Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2008-37

Uniform Convergence Rates of Kernel Estimators with Heterogenous, Dependent Data

Dennis Kristensen

School of Economics and Management
University of Aarhus
Building 1322, DK-8000 Aarhus C
Denmark

Aarhus School of Business
University of Aarhus

Handelshøjskolen
Aarhus Universitet

UNIVERSITY OF COPENHAGEN

Uniform Convergence Rates of Kernel Estimators with Heterogenous, Dependent Data*

Dennis Kristensen†
Columbia University and CREATES‡

May 2008

Abstract

The main uniform convergence results of Hansen (2008) are generalized in two directions: Data is allowed to (i) be heterogenously dependent and (ii) depend on a (possibly unbounded) parameter. These results are useful in semiparametric estimation problems involving time-inhomogenous models and/or sampling of continuous-time processes. The usefulness of these results are 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 intialized at its stationary distribution.

Keywords: Nonparametric estimation; uniform consistency; kernel estimation; density estimation; heterogeneous time series.

JEL-Classification: C14, C32.

*I wish to thank Bruce Hansen for helpful discussions and suggestions.

†E-mail: dk2313@columbia.edu.

‡Center for Research in Econometric Analysis of Time Series, CREATES, is funded by The Danish National Research Foundation.

1 Motivation and Main Results

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 is stationary and strong mixing. In this note, we extend his results in two directions: First, we allow for heterogenous data where the random variables are nonidentically distributed but still mixing. Second, data can potentially depend on a parameter and we show uniform convergence also over both the parameter set. The main conclusion of this note is that as long as the mixing coefficients and suitably 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 is non-stationary (but mixing) and the distributions potentially vary with the sample size. Three specific examples where this situation arises are (i) time-varying models (Dahlhaus et al., 1999; Cai, 2007; Kristensen, 2008a); (ii) continuous-time stochastic processes sampled at discrete time points where time between observations shrink (Bandi and Phillips, 2003; Kristensen, 2008b); and (iii) Markov chains that have not been intialised 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 index models (Li and Wooldridge, 2002). Another application can be found in the recent literature on nonparametric simulated maximum-likelihood estimation (Fermanian and Salanie, 2004; Kristensen and Shin, 2008).

In this Section, we present the main results. In Section 2, the usefulness of these is demonstrated by two examples: The first is to kernel regression estimation in the locally stationary AR(1) model, and the second is kernel estimation of the stationary density of a homogenous Markov chain that has not been intialised in the stationary distribution. All proofs and lemmas can be found in Section 3.

Let $(Y_{n,i}(\gamma), X_{n,i}(\gamma)) \in \mathbb{R}\times\mathbb{R}^d$, $i = 1,...,n$, $n = 1,2,....$, be a triangular array of random variables and $\mathcal{F}^k_{n,i}(\gamma) = \mathcal{F}(Y_{n,i}(\gamma), X_{n,i}(\gamma), ..., Y_{n,k}(\gamma), X_{n,k}(\gamma))$ an associated sigma-algebra. The random variables 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. Let $f_{n,i}(x;\gamma)$ and $f_{n,ij}(x, y;\gamma)$ denote the densities of $X_{n,i}$ and $(X_{n,i}, X_{n,j})$. We then define for any triangular array $W_{n,i}(\gamma)$ and for some $M \geq 0$:

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

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

We are now ready to state the following conditions on the data where we write $\sup_{i,n}$ for $\sup_{n\geq 1}\sup_{1\leq i\leq n}$:

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

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

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 functions $\gamma \mapsto Y_{n,i}(\gamma)$ and $\gamma \mapsto X_{n,i}(\gamma)$ are 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,

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

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

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

A.4 The functions defined in Eq. (1)-(2) satisfy

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

for some constants $\bar{B}_0, \bar{B}_{W,3}, \lambda \geq 0$ with $W = Y$, $W = \dot{Y}$ and $Y\dot{X}$.

