DYNAMIC MISSPECIFICATION IN NONPARAMETRIC COINTEGRATING REGRESSION

By

Ioannis Kasparis and Peter C.B. Phillips

June 2009

COWLES FOUNDATION DISCUSSION PAPER NO. 1700

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

http://cowles.econ.yale.edu/

Dynamic Misspecification in Nonparametric Cointegrating Regression

Ioannis Kasparis
University of Cyprus

and

Peter C. B. Phillips*

Yale University, University of Auckland
University of York & Singapore Management University

March, 2009

Abstract

Linear cointegration is known to have the important property of invariance under temporal translation. The same property is shown not to apply for nonlinear cointegration. The requisite limit theory involves sample covariances of integrable transformations of non-stationary sequences and time translated sequences, allowing for the presence of a bandwidth parameter so as to accommodate kernel regression. The theory is an extension of Wang and Phillips (2008) and is useful for the analysis of nonparametric regression models with a misspecified lag structure and in situations where temporal aggregation issues arise. The limit properties of the Nadaraya-Watson (NW) estimator for cointegrating regression under misspecified lag structure are derived, showing the NW estimator to be inconsistent with a "pseudo-true function" limit that is a local average of the true regression function. In this respect nonlinear cointegrating regression differs importantly from conventional linear cointegration which is invariant to time translation. When centred on the pseudo-function and appropriately scaled, the NW estimator still has a mixed Gaussian limit distribution. The convergence rates are the same as those obtained under correct specification but the variance of the limit distribution is larger. Some applications of the limit theory to non-linear distributed lag cointegrating regression are given and the practical import of the results for index models, functional regression models, and temporal aggregation are discussed.

Keywords: Dynamic misspecification, Functional regression, Integrable function, Integrated process, Local time, Misspecification, Mixed normality, Nonlinear cointegration, Nonparametric regression.

JEL classification: C22, C32

*Partial support from NSF Grant #SES 06-47086 is acknowledged.

1 Introduction

Cointegration methods have been highly popular for more than two decades in the empirical time series literature, particularly in macroeconomics and international finance. The standard cointegrating model used in these empirical studies is linear, usually a parametric vector autoregression (VAR) with reduced rank structure intended to capture the long run relations and with a lag structure designed to deal with transient dynamics. Recent work has begun to consider modifications to these models that introduce a variety of nonlinear specifications. For example, Corradi, Swanson and White (2000), Teräsvirta and Ellianson (2001) and others have introduced nonlinear short-run dynamics into vector error correction models (VECMs) and sought to allow for nonlinear transition mechanisms. But the possibility of nonlinear long-run dynamics has received much less attention.

Park and Phillips (1999, 2001) developed a limit theory for nonlinear transformations of unit root processes that provides a theoretical base for modeling nonlinear long-run relations in a parametric framework (see also Chang, Park and Phillips (2001)). Other recent work (Guerre, 2004; Karlsen, Mykelbust and Tjøstheim, 2007; Schienle, 2008; Wang and Phillips, 2008, 2009;) has provided a limit theory for nonparametric cointegrating regression using Markov chain and local time asymptotics. The current paper takes the Wang and Phillips (2009; hereafter WP) framework and analyzes the effects of misspecification relating to the lag structure of the model. This kind of misspecification is potentially relevant in a variety of contexts and can be especially relevant in situations in which temporal aggregation issues arise.

As shown in Park and Phillips (1999, 2001), the limit theory for nonlinear transformations of integrated processes can be quite different than that which is well known for linear models. Park and Phillips consider two families of nonlinear functions of unit root processes: locally integrable (LI) functions and integrable (I) functions. The linear cointegrating model, for instance, is locally integrable and well studied. Correspondingly, the limit theory for smooth locally integrable models tends to be similar to that of standard cointegrating models. On the other hand the limit theory for integrable models is very different. Sample averages of integrable transformations of unit root time series exhibit a form of weak intensity – even weaker than that of an i.i.d. or stationary time series, which typically carry a signal that is of the same order of magnitude as the sample size n. The explanation for this reduction in intensity is that integrable functions attenuate the effects of large deviations of the process from the origin. Since nonstationary time series like random walks spend much of their time away from the origin, this attenuation leads to an overall reduction in the sample intensity of such functions. In addition, for integrable functions, the limit theory is determined by the local time of the limit process of the standardized time series at some point like the origin, and not by the local time averaged over the whole real line, as in the case of sample functions in the LI family. A typical example of the latter is the sample variance of a unit root process whose limit behavior takes the form of a quadratic functional of Brownian motion which can be rewritten as a spatial integral (a spatial sample variance, in fact) over the whole real line weighted by the local time density process, as explained in Phillips (2001).

In this paper we stress another difference between the two families. LI models are typically invariant to finite lags, at least as far as asymptotic properties are concerned. In other words, cointegrating relations persist across finite temporal shifts in the observations and consistent estimation of these relations applies in the usual way. On the other hand I transformations are not invariant to finite lags. This fact has the following important implication. Contrary to LI models, misspecifing the lag structure in an I regression, can lead to inconsistent estimation. For instance, suppose that the true model is the simple linear in parameters nonlinear cointegrated system

$$y_t = \theta g(x_t) + u_t, \tag{1}$$

where θ is an unknown parameter, Δx_t is $iid\,(0, \sigma_x^2)$ and u_t is some independent $iid\,(0, \sigma_u^2)$ error. In place of (1), suppose that the following dynamically misspecified model is estimated by least squares (LS):

$$y_t = \hat{\theta} g(x_{t-1}) + \hat{u}_t.$$

If the regression function g is continuous and locally integrable it can be shown easily (see, for example, Kasparis 2008, Lemma A1(b)) that the LS estimator in this case

$$\hat{\theta} = \theta \frac{\sum_{t=1}^{n} g(x_t) g(x_{t-1})}{\sum_{t=1}^{n} g(x_{t-1})^2} + o_p(1) = \theta \frac{\sum_{t=1}^{n} g(x_{t-1})^2}{\sum_{t=1}^{n} g(x_{t-1})^2} + o_p(1) = \theta + o_p(1),$$

and so $\hat{\theta}$ is consistent for θ in spite of the lag misspecification, just as in conventional linear cointegrating regression. On the other hand, if the regression function f is integrable then it follows directly from the limit theory of Kasparis, Phillips and Magdalinos (2008) (see also Theorem 1 below) that

$$\hat{\theta} = \theta \frac{\sum_{t=1}^{n} g(x_t) g(x_{t-1})}{\sum_{t=1}^{n} g(x_{t-1})^2} + o_p(1) = \theta \frac{\mathbf{E} \int_{-\infty}^{\infty} g(s) g(s + \Delta x_t) ds}{\int_{-\infty}^{\infty} g(s)^2 ds} + o_p(1),$$

and $\hat{\theta}$ is inconsistent. Thus, small issues of lag specification and timing do matter in nonlinear nonstationary regression.

One of the main results of the present paper is to show that the Nadaraya-Watson (NW) kernel estimator $\hat{f}(x)$ of $f(x) = \theta g(x)$ exhibits this kind of inconsistency due to the use of integrable functions in the construction of the kernel regression function. In fact, it will be shown that, under certain regularity conditions and this type of dynamic mistiming, the NW estimator converges to a pseudo-true function of the following form

$$\hat{f}(x) \xrightarrow{p} \mathbf{E} f(x + \Delta x_t),$$

involving a functional of f (Theorem 2 and (13) below). Thus, the effect of the lag misspecification is to induce a shift in the limit, based on a local average of the function around the regression point x. In addition, the NW estimator, when centred on the pseudo-true function and appropriately scaled, has a mixed Gaussian limit distribution. The convergence rates are the same as those reported by WP. Nevertheless, the variance of the limit distribution is larger than that obtained under correct specification.

This kind of dynamic induced inconsistency arises in many other cases where the model and estimation procedure involves integrable functions and timing issues are relevant in specification. For example, the maximum likelihood estimator of discrete choice models involves integrable functions (see Park and Phillips, 2000) and will be similarly subject to the effects of dynamic specification error. Issues of timing in dynamic specification are likely to be particularly important in market intervention models of the type studied in Hu and Phillips (2004).

