}\right))-(1/T)\sum_{s=1}^{T}u_{i,s}$, it can be verified that

$$\mathbb{P}\left(\left\|\sum_{j=0}^{J}\frac{w_j}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}g_K^0\left(y_{i,t+j}\right)\left(\widehat{u}_{i,t}^0-u_{i,t}\right)\right\|>\varepsilon\right)$$
$$\leq\mathbb{P}\left(\left\|\sum_{j=0}^{J}\frac{w_j}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\underline{g}_K\left(y_{i,t+j}\right)\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right\|>\frac{\varepsilon}{2}\right)\quad\textbf{(A.8)}$$
$$+\mathbb{P}\left(\left\|\sum_{j=0}^{J}\frac{w_j}{NT(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\sum_{s=1}^{T}\underline{g}_K\left(y_{i,s+j}\right)\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right\|>\frac{\varepsilon}{2}\right)\textbf{(A.9)}$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

for any $\varepsilon > 0$. Similarly as the previous cases, the first term (A.8) satisfies

$$\begin{aligned}
&\mathbb{P}\left(\left\|\sum_{j=0}^{J}\frac{w_j}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\underline{g}_K\left(y_{i,t+j}\right)\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right\| > \frac{\varepsilon}{2}\right)\\
&\quad\le \sum_{j=0}^{J}\mathbb{P}\left(\left\|\frac{1}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\underline{g}_K\left(y_{i,t+j}\right)\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right\| > \frac{\varepsilon}{2C_w J}\right)\\
&\quad\le \sum_{j=0}^{J}\left(\frac{2C_w J}{\varepsilon}\right)^2\left(\frac{T}{T-j}\right)^2\mathbb{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\underline{g}_K\left(y_{i,t+j}\right)\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right\|^2\\
&\quad= \frac{4C_w^2J^3}{\varepsilon^2}\times O_p\left(\frac{K}{N}\left(\frac{K}{NT}+K^{-2\delta}+\frac{K}{T^2}\right)\right)\to 0
\end{aligned}$$

since $J^3K^2/N^2T = (J^3/T)(K^2/NT)(T/N)\to 0$, $J^3K^{1-2\delta}/N = (J^3/T)(K/NT)(\sqrt{NT}K^{-\delta})^2(T/N)\to 0$, and $J^3K^2/NT^2 = (J^3/T)(K^2/NT)\to 0$. Note that for each j

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\underline{g}_K\left(y_{i,t+j}\right)\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right\|^2\\
&\quad= \frac{1}{NT^2}\sum_{k=1}^{K}\mathbb{E}\left\{\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\left[\sum_{t=1}^{T-j}\underline{g}_{Kk}\left(y_{i,t+j}\right)\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right]^2\right\}\\
&\quad= \frac{K}{NT^2}\times O\left(T^2\left(\frac{K}{NT}+K^{-2\delta}+\frac{K}{T^2}\right)\right),
\end{aligned}$$

in which the generalized-C_r and the Cauchy–Schwarz inequalities in turn yield

$$\begin{aligned}
&\mathbb{E}_i\left[\sum_{t=1}^{T-j}\underline{g}_{Kk}\left(y_{i,t+j}\right)\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right]^2\\
&\quad\le (T-j)\sum_{t=1}^{T-j}\mathbb{E}_i\left[\underline{g}_{Kk}\left(y_{i,t+j}\right)\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right]^2\\
&\quad\le (T-j)\sum_{t=1}^{T-j}\left[\mathbb{E}_i\left|\underline{g}_{Kk}\left(y_{i,t+j}\right)\right|^4\mathbb{E}_i\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)^4\right]^{1/2}\\
&\quad\le (T-j)^2\zeta^{1/2}\left[\mathbb{E}_i\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)^4\right]^{1/2} = O\left(T^2\left(\frac{K}{NT}+K^{-2\delta}+\frac{K}{T^2}\right)\right)
\end{aligned}$$

from Assumption 7-(iv) and Theorem 3.1. It can also be verified that (A.8) goes to zero as $N, T, J\to\infty$ using the same argument since