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

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

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 (A.2) and then require that certain moments of the data and their derivatives exist in (A.3). We allow the parameter space Γ to be unbounded and the functions $B_k(\gamma)$ in (A.4) to be unbounded but with at most polynomial growth. This is particularly relevant in the case of NPSMLE where γ will contain past values of the observed process and one is not willing to assume a compact support, see Kristensen and Shin (2008). When Γ is unbounded we will only be able to give uniform convergence over a (growing compact) subset of Γ which will depend on the polynomial bound λ. The polynomial bounds on the functions $B_0(\gamma)$ and $B_{W,k}(\gamma)$, $k = 1, 2, 3$, are imposed to simplify the uniform convergence results. They can be exchanged for other bounds as discussed below. Finally, (A.5) is a strengthening of the condition on $B_{2,W}(\gamma)$ in (A.4), and corresponds to the condition imposed in Hansen (2008, Eq. 21), but is here required to hold for $W = \dot{Y}$ and $Y\dot{X}$ in addition to $W = Y$.

In the case, where $(Y_{n,i}(\gamma), X_{n,i}(\gamma)) = (Y_i, X_i)$, (A.1)-(A.4) collapse to Assumption 1-2 of Hansen (2008). In particular, we have $\dot{Y} = \dot{X} = \lambda = 0$. However, we 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 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.

Due to heterogeneity, $E[\hat{\Psi}(x;\gamma)]$ does not necessarily have a well-defined limit as $n \to \infty$ under (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

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

as $n \to \infty$, where $m_{n,i}(x;\gamma) = E[Y_{n,i}(\gamma)\,|X_{n,i}(\gamma) = x]$. 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 $E[\hat{\Psi}(x;\gamma)] = f(x;\gamma)\, m(x;\gamma) + o(1)$. The existence of such a limit of $E[\hat{\Psi}(x;\gamma)]$ 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, since the main objective of this note is to bound the variance of $\hat{\Psi}(x;\gamma)$. For that purpose, (A.1)-(A.5) suffice. Similar type of 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.

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 (i) $K(u) = 0$ for $\|u\| > L$ and $|K(u) - K(u')| \leq \Lambda_1 \|u - u'\|$, or (ii) $K(u)$ 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 $\|u\| \geq L$.

Assumption (A.6.1) and (A.6.2) are identical to Hansen (2008, Assumption 3) and Hansen (2008, Eq. 22) respectively.

Theorem 1 *Assume that (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*

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

(i)

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

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

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

Remark 2

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

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

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

3. *If* $B_{k,W}(\gamma)$, $k = 0,1,2,3$, *are all uniformly bounded, then Theorem 1(ii) can be shown to hold uniformly over* $\gamma \in \Gamma$.

2 Two Examples

We here present two applications of Theorem 1 when data is not parameter-dependent. For an application of Theorem 1 with parameter-dependence, we refer to Kristensen and Shin (2008). We here implicitly assume that the kernel satisfies (A.6.1)-(A.6.2) and is of order $r \ge 1$: $\int K(z) z^i dz = 0$, $i = 1, ..., r-1$, and $\int K(z) |z|^r dz < \infty$.

The first example where Theorem 1 becomes relevant is in the nonparametric estimation of time-varying regression models. We here consider the following sequence of time-varying AR(1) models,

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

where ε_i are assumed to be i.i.d.$(0, \sigma^2)$ with $E[|\varepsilon_t|^s] < \infty$, $s > 4$. Here, the distribution of $W_{n,i}$ obviously both depends on i and n. Orbe et al. (2007, Lemma 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 (A.1) and $\sup_{i,n} E\left[|W_{n,i}|^s\right] < \infty$.

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 $X_{n,i} = i/n$ (which for each $n \geq 1$ corresponds to i.i.d. draws from $U[0,1]$), we can now apply Theorem 1 to both $\hat{\Psi}_1(\tau)$ and $\hat{\Psi}_2(\tau)$. However, due to boundary problems, the above estimator will be asymptotically biased at each of the two endpoints of the time series, $\tau = 0$ and $\tau = 1$. We handle this by 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 since these will not suffer from boundary biases, c.f. Cai (2007) and Chen (2000) respectively, but we here maintain the above estimator for simplicity.