We start the analysis by providing a basic limit result, useful for the analysis of misspecified non-parametric models. We consider sample covariances of functions of non-stationary sequences and non-contemporaneous integrable functions of such sequences. A bandwidth parameter is permitted in the integrable functions, thereby making the resultant limit theory relevant in non parametric estimation. The limit result given here extends some of the theory of WP and makes substantial use of that framework. WP consider sample sums of integrable tranformations of non-stationary time series that involve a bandwidth sequence and apply their theory to nonparametric nonstationary regression with correctly specified lag structure. Our work is also related to Kasparis, Phillips and Magdalinos (2008), who consider parametric IV estimation of models with integrable functions where no bandwidth elements are involved.

The WP limit theory has also been extended by Phillips (2009) in a different direction where the focus is spurious non-parametric regression. That work provides a limit theory for the sample covariance of a non-stationary sequence and a kernel function of another (and possibly unrelated) nonstationary sequence. It is indirectly related to the current paper because some similar sample covariances arise in the limit theory.

The remainder of the paper is organized as follows. Section 2 provides the model framework, assumptions and some preliminary theory. Section 3 gives the main results. Section 4 provides some applications in contexts of interest for applied work, and Section 5 concludes. Technical results and proofs are given in the Appendices.

2 Theoretical framework and preliminary results

We assume that the time series $\{y_t\}_{t=1}^n$ is generated by the model:

$$y_t = f(x_{t-r}) + u_t, \text{ for some integer lag } r \geq 0. \tag{2}$$

where f satisfies certain convolution integrability conditions given later (in particular, Assumption 2.1(c) below). The regressor x_t is a nonstationary process and u_t is a martingale difference sequence, respectively, both defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. For example, in many applications it will be sufficient for $\{x_t\}_{t=1}^n$ to be generated as a unit root process or as a near integrated array of the commonly used form

$$x_t = \rho_n x_{t-1} + v_t, \; x_0 = O_p(1), \tag{3}$$

with $\rho_n = 1 - \frac{c}{n}$ for some constant c. To avoid unnecessary triangular array complications in the development that follows we focus on the unit root generating model for x_t, although our main results continue to hold with minor changes under (3).

We concentrate on the case where a version of (2) is fitted by nonparametric kernel regression. However, the fitted model involves a lag misspecification resulting from incorrect timing, so that the fitted model has the (lag misspecified) form

$$y_t = \hat{f}(x_{t-s}) + \hat{u}_t, \text{ for some fixed integer lag } s \geq 0,\ r \neq s, \tag{4}$$

where $\hat{f}$ is the NW regression estimator defined by

$$\hat{f}(x) = \frac{\sum_{t=s}^{n} K\left(\frac{x_{t-s}-x}{h}\right) y_t}{\sum_{t=s}^{n} K\left(\frac{x_{t-s}-x}{h}\right)}, \tag{5}$$

for some kernel function K.

In order to develop a limit theory for $\hat{f}(x)$ we need to be more specific about the model (2) and its components. The assumptions below are largely based on WP. We start by introducing the following notation used in that work. First, c_n and d_n are sequences of real numbers satisfying $c_n, d_n \to \infty$. The sequence d_n provides a standardization for the nonstationary regressor x_t and is commonly just $d_n = \sqrt{n}$, as in the case of (3). Then, $x_{t,n} = x_t/d_n, 0 \leq t \leq n, n \geq 1$ is a triangular array and the standardization ensures that $x_{t,n}$ has a limit distribution. We also introduce a sequence of real numbers $d_{l,k,n}$ for which $(x_{l,n} - x_{k,n})/d_{l,k,n}$ has a limit distribution as $l - k \to \infty$. When $d_n = \sqrt{n}$, then $d_{l,k,n} = \sqrt{l-k}/\sqrt{n}$, as in WP. The sequence c_n is a secondary sequence which differs from d_n by a bandwidth factor, so that we usually have $c_n = d_n/h_n = \sqrt{n}/h_n$ for some bandwidth sequence $h_n \to 0$ arising in the kernel estimation. As in WP, it is convenient also to use the set notation.

$$\Omega_n(\eta) = \{(l,k) : \eta n \leq k \leq (1-\eta)n,\ k + \eta n \leq l \leq n\}, 0 < \eta < 1.$$

Assumption 2.1

For all $0 \leq k < l \leq n, n \geq 1$, there exist a sequence of constants $d_{l,k,n}$ and a sequence of σ-fields $\mathcal{F}_{k,n}$ (define $\mathcal{F}_{0,n} = \sigma\{\varnothing, \Omega\}$, the trivial σ-field) such that,

(a) for some $p_0 > 0$ and $C > 0$, $\inf_{(l,k)\in\Omega_n(\eta)} d_{l,k,n} \geq \eta^{p_0}/C$ as $n \to \infty$,

$$\lim_{\eta\to 0}\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{[\eta n]}(d_{l,0,n})^{-1} = 0, \tag{6}$$

$$\lim_{\eta\to 0}\lim_{n\to\infty}\frac{1}{n}\sum_{l=[(1-\eta)\,n]}^{n}(d_{l,0,n})^{-1} = 0, \tag{7}$$

$$\lim_{\eta\to 0}\lim_{n\to\infty}\frac{1}{n}\max_{0\leq k\leq(1-\eta)\,n}\sum_{l=k+1}^{k+[\eta\,n]}(d_{l,k,n})^{-1} = 0, \tag{8}$$

$$\limsup_{n\to\infty}\frac{1}{n}\max_{0\leq k\leq n-1}\sum_{l=k+1}^{n}\left(d_{l,k,n}\right)^{-1} < \infty; \tag{9}$$

(b) *x_k is adapted to $\mathcal{F}_{n,k-1}$ and conditional on $\mathcal{F}_{n,k-1}$, $(x_{l,n} - x_{k,n})/d_{l,k,n}$ has density function $h_{l,k,n}(x)$ such that*

(i) $\sup_{l,k}\sup_x h_{l,k,n}(x) = C < \infty$

(ii) *for some* $k_0 > 0$,

$$\sup_{(l,k)\in\Omega_n\left(\delta^{1/(2k_0)}\right)} \sup_{|x|\leq\delta} |h_{l,k,n}(x) - h_{l,k,n}(0)| = o_p(1),$$

when $n \to \infty$ *first and then* $\delta \to 0$.

(c) *Conditional on* $\mathcal{F}_{n,(r\wedge s)-1}$, $x_r - x_s$ *has density function* $p_{r-s}(v)$, *such that*

$$\int_{-\infty}^{\infty} |f(x+v)|\, p_{r-s}(v)dv < \infty,$$

for each $x \in \mathbb{R}$.

Remark. Conditions (6)-(9) hold when x_t is a unit root or a near unit root process. In that case the sequence $d_{l,0,n} = \sqrt{l/n}$. Then, Euler summation gives

$$\lim_{\eta\to 0}\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{[\eta n]}(l/n)^{-1/2} = \lim_{\eta\to 0} 2\int_0^{\eta} s^{1/2}ds = 0,$$

and this establishes (6). Similar arguments validate (7) to (9).

Assumption 2.2.

(a) *There is a sequence of real numbers* $d_n \to \infty$ *such that the process* $x_{[nt],n} := x_{[nt]}/d_n$ *on the Skorohod space* $D[0,1]$, *converges weakly to a Gaussian process* $G(t)$ *that has a continuous local time process* $L_G(t,s)$.

(b) *On a suitable probability space* $\sup_{0\leq t\leq 1}\left|x_{[nt],n} - G(t)\right| = o_p(1)$.

Assumption 2.3. *Set* $0 < \gamma \leq 1$.

(a) $\lim_{n\to\infty} d_n/c_n = 0$, *where* $d_n \to \infty$ is the sequence in Assumption 2.2 (b) and c_n also satisfies $c_n \to \infty$;

(b) *For* n *large enough,* $\left|f\left(\frac{d_n}{c_n}z + x - v\right) - f(x-v)\right| \leq (d_n/c_n)^{\gamma} f_1(z,x,v)$ *with* $\int_v\int_z f_1(z,x,v)\,|g(z)|\,p(v)dzdv < \infty$, *for each* x.