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

$$
\begin{aligned}
&\mathbb{E}\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T-j}\sum_{s=1}^{T}\underline{g}_K\left(y_{i,s+j}\right)\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right\|^2\\
&\quad=\frac{1}{NT^4}\sum_{k=1}^{K}\mathbb{E}\left\{\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\left(\sum_{s=1}^{T}\underline{g}_{Kk}\left(y_{i,s+j}\right)\right)^2\left(\sum_{t=1}^{T-j}\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right)^2\right\}\\
&\quad\le\frac{1}{NT^4}\sum_{k=1}^{K}\mathbb{E}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\mathbb{E}_i\left(\sum_{s=1}^{T}\underline{g}_{Kk}\left(y_{i,s+j}\right)\right)^4\right.\right.\\
&\qquad\left.\left.\times\,\mathbb{E}_i\left(\sum_{t=1}^{T-j}\left(m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right)\right)^4\right]^{1/2}\right\}\\
&\quad\le\frac{1}{NT^4}\sum_{k=1}^{K}\mathbb{E}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[T^3\sum_{s=1}^{T}\mathbb{E}_i\left|\underline{g}_{Kk}\left(y_{i,s+j}\right)\right|^4\right.\right.\\
&\qquad\left.\left.\times(T-j)^3\sum_{t=1}^{T-j}\mathbb{E}_i\left|m\left(y_{i,t-1}\right)-\widehat{m}\left(y_{i,t-1}\right)\right|^4\right]^{1/2}\right\}\\
&\quad=\frac{K}{NT^4}\times O\left(T^4\left(\frac{K}{NT}+K^{-2\delta}+\frac{K}{T^2}\right)\right).
\end{aligned}
$$

■

LEMMA A1.7. *Under Assumptions 1, 2, 7, and 6,* $\widehat{\sigma}^2=(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widehat{u}_{i,t}^0)^2\to_p\sigma^2$ *as* $N,T\to\infty$ *jointly, where* $\widehat{u}_{i,t}^0=y_{i,t}^0-\widehat{\theta}_K'g_K^0\left(y_{i,t-1}\right)$.

Proof of Lemma A1.7. Letting $\widehat{m}^0\left(y_{i,t-1}\right)=\widehat{\theta}_K'g_K^0\left(y_{i,t-1}\right)$, we have

$$
\begin{aligned}
\left|\widehat{\sigma}^2-\sigma^2\right|&\le\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t}^0-m^0\left(y_{i,t-1}\right)\right)^2-\sigma^2\right|\\
&\quad+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widehat{m}^0\left(y_{i,t-1}\right)-m^0\left(y_{i,t-1}\right)\right)^2\\
&\quad+2\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t}^0-m^0\left(y_{i,t-1}\right)\right)\left(\widehat{m}^0\left(y_{i,t-1}\right)-m^0\left(y_{i,t-1}\right)\right)\right|\\
&\equiv B_1(N,T)+B_2(N,T)+B_3(N,T).
\end{aligned}
$$

We first observe that since $y_{i,t}^0-m^0\left(y_{i,t-1}\right)=u_{i,t}^0=u_{i,t}-(1/T)\sum_{s=1}^{T}u_{i,s}$,

$$
B_1(N,T)\le\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(u_{i,t}^2-\sigma^2\right)\right|+\frac{1}{NT^2}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}u_{i,t}\right)^2\equiv B_{11}(N,T)+B_{12}(N,T),
$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

where $\mathbb{E}B_{11}(N,T)^2 = (1/NT)\mathbb{E}\left(u_{i,t}^2 - \sigma^2\right)^2 = (1/NT)\left\{\mathbb{E}u_{i,t}^4 - \sigma^4\right\} = O(1/NT)$ since $\mathbb{E}u_{i,t}^4 < \infty$ from Assumption 1 and $u_{i,t}$ is *i.i.d.* with mean zero and variance σ^2. Moreover, it is straightforward that $\mathbb{E}B_{12}(N,T) = O(1/T)$. Therefore, $B_1(N,T) \to_p 0$ as $N,T \to \infty$. Now, from Theorem 3.1, it follows that $\mathbb{E}B_2(N,T) = O\left(K/NT + K^{-2\delta} + K/T^2\right) \to 0$ as $N,T \to \infty$. Similarly, by Cauchy–Schwarz, Theorem 3.1 yields

$$B_3(N,T)^2 \leq 4\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(u_{i,t}^0\right)^2\right)\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widehat{m}^0\left(y_{i,t-1}\right) - m^0\left(y_{i,t-1}\right)\right)^2\right)$$
$$\leq 4\sigma^2 O_p\left(\frac{K}{NT} + K^{-2\delta} + \frac{K}{T^2}\right) \to 0,$$

which completes the proof. ■

A.2 Proofs of the Main Results

Proof of Proposition 2.1. We let $\varphi(y) = |y|$, which is a locally bounded and nonnegative measurable function. Then, for $|y| \geq \overline{y}$, (1) and Assumptions 1 and 2-(iii) imply

$$\mathbb{E}[\varphi(y_{i,t+1})|y_{i,t} = y, \mu_i = \mu] \leq |m(y)| + |\mu| + \mathbb{E}|u_{i,t+1}| \leq c_1|y| + c_0 + |\mu| + \mathbb{E}|u_{i,t+1}|.$$

