Econometric Theory, **25**, 2009, 1433–1445. Printed in the United States of America.
doi:10.1017/S0266466609090744

UNIFORM CONVERGENCE RATES OF KERNEL ESTIMATORS WITH HETEROGENEOUS DEPENDENT DATA

DENNIS KRISTENSEN
Columbia University
and
CREATES

The main uniform convergence results of Hansen (2008, *Econometric Theory* 24, 726–748) are generalized in two directions: Data are allowed to (a) be heterogeneously dependent and (b) depend on a (possibly unbounded) parameter. These results are useful in semiparametric estimation problems involving time-inhomogeneous models and/or sampling of continuous-time processes. The usefulness of these results is demonstrated by two applications: kernel regression estimation of a time-varying AR(1) model and the kernel density estimation of a Markov chain that has not been initialized at its stationary distribution.

1. INTRODUCTION

Uniform convergence rates of kernel estimators have many useful applications, in particular in semiparametric estimation problems. Recently, Hansen (2008) provided a set of strong results for the case where data are stationary and strong mixing. In this note, we extend his results in two directions: First, we allow for heterogeneous data where the random variables are nonidentically distributed but still mixing. Second, data can potentially depend on a parameter, and we show uniform convergence over both the domain of the data variables and the parameter set. The main conclusion of this note is that as long as the mixing coefficients and suitable moments of data are uniformly bounded as functions of the sample size and the parameter, the results of Hansen (2008) still go through.

The first extension is useful in situations where data are nonstationary (but mixing) and the distributions potentially vary with the sample size. Three specific examples where this situation arises are (a) time-varying models (Dahlhaus,

I thank Bruce Hansen for helpful discussions and suggestions. Center for Research in Econometric Analysis of Time Series (CREATES) is funded by the Danish National Research Foundation. Address correspondence to Dennis Kristensen, Economics Department, Columbia University, 1018 International Affairs Building, MC 3308, 420 West 118th Street, New York, NY 10027, USA; e-mail: dk2313@columbia.edu.

© 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

Neumann, and von Sachs, 1999; Cai, 2007; Kristensen, 2008), (b) continuous-time stochastic processes sampled at discrete time points where time between observations shrinks (Bandi and Phillips, 2003; Kristensen, 2009), and (c) Markov chains that have not been initialized at their stationary distribution (Kim and Lee, 2004; Yu, 1993).

The second extension addresses a situation that often appears in semiparametric problems where kernel estimators depend on the finite-dimensional parameter of interest, for example, single-index models (Xia and Härdle, 2006) and models with generated regressors (Li and Wooldridge, 2002). Another important application can be found in the recent literature on nonparametric simulated maximum-likelihood and generalized method of moments (GMM) estimation (Altissimo and Mele, 2008; Brownlees, Kristensen, and Shin, 2009; Creel and Kristensen, 2009; Fermanian and Salanié, 2004; Kristensen and Shin, 2008).

Finally, we note that our results could probably be combined with those of Kanaya (2008) to obtain uniform results under weaker moment conditions than stated here.

In the following section, we present the main results. In Section 3, the usefulness of these results is demonstrated by two examples. The first is kernel regression estimation in the locally stationary AR(1) model, and the second is kernel estimation of the stationary density of a homogeneous Markov chain that has not been initialized in the stationary distribution. Both of these results appear to be new. All proofs and lemmas can be found in the Appendix.

2. MAIN RESULT

Let $\left(Y_{n,i}(\gamma), X_{n,i}(\gamma)\right) \in \mathbb{R}\times\mathbb{R}^d, i = 1,\dots,n, n = 1,2,\dots$, be a triangular array of random variables that depend on a parameter $\gamma \in \Gamma \subseteq \mathbb{R}^k$. We are then interested in deriving uniform convergence rates of the following sample average:

$$\hat{\Psi}(x;\gamma) = \frac{1}{nh^d}\sum_{i=1}^{n} Y_{n,i}(\gamma) K\left(\frac{X_{n,i}(\gamma)-x}{h}\right),$$

where $K : \mathbb{R}^d \mapsto \mathbb{R}$ is some weight function and $h > 0$ is a bandwidth.

To state the conditions that will be imposed on the data, we first introduce some additional notation: Let $f_{n,i}(x;\gamma)$ and $f_{n,ij}(x,y;\gamma)$ denote the densities of $X_{n,i}$ and $(X_{n,i}, X_{n,j})$ and define $\mathcal{F}_{n,i}^{j}(\gamma) = \mathcal{F}(Y_{n,i}(\gamma), X_{n,i}(\gamma),\dots,Y_{n,j}(\gamma), X_{n,j}(\gamma))$, $i < j$, as a sequence of sigma algebras associated with the observed sequences. Furthermore, for any triangular array $W_{n,i}(\gamma)$ and for some $M \geq 0$, define the following functions of γ:

$$\begin{aligned} B_0(\gamma) &= \sup_{i,n}\sup_{x\in\mathbb{R}^d} f_{n,i}(x;\gamma), \\ B_{W,1}(\gamma) &= \sup_{i,n}\sup_{x} \mathrm{E}\left[\left|W_{n,i}(\gamma)\right| \middle| X_{n,i}(\gamma) = x\right] f_{n,i}(x;\gamma), \end{aligned} \tag{1}$$

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

$$B_{W,2}(\gamma) = \sup_n \sup_{|i-j|\geq M} \sup_{x,y} \mathrm{E}\left[\left|W_{n,i}(\gamma) W_{n,j}(\gamma)\right| \left| X_{n,i}(\gamma) = x, X_{n,j}(\gamma) = y\right.\right] \times f_{n,ij}(x, y; \gamma), \quad (2)$$

where $\sup_{i,n}$ is short for $\sup_{n\geq 1} \sup_{1\leq i\leq n}$. We then impose the following conditions on the data.