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

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

The second example is where $\{X_i\}$ is a d-dimensional, time-homogenous Markov chain with transition density $p(y|x)$, $P(X_{i+1} \in A | X_i) = \int_A p(y|x)\, dy$. We assume that the Markov chain is mixing such that a stationary marginal density, $f(x)$, exists. We then wish to estimate $f(x)$. If the observed sequence was not intialised at $f(x)$, then it is non-stationary. For example, if $X_0 = x$ for some given x then $P(X_i \in A) = \int_A p_i(y|x)\, dy$, $i = 1, 2, \ldots$, where

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

with $p_1(y|x) = p(y|x)$. However, the distribution will converge towards the stationary one in the total variation norm, $\sup_A \left|\int_A p_i(y|x)\, dy - \int_A f(x)\, dx\right| \to 0$, $i \to \infty$, with geometric rate, c.f. Meyn and Tweedie (1993). So asymptotically we are able to recover the stationary density pointwise by

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

To obtain uniform convergence of the transition density (and thereby uniform convergence of $\hat{f}$) towards f however, convergence in the total variation norm does not suffice. To obtain this, we here impose the Strong Doeblin Condition, see Holden (2000):

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

Theorem 4 *Assume that the Markov chain $\{X_i\}$ satisfies the strong Doeblin condition (5), and its transition density $y \mapsto p(y|x)$ is $r \geq 1$ times differentiable $\partial^r p(y|x)/\partial y^r$ being uniformly continuous for all x. Also, $\|x\|^q f(x)$ is bounded for some $q \geq d$. Then:*

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

3 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, Theorem 1):

Lemma 5 *Under (A.1)-(A.4) and (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 $\bar{\Theta} < \infty$.*

Proof. We have

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

where

$$Z_{n,i}(x;\gamma) := K\left(\frac{X_{n,i}(\gamma) - x}{h}\right)Y_{n,i}(\gamma) - E\left[K\left(\frac{X_{n,i}(\gamma) - x}{h}\right)Y_{n,i}(\gamma)\right]. \qquad (6)$$

Thus,

$$\begin{aligned} nh^d \mathrm{Var}(\hat{\Psi}(x;\gamma)) &= nh^d E\left[\left(\frac{1}{nh^d}\sum\nolimits_{i=1}^n Z_{n,i}(x;\gamma)\right)^2\right] \\ &= \frac{1}{nh^d}E\left[\left(\sum\nolimits_{i=1}^n Z_{n,i}(x;\gamma)\right)^2\right] \\ &\leq \frac{1}{nh^d}\sum_{i=1}^n\sum_{j=1}^n |E[Z_{n,i}(x;\gamma)Z_{n,j}(x;\gamma)]|, \qquad (7) \end{aligned}$$

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

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

where $\bar{\mu}(\gamma) = \bar{K}^{s-1}\mu B_0(\gamma)B_{1,Y}(\gamma)$. Plugging these bounds into the right hand side of Eq. (7) establishes the result. ■

Next, we state a triangular version of Liebscher (1996, Theorem 2.1):

Lemma 6 *Let $Z_{n,i}$ be a zero-mean triangular array such that $|Z_{n,i}| \leq b_n$ with strong mixing coefficients $\alpha_n(i)$. Then for any $\varepsilon > 0$ and $m \leq n$ with $m < \varepsilon b_n/4$:*

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

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

In Hansen (2008, p. 739), Lemma 6 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

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

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

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

where the second term satisfies

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

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 will in the following 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

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

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

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

where

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

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

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

Defining

$$\begin{aligned}
\hat{\Psi}_1^*(x;\gamma) &= \left(nh^d\right)^{-1} \sum_i Y_{n,i}(\gamma) \dot{X}_{n,i}(\gamma_j) K^*\left(\frac{X_{n,i}(\gamma)-x}{h}\right), \\
\hat{\Psi}_2^*(x;\gamma) &= \left(nh^d\right)^{-1} \sum_i Y_{n,i}(\gamma) K^*\left(\frac{X_{n,i}(\gamma)-x}{h}\right), \\
\hat{\Psi}^{**}(x;\gamma) &= \left(nh^d\right)^{-1} \sum_i \dot{Y}_{n,i}(\gamma_j) K^{**}\left(\frac{X_{n,i}(\gamma)-x}{h}\right),
\end{aligned}$$

we therefore obtain by the same arguments as in Hansen (2008, Proof of Theorem 2),