Since it is assumed that μ_i has a compact support and $\mathbb{E}|u_{i,t}| < \infty$, it thus follows that there exists some finite real constant $\epsilon = \epsilon(\mu)$ such that $\mathbb{E}[\varphi(y_{i,t+1})|y_{i,t} = y, \mu_i = \mu] \leq c_1\varphi(y) - \epsilon$ with $0 < c_1 < 1$, which satisfies the Lyapunov function test criterion (e.g., Tweedie, 1983; Nummelin, 1984). Therefore, by Theorem 4.1 in Tjøstheim (1990), Assumptions 1 and 2 imply that the homogeneous Markov Chain $\{y_{i,t}\}$ satisfying the nonlinear autoregression (1) is geometrically (Harris) ergodic over t for each i. Furthermore, since the initial value $y_{i,0}$ is drawn from an invariant measure π_i, geometric ergodicity ensures that there exists a unique invariant probability measure π_i and the Markov chain is stationary for each i (e.g., Thm. 7 in Doukhan, 1994, Sect. 2.4.2; Prop. 1 in Bhattacharya and Lee, 1995; Prop. 1 in Chen and Shen, 1998). Finally, since a homogeneous Markov chain is β-mixing with mixing coefficients tending to zero at an exponential rate if it is geometrically ergodic (e.g., Meyn and Tweedie, 1993; Doukhan, 1994; Liebscher, 2005), we can conclude that conditional on μ_i, the homogeneous Markov process $\{y_{i,t}\}$ is stationary and β-mixing with exponential decay for each i. ■

Proof of Theorem 3.1. We define $NT \times K$ matrix $\mathbf{g}_K = (\underline{g}_K(y_{1,0}), \cdots, \underline{g}_K(y_{N,T-1}))'$, and $NT \times 1$ vectors $\mathbf{u} = (u_{1,1}, \cdots, u_{N,T})'$, $\mathbf{m} = (m(y_{1,0}), \cdots, m(y_{N,T-1}))'$. We also define $\mathbf{g}_K^0$, $\mathbf{u}^0$, and $\mathbf{m}^0$ similarly: $\mathbf{g}_K^0 = \left(g_K^0(y_{1,0}), \cdots, g_K^0(y_{N,T-1})\right)'$, $\mathbf{u}^0 = \left(u_{1,1}^0, \cdots, u_{N,T}^0\right)'$, $\mathbf{m}^0 = \left(m^0(y_{1,0}), \cdots, m^0(y_{N,T-1})\right)'$. We write

$$\widehat{\theta}_K - \theta_K = \left(\mathbf{g}_K^{0\prime}\mathbf{g}_K^0\right)^{-1}\mathbf{g}_K^{0\prime}\left(\mathbf{m}^0 - \mathbf{g}_K^0\theta_K\right) + \left(\mathbf{g}_K^{0\prime}\mathbf{g}_K^0\right)^{-1}\mathbf{g}_K^{0\prime}\mathbf{u}^0. \quad \textbf{(A.10)}$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

First note that since $\widehat{\Gamma}_K = \mathbf{g}_K^{0\prime}\mathbf{g}_K^0/NT$, Lemma A1.1 implies that $||\widehat{\Gamma}_K - \Gamma_K|| = O_p(K/\sqrt{NT})$ that goes to zero as $K/\sqrt{NT} \to 0$. Similarly as Newey (1997), we define $\mathbb{I}_{NT} = 1\{\lambda_{\min}(\widehat{\Gamma}_K) > \lambda_{\min}(\Gamma_K)/2\}$ so that $\mathbb{P}(\mathbb{I}_{NT} = 1) \to 1$ as $N, T \to \infty$. Then since $\mathbf{g}_K^0\left(\mathbf{g}_K^{0\prime}\mathbf{g}_K^0\right)^{-1}\mathbf{g}_K^{0\prime}$ is idempotent,