A.1. For all $\gamma \in \Gamma$: The triangular array $\{(Y_{n,i}(\gamma), X_{n,i}(\gamma)) : i = 1, \ldots, n, \quad n \geq 1\}$ is strongly mixing, and its mixing coefficients,

$$\alpha_{n,i}(\gamma) = \sup_{-n\leq j\leq n} \sup_{A\in\mathcal{F}^k_{n,-\infty}(\gamma), B\in\mathcal{F}^\infty_{n,n+i}(\gamma)} |P(A\cap B) - P(A)P(B)|,$$

satisfy $\alpha_{n,i}(\gamma) \leq Ai^{-\beta}$ for some $0 < A, \beta < \infty$ (which do not depend on n and γ).

A.2. The random functions $\gamma \mapsto Y_{n,i}(\gamma)$ and $\gamma \mapsto X_{n,i}(\gamma)$ are almost surely differentiable with derivatives $\dot{Y}_{n,i}(\gamma)$ and $\dot{X}_{n,i}(\gamma)$.

A.3. For some $s > 2$, uniformly over n and i,

$$\mathrm{E}\left[\left|Y_{n,i}(\gamma)\right|^s\right] < \infty, \quad \mathrm{E}\left[\left\|\dot{Y}_{n,i}(\gamma)\right\|^s\right] < \infty, \quad \mathrm{E}\left[\left|Y_{n,i}(\gamma)\right|^s \left\|\dot{X}_{n,i}(\gamma)\right\|^s\right] < \infty.$$

With $\bar{d} = d + k$ and for some $q > 0$, the mixing exponent β in Assumption A.1 satisfies

$$\beta > \frac{1 + (s-1)\left(1 + \bar{d}/q + \bar{d}\right)}{s-2}.$$

A.4. The functions defined in equations (1) and (2) satisfy

$$B_0(\gamma) \leq \bar{B}_0\left(1 + \|\gamma\|^\lambda\right), \qquad B_{W,j}(\gamma) \leq \bar{B}_{W,j}\left(1 + \|\gamma\|^\lambda\right), \qquad j = 1, 2,$$

for some constants $\bar{B}_0$, $\bar{B}_{W,1}$, $\bar{B}_{W,2}, \lambda \geq 0$ with $W = Y$, $W = \dot{Y}$ and $Y\dot{X}$.

A.5. There exist constants $q \geq d$ and $\bar{B}_{W,3} \geq 0$ such that

$$B_{W,3}(\gamma) = \sup_{i,n} \sup_x \|x\|^q \mathrm{E}\left[\left|W_{n,i}(\gamma)\right| \left| X_{n,i}(\gamma) = x\right.\right] f_{n,i}(x; \gamma)$$

satisfies $B_{W,3}(\gamma) \leq \bar{B}_{W,3}\left(1 + \|\gamma\|^\lambda\right)$ for $W = Y$, $W = \dot{Y}$ and $Y\dot{X}$.

Assumption A.1 restricts the data to be strongly mixing and imposes uniform bounds on the mixing coefficients as functions of n and γ. We assume differentiability in Assumption A.2 and then require that certain moments of the data and their derivatives exist in Assumption A.3. We allow the parameter space Γ to be unbounded and the functions $B_{W,j}(\gamma)$ in Assumption A.4 to be unbounded but with at most polynomial growth. This is particularly relevant in the case of simulation-based estimators using kernel methods where γ may contain past

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

values of the observed process; in this case, one is normally not willing to assume a compact support; see Brownlees et al. (2009) and Kristensen and Shin (2008). When Γ is unbounded we will only be able to give uniform convergence over a (growing compact) subset of Γ that will depend on the polynomial bound λ of the functions $B_0(\gamma)$ and $B_{W,j}(\gamma)$, $j = 1, 2, 3$. Polynomial bounds are imposed to simplify the uniform convergence results and can be exchanged for other bounds as discussed later. Finally, Assumption A.5 is a strengthening of the bound imposed on $B_{W,1}(\gamma)$ in Assumption A.4 and corresponds to the condition imposed in Hansen (2008, eqn. (21)) but is here required to hold for $W = \dot{Y}$ and $Y\dot{X}$ in addition to $W = Y$.

In the case where data are homogeneous and parameter independent, $(Y_{n,i}(\gamma), X_{n,i}(\gamma)) = (Y_i, X_i)$, Assumptions A.1–A.4 collapse to Assumptions 1 and 2 of Hansen (2008). In particular, we have $\dot{Y} = \dot{X} = \lambda = 0$. However, in contrast to Hansen (2008), we do not impose stationarity, meaning, for example, that the density of X_i may depend on i, $f_i(x)$. More general assumptions regarding the dependence of the data can be found in the paper by Andrews (1995), who allows for near-epoch dependence. On the other hand, Andrews (1995) does not allow for the distribution to depend on sample size and obtains less precise rates compared to the ones stated here.

Because of heterogeneity, $\mathrm{E}[\hat{\Psi}(x;\gamma)]$ does not necessarily have a well-defined limit as $n \to \infty$ under Assumptions A.1–A.5. For example, in the case where K is a standard kernel with $\int K(z)\,dz = 1$, in great generality it will hold that

$$\mathrm{E}[\hat{\Psi}(x;\gamma)] = n^{-1}\sum_{i=1}^{n} f_{n,i}(x;\gamma)\, m_{n,i}(x;\gamma) + o(1),$$

as $n \to \infty$, where $m_{n,i}(x;\gamma) = \mathrm{E}[Y_{n,i}(\gamma)|X_{n,i}(\gamma) = x]$. To ensure that the sum on the right-hand side has a well-defined limit, further restrictions have to be imposed. In most applications, $f_{n,i}(x;\gamma) = f(x;\gamma) + o(1)$ and $m_{n,i}(x;\gamma) = m(x;\gamma) + o(1)$, in which case $\mathrm{E}[\hat{\Psi}(x;\gamma)] = f(x;\gamma)m(x;\gamma) + o(1)$. The existence of such a limit of $\mathrm{E}[\hat{\Psi}(x;\gamma)]$ is something that should be verified on a case-by-case basis and, for example, holds under stationarity. However, we will not impose any such restrictions here, because the main objective of this note is to bound the variance of $\hat{\Psi}(x;\gamma)$. For that purpose, Assumptions A.1–A.5 suffice. Similar results can be found in Andrews (1995), where the object of interest is a sum of densities/regression functions and no restrictions are imposed to ensure that this sum converges. In the two examples in Section 3, we verify that $\mathrm{E}[\hat{\Psi}(x;\gamma)]$ has a well-defined limit as $n \to \infty$.