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

The last terms is of order $O\left(a_n d_n^{\lambda}\right)$ since, by using the same arguments as in the Proof of Lemma 5, $E[\bar{\Psi}_k^*(x;\gamma)] \leq \sqrt{\Theta_k^*(\gamma)/(nh^d)}$ and $E[\bar{\Psi}^{**}(x;\gamma)] \leq \sqrt{\Theta^{**}(\gamma)/(nh^d)}$ where $\Theta_k^*(\gamma) = O\left(d_n^{2\lambda}\right)$ and $\Theta^{**}(\gamma) = O\left(d_n^{2\lambda}\right)$. We now show that the first term converges with the claimed rate; the three other ones involving $\bar{\Psi}_1^*$, $\bar{\Psi}_2^*$ and $\bar{\Psi}^{**}$ are treated in the same way since $W = Y\dot{X}$ and $\dot{Y}$ satisfy the same conditions as $W = Y$. With $Z_{n,i}(x;\gamma)$ defined in Lemma 5,

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

such that, with $m = a_n^{-1}\tau_n^{-1}$, Lemma 6 yields:

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

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

$$\sup_{\|\gamma\|\leq d_n}\frac{\Theta(\gamma)}{d_n^{2\lambda}} \leq \bar{\Theta}\sup_{\|\gamma\|\leq d_n}\frac{B_0^2(\gamma)\, B_1^2(\gamma) + 1}{d_n^{2\lambda}} \leq \bar{\Theta}\sup_{\|\gamma\|\leq d_n}\frac{\bar{B}_0^2\bar{B}_1^2\|\gamma\|^{2\lambda} + 1}{d_n^{2\lambda}} = O(1), \quad n\to\infty.$$

In total,

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

where

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

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

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

Proof of Theorem 3. We first verify that (A.1)-(A.5) hold. By Orbe et al. (2004, Lemma A.4), $\{W_{n,i}\}$ is strongly mixing with geometrically decreasing mixing coefficients. Thus, (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 $X_{n,i}$ ~i.i.d. $U[0,1]$ with $\beta = +\infty$. (A.2) is satisfied with $\lambda = 0$ and any $s > 2$ as long as $E\left[|\varepsilon_i|^{2s}\right] < \infty$. The densities $f_{n,i}(x) = \mathbb{I}\{0 \leq x \leq 1\}$ and $f_{n,ij}(x,y) = \mathbb{I}\{0 \leq x \leq 1, 0 \leq y \leq 1\}$, and

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

which are both bounded such that (A.4)-(A.5) hold. From Theorem 1, we then obtain $\sup_{\tau\in[0,1]}\left|\hat{\Psi}_k(\tau)-E[\hat{\Psi}_k(\tau)]\right| = O_P(\log(n)/(nh))$, $k=1,2$. Next, by Dahlhaus (1996a), $\Lambda_{n,i} = E\left[W_{n,i}^2\right] = \Lambda(i/n)+o\left(n^{-1}\right)$, where $\Lambda(i/n) := \sigma^2/\left(1-a^2(i/n)\right)$. We therefore obtain, using the same arguments as in e.g. Cai (2007), that uniformly over the interval $\tau\in[b,1-b]$,

$$
\begin{aligned}
E[\hat{\Psi}_1(\tau)] &= \sum_{i=1}^{n} K\left(\frac{i/n-\tau}{h}\right) a(i/n)\left[\Lambda(i/n)+o\left(n^{-1}\right)\right] = \Psi_1(\tau)+O(h^r), \\
E[\hat{\Psi}_2(\tau)] &= \sum_{i=1}^{n} K\left(\frac{i/n-\tau}{h}\right)\left[\Lambda(i/n)+o\left(n^{-1}\right)\right] = \Psi_2(\tau)+O(h^r),
\end{aligned}
$$

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

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

where $\bar{\Psi}_k(\tau)\in[\Psi_k(\tau),\hat{\Psi}_k(\tau)]$. Since $\tau\mapsto\Psi_2(\tau)$ 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, $\left|\bar{\Psi}_1(\tau)\right|\leq\sup_{\tau\in[0,1]}|\Psi_1(\tau)|/2$ almost surely. This proves the result. ■

Proof of Theorem 4. It is easily checked that (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, c.f. 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, $\|x\|^q f(x)$ being bounded implies that $\|x\|^q p_i(x|y)$ is bounded for all $i\geq 1$ and y. This gives us the desired bound for the variance component. Next, by standard arguments,