$$\begin{aligned}
&\mathbb{I}_{NT}\left\|\widehat{\Gamma}_K^{-1}(\mathbf{g}_K^{0\prime}\left(\mathbf{m}^0 - \mathbf{g}_K^0\theta_K\right)/NT)\right\|^2 \\
&\quad \le \mathbb{I}_{NT}\left\|\widehat{\Gamma}_K^{-1/2}\right\|_1^2\left\|(\mathbf{g}_K^{0\prime}\mathbf{g}_K^0/NT)^{-1/2}(\mathbf{g}_K^{0\prime}\left(\mathbf{m}^0 - \mathbf{g}_K^0\theta_K\right)/NT)\right\|^2 \\
&\quad = \mathbb{I}_{NT}\lambda_{\max}^2(\widehat{\Gamma}_K^{-1/2})\left[\left(\mathbf{m}^0 - \mathbf{g}_K^0\theta_K\right)'\mathbf{g}_K^0\left(\mathbf{g}_K^{0\prime}\mathbf{g}_K^0\right)^{-1}\mathbf{g}_K^{0\prime}\left(\mathbf{m}^0 - \mathbf{g}_K^0\theta_K\right)/NT\right] \\
&\quad \le \mathbb{I}_{NT}\lambda_{\min}\left(\widehat{\Gamma}_K\right)\left[\left(\mathbf{m}^0 - \mathbf{g}_K^0\theta_K\right)'\left(\mathbf{m}^0 - \mathbf{g}_K^0\theta_K\right)/NT\right] = O\left(K^{-2\delta}\right) \qquad \textbf{(A.11)}
\end{aligned}$$

from Lemma A1.4, where $\lambda_{\max}(\widehat{\Gamma}_K^{-1/2}) = [\lambda_{\min}(\widehat{\Gamma}_K)]^{-1/2}$ and $\mathbb{I}_{NT}\lambda_{\min}(\widehat{\Gamma}_K) = O_p(1)$. Note that the first inequality is from $||AB|| \le ||A||_1||B||$ for symmetric and positive definite matrix A. Furthermore, we observe that $||\widehat{\Gamma}_K^{-1}(\underline{\mathbf{g}}_K^{0\prime}\mathbf{u}^0/NT)||^2 \le ||\widehat{\Gamma}_K^{-1}||_1^2||\underline{\mathbf{g}}_K^{0\prime}\mathbf{u}^0/NT||^2$, where $||\widehat{\Gamma}_K^{-1}||_1 = O_p(1)$ from Lemma A1.2, and

$$\left\|\frac{1}{NT}\underline{\mathbf{g}}_K^{0\prime}\mathbf{u}^0\right\|^2 \le \left\|\frac{1}{NT}\underline{\mathbf{g}}_K'\mathbf{u}\right\|^2 + \left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\underline{g}_K\left(y_{i,t-1}\right)u_{i,s}\right\|^2 = O_p\left(\frac{K}{NT} + \frac{K}{T^2}\right)$$

by Lemma A1.3. It thus follows that

$$\mathbb{I}_{NT}\left\|\widehat{\Gamma}_K^{-1}\left(\mathbf{g}_K^{0\prime}\mathbf{u}^0/NT\right)\right\|^2 = O_p\left(\frac{K}{NT} + \frac{K}{T^2}\right). \qquad \textbf{(A.12)}$$

Therefore, by combining (A.11) and (A.12), we have

$$\begin{aligned}
\mathbb{I}_{NT}\left\|\widehat{\theta}_K - \theta_K\right\|^2 &\le 2\mathbb{I}_{NT}\left\|\widehat{\Gamma}_K^{-1}\left(\mathbf{g}_K^{0\prime}\mathbf{u}^0/NT\right)\right\|^2 + 2\mathbb{I}_{NT}\left\|\widehat{\Gamma}_K^{-1}\left(\mathbf{g}_K^{0\prime}\left(\mathbf{m}^0 - \mathbf{g}_K^0\theta_K\right)/NT\right)\right\|^2 \\
&= O_p\left(K^{-2\delta} + K/NT + K/T^2\right).
\end{aligned}$$