Next, we impose regularity conditions on the kernel.

A.6.1. The function K satisfies $|K(u)| \leq \bar{K} < \infty$ and $\int |K(u)|\,du \leq \mu < \infty$. There exists $\Lambda_1, L < \infty$ such that either (a) $K(u) = 0$ for $\|u\| > L$ and $|K(u) - K(u')| \leq \Lambda_1 \left\|u - u'\right\|$ or (b) $K(u)$ is differentiable with $|\partial K(u)/\partial u| \leq \Lambda_1$ and, for some $\nu > 1$, $|\partial K(u)/\partial u| \leq \Lambda_1 \|u\|^{-\nu}$ for $\|u\| \geq L$.

A.6.2. For some $\Lambda_2 < \infty$, $|K(u)| \leq \Lambda_2 \|u\|^{-\nu}$ for $\|u\| \geq L$.

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

Assumptions A.6.1 and A.6.2 are identical to Hansen (2008, Assump. 3) and Hansen (2008, eqn. (22)), respectively. We are now ready to state the main result.

THEOREM 1. *Assume that Assumptions A.1–A.4 and A.6.1 hold and with*

$$\theta = \frac{\beta - 1 - \bar{d} - \bar{d}/q - (1+\beta)/(\bar{s}-1)}{\beta + 3 - \bar{d} - (1+\beta)/(\bar{s}-1)}$$

the bandwidth satisfies $\log(n)/(n^{\theta}h^{d}) \to 0$. *Then with*

$$c_n = O\left(\log(n)^{1/2} n^{1/2q}\right), \qquad d_n = O\left(\log(n)^{1/2} n^{1/2q}\right):$$

(i) $$\sup_{\|x\| \le c_n} \sup_{\|\gamma\| \le d_n} |\hat{\Psi}(x;\gamma) - \mathrm{E}[\hat{\Psi}(x;\gamma)]| = O_P\left(d_n^{\lambda}\sqrt{\log(n)/\left(nh^{d}\right)}\right).$$

(ii) If additionally Assumptions A.5 and A.6.2 hold, then

$$\sup_{x \in \mathbb{R}^d} \sup_{\|\gamma\| \le d_n} |\hat{\Psi}(x;\gamma) - \mathrm{E}[\hat{\Psi}(x;\gamma)]| = O_P\left(d_n^{\lambda}\sqrt{\log(n)/(nh^{d})}\right).$$

Remark 1.

(a) The results can be extended to uniform almost sure convergence by changing the restrictions on β, θ, and c_n as in Hansen (2008, Thms. 3 and 5).

(b) The polynomial bounds imposed on $B_{W,j}(\gamma)$, $j = 0, 1, 2, 3$, can be removed and the preceding results will still go through with rate $O_P\left(\bar{B}_n\sqrt{\log(n)/(nh^{d})}\right)$ where

$$\bar{B}_n = \max_{W = W, \dot{Y}, Y\dot{X}} \max_{j=0,1,2,3} \sup_{\|\gamma\| \le d_n} B_{W,j}(\gamma).$$

(c) If $B_{W,j}(\gamma)$, $j = 0, 1, 2, 3$, are all uniformly bounded, then Theorem 1(ii) holds uniformly over $\gamma \in \Gamma$.

3. TWO EXAMPLES

We here present two applications of Theorem 1 to discrete-time models where data are not parameter-dependent. For applications of Theorem 1 with parameter dependence, we refer to Brownlees et al. (2009), Creel and Kristensen (2009), and Kristensen and Shin (2008), whereas Kristensen (2009) applies the result to a continuous-time model.

We will in the following discussion implicitly assume that the kernel satisfies Assumptions A.6.1 and A.6.2 and is of order $r \ge 1$: $\int K(z) z^{i} dz = 0$, $i = 1, \ldots, r-1$, and $\int K(z)|z|^{r} dz < \infty$.

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

The first example where Theorem 1 becomes relevant is in the nonparametric estimation of time-varying regression models. Suppose that we have observed $W_{n,i}$, $i = 0, 1, \ldots, n$, from the following sequence of time-varying AR(1) models:

$$W_{n,i} = a\left(\frac{i}{n}\right) W_{n,i-1} + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where the sequence $\{\varepsilon_i\}$ is assumed to be i.i.d.$(0, \sigma^2)$ with $\mathrm{E}[|\varepsilon_t|^s] < \infty$, $s > 4$. The assumption of independent and identically distributed (i.i.d.) errors is imposed for simplicity and can be relaxed to allow for dependent and mixing errors under suitable conditions on the associated mixing coefficients.

The distribution of $W_{n,i}$ obviously depends on both i and n and will in general be nonstationary. However, Orbe, Ferreira, and Rodriguez-Poo (2005, Lem. A.4) show that when the autoregressive coefficient function $a(s)$ is uniformly bounded below one,

$$a_{\max} = \sup_{0 \le s \le 1} |a(s)| < 1, \tag{3}$$

then $\{W_{n,i}\}$ satisfies the strong mixing condition in Assumption A.1 and $\sup_{i,n} \mathrm{E}[|W_{n,i}|^s] < \infty$.