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

uniformly over x and for any given initial value y. Due to the Doeblin condition, there exists constants $M<\infty$ and $\rho<1$ such that uniformly in x,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}|p_i(x|y)-f(x)| &\leq \sup_z f(z)\times\frac{1}{n}\sum_{i=1}^{n}\frac{|p_i(x|y)-f(x)|}{f(x)} \leq \sup_z f(z)\times\frac{M}{n}\sum_{i=1}^{n}(1-\rho)^{-i} \\
&= O(1/n),
\end{aligned}
$$

where the last inequality follows by Holden (2000, Theorem 1). ■

References

Andrews, D.W.K. (1995) Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory* 11, 560-596.

Bandi, F. and P.C.B. Phillips (2003) Fully Nonparametric Estimation of Scalar Diffusion Models. *Econometrica* 71, 241–83.

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.

Dahlhaus, R., M.H. Neumann and R. von Sachs (1999) Nonlinear Wavelet Estimation of Timevarying Autoregressive Processes. *Bernoulli* 5, 873-906.

Fermanian, J.-D. and 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.

Kim, T.Y. and S. Lee (2004) Kernel Density Estimator for Strong Mixing Processes. *Journal of Statistical Planning and Inference* 133, 273-284.

Kristensen, D. (2008a) Estimation and Testing in Non- and Semiparametric Time-varying Regressions. Manuscript, Department of Economics, Columbia University.

Kristensen, D. (2008b) Nonparametric Filtering of the Realised Spot Volatility: A Kernel-based Approach. Manuscript, Department of Economics, Columbia University.

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

Li, Q. and 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. and R.L. Tweedie (1993) *Markov Chains and Stochastic Stability*. New York: Springer-Verlag.

Orbe, S., E. Ferreira and J. Rodriguez-Poo (2004) Nonparametric Estimation of Time Varying Parameters Under Shape Restrictions. *Journal of Econometrics* 126, 53-77.

Yu, B. (1993) Density Estimation in the L^∞ Norm for Dependent Data with Applications to the Gibbs Sampler. *Annals of Statistics* 21, 711-735.

Research Papers 2008

CREATES
Center for Research in Econometric Analysis of Time Series

2008-23:	Silja Kinnebrock and Mark Podolskij: An Econometric Analysis of Modulated Realised Covariance, Regression and Correlation in Noisy Diffusion Models
2008-24:	Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Small Bandwidth Asymptotics for Density-Weighted Average Derivatives
2008-25:	Mark Podolskij and Mathias Vetter: Bipower-type estimation in a noisy diffusion setting
2008-26:	Martin Møller Andreasen: Ensuring the Validity of the Micro Foundation in DSGE Models
2008-27:	Tom Engsted and Thomas Q. Pedersen: Return predictability and intertemporal asset allocation: Evidence from a bias-adjusted VAR model
2008-28:	Frank S. Nielsen: Local polynomial Whittle estimation covering non-stationary fractional processes
2008-29:	Per Frederiksen, Frank S. Nielsen and Morten Ørregaard Nielsen: Local polynomial Whittle estimation of perturbed fractional processes
2008-30:	Mika Meitz and Pentti Saikkonen: Parameter estimation in nonlinear AR-GARCH models
2008-31:	Ingmar Nolte and Valeri Voev: Estimating High-Frequency Based (Co-) Variances: A Unified Approach
2008-32:	Martin Møller Andreasen: How to Maximize the Likelihood Function for a DSGE Model
2008-33:	Martin Møller Andreasen: Non-linear DSGE Models, The Central Difference Kalman Filter, and The Mean Shifted Particle Filter
2008-34:	Mark Podolskij and Daniel Ziggel: New tests for jumps: a threshold-based approach
2008-35:	Per Frederiksen and Morten Ørregaard Nielsen: Bias-reduced estimation of long memory stochastic volatility
2008-36:	Morten Ørregaard Nielsen: A Powerful Test of the Autoregressive Unit Root Hypothesis Based on a Tuning Parameter Free Statistic
2008-37:	Dennis Kristensen: Uniform Convergence Rates of Kernel Estimators with Heterogenous, Dependent Data