Next, by the triangular inequality like Newey (1997), Assumptions 6 and 7 imply

$$\begin{aligned}
&\mathbb{E}\mathbb{I}_{NT}\int\left[\widehat{m}\left(y\right) - m\left(y\right)\right]^2\pi_i\left(dy\right) \\
&\quad = \mathbb{E}\mathbb{I}_{NT}\int\left[g_K\left(y\right)'\left(\widehat{\theta}_K - \theta_K\right) + \left(g_K\left(y\right)'\theta_K - m\left(y\right)\right)\right]^2\pi_i\left(dy\right) \\
&\quad \le 2\mathbb{E}\mathbb{I}_{NT}\int\left(g_K\left(y\right)'\left(\widehat{\theta}_K - \theta_K\right)\right)^2\pi_i\left(dy\right) + 2\mathbb{E}\mathbb{I}_{NT}\int\left(m\left(y\right) - g_K\left(y\right)'\theta_K\right)^2\pi_i\left(dy\right) \\
&\quad \le 2\mathbb{E}\mathbb{I}_{NT}C\left\|\widehat{\theta}_K - \theta_K\right\|^2 + 2\left\|m\left(y\right) - g_K\left(y\right)'\theta_K\right\|_{\infty,\omega}^2\mathbb{E}\mathbb{I}_{NT}\int\left(1 + \|y\|^2\right)^{\omega}\pi_i\left(dy\right) \\
&\quad = O\left(K^{-2\delta} + K/NT + K/T^2\right) + O\left(K^{-2\delta}\right) \\
&\quad = O\left(K^{-2\delta} + K/NT + K/T^2\right)
\end{aligned}$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

for some positive constant $C < \infty$. Note that the second inequality is because we can assume that $\int g_K(y) g_K(y)' \pi_i(dy) \leq C I_K < \infty$ without loss of generality from Assumption 7 once we consider orthonormalized basis functions $g_{Kk}(\cdot)$ for $k = 1, \cdots, K$. ■

Proof of Theorem 3.2. Since $\widehat{m}(y) - m(y) = g_K(y)'(\widehat{\theta}_K - \theta_K) - \left(m(y) - g_K(y)'\theta_K\right)$, we write

$$\frac{\sqrt{NT}\left(\widehat{m}(y) - m(y) + (1/T) g_K(y)' \Gamma_K^{-1} \Phi_K\right)}{\sqrt{g_K(y)' \Gamma_K^{-1} g_K(y)}}$$
$$= \frac{g_K(y)' \sqrt{NT}\left(\widehat{\theta}_K - \theta_K + (1/T)\Gamma_K^{-1}\Phi_K\right)}{\sqrt{g_K(y)' \Gamma_K^{-1} g_K(y)}} - \frac{\sqrt{NT}\left(m(y) - g_K(y)'\theta_K\right)}{\sqrt{g_K(y)' \Gamma_K^{-1} g_K(y)}}. \quad \textbf{(A.13)}$$

The second term in (A.13) is $o_p(1)$ since $\sqrt{NT}\left|m(y) - g_K(y)'\theta_K\right| = O(K^{-\delta}\sqrt{NT}) \to 0$ from Lemma A1.4, and $g_K(y)' \Gamma_K^{-1} g_K(y) > 0$ for any y from Assumption 7-(ii). Therefore, the asymptotic distribution of $\widehat{m}(\cdot)$ is determined by the asymptotic behavior of the first term in (A.13), which can be decomposed as[10]

$$g_K(y)' \sqrt{NT}\left(\widehat{\theta}_K - \theta_K + (1/T)\Gamma_K^{-1}\Phi_K\right) / \sqrt{g_K(y)' \Gamma_K^{-1} g_K(y)}$$
$$= \widehat{\rho}(y)' \widehat{\Gamma}_K^{-1/2} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N}\sum_{t=1}^{T} \underline{g}_K\left(y_{i,t-1}\right) u_{i,t}$$
$$- \widehat{\rho}(y)' \widehat{\Gamma}_K^{-1/2} \sqrt{\frac{N}{T}} \left(\frac{1}{NT} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \underline{g}_K\left(y_{i,t-1}\right) u_{i,s} - \widehat{\Phi}_K \right)$$
$$+ \widehat{\rho}(y)' \widehat{\Gamma}_K^{-1/2} \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N}\sum_{t=1}^{T} g_K^0\left(y_{i,t-1}\right) \left\{ m^0\left(y_{i,t-1}\right) - g_K^0\left(y_{i,t-1}\right)'\theta_K \right\} \right)$$
$$\equiv D_1(N,T,K) - D_2(N,T,K) + D_3(N,T,K), \quad \textbf{(A.14)}$$

where $\widehat{\rho}(y) = \widehat{\Gamma}_K^{-1/2} g_K(y) / \sqrt{g_K(y)' \Gamma_K^{-1} g_K(y)}$. Note that, for each y, $\widehat{\rho}(y)$ satisfies $\|\widehat{\rho}(y)\|^2 = [g_K(y)'(\widehat{\Gamma}_K^{-1} - \Gamma_K^{-1}) g_K(y) / g_K(y)' \Gamma_K^{-1} g_K(y)] + 1 \to_p 1$ from Lemma A1.2.