We consider the following kernel estimator of the function a:

$$\hat{a}(\tau) = \frac{\hat{\Psi}_1(\tau)}{\hat{\Psi}_2(\tau)} = \frac{\sum_{i=1}^n K\left(\frac{i/n-\tau}{h}\right) W_{n,i} W_{n,i-1}}{\sum_{i=1}^n K\left(\frac{i/n-\tau}{h}\right) W_{n,i-1}^2}.$$

By defining $X_{n,i} = i/n \in \mathbb{R}$ (which for each $n \ge 1$ corresponds to i.i.d. draws from a uniform distribution on $[0, 1]$), we can now apply Theorem 1 to both $\hat{\Psi}_1(\tau)$ and $\hat{\Psi}_2(\tau)$. However, because of boundary problems, the preceding estimator will be asymptotically biased at each of the two endpoints of the time series, $\tau = 0$ and $\tau = 1$; see, for example, Kristensen (2008, 2009). We circumvent this problem by only showing uniform convergence over the expanding interval $\tau \in [b, 1-b]$ with $b \to 0$. Alternatively, one could instead use a local linear estimator or a boundary kernel because these will not suffer from boundary biases (cf., respectively, Cai, 2007; and Chen, 2000), but we here maintain the preceding estimator for simplicity.

THEOREM 2. *Assume that (i) $a : [0, 1] \mapsto \mathbb{R}$ is $r \ge 1$ times continuously differentiable and satisfies equation (3) and (ii) $\mathrm{E}[|\varepsilon_t|^s] < \infty$ for some $s > 4$. Then for any sequence $b \to 0$ satisfying $h/b \to 0$:*

$$\sup_{\tau \in [b, 1-b]} \left|\hat{a}(\tau) - a(\tau)\right| = O_P\left(h^r\right) + O_P\left(\sqrt{\log(n)/(nh)}\right).$$

The second example considers the case where we have observed a d-dimensional, time-homogeneous Markov chain $\{X_i\}$. We assume that the Markov

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

chain is geometrically ergodic such that a stationary marginal density, $f(x)$, exists. We then wish to estimate $f(x)$. If the observed sequence is not initialized at $f(x)$, then it is nonstationary. For example, if the chain is initialized at $X_0 = x$ for some given $x \in \mathbb{R}^d$ then $P(X_i \in A) = \int_A p_i(y|x)\,dy$, $i = 1, 2, \ldots$, where $p_i(y|x)$ is the i-step transition density,

$$p_i(y|x) = \int_{\mathbb{R}^d} p(y|z)\, p_{i-1}(z|x)\, dz, \tag{4}$$

with $p_1(y|x) := p(y|x)$, and $p(y|x)$ being the transition density, $P(X_{i+1} \in A|X_i) = \int_A p(y|x)\,dy$. However, the distribution will converge toward the stationary one in the total variation norm, $\sup_A \left|\int_A p_i(y|x)\,dy - \int_A f(y)\,dy\right| \to 0$ as $i \to \infty$, with geometric rate (cf. Meyn and Tweedie, 1993). So we expect that

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$

will be pointwise consistent. However, for uniform convergence of $\hat{f}$ toward f to hold convergence of $p_i(y|x)$ toward $f(y)$ in the total variation norm does not suffice. To obtain this, we have to strengthen the convergence to be uniform. This is here done by imposing the so-called strong Doeblin condition:

$$\exists s \geq 1 \exists \rho \in (0,1) : p_s(y|x) \geq \rho f(y), \tag{5}$$

which implies that there exists a constant $M > 0$ such that

$$|p_{s+i}(y|x) - f(y)| \leq M(1-\rho)^{-i} f(y), \tag{6}$$

for all $i > 0$; see Holden (2000, Thm. 1). The preceding uniform convergence of the transition densities as $i \to \infty$ will ensure uniform convergence of $\hat{f}$ toward f.

THEOREM 3. *Assume that (i) the Markov chain $\{X_i\}$ satisfies the strong Doeblin condition (5), (ii) $y \mapsto p(y|x)$ is $r \geq 1$ times differentiable with $\partial^r p(y|x)/\partial y^r$ being uniformly continuous for all x, and (iii) $\|y\|^q f(y)$ is bounded for some $q \geq d$. Then*

$$\sup_{x \in \mathbb{R}^d} \left|\hat{f}(x) - f(x)\right| = O_P\left(h^r\right) + O_P\left(\sqrt{\log(n)/\left(nh^d\right)}\right).$$

Results similar to Theorem 3 can be obtained for kernel estimators of, for example, the transition density, $p(y|x)$, or conditional moments, $\mathrm{E}[g(X_i)|X_{i-1} = x]$.

REFERENCES

Altissimo, F. & A. Mele (2009) Simulated nonparametric estimation of dynamic models. *Review of Economic Studies* 76, 413–450.

Andrews, D.W.K. (1995) Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560–596.

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

Bandi, F. & P.C.B. Phillips (2003) Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 241–283.

Brownlees, C.T., D. Kristensen, & Y. Shin (2009) Nonparametric Simulated Maximum Likelihood Estimation of Dynamic Latent Variable Models. Manuscript, Department of Economics, Columbia University.

Cai, Z. (2007) Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics* 136, 163–188.

Chen, S.X. (2000) Beta kernel smoothers for regression curves. *Statistica Sinica* 10, 73–91.

Creel, M. & D. Kristensen (2009) Estimation of Dynamic Latent Variable Models Using Simulated Nonparametric Moments. Manuscript, Department of Economics, Universitat Autònoma de Barcelona.

Dahlhaus, R., M.H. Neumann, & R. von Sachs (1999) Nonlinear wavelet estimation of time-varying autoregressive processes. *Bernoulli* 5, 873–906.

Fermanian, J.-D. & B. Salanié (2004) A nonparametric simulated maximum likelihood estimation method. *Econometric Theory* 20, 701–734.