(c) $\int_z |z|\,|g(z)|\,dz$ *and* $\int_v |f(x-v)|\,p_{r-s}(v)\,|v|\,dv < \infty$ *for all* x.

Assumption 2.3*. *Set* $0 < \gamma \leq 1$.

(a) $\lim_{n\to\infty} d_n/c_n = m_0 > 0$,

(b) *for* n *large enough,* $\left|f\left(\frac{d_n}{c_n}z + x - v\right) - f(m_0 z + x - v)\right| \leq \left|\frac{d_n}{c_n} - m_0\right|^{\gamma} f_1(z,x,v)$ *with* $\int_v\int_z f_1(z,x,v)\,|g(z)|\,p(v)dzdv < \infty$, *for each* x.

(c) $\int_s\int_v |f(m_0 z + x - v)\,g(z)|\,p_{r-s}(v)\,(|v| + |z|)\,dvdz < \infty$, *for each* x *and* $m_0 \geq 0$.

Assumptions 2.2 (a) and (b) are the same as Assumptions 2.2 and 2.3 in WP, and Assumptions 2.1 (a) and (b) are similar to Assumption 2.3 of WP. Assumption 2.1 (c) is a simple convolution integrability condition, which is clearly satisfied under suitable majorization, for example whenever the density p_{r-s} is bounded and f is integrable. When $d_n = \sqrt{n}$ and $c_n = \sqrt{n}/h$, Assumption 2.3 (a) requires that the bandwidth sequence $h \to 0$ as $n \to \infty$. By contrast, Assumption 2.3* (a) corresponds in this case to fixed h. When $m_0 = 1$, this reduces to a condition relevant to a parametric estimation problem. The remaining parts of Assumptions 2.3 and 2.3* impose Lipschitz and integrability conditions on f, which are useful technical conditions.

The following result provides a limit theory for functionals of the following form

$$\frac{c_n}{n}\sum_{t=1}^{[n\eta]} f\left(x_{t-r}\right) g\left[c_n\left(x_{t-s,n}-\frac{x}{d_n}\right)\right]. \tag{10}$$

The result is therefore an extension of Theorem 1 of WP and relates also to Theorem 1 of Phillips (2009), although neither of the earlier results involved an additional integrable function f in the sample function, as occurs in (10). The scale constant τ in the limit results (11) and (12) similarly involves the function f, whereas in WP, τ is the energy functional $\tau = \int_{-\infty}^{\infty} g\left(z\right)dz$ involving only g.

In what follows it will be convenient to use the notation[1]

$$\sum_{rs} v_i = \mathbf{1}\left(s>r\right)\sum_{i=r+1}^{s} v_i - \mathbf{1}\left(r>s\right)\sum_{i=s+1}^{r} v_i.$$

Theorem 1. *Suppose that Assumption 2.1 and the following conditions hold:*

(a) $\left|f\left(\frac{d_n}{c_n}z+x-v\right)\right| \leq f_0\left(z,x,v\right)$ *for n large enough, with* $\int_v\int_z f_0(z,x,v)\left|g(z)\right| p_{r-s}(v)dzdv < \infty$,
$\int_v\left\{\int_z\left|f_0\left(z,x,v\right)\right|\left|g\left(z\right)\right|dz\right\}^2 p_{r-s}(v)dv < \infty$ *and* $\int_v\int_z f_0^2(z,x,v)g^2(z)p_{r-s}(v)dzdv < \infty$, *for and each $x\in\mathbb{R}$, and $r,s\in\mathbb{N}$;*

(b) *Assumption 2.3 holds and*

$$\tau := \mathbf{E}f\left(x+\sum_{rs} v_i\right)\int_{-\infty}^{\infty} g\left(z\right)dz;$$

or

(c) *Assumption 2.3* holds and*

$$\tau := \mathbf{E}\int_{-\infty}^{\infty} f\left(m_0 z + x + \sum_{rs} v_i\right) g(z)dz.$$

We have the following:

---

[1] Observe that for $s>r$ we have

$$x_{t-r}-x_{t-s} = \sum_{j=1}^{s-r} v_{t-s+j} =_d \sum_{j=1}^{s-r} v_j =_d \sum_{j=r+1}^{s} v_j,$$

by stationarity and similarly for $s<r$

$$x_{t-r}-x_{t-s} =_d -\sum_{j=s+1}^{r} v_j.$$

(i) *If Assumption 2.2(a) holds, then, as* $n \to \infty$

$$\frac{c_n}{n} \sum_{t=1}^{[n\eta]} f\left(x_{t-r}\right) g\left[c_n\left(x_{t-s,n} - \frac{x}{d_n}\right)\right] \xrightarrow{d} \tau L(\eta, 0). \tag{11}$$

(ii) *If Assumption 2.2(b) holds, then, as* $n \to \infty$

$$\sup_{0 \le \eta \le 1} \left| \frac{c_n}{n} \sum_{t=1}^{[n\eta]} f\left(x_{t-r}\right) g\left[c_n\left(x_{t-s,n} - \frac{x}{d_n}\right)\right] - \tau L(\eta, 0) \right| \xrightarrow{p} 0. \tag{12}$$

When $f = 1$, (11) reduces to

$$\frac{c_n}{n} \sum_{t=1}^{[n\eta]} g\left[c_n\left(x_{t-s,n} - \frac{x}{d_n}\right)\right] \xrightarrow{d} \left(\int_{-\infty}^{\infty} g\left(z\right) dz\right) L(\eta, 0),$$

corresponding to theorem 1 in WP. When $m_0 = 1$, $x = 0$, $r = s$ and $c_n \sim d_n$, the sample function effectively becomes $\frac{d_n}{n} \sum_{t=1}^{[n\eta]} f\left(x_{t-r}\right) g\left(x_{t-r}\right)$ and we have the conventional limit theory

$$\frac{d_n}{n} \sum_{t=1}^{[n\eta]} f\left(x_{t-r}\right) g\left(x_{t-r}\right) \xrightarrow{d} \left(\int_{-\infty}^{\infty} f\left(z\right) g\left(z\right) dz\right) L(\eta, 0)$$

for integrable fg, as given in Park and Phillips (1999).

3 Kernel regression under dynamic misspecificaion

We now proceed to develop a limit theory for the NW kernel regression estimator (5) in the case of dynamic misspecification of the form (4). We start with the following regularity conditions on the kernel and regression function, which are similar to those used in WP.

Assumption 3.1. *The kernel* K *satisfies* $\int_{-\infty}^{\infty} K(s)ds = 1$ *and* $\sup_s |K(s)| < \infty$.

Assumption 3.2. *For given* x, *there exists a real function* $f_1(s, x)$ *such that, when* h *is sufficently small,* $|\mathbf{E}f\left(hy + x + \sum_{rs} v_i\right) - \mathbf{E}f\left(x + \sum_{rs} v_i\right)| \le h^\gamma f_1(y, x)$ *with* $0 < \gamma \le 1$, *for all* $y \in \mathbb{R}$ *and* $\int_{-\infty}^{\infty} K(s) f_1(s, x)ds < \infty$. *Further,* $\mathbf{E}f\left(x + \sum_{rs} v_i\right)^2 < \infty$.

Assumption 3.3. $(u_t, \mathcal{F}_{n,t})$ *is a martingale difference sequence with* $\mathbf{E}(u_t^2|\mathcal{F}_{n,t-1}) = \sigma_u^2 < \infty$ *a.s.*

Assumption 3.4. $\sup_{1 \le t \le n} \mathbf{E}(u_t^4|\mathcal{F}_{n,t-1}) < \infty$ *a.s.*

The following result gives the probability limit and limit distribution of $\hat{f}(x)$, showing the effect of dynamic misspecification.

Theorem 2. *Suppose that:*

(a) *Assumptions 3.1-3.3 hold.*

(b) *The bandwidth* h *satisfies* $nh/d_n \to \infty$ *and* $h \to 0$ *as* $n \to \infty$.