We first show that $D_2(N,T,K)$ and $D_3(N,T,K)$ are asymptotically negligible. Since Lemma A1.6 yields

$$\left\| \frac{1}{NT} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \underline{g}_K\left(y_{i,t-1}\right) u_{i,s} - \widehat{\Phi}_K \right\| \leq \left\| \frac{1}{NT} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \underline{g}_K\left(y_{i,t-1}\right) u_{i,s} - \Phi_K \right\|$$
$$+ \left\| \Phi_K - \widehat{\Phi}_K \right\| \to_p 0,$$

we have

$$\|D_2(N,T,K)\| \leq \|\widehat{\rho}(y)\| \left\| \widehat{\Gamma}_K^{-1/2} \right\|_1 \sqrt{\frac{N}{T}} \left\| \frac{1}{NT} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \underline{g}_K\left(y_{i,t-1}\right) u_{i,s} - \widehat{\Phi}_K \right\| \to_p 0$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

by Assumption 4 and Lemmas A1.2 and A1.6. As in the proof of Theorem 3.1, we also have

$$\begin{aligned}\|D_3(N,T,K)\|^2 &= \left\|\widehat{\rho}(y)'\left(\mathbf{g}_K^{0\prime}\mathbf{g}_K^0\right)^{-1/2}\mathbf{g}_K^{0\prime}\left(\mathbf{m}^0-\mathbf{g}_K^0\theta_K\right)\right\|^2 \\ &\le \|\widehat{\rho}(y)\|^2\left\{\left(\mathbf{m}^0-\mathbf{g}_K^0\theta_K\right)'\mathbf{g}_K^0\left(\mathbf{g}_K^{0\prime}\mathbf{g}_K^0\right)^{-1}\mathbf{g}_K^{0\prime}\left(\mathbf{m}^0-\mathbf{g}_K^0\theta_K\right)\right\} \\ &\le \|\widehat{\rho}(y)\|^2\left\{\left(\mathbf{m}^0-\mathbf{g}_K^0\theta_K\right)'\left(\mathbf{m}^0-\mathbf{g}_K^0\theta_K\right)\right\} = O_p(K^{-2\delta})\to 0.\end{aligned}$$

For $D_1(N,T,K)$, we write

$$\begin{aligned}D_1(N,T,K) &= \rho(y)'\Gamma_K^{-1/2}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \underline{g}_K\left(y_{i,t-1}\right)u_{i,t} \\ &\quad+\rho(y)'\left(\Gamma_K^{1/2}\widehat{\Gamma}_K^{-1}-\Gamma_K^{-1/2}\right)\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \underline{g}_K\left(y_{i,t-1}\right)u_{i,t} \\ &\equiv D_{11}(N,T,K)+D_{12}(N,T,K),\end{aligned}$$

where $\rho(y) = \Gamma_K^{-1/2}g_K(y)/\sqrt{g_K(y)'\Gamma_K^{-1}g_K(y)}$. The first term satisfies $D_{11}(N,T,K)\to_d \mathcal{N}\left(0,\sigma^2\right)$ as $N,T\to\infty$ by Lemma A1.5. The second term is negligible since similarly as Newey (1997, p. 164)