Hansen, B.E. (2008) Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726–748.

Holden, L. (2000) Convergence of markov chains in the relative supremum norm. *Journal of Applied Probability* 37, 1074–1083.

Kanaya, S. (2008) Uniform Convergence of Kernel-Based Nonparametric Estimators for Continuous and Discrete Time Processes: A Damping Approach. Manuscript, University of Oxford.

Kim, T.Y. & S. Lee (2004) Kernel density estimator for strong mixing processes. *Journal of Statistical Planning and Inference* 133, 273–284.

Kristensen, D. (2008) Semi-Nonparametric Estimation and Testing of Time-Varying Regressions. Manuscript, Department of Economics, Columbia University.

Kristensen, D. (2009) Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory* 25, forthcoming.

Kristensen, D. & Y. Shin (2008) Estimation of Dynamic Models with Nonparametric Simulated Maximum Likelihood. Manuscript, Department of Economics, Columbia University.

Li, Q. & J.M. Wooldridge (2002) Semiparametric estimation of partially linear models for dependent data with generated regressors. *Econometric Theory* 18, 625–645.

Liebscher, E. (1996) Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications* 65, 69–80.

Meyn, S.P. & R.L. Tweedie (1993) *Markov Chains and Stochastic Stability*. Springer-Verlag.

Orbe, S., E. Ferreira, & J. Rodriguez-Poo (2005) Nonparametric estimation of time varying parameters under shape restrictions. *Journal of Econometrics* 126, 53–77.

Xia, Y. & W. Härdle (2006) Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis* 97, 1162–1184.

Yu, B. (1993) Density estimation in the L^∞ norm for dependent data with applications to the Gibbs sampler. *Annals of Statistics* 21, 711–735.

APPENDIX: Proofs and Lemmas

We first state two lemmas that will be used to prove Theorem 1. The first lemma is an extension of Hansen (2008, Thm. 1).

LEMMA A.1. *Under Assumptions A.1–A.4 and Assumption A.6.1,* $\mathrm{Var}(\hat{\Psi}(x;\gamma)) \leq \Theta(\gamma)/(nh^d)$ *as* $n \to \infty$ *where* $\Theta(\gamma) = \bar{\Theta}\left[B_0^2(\gamma)B_{1,Y}^2(\gamma) + B_{2,Y}(\gamma) + 1\right] < \infty$ *for some constant* $\bar{\Theta} < \infty$.

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

Proof. We have

$$\hat{\Psi}(x;\gamma) - \mathrm{E}[\hat{\Psi}(x;\gamma)] = \frac{1}{nh^d}\sum_{i=1}^{n} Z_{n,i}(x,\gamma),$$

where

$$Z_{n,i}(x;\gamma) := K\left(\frac{X_{n,i}(\gamma)-x}{h}\right) Y_{n,i}(\gamma) - \mathrm{E}\left[K\left(\frac{X_{n,i}(\gamma)-x}{h}\right) Y_{n,i}(\gamma)\right]. \tag{A.1}$$

Thus,

$$\begin{aligned} nh^d \mathrm{Var}(\hat{\Psi}(x;\gamma)) &= nh^d \,\mathrm{E}\left[\left(\frac{1}{nh^d}\sum_{i=1}^{n} Z_{n,i}(x;\gamma)\right)^2\right] \\ &= \frac{1}{nh^d}\,\mathrm{E}\left[\left(\sum_{i=1}^{n} Z_{n,i}(x;\gamma)\right)^2\right] \\ &\le \frac{1}{nh^d}\sum_{i=1}^{n}\sum_{j=1}^{n} \left|\mathrm{E}\left[Z_{n,i}(x;\gamma) Z_{n,j}(x;\gamma)\right]\right|. \end{aligned} \tag{A.2}$$

By following the arguments of Hansen (2008), we establish that

$$\left|\mathrm{E}\left[Z_{n,i}(x;\gamma) Z_{n,j}(x;\gamma)\right]\right| \le \begin{cases} \bar{\mu}(\gamma) h^d, & |i-j| \le M \\ \left(\mu^2 B_{Y,2}(\gamma) + \bar{\mu}^2(\gamma)\right) h^{2d}, & M < |i-j| \le h^{-d} \\ 6A\bar{\mu}^{2/s}(\gamma)\,|i-j|^{-(2-2/s)}\, h^{2d/s}, & h^d < |i-j| \end{cases},$$

where $\bar{\mu}(\gamma) = \bar{K}^{s-1}\mu B_0(\gamma) B_{Y,1}(\gamma)$. Plugging these bounds into the right-hand side of equation (A.2) establishes the result. ■

Next, we state a triangular version of Liebscher (1996, Thm. 2.1).

LEMMA A.2. *Let $Z_{n,i}$ be a zero-mean triangular array such that $|Z_{n,i}| \le b_n$ with strong mixing coefficients $\alpha_n(i)$. Then for any $\varepsilon > 0$ and $m \le n$ with $m < \varepsilon/(4b_n)$,*

$$P\left(\left|\sum_{i=1}^{n} Z_{n,i}\right| > \varepsilon\right) \le 4\exp\left[\frac{\varepsilon^2}{64\sigma_{n,m}^2 n/m + 8/3\varepsilon b_n m}\right] + 4\frac{n}{m}\alpha_n(m),$$

where $\sigma_{n,m}^2 = \mathrm{E}\left[\sum_{i=1}^{n} Z_{n,i}\right]$.

In Hansen (2008, p. 739), Lemma A.2 is stated for the case of a stationary sequence Z_i, but the original result holds for triangular mixing arrays without imposing stationarity.