Then, as $n \to \infty$,

$$\hat{f}(x) \xrightarrow{p} \mathbf{E} f\left(x + \sum_{rs} v_i\right). \tag{13}$$

In addition, suppose the following hold:

(c) *Assumption 3.4 holds.*

(d) *Set* $c_n := d_n/h$. *The component functions* $\{f^2, f^4\}$ *and the power kernel functions* $\{K^2, K^4\}$ *in the sample quantities* $\frac{c_n}{n}\sum_{t=1}^n f^2(d_n x_{t-r,n}) K^2\left[c_n\left(x_{t-s,n} - \frac{x}{d_n}\right)\right]$ *and* $\frac{c_n}{n}\sum_{t=1}^n f^4(d_n x_{t-r,n}) K^4\left[c_n\left(x_{t-s,n} - \frac{x}{d_n}\right)\right]$ *both satisfy the conditions of Theorem 1.*

(e) *The bandwidth parameter* h *satisfies* $nh^{1+2\gamma}/d_n \to \infty$.

Then, as $n \to \infty$,

$$\left(\sum_{t=1}^n K\left(\frac{x_{t-s} - x}{h}\right)\right)^{1/2}\left(\hat{f}(x) - \mathbf{E} f\left(x + \sum_{rs} v_i\right)\right) \xrightarrow{d} N\left(0, \sigma^2\right), \tag{14}$$

where $\sigma^2 = \left[\sigma_u^2 + \mathbf{Var}\left\{f\left(x + \sum_{rs} v_i\right)\right\}\right]\int_{-\infty}^{\infty} K(s)^2 ds$.

The probability limit of the NW kernel estimator $\hat{f}(x)$ is

$$\mathbf{E} f\left(x + \sum_{rs} v_i\right) = \int f(x + w)\, p_{r-s}(w)\, dw, \tag{15}$$

where $\sum_{rs} v_i$ has density $p_{r-s}(w)$.[2] The limit (15) is an average of f taken around the value at x with respect to this density. For instance, when $s > r$ we have

$$x_{t-r} - x_{t-s} = \sum_{i=1}^{s-r} v_{t-s+i} =_d \sum_{i=1}^{s-r} v_i =_d \sum_{i=r+1}^{s} v_i,$$

under stationarity. If $r = s$ then there is no dynamic misspecification in the fitted equation and the estimate is consistent so that $\hat{f}(x) \to_p f(x)$ with a limit distribution

$$\left(\sum_{t=1}^n K\left(\frac{x_{t-s} - x}{h}\right)\right)^{1/2}\left(\hat{f}(x) - f(x)\right) \xrightarrow{d} N\left(0, \sigma_u^2 \int_{-\infty}^{\infty} K(s)^2 ds\right), \tag{16}$$

as in WP under suitable undersmoothing or choice of h in the regression. Both (16) and (14) may be adjusted to account for a bias term of $O(h^2)$ in the limit theory, as shown in

[2] As in footnote 1 we have

$$\sum_{rs} v_i = \mathbf{1}(s > r) \sum_{i=r+1}^{s} v_i - \mathbf{1}(r > s) \sum_{i=s+1}^{r} v_i$$

Then, for $s > r$, $p_{r-s}(w)$ is the density of $x_{t-r} - x_{t-s} =_d \sum_{i=r+1}^{s} v_i$, and if $s < r$, $p_{r-s}(w)$ is the density of $x_{t-r} - x_{t-s} =_d -\sum_{i=s+1}^{r} v_i$. So $\sum_{rs} v_i$ has density $p_{r-s}(w)$.

Wang and Phillips (2009), but in view of the inconsistency already present in (14) there is little reason to provide that development in the case of misspecification.

The limit distributions (14) and (16) differ in terms of both centering and variance. The centering is explained by the inconsistency (13) under mistiming ($r \neq s$) of the lagged relationship. The additional variance in the limit distribution (14) occurs because

$$\sigma_u^2 + \mathbf{Var}\left\{ f\left(x + \sum_{rs} v_i \right) \right\} > \sigma_u^2$$

whenever $r \neq s$. The extra component in the variance is $\mathbf{Var}\left\{ f\left(x + \sum_{rs} v_i \right) \right\}$, which arises as in (11) of Theorem 1 because the limit of the average conditional variance involves averaging over the distribution of $\sum_{rs} v_i$, just as it does in the case of the first moment. In consequence, lag misspecification in the fitted nonparametric cointegrating relation (4) produces both inconsistency and a reduction in precision in the limit theory for the NW estimator.

In the special case of linear cointegration with $f(x_t) = \theta x_t$, we have from (13)

$$\mathbf{E}f\left(x + \sum_{rs} v_i \right) = \theta x + \sum_{rs} \mathbf{E}v_i = \theta x,$$

so that kernel regression is consistent under lag misspecification, corresponding to the temporal invariance of linear cointegrating regression. In this case, (14) becomes

$$\left(\sum_{t=1}^{n} K\left(\frac{x_{t-s} - x}{h} \right) \right)^{1/2} \left(\hat{f}(x) - f(x) \right) \xrightarrow{d} N\left(0, \sigma^2\right),$$

with

$$\sigma^2 = \left[\sigma_u^2 + |s-r|\, \sigma_v^2 \right] \int_{-\infty}^{\infty} K(s)^2 ds > \sigma_u^2 \int_{-\infty}^{\infty} K(s)^2 ds,$$

since $\mathbf{Var}\{\sum_{rs} v_i\} = |s-r|\, \sigma_v^2$. Hence, lag shifts in a linear cointegrating regression do impact the variance of the limit distribution in kernel regression. The same is true, of course, for linear parametric cointegrating regression.

It is interesting to compare the limit results given in Theorem 2 with those of a stationary time series regression. Suppose model (2) is the true model and (4) is the fitted model, as above, but that x_t is a stationary time series satisfying certain asymptotic dependence or mixing conditions that validate nonparametric regression (see for example Li and Racine, 2007). This type of situation seems not to have been analyzed in the literature. However, it is readily shown by conventional methods for stationary nonparametric regression that under suitable regularity and mixing conditions

$$\hat{f}(x) \xrightarrow{p} \mathbf{E}f\left(x_{t-r} | x_{t-s} = x \right), \tag{17}$$

which is the analogue for the stationary time series x_t of the inconsistency shown in (13). For when x_t follows a unit root process, we have $x_{t-r} = x_{t-s} + \sum_{i=1}^{s-r} v_{t-s+i}$ for $s > r$. Then, when we condition on $x_{t-s} = x$ for this nonstationary data generating process, the right side of (17) may be written in the form

$$\mathbf{E}f\left(x_{t-s} + \sum_{i=1}^{s-r} v_{t-s+i} | x_{t-s} = x \right) = \mathbf{E}f\left(x + \sum_{i=r+1}^{s} v_i \right),$$

which corresponds precisely to the limit in (13) because $\sum_{rs} v_i = \sum_{i=r+1}^{s} v_i$ when $s > r$ by definition. Thus, the effect of dynamic misspecification on inconsistency in nonparametric regression is the same for nonstationary time series as it is for stationary time series.

For specification testing purposes it is useful to have an error variance estimator. We consider the following estimator

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{n} \left[y_t - \hat{f}(x)\right]^2 K_h(x_{t-s} - x)}{\sum_{t=1}^{n} K_h(x_{t-s} - x)}$$

Under correct specification and a constant error variance σ_u^2, we know from Wang and Phillips (2008) that $\hat{\sigma}^2 = \sigma_u^2 + o_p(1)$. Under dynamic misspecification, it turns out that $\hat{\sigma}^2$ estimates consistently the component that determines the limit variance under misspecification. This is demonstrated in the following result.

Theorem 3. *Suppose that the conditions of Theorem 2 hold. Then, as $n \to \infty$,*

$$\hat{\sigma}^2 \xrightarrow{p} \sigma_u^2 + \mathbf{Var}\left\{ f\left(x + \sum_{rs} v_i \right) \right\}.$$

Moreover, under linearity where $f(x) = \theta x$ we have

$$\hat{t}(x, \theta) := \left(\frac{\sum_{t=1}^{n} K\left(\frac{x_{t-s}-x}{h}\right)}{\hat{\sigma}^2 \int_{-\infty}^{\infty} K(s)^2 ds} \right)^{1/2} \left(\hat{f}(x) - \theta x \right) \xrightarrow{d} N(0,1),$$

as $n \to \infty$.