$$\begin{aligned}&\|D_{12}(N,T,K)\|^2 \\ &\quad=\rho(y)'\left(\Gamma_K^{1/2}\widehat{\Gamma}_K^{-1}-\Gamma_K^{-1/2}\right)\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T u_{i,t}u_{j,s}\underline{g}_K\left(y_{i,t-1}\right)\underline{g}_K\left(y_{j,s-1}\right)'\right\} \\ &\qquad\times\left(\Gamma_K^{1/2}\widehat{\Gamma}_K^{-1}-\Gamma_K^{-1/2}\right)'\rho(y) \\ &\quad=\rho(y)'\left(\Gamma_K^{1/2}\widehat{\Gamma}_K^{-1}-\Gamma_K^{-1/2}\right)\left(\sigma^2\Gamma_K\right)\left(\Gamma_K^{1/2}\widehat{\Gamma}_K^{-1}-\Gamma_K^{-1/2}\right)'\rho(y)+\rho(y)' \\ &\qquad\times\left(\Gamma_K^{1/2}\widehat{\Gamma}_K^{-1}-\Gamma_K^{-1/2}\right) \\ &\qquad\times\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T u_{i,t}u_{j,s}\underline{g}_K\left(y_{i,t-1}\right)\underline{g}_K\left(y_{j,s-1}\right)'-\sigma^2\Gamma_K\right\} \\ &\qquad\times\left(\Gamma_K^{1/2}\widehat{\Gamma}_K^{-1}-\Gamma_K^{-1/2}\right)'\rho(y) \\ &\quad=\sigma^2\left\|\rho(y)'\left(\Gamma_K^{1/2}\widehat{\Gamma}_K^{-1}-\Gamma_K^{-1/2}\right)\Gamma_K^{1/2}\right\|^2+o_p(1)\left\|\rho(y)'\left(\Gamma_K^{1/2}\widehat{\Gamma}_K^{-1}-\Gamma_K^{-1/2}\right)\right\|^2 \\ &\quad\le\|\rho(y)\|^2\left\|\Gamma_K^{1/2}\right\|_1^2\left\|\widehat{\Gamma}_K^{-1}-\Gamma_K^{-1}\right\|_1^2\left\{\sigma^2\left\|\Gamma_K^{1/2}\right\|_1^2+o_p(1)\right\}\to_p 0\end{aligned}$$

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press

as $N, T \to \infty$ by Lemmas A1.2, where

$$\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{i,t}u_{j,s}\underline{g}_K\left(y_{i,t-1}\right)\underline{g}_K\left(y_{j,s-1}\right)'\right]$$
$$=\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\left(u_{i,t}^2\underline{g}_K^2\left(y_{i,t-1}\right)\right)\right]=\sigma^2\Gamma_K$$

from Lemma A1.3. Therefore, $D_1(N,T,K)\to_d \mathcal{N}\left(0,\sigma^2\right)$ as $N, T \to \infty$. It follows that, from (A.13) and (A.14),

$$v_K(y)^{-1/2}\left(\widehat{m}(y)-m(y)+(1/T)\widehat{b}_K(y)\right)\to_d \mathcal{N}(0,1) \text{ for } y\in\mathcal{Y},$$

where $v_K(y)=\sigma^2 g_K(y)'\Gamma_K^{-1}g_K(y)/NT$ and $\widehat{b}_K(y)=g_K(y)'\widehat{\Gamma}_K^{-1}\widehat{\Phi}_K$.

Using $\widehat{v}_K(y)=\widehat{\sigma}^2 g_K(y)'\widehat{\Gamma}_K^{-1}g_K(y)/NT$ yields the identical asymptotic normality result since $\widehat{\sigma}^2\to_p\sigma^2$ by Lemma A1.7 and $||\widehat{\Gamma}_K^{-1}-\Gamma_K^{-1}||\to_p 0$ by Lemma A1.2 as $N, T\to\infty$. Note that (A.13) can be rewritten as

$$\frac{\sqrt{NT}\left(\widehat{m}(y)-m(y)+(1/T)g_K(y)'\widehat{\Gamma}_K^{-1}\widehat{\Phi}_K\right)}{\sqrt{g_K(y)'\widehat{\Gamma}_K^{-1}g_K(y)}}$$
$$=\frac{g_K(y)'\sqrt{NT}\left(\widehat{\theta}_K-\theta_K+(1/T)\widehat{\Gamma}_K^{-1}\widehat{\Phi}_K\right)}{\sqrt{g_K(y)'\Gamma_K^{-1}g_K(y)}}\sqrt{\frac{g_K(y)'\Gamma_K^{-1}g_K(y)}{g_K(y)'\widehat{\Gamma}_K^{-1}g_K(y)}}$$

and $g_K(y)'\Gamma_K^{-1}g_K(y)/g_K(y)'\widehat{\Gamma}_K^{-1}g_K(y)\to_p 1$ as $N, T\to\infty$. ■

https://doi.org/10.1017/S0266466614000188 Published online by Cambridge University Press