Proof of Theorem 1. Define

$$a_n = \sqrt{\frac{\log(n)}{nh^d}}, \qquad \tau_n = a_n^{-1/(s-1)}, \qquad A_n = \left\{(x,\gamma) \in \mathbb{R}^d \times \Gamma : \|x\| \le c_n, \|\gamma\| \le d_n\right\}.$$

We follow the same steps as in Hansen (2008, proof of Thm. 2). Write

$$\begin{aligned} \hat{\Psi}(x;\gamma) &= \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{X_{n,i}(\gamma)-x}{h}\right) Y_{n,i}(\gamma)\,\mathbb{I}\left\{\left|Y_{n,i}(\gamma)\right| \le \tau_n\right\} \\ &\quad + \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{X_{n,i}(\gamma)-x}{h}\right) Y_{n,i}(\gamma)\mathbb{I}\left\{\left|Y_{n,i}(\gamma)\right| > \tau_n\right\} \\ &= \hat{\Psi}_1(x;\gamma) + \hat{\Psi}_2(x;\gamma), \end{aligned}$$

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

where the second term satisfies

$$\begin{aligned}
\mathrm{E}\left[\left|\hat{\Psi}_2(x;\gamma)\right|\right] &\leq \tau_n^{-(s-1)} \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^d} |K(u)| \sup_{A_n} \mathrm{E}\left[\left|Y_{n,i}(\gamma)\right|^s | X_{n,i}(\gamma) = x - hu\right] \\
&\quad \times f_{n,i}(x-hu;\gamma)\,du \\
&\leq \tau_n^{-(s-1)} \int_{\mathbb{R}^d} |K(u)|\,du \\
&\quad \times \left\{ \sup_{1\leq i\leq n} \sup_{A_n} \mathrm{E}\left[\left|Y_{n,i}(\gamma)\right|^s | X_{n,i}(\gamma) = z\right] f_{n,i}(z;\gamma) \right\} \\
&\leq \mu\,\tau_n^{-(s-1)} \sup_{\|\gamma\|\leq d_n} B_{Y,1}(\gamma) \\
&= O\left(a_n d_n^{\lambda}\right)
\end{aligned}$$

such that by Markov's inequality, $\sup_{x,|\gamma|\leq d_n} \left|\hat{\Psi}_2(x;\gamma)\right| = O_P(a_n d_n^{\lambda})$. We can therefore restrict our attention to $\hat{\Psi}_1(x)$, and we will in what follows call this $\hat{\Psi}(x;\gamma)$ with $Y_{n,i}(\gamma)$ bounded.

We split up the set A_n into $N \leq c_n^d d_n^k h^{-d-k} a_n^{-d-k}$ balls of the form

$$A_{n,j} = \left\{(x,\gamma) : \|x - x_j\| \leq h a_n, \qquad \|\gamma - \gamma_j\| \leq h a_n\right\}.$$

Define $K^*(z)$ as in Hansen (2008, proof of Thm. 2) and $K^{**}(z) := |K(z)|$. Then

$$\begin{aligned}
&\left| Y_{n,i}(\gamma) K\left(\frac{X_{n,i}(\gamma) - x}{h}\right) - Y_{n,i}(\gamma_j) K\left(\frac{X_{n,i}(\gamma_j) - x_j}{h}\right)\right| \\
&\quad \leq \left| Y_{n,i}(\gamma_j) K\left(\frac{X_{n,i}(\gamma) - x}{h}\right) - Y_{n,i}(\gamma_j) K\left(\frac{X_{n,i}(\gamma_j) - x}{h}\right)\right| \\
&\qquad + \left| Y_{n,i}(\gamma) K\left(\frac{X_{n,i}(\gamma_j) - x}{h}\right) - Y_{n,i}(\gamma_j) K\left(\frac{X_{n,i}(\gamma_j) - x}{h}\right)\right| \\
&\qquad + \left| Y_{n,i}(\gamma_j) K\left(\frac{X_{n,i}(\gamma_j) - x}{h}\right) - Y_{n,i}(\gamma_j) K\left(\frac{X_{n,i}(\gamma_j) - x_j}{h}\right)\right| \\
&\quad =: S_1 + S_2 + S_3,
\end{aligned}$$

where

$$\begin{aligned}
S_1 &\leq \left|Y_{n,i}(\gamma_j)\right| \left\|X_{n,i}(\gamma) - X_{n,i}(\gamma_j)\right\| \frac{1}{h} K^*\left(\frac{X_{n,i}(\gamma_j) - x}{h}\right) \\
&\leq \frac{\|\gamma - \gamma_j\|}{h} \left|Y_{n,i}(\gamma_j)\right| \left\|\dot{X}_{n,i}(\gamma_j)\right\| K^*\left(\frac{X_{n,i}(\gamma_j) - x}{h}\right),
\end{aligned}$$

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

$$S_2 \leq \left\| Y_{n,i}(\gamma) - Y_{n,i}(\gamma_j) \right\| K^{**}\left(\frac{X_{n,i}(\gamma_j) - x}{h}\right)$$

$$\leq \left\| \gamma - \gamma_j \right\| \left\| \dot{Y}_{n,i}(\gamma_j) \right\| K^{**}\left(\frac{X_{n,i}(\gamma_j) - x}{h}\right),$$

$$S_3 \leq \frac{\left\| x - x' \right\|}{h} \left| Y_{n,i}(\gamma_j) \right| K^{*}\left(\frac{X_{n,i}(\gamma_j) - x_j}{h}\right).$$