Remarks.

(a) Theorem 3 shows that under linearity the t statistic $\hat{t}(x,\theta) \xrightarrow{d} N(0,1)$ under both correct and incorrect dynamic specification. The statistic may therefore form the basis of a linearity test that is robust to dynamic misspecification, as we now discuss.

(b) Let $\hat{\theta}$ be the least squares estimator $\hat{\theta} = \sum_{t=1}^{n} x_t y_t / \sum_{t=1}^{n} x_t^2$. Since $\hat{\theta}$ is $O(n)$ consistent for θ under linearity, we have

$$\hat{t}(x, \hat{\theta}) \xrightarrow{d} N(0,1). \tag{18}$$

Under the alternative specification of (smooth) non-linear asymptotically homogeneous $f(x)$ we find that

$$\hat{t}(x, \hat{\theta}) \sim \left(\frac{nh}{d_n}\right)^{1/2} \left\{ \mathbf{E} f\left(x + \sum_{rs} v_i \right) + \frac{\kappa_f(\sqrt{n}) \int_{-\infty}^{\infty} s h_f(s) L_G(1,s) ds}{\sqrt{n} \int_{-\infty}^{\infty} s^2 L_G(1,s) ds} x \right\}, \tag{19}$$

where h_f and κ_f are the limit homogeneous function and asymptotic order of f respectively (see Park and Phillips, 2001, for full definitions). Under the alternative specification of integrable $f(x)$ (and $xf(x)$) we find that

$$\hat{t}(x, \hat{\theta}) \sim \left(\frac{nh}{d_n}\right)^{1/2} \left\{ \mathbf{E} f\left(x + \sum_{rs} v_i \right) + \frac{\int_{-\infty}^{\infty} s f(s) ds L_G(1,0)}{n^{3/2} \int_{-\infty}^{\infty} s^2 L_G(1,s) ds} x \right\}. \tag{20}$$

Results (19) and (20) show that the simple linearity test statistic $\hat{t}(x, \hat{\theta})$ in (18) has power against both homogeneous and integrable nonlinear functions and is robust to dynamic specification.

4 Some Practical Applications

Example 1. (Single index model) Suppose that y_t is generated by the single index model:

$$y_t = f(\lambda x_t + (1-\lambda)x_{t-1}) + u_t,\ 0 \le \lambda \le 1,$$

where the regressor x_t satisfies Assumptions 2.1 and 2.2 and u_t is a martingale difference sequence satisfying Assumptions 3.3 and 3.4. The fitted model takes the following form

$$y_t = \hat{f}(x_t) + \hat{u}_t,$$

omitting the indexed regressor and therefore misspecifying the lagged dependence in the relationship. When x_t is an integrated process,

$$\lambda x_t + (1-\lambda)x_{t-1} = x_{t-1} + \lambda v_t = x_t - (1-\lambda)v_t,$$

and then

$$\hat{f}(x) \xrightarrow{p} \mathbf{E}f(x-(1-\lambda)v_t),$$

as in Theorem 2 (b). Thus, indexing effects are important in nonlinear models of cointegration, in contrast to linear models where the temporal invariance of long run linear relations means that they can be safely ignored.

Example 2. (Temporal aggregation) When a regressor x_t is sampled (two times) more frequently than y_t, Ghysels, Santa-Clara and Valkanov (2004, 2006) propose mixed data sampling (MIDAS) regression models in which the conditional expectation of the dependent variable y_t is a distributed lag of the regressor, which may be recorded at a higher frequency. A simple example of such a regression arises in the case of temporal aggregation where the model takes the form

$$y_t = \lambda f(x_t) + (1-\lambda)f(x_{t-1}) + u_t,\ 0 \le \lambda \le 1, \tag{21}$$

and where x_t and u_t are as in Example 1. If the fitted model ignores the temporal aggregation in (21) and is a simple nonparametric regression of the form

$$y_t = \hat{f}(x_t) + \hat{u}_t,$$

then Theorem 2 shows that

$$\hat{f}(x) \xrightarrow{p} \lambda f(x) + (1-\lambda)\mathbf{E}f(x-v_t).$$

Thus, in the same way as indexing, temporal aggregation has important effects in nonlinear cointegration models.

Example 3 (Nonparametric unit root autoregression) Suppose that the true model is given by the autoregression

$$x_t = f(x_{t-1}) + u_t, \tag{22}$$

with $f(x) = x$, although the linear form of the autoregression is unknown to the econometrician, and where u_t is $iid\,(0,\sigma^2)$. The fitted model involves a longer lag and has the form

$$x_t = \hat{f}(x_{t-2}) + \hat{u}_t. \tag{23}$$

Under the true model (22) Assumption 2.2 holds with $x_{[nt],n} = \frac{1}{\sqrt{n}} \sum_{t=3}^{[nt]-2} u_t \xrightarrow{d} G(t)$, where $G(t)$ is Brownian motion. In view of Theorem 2 we get

$$\left(\sum_{t=1}^{n} K\left(\frac{x_{t-2}-x}{h} \right) \right)^{1/2} \left(\hat{f}(x) - x \right) \xrightarrow{d} N\left(0, 2\sigma_u^2 \int_{-\infty}^{\infty} K(s)^2 ds \right).$$

Note that the NW nonparametric estimator is consistent because $f(x)$ is a linear function. Nevertheless, there is a reduction in accuracy of $\hat{f}(x)$ due to the additional component σ_u^2 in the asymptotic variance. Similar effects occur in the case of linear unit root estimation. In particular, if (23) is estimated by linear regression in the form

$$x_t = \hat{\rho} x_{t-2} + \hat{u}_t,$$

then conventional weak convergence methods show that

$$n(\hat{\rho} - 1) \xrightarrow{d} \frac{2\int_0^1 W dW}{\int_0^1 W^2},$$

so that the limit distribution of the parametric estimator is rescaled by 2.

Example 4. (Functional coefficient regression models) Cai, Li and Park (2009, hereafter CLP) recently considered functional coefficient regression models with possibly nonstationary covariates that determine the functional regression coefficients. The model in CLP has the form

$$y_t = \beta(z_t)' x_t + \varepsilon_t, \quad t = 1, ..., n \tag{24}$$

where y_t and z_t are scalar, z_t is an I(1) process, x_t is stationary, and ε_t is a martingale difference sequence with constant conditional variance σ^2 and finite fourth moments. The functional coefficient $\beta(\cdot)$ is twice continuously differentiable and is the object of nonparametric estimation interest. CLP consider the local linear nonparametric estimator $\hat{\beta}(z)$ of $\beta(z)$. Under regularity conditions and using methods closely related to those of Wang and Phillips (2008), CLP showed that for any fixed z

$$\sqrt{n^{1/2}h}\left(\hat{\beta}(z) - \beta(z) - \frac{h^2}{2} B_\beta(z) \right) \xrightarrow{d} MN\left(0, \frac{\sigma_\varepsilon^2 \nu_0(K)}{L_{W_z}(1,0)} \left[E(x_t x_t') \right]^{-1} \right), \tag{25}$$

where the bias function $B_\beta(z) = \mu_2(K)\beta''(z)$, $L_{W_z}(1,0)$ is the local time of the limit Brownian motion process $W_z(r)$ for which $n^{-1/2} z_{[nr]} \xrightarrow{d} W_z(r)$, the constants have the usual form $\mu_2(K) = \int s^2 K(s)\, ds$, $\nu_0(K) = \int K(s)^2 ds$, and MN signifies mixed normality. In practice, undersmoothing will typically be employed (in this case requiring that $n^{1/10}h \to 0$), leading to the following useable limit result

$$\sqrt{n^{1/2}h}\left(\hat{\beta}(z) - \beta(z) \right) \xrightarrow{d} MN\left(0, \frac{\sigma_\varepsilon^2 \nu_0(K)}{L_{W_z}(1,0)} \left[E(x_t x_t') \right]^{-1} \right). \tag{26}$$

It will often be appropriate in empirical work to introduce lags into the specification (24). For example, the functional response function in (24) may take the form $\beta(z_{t-r})$ for some suitable integer $r > 0$ representing a delay in the impact of z_t on the functional regression response. In general, of course, the correct lag response will be unknown and any specification will only be approximate. The present paper shows that such specification issues are important in the nonstationary regression contest. For example, if (24) is estimated when the true response function is $\beta(z_{t-1})$, the methods of the present paper may be used to show that the nonparametric estimate $\hat{\beta}(z)$ has the following limit theory

$$\sqrt{n^{1/2}h}\left(\hat{\beta}(z) - E\{\beta(z-\Delta z_t)\}\right) \xrightarrow{d} MN\left(0, \frac{\{\sigma_\varepsilon^2 + \mathbf{Var}[\beta(z-\Delta z_t)]\}\nu_0(K)}{L_{W_z}(1,0)}[E(x_t x_t')]^{-1}\right).$$

Misspecification of functional regression therefore leads to inconsistency and an increase in limiting variance. These results hold for local level and local linear nonparametric regression procedures. Similar results also apply in the case of functional coefficient cointegrating regressions, which have recently been investigated by Xiao (2009) in the case of stationary covariates. A detailed analysis of these models will be reported elsewhere.

Example 5. (Parametric distributed lag cointegrating regression) Suppose that f_1 and f_2 are integrable functions and that a nonlinear cointegrating relationship between y_t and an integrated process x_t takes the following distributed lag form

$$y_t = \theta_1 f_1(x_t) + \theta_2 f_2(x_{t-1}) + u_t, \tag{27}$$

where x_t and u_t are again as in Example 1. Let $f_t = (f_1(x_t),\ f_2(x_{t-1}))'$, $\theta = (\theta_1, \theta_2)'$ and $\hat{\theta}$ be the least squares estimator of θ in (27). Applying Theorem 1 gives

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_t f_t' \xrightarrow{d} L(1,0)V,$$

where

$$V := \begin{bmatrix} \int_{-\infty}^{\infty} f_1(s)^2 ds & \mathbf{E}\int_{-\infty}^{\infty} f_1(s+v_t) f_2(s) ds \\ \mathbf{E}\int_{-\infty}^{\infty} f_1(s+v_t) f_2(s) ds & \int_{-\infty}^{\infty} f_2(s)^2 ds \end{bmatrix}.$$

Since V is positive definite in general, there is no asymptotic collinearity among the regressors in (27) at this level of intensity, which contrasts with the linear case where x_t and x_{t-1} are, of course, trivially cointegrated. In view of the above and the martingale central limit theorem (e.g. Kasparis, Phillips and Magdalinos, 2008) we have the following limit theory in this case:

$$\sqrt[4]{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} \sigma_u L_G(1,0)^{-1/2} V^{-1/2} Z, \tag{28}$$

where Z is standard bivariate normal. Thus, $\hat{\theta}$ is consistent and asymptotically mixed normally distributed with the usual $n^{1/4}$ rate of convergence that applies for regressors that are integrable functions of a unit root process (Park and Phillips, 1999, 2001). Unlike the linear case where the regressors are trivially cointegrated and the limit theory is degenerate, there is no degeneracy in the limit distribution (28).

5 Concluding Discussion

The results presented here show that the temporal invariance of linear cointegrating relations fails in the nonlinear case and mistiming of the regression function results in inconsistency in kernel regresion. In consequence, correct dynamic specification takes on new significance in nonlinear cointegrating systems. Specification tests for nonlinear cointegraton therefore need to take lag distribution and timing effects specifically into account.

The nonlinear setting clearly opens up many new possibilities for specification testing, including testing functional form in a particular locality corresponding to the kernel regression, allowance for short memory in the regression equation errors and endogeneity in the regressors. The differing effect on nonstationarity of various nonlinear functional forms in regression also means that simple residual based tests for stationarity, such as KPSS (1971) tests, may be misleading in the nonlinear context. Indeed, the long run and memory properties of the regressor may be substantially altered through nonlinear filtering. Since nonlinear functionals can change the integration order, the dependent variable in a nonlinear model may well have less memory than the regressor, meaning that misspecification may be harder to detect than it is in linear models. Specification tests for cointegration models where there is nonlinearity of unknown form are therefore likely to present far greater challenges than in the case of parametric linear cointegration.

6 Appendix A: Supporting Results

The following lemmas are largely based on WP, extending that framework as needed to accommodate sample covariances of convolution integrable functions (f) and integrable kernels (g) involving x_t. It will be convenient to use notation $\phi_\epsilon(x) = (2\pi\epsilon^2)^{-1/2}\exp(-x^2/2\epsilon^2)$ and $\phi(x) = \phi_1(x)$. We also often write the density $p_1(v)$ as $p(v)$.

Lemma 1. *Suppose that*

(a) *Assumption 2.1 holds.*

(b) $\left|f\left(\frac{d_n}{c_n}z + x - v\right)\right| \le f_0(z,x,v)$, *for n large enough and*

(i) $\int_v\int_z f_0(z,x,v)\,|g(z)|\,p_{r-s}(v)dzdv < \infty$,

(ii) $\int_v\left\{\int_z |f_0(z,x,v)|\,|g(z)|\,dz\right\}^2 p_{r-s}(v)dv < \infty$ *and*

(iii) $\int_v\int_z f_0^2(z,x,v)g^2(z)p_{r-s}(v)dzdv < \infty$,

for $r,s \in \mathbb{N}$ and $x \in \mathbb{R}$.

Let

$$L_{n,\epsilon}(\eta) := \frac{c_n}{n}\sum_{t=1}^{[n\eta]}\int_{-\infty}^{\infty} f\left(d_n\left(x_{t-r,n} + z\epsilon\right)\right) g\left(c_n\left(x_{t-s,n} - \frac{x}{d_n} + z\epsilon\right)\right)\phi(z)\,dz$$

Then

$$L_{n,\epsilon}(\eta) = \frac{c_n}{n}\sum_{t=1}^{[n\eta]}\mathbf{E}_{t-(r\vee s)-1}\int_{-\infty}^{\infty} f\left(d_n\left(x_{t-r,n} + z\epsilon\right)\right) g\left(c_n\left(x_{t-s,n} - \frac{x}{d_n} + z\epsilon\right)\right)\phi(z)\,dz + o_p(1),$$

uniformly in η.

Proof of Lemma 1: Without loss of generality, we shall assume that $r = 1$ and $s = 0$. The proof for the general case is identical but requires more complicated notation. Consider

$$
\begin{aligned}
L_{n,\epsilon}(\eta) &= \frac{c_n}{n} \sum_{t=1}^{[n\eta]} \underbrace{\int_{-\infty}^{\infty} f\left(d_n\left(x_{t-1,n} + z\epsilon\right)\right) g\left(c_n\left(x_{t,n} - \frac{x}{d_n} + z\epsilon\right)\right) \phi\left(z\right) dz}_{:=z_t} \\
&= \frac{c_n}{n} \sum_{t=1}^{[n\eta]} \mathbf{E}_{t-2} z_t + \frac{c_n}{n} \sum_{t=1}^{[n\eta]} \left(z_t - \mathbf{E}_{t-2} z_t\right). \qquad (29)
\end{aligned}
$$

We show that the second term in (29) is $o_p(1)$. Notice that $\{(z_t - \mathbf{E}_{t-2}z_t), \mathcal{F}_{t-1}\}$ is a martingale difference sequence. Hence,

$$
\begin{aligned}
\mathbf{E}\mathbf{E}_{t-2}\left(\frac{c_n}{n} \sum_{t=1}^{[n\eta]} \left(z_t - \mathbf{E}_{t-2} z_t\right)\right)^2 &= \left(\frac{c_n}{n}\right)^2 \mathbf{E}\mathbf{E}_{t-2} \sum_{t=1}^{[n\eta]} \left(z_t - \mathbf{E}_{t-2} z_t\right)^2 \\
&= \left(\frac{c_n}{n}\right)^2 \mathbf{E}\left\{\sum_{t=1}^{[n\eta]} \mathbf{E}_{t-2} z_t^2 - \sum_{t=1}^{[n\eta]} \left(\mathbf{E}_{t-2} z_t\right)^2\right\}.
\end{aligned}
$$

Consider

$$
\begin{aligned}
& \left(\frac{c_n}{n}\right)^2 \sum_{t=1}^{[n\eta]} \mathbf{E}\mathbf{E}_{t-2} z_t^2 \\
= & \left(\frac{c_n}{n}\right)^2 \sum_{t=1}^{[n\eta]} \mathbf{E}\mathbf{E}_{t-2} \left\{\int_{-\infty}^{\infty} f\left(d_n\left(x_{t-1,n} + z\epsilon\right)\right) g\left(c_n\left(x_{t,n} - \frac{x}{d_n} + z\epsilon\right)\right) \phi\left(z\right) dz\right\}^2 \\
= & \left(\frac{c_n}{n}\right)^2 \sum_{t=1}^{[n\eta]} \mathbf{E} \int_v \left\{\int_l f\left(d_n l + x - v\right) g\left(c_n l\right) \phi_\epsilon\left(l - x_{t-1,n} - \frac{v}{d_n} + \frac{x}{d_n}\right) dl\right\}^2 p(v) dv \\
\leq & \phi_\epsilon^2\left(0\right) \frac{c_n^2}{n} \int_v \left\{\int_l f\left(d_n l + x - v\right) g\left(c_n l\right) dl\right\}^2 p(v) dv \\
\leq & \phi_\epsilon^2\left(0\right) \frac{c_n}{n} \int_v \left\{\int_m \left| f\left(\frac{d_n}{c_n} m + x - v\right)\right| \left|g\left(m\right)\right| dm\right\}^2 p(v) dv \\
\leq & \phi_\epsilon^2\left(0\right) \frac{c_n}{n} \int_v \left\{\int_m \left|f_0\left(m, x, v\right)\right| \left|g\left(m\right)\right| dm\right\}^2 p(v) dv \to 0,
\end{aligned}
$$

where the last inequality holds for n large enough. Next consider

$$
\begin{aligned}
& \left(\frac{c_n}{n}\right)^2 \mathbf{E} \sum_{t=1}^{[n\eta]} \left(\mathbf{E}_{t-2} z_t\right)^2 \\
= & \left(\frac{c_n}{n}\right)^2 \mathbf{E} \sum_{t=1}^{[n\eta]} \left\{ \mathbf{E}_{t-2} \int_{-\infty}^{\infty} f\left[d_n\left(x_{t-1,n} + z\epsilon\right)\right] g\left(c_n\left(x_{t,n} - \frac{x}{d_n} + z\epsilon\right)\right) \phi\left(z\right) dz \right\}^2 \\
= & \left(\frac{c_n}{n}\right)^2 \sum_{t=1}^{[n\eta]} \mathbf{E} \left\{ \int_v \int_z f\left[d_n\left(x_{t-1,n} + z\epsilon\right)\right] g\left[c_n\left(x_{t-1,n} + \frac{v}{d_n} - \frac{x}{d_n} + z\epsilon\right)\right] \phi\left(z\right) p(v) dz dv \right\}^2 \\
\leq & \phi_\epsilon^2\left(0\right) \frac{1}{n} \left\{ \int_v \int_l \left| f\left(\frac{d_n}{c_n} l + x - v\right) g\left(l\right) \right| p(v) dv dl \right\}^2 \\
\leq & \phi_\epsilon^2\left(0\right) \frac{1}{n} \left\{ \int_v \int_l f_0\left(l, x, v\right) \left|g\left(l\right)\right| p(v) dv dl \right\}^2 \to 0,
\end{aligned}
$$

as required. ■

Lemma 2. *Suppose that Assumption 2.3 or Assumption 2.3* holds. Set*

$$
L_{n,\epsilon}(\eta) = \frac{c_n}{n} \sum_{t=1}^{[n\eta]} \mathbf{E}_{t-(r\vee s)-1} \int_{-\infty}^{\infty} f\left[d_n\left(x_{t-r,n} + \epsilon z\right)\right] g\left[c_n\left(x_{t-s,n} - \frac{x}{d_n} + \epsilon z\right)\right] \phi(z) dz.
$$

Then

$$
\lim_{n\to\infty} \sup_{0\leq\eta\leq1} \left| L_{n,\epsilon}(\eta) - \tau \sum_{t=1}^{[n\eta]} \phi_\epsilon(x_{t-(r\vee s),n}) \right| = 0,
$$

where $\tau := \begin{cases} \mathbf{E} f\left(x + \sum_{rs} v_i\right) \int_{-\infty}^{\infty} g\left(z\right) dz, \text{ if Assumption 2.3 holds} \\ \mathbf{E} \int_{-\infty}^{\infty} f\left(m_o z + x + \sum_{rs} v_i\right) g(z) dz, \text{ if Assumption 2.3* holds.} \end{cases}$

Proof of Lemma 2: Without loss of generality, assume that $r = 1$ and $s = 0$.

(a) We first show the result under Assumption 2.3. Consider

$$
\begin{aligned}
& \frac{c_n}{n} \sum_{t=1}^{[n\eta]} \mathbf{E}_{t-2} \int_{-\infty}^{\infty} f\left(d_n\left\{x_{t-1,n} + \epsilon z\right\}\right) g\left(c_n\left\{x_{t-1,n} + \frac{v_t}{d_n} - \frac{x}{d_n} + \epsilon z\right\}\right) \phi(z) dz \\
= & \frac{c_n}{n} \sum_{t=1}^{[nr]} \int_v \int_z f\left(d_n\left\{x_{t-1,n} + \epsilon z\right\}\right) g\left(c_n\left\{x_{t-1,n} + \frac{v}{d_n} - \frac{x}{d_n} + \epsilon z\right\}\right) \phi_\epsilon(z) p(v) dz dv \\
= & \frac{1}{n} \sum_{t=1}^{[n\eta]} \int_v \int_z f\left(d_n\left\{\frac{z}{c_n} + \frac{x}{d_n} - \frac{v}{d_n}\right\}\right) g\left(z\right) \phi_\epsilon\left(\frac{z}{c_n} - x_{t-1,n} + \frac{x}{d_n} - \frac{v}{d_n}\right) p(v) dz du \\
: & = T_n(\eta)
\end{aligned}
$$

Notice that by Assumption 2.3(b) and the Lipschitz continuity of ϕ_ϵ we get

$$\begin{aligned}
&\left| f\left(\frac{d_n}{c_n}z + x - v\right)\phi_\epsilon\left(\frac{z}{c_n} - x_{t-1,n} + \frac{v}{d_n} - \frac{x}{d_n}\right) - f(x-v)\,\phi_\epsilon(x_{t-1,n})\right| \\
\leq\ & |\phi_\epsilon(0)| \left| f\left(\frac{d_n}{c_n}z + x - v\right) - f(x-v)\right| \\
&+ |f(x-v)| \left|\phi_\epsilon\left(\frac{z}{c_n} - x_{t-1,n} + \frac{v}{d_n} - \frac{x}{d_n}\right) - \phi_\epsilon(x_{t-1,n})\right| \\
\leq\ & |\phi_\epsilon(0)| \left(\frac{d_n}{c_n}\right)^\gamma f_0(z,v,x) + |f(x-v)|\, C\left|\frac{z}{c_n} + \frac{v}{d_n} - \frac{x}{d_n}\right|,
\end{aligned}$$

where C is a Lipschitz constant. Therefore,

$$\begin{aligned}
&\left| T_n(\eta) - \frac{1}{n}\sum_{t=1}^{[n\eta]} \int_v\int_z f(x-v)\,\phi_\epsilon(x_{t-1,n})\, g(z)p(v)dzdv \right| \\
\leq\ & \left(\frac{d_n}{c_n}\right)^\gamma |\phi(0)| \int_u\int_z f_0(z,v,x)\,|g(z)|\, p(v)dzdv \\
&+ C\int_v\int_z |f(x-v)|\, p(v)\left|\frac{z}{c_n} + \frac{v}{d_n} - \frac{x}{d_n}\right| |g(z)|\, dzdv \to 0
\end{aligned}$$

as required.

(b) Suppose that Assumption 2.3* holds. Consider,

$$\begin{aligned}
&\left| f\left(\frac{d_n}{c_n}z + x - v\right)\phi\left(\frac{z}{c_n} - x_{t-1,n} + \frac{v}{d_n} - \frac{x}{d_n}\right) - f(m_o z + x - v)\,\phi_\epsilon(x_{t-1,n})\right| \\
\leq\ & |\phi_\epsilon(0)| \left|\frac{d_n}{c_n} - m_o\right|^\gamma f_0(z,v,x) + |f(m_o z + x - v)|\, C\left|\frac{z}{c_n} + \frac{v}{d_n} - \frac{x}{d_n}\right| \to 0,
\end{aligned}$$

as $n \to \infty$. In view of the above, the result can be shown using the same arguments as those in part (a). ■

Lemma 3. *Suppose that*

(a) *Assumption 2.2 holds.*

(b) $\left| f\left(\frac{d_n}{c_n}z + x - v\right)\right| \leq f_0(z,x,v)$ *for* n large enough *with* $\int_v\int_z f_0(z,x,v)\,|g(z)|\, p_{r-s}(v)dzdv < \infty$, *for each* $x \in \mathbb{R}$, *and* $r > s \in \mathbb{N}$.

(c) $\sup_s |g(s)| < \infty$

Let $q \in \mathbb{N}$ *with* $q > 1$. *We have*

$$M_n(\eta) := \frac{c_n}{n}\sum_{t=1}^{[n\eta]} f(d_n x_{t-r,n}) \left\{ \mathbf{E}_{t-r-1} g\left[c_n\left(x_{t-s,n} - \frac{x}{d_n}\right)\right]\right\}^q = o_p(1)$$

uniformly in η.

Proof of Lemma 3: Without loss of generality, assume that $r = 1$ and $s = 0$. We have

$$\begin{aligned}
& \mathbf{E}\left|M_n(\eta)\right| \\
= \; & \frac{c_n}{n} \int_s \left| \sum_{t=1}^{[n\eta]} \int_{v_1} \cdots \int_{v_q} f\left(d_n d_{t-1,0,n} l\right) g\left[c_n\left(d_{t-1,0,n} l + \frac{v_1}{d_n}\right)\right] \ldots g\left[c_n\left(d_{t-1,0,n} l + \frac{v_q}{d_n}\right)\right] \right. \\
& \left. \times\, p(v_1) p(v_q) dv_1 \ldots dv_q \right| h_{t-1,0,n}(l)\, dl \\
\leq \; & \frac{1}{n} \sum_{t=1}^{n} \frac{1}{d_{t-1,0,n}} \int_m \int_{v_1} \cdots \int_{v_q} \left| f\left(\frac{d_n}{c_n} m - v_1\right)\right| \left|g(m)\right| \left| \prod_{i=2}^{q} g\left(m + \frac{c_n}{d_n}(v_i - v_1)\right)\right| \\
& \times p(v_1) \ldots p(v_q) dv_1 \ldots dv_q dm \\
\leq \; & \frac{1}{n} \sum_{t=1}^{n} \frac{1}{d_{t-1,0,n}} \int_m \int_{v_1} \cdots \int_{v_q} f_0(m, v_1) \left|g(m)\right| \\
& \times \left| \prod_{i=2}^{q} g\left(m + \frac{c_n}{d_n}(v_i - v_1)\right)\right| p(v_1) \ldots p(v_q) dv_1 \ldots dv_q dm \text{ (for } n \text{ large enough)} \\
\leq \; & A \int_m \int_{v_1} \cdots \int_{v_q} f_0(m, v_1) \left|g(m)\right| \left| \prod_{i=2}^{q} g\left(m + \frac{c_n}{d_n}(v_i - v_1)\right)\right| p(v_1) \ldots p(v_q) dv_1 \ldots dv_q dm \to 0,
\end{aligned}$$

as $n \to \infty$, by dominated convergence since $g\left(m + \frac{c_n}{d_n}(v_{i+1} - v_1)\right) \to 0$ everywhere, $\int_{v_1} \int_m f_0(m, v_1) \left|g(m)\right| p(v_1) dv_1 dm < \infty$, and $\sup_s |g(s)| < \infty$. ■

Lemma 4. *Suppose that*

(a) *Assumption 2.2 holds.*

(b) $\left| f\left(\frac{d_n}{c_n} z + x - v\right)\right| \leq f_0(z, x, v)$ *for* n large enough *with* $\int_v \int_z f_0(z, x, v) \left|g(z)\right| p_{r-s}(v) dz dv < \infty$*, for each* $x \in \mathbb{R}$ *and* $r > s \in \mathbb{N}$.

Set

$$M_n(\eta) := \frac{c_n}{n} \sum_{t=1}^{[n\eta]} f\left(d_n x_{t-r,n}\right) \mathbf{E}_{t-r-1} g\left[c_n\left(x_{t-s,n} - \frac{x}{d_n}\right)\right]$$

Then

$$\sup_n \sup_{0 \leq \eta \leq 1} \mathbf{E}\left|M_n(\eta)\right| < \infty.$$

Proof of Lemma 4: Without loss of generality, assume that $r = 1$ and $s = 0$. We have

$$\begin{aligned}
\mathbf{E}\left|M_n(\eta)\right| \; = \; & \left(\frac{c_n}{n}\right) \mathbf{E} \sum_{t=1}^{[n\eta]} \int_v \left| f\left(d_n x_{t,n-1}\right) g\left[c_n\left(x_{t-1,n} + \frac{v}{d_n} - \frac{x}{d_n}\right)\right]\right| p(v) dv \\
= \; & \left(\frac{c_n}{n}\right) \mathbf{E} \sum_{t=1}^{[n\eta]} \int_s \int_v \left| f\left(d_n d_{t-1,0,n} s\right)\right) g\left[c_n\left(d_{t-1,0,n} s + \frac{v}{d_n} - \frac{x}{d_n}\right)\right]\Big| p(v) h_{t-1,0,n}(s)\, dv ds \\
\leq \; & \frac{1}{n} \sum_{t=1}^{[n\eta]} \frac{1}{d_{t-1,0,n}} \int_s \int_v \left| f\left(\frac{d_n}{c_n} m + x - v\right) g(m)\right| p(v) dv dm \\
\leq \; & A \int_s \int_v \left| f_0(m, v, x)\right) g(m) \big| p(v) dv dm < \infty,
\end{aligned}$$

as required. ■

Lemma 5. *Suppose that Assumptions 2.1-2.3 and the conditions of Theomem 1 hold. Let $q, r, s \in \mathbb{N}$ with $q > 1$ and $r < s$. Then*

$$\sup_{0\leq\eta\leq1}\left|\frac{c_n}{n}\sum_{t=1}^{[nr]}\int_{-\infty}^{\infty}\{\mathbf{E}_{t-s-1}f\left[(x_{t-r})\right]\}^q g\left[c_n\left(x_{t-s,n}-\frac{x}{d_n}\right)\right]-\tau L(\eta,0)\right|\xrightarrow{p}0$$

where $\tau := \{\mathbf{E}f\left(x+\sum_{rs}v_i\right)\}^q\int_{-\infty}^{\infty}g\left(z\right)dz$.

Proof of Lemma 5: Set

$$L_{n,\epsilon}(\eta)=\frac{c_n}{n}\sum_{t=1}^{[n\eta]}\int_{-\infty}^{\infty}\{\mathbf{E}_{t-s-1}f\left[d_n\left(x_{t-r,n}+\epsilon z\right)\right]\}^q g\left[c_n\left(x_{t-s,n}-\frac{x}{d_n}+\epsilon z\right)\right]\phi(z)dz.$$

and

$$L_n(\eta)=\frac{c_n}{n}\sum_{t=1}^{[n\eta]}\{\mathbf{E}_{t-s-1}f\left(d_n x_{t-r,n}\right)\}^q g\left[c_n\left(x_{t-s,n}-\frac{x}{d_n}\right)\right].$$

It can be shown along the lines of Lemma 2 that

$$\lim_{n\to\infty}\sup_{0\leq\eta\leq1}\left|L_{n,\epsilon}