Defining

$$\hat{\Psi}_1^*(x;\gamma) = \left(nh^d\right)^{-1} \sum_i Y_{n,i}(\gamma) \dot{X}_{n,i}(\gamma_j) K^*\left(\frac{X_{n,i}(\gamma) - x}{h}\right),$$

$$\hat{\Psi}_2^*(x;\gamma) = \left(nh^d\right)^{-1} \sum_i Y_{n,i}(\gamma) K^*\left(\frac{X_{n,i}(\gamma) - x}{h}\right),$$

$$\hat{\Psi}^{**}(x;\gamma) = \left(nh^d\right)^{-1} \sum_i \dot{Y}_{n,i}(\gamma_j) K^{**}\left(\frac{X_{n,i}(\gamma) - x}{h}\right),$$

we therefore obtain by the same arguments as in Hansen (2008, proof of Thm. 2)

$$\begin{aligned}\sup_{A_{n,j}} \left| \hat{\Psi}(x,\gamma) - \hat{\Psi}(x_j,\gamma_j) \right| \leq{}& \left| \hat{\Psi}(x;\gamma_j) - \mathrm{E}[\hat{\Psi}(x;\gamma_j)] \right| + \left| \hat{\Psi}_1^*(x,\gamma_j) - \mathrm{E}[\hat{\Psi}_1^*(x;\gamma_j)] \right| \\ &+ \left| \hat{\Psi}_2^*(x,\gamma_j) - \mathrm{E}[\hat{\Psi}_2^*(x;\gamma_j)] \right| \\ &+ \left| \hat{\Psi}^{**}(x;\gamma_j) - \mathrm{E}[\hat{\Psi}^{**}(x;\gamma_j)] \right| \\ &+ 2a_n \left(\mathrm{E}[\hat{\Psi}_1^*(x;\gamma_j)] + \mathrm{E}[\hat{\Psi}_2^*(x;\gamma_j)] + \mathrm{E}[\hat{\Psi}^{**}(x;\gamma_j)] \right).\end{aligned}$$

The last term is of order $O(a_n d_n^{\lambda})$ because, by using the same arguments as in the proof of Lemma A.1, $\mathrm{E}[\bar{\Psi}_j^*(x;\gamma)] \leq \sqrt{\Theta_j^*(\gamma)/(nh^d)}$ and $\mathrm{E}[\bar{\Psi}^{**}(x;\gamma)] \leq \sqrt{\Theta^{**}(\gamma)/(nh^d)}$, where $\Theta_j^*(\gamma) = O(d_n^{2\lambda})$ and $\Theta^{**}(\gamma) = O(d_n^{2\lambda})$. We now show that the first term converges with the claimed rate; the three other terms involving $\bar{\Psi}_1^*$, $\bar{\Psi}_2^*$, and $\bar{\Psi}^{**}$ are treated in the same way because $W = Y\dot{X}$ and $\dot{Y}$ satisfy the same conditions as $W = Y$. With $Z_{n,i}(x;\gamma)$ defined in Lemma A.1,

$$\left| Z_{n,i}(x,\gamma) \right| \leq b_n := C\tau_n, \qquad \mathrm{E}\left[\left(\sum_{i=1}^n Z_{n,i}(x,\gamma) \right)^2 \right] \leq \Theta(\gamma) n h^d$$

such that, with $m = a_n^{-1}\tau_n^{-1}$, Lemma A.2 yields

$$\begin{aligned}P\left(\left| \sum_{i=1}^n Z_{n,i}(x;\gamma) \right| > M a_n d_n^{\lambda} n h^d \right) &\leq 4\exp\left[\frac{M^2 \log(n) d_n^{2\lambda}}{64\Theta(\gamma) + 6\bar{K} M d_n^{\lambda}} \right] + 4\frac{n}{m}\alpha_n(m) \\ &\leq 4n^{-M(64+6\bar{K})} + 4An a_n^{1+\beta}\tau_n^{1+\beta},\end{aligned}$$

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

where the second inequality follows by choosing $M \geq \Theta(\gamma)/d_n^{2\lambda}$ where

$$\sup_{\|\gamma\| \leq d_n} \frac{\Theta(\gamma)}{d_n^{2\lambda}} \leq \bar{\Theta} \sup_{\|\gamma\| \leq d_n} \frac{B_0^2(\gamma) B_{Y,1}^2(\gamma) + 1}{d_n^{2\lambda}}$$

$$\leq \bar{\Theta} \sup_{\|\gamma\| \leq d_n} \frac{\bar{B}_0^2 \bar{B}_{Y,1}^2 \|\gamma\|^{2\lambda} + 1}{d_n^{2\lambda}} = O(1), \qquad n \to \infty.$$

In total,

$$P\left(\sup_{A_n} \left| \hat{\Psi}(x;\gamma) - \mathrm{E}\left[\hat{\Psi}(x;\gamma) \right] \right| > 3Ma_n \right) = O(T_1) + O(T_2),$$

where

$$T_1 = c_n^d d_n^k h^{-(d+k)} a_n^{-(d+k)} n^{-M(64+6\bar{K})}, \qquad T_2 = c_n^d d_n^k h^{-(d+k)} a_n^{1+\beta-(d+k)} \tau_n^{1+\beta} n.$$

We can now follow the same arguments as in Hansen (2008, proof of Thm. 2) to obtain that both of these are $o(1)$, because we have specified c_n and d_n to have the same order and θ has been chosen accordingly.

The second part follows along the same lines as Hansen (2008, proof of Thm. 3) by extending his arguments in the same manner as we have extended those of Hansen (2008, proof of Thm. 2). ■

Proof of Theorem 2. We first verify that Assumptions A.1–A.5 hold. By Orbe et al. (2005, Lem. A.4), $\{W_{n,i}\}$ is strongly mixing with geometrically decreasing mixing coefficients. Thus, Assumption A.1 holds with $Y_{n,i} = W_{n,i} W_{n,i-1}$ and $Y_{n,i} = W_{n,i-1}^2$, respectively, and $X_{n,i} \sim$ i.i.d. $U[0,1]$ with $\beta = +\infty$. Assumption A.2 is satisfied with $\lambda = 0$ and any $s > 2$ as long as $\mathrm{E}\left[|\varepsilon_i|^{2s} \right] < \infty$. The densities $f_{n,i}(x) = \mathbb{I}\{0 \leq x \leq 1\}$ and $f_{n,ij}(x,y) = \mathbb{I}\{0 \leq x \leq 1, \quad 0 \leq y \leq 1\}$, and

$$\mathrm{E}\left[\left| W_{n,i} W_{n,i-1} \right| \left| X_{n,i} = x \right. \right] = \left| a(x) \right| \mathrm{E}\left[W_{n,i-1}^2 \right],$$

$$\mathrm{E}\left[\left(W_{n,i} W_{n,i-1} \right) \left(W_{n,j} W_{n,j-1} \right) \left| X_{n,i} = x, X_{n,j} = y \right. \right]$$

$$= \left| a(x) \right| \left| a(y) \right| \mathrm{E}\left[W_{n,i-1}^2 W_{n,j-1} \right],$$

which are both bounded such that Assumptions A.4 and A.5 hold. From Theorem 1, we then obtain $\sup_{\tau \in [0,1]} \left| \hat{\Psi}_k(\tau) - \mathrm{E}[\hat{\Psi}_k(\tau)] \right| = O_P(\log(n)/(nh))$, $k = 1, 2$. Next, by Dahlhaus (1999), $\Lambda_{n,i} = \mathrm{E}\left[W_{n,i}^2 \right] = \Lambda(i/n) + o(n^{-1})$, where $\Lambda(i/n) := \sigma^2/(1 - a^2(i/n))$. We therefore obtain, using the same arguments as in, for example, Cai (2007), that uniformly over the interval $\tau \in [b, 1-b]$,

$$\mathrm{E}[\hat{\Psi}_1(\tau)] = \sum_{i=1}^{n} K\left(\frac{i/n - \tau}{h} \right) a(i/n) \left[\Lambda(i/n) + o\left(n^{-1} \right) \right] = \Psi_1(\tau) + O\left(h^r \right),$$

$$\mathrm{E}[\hat{\Psi}_2(\tau)] = \sum_{i=1}^{n} K\left(\frac{i/n - \tau}{h} \right) \left[\Lambda(i/n) + o\left(n^{-1} \right) \right] = \Psi_2(\tau) + O\left(h^r \right),$$

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press

where $\Psi_1(\tau) = a(\tau)\Lambda(\tau)$ and $\Psi_k(\tau) = \Lambda(\tau)$. By the mean-value theorem,

$$\left|\hat{a}(\tau) - a(\tau)\right| \leq \frac{\left|\hat{\Psi}_1(\tau) - \Psi_1(\tau)\right|}{\bar{\Psi}_2(\tau)} + \frac{\left|\bar{\Psi}_1(\tau)\right|}{\bar{\Psi}_2^2(\tau)}\left|\hat{\Psi}_2(\tau) - \Psi_2(\tau)\right|,$$

where $\bar{\Psi}_k(\tau) \in [\Psi_k(\tau), \hat{\Psi}_k(\tau)]$. Because $\tau \mapsto \Psi_2(\tau)$ is continuous and positive, $\Psi_{2,\min} := \inf_{\tau\in[0,1]}\Psi_2(\tau) > 0$, and $\inf_{\tau\in[b,1-b]}|\hat{\Psi}_2(\tau) - \Psi_2(\tau)| = O_P(1)$, $\bar{\Psi}_2(\tau) \geq \Psi_{2,\min}/2$ almost surely as $n \to \infty$. Similarly, $\left|\bar{\Psi}_1(\tau)\right| \leq \sup_{\tau\in[0,1]}\left|\Psi_1(\tau)\right|/2$ almost surely. This proves the result. ■

Proof of Theorem 3. We check that Assumptions A.1–A.5 hold with $Y_{n,i} = 1$ under the conditions imposed on the Markov chain: The Doeblin condition implies strong geometric mixing (cf. Holden, 2000), and the uniform continuity of $\partial^r p(y|x)/\partial y^r$ implies the same property of $\partial^r p_i(y|x)/\partial y^r$ by the recursion formula (4). Finally, $\|y\|^q f(y)$ being bounded at all y implies that $\|y\|^q p_{s+i}(y|x)$ is bounded uniformly over $i \geq 1$ and x, where s is defined in equation (5):

$$\begin{aligned}\|y\|^q p_{s+i}(y|x) &\leq \|y\|^q \left|p_{s+i}(y|x) - f(y)\right| + \|y\|^q f(y)\\ &\leq \left\{M(1-\rho)^{1-i} + 1\right\}\|y\|^q f(y),\end{aligned}$$

where the second inequality follows by equation (6). We can here disregard the first s transition densities because these will not have an impact asymptotically as $n \to \infty$. Theorem 1 now supplies us with the desired bound for the variance component.

Next, by standard arguments,

$$\mathrm{E}[\hat{f}(x)] = \frac{1}{n}\sum_{i=1}^{n}\int K(z)\, p_i(x + zh|y)\, dz = \frac{1}{n}\sum_{i=1}^{n} p_i(x|y) + O\left(h^r\right)$$

uniformly over x and for any given initial value y. Because of the Doeblin condition, there exist constants $M < \infty$ and $\rho < 1$ such that uniformly over x,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}|p_i(x|y) - f(x)| &\leq \frac{1}{n}\sum_{i=1}^{s}|p_i(x|y) - f(x)| + \sup_{z\in\mathbb{R}^d} f(z)\\ &\quad \times \frac{1}{n}\sum_{i=1}^{n-s}\frac{\left|p_{s+i}(x|y) - f(x)\right|}{f(x)}\\ &\leq \frac{1}{n}\sum_{i=1}^{s}|p_i(x|y) - f(x)| + \sup_{z\in\mathbb{R}^d} f(z) \times \frac{M}{n}\sum_{i=1}^{n-s}(1-\rho)^{-i}\\ &= O(1/n),\end{aligned}$$

where the second inequality again follows by equation (6). ■

https://doi.org/10.1017/S0266466609090744 Published online by Cambridge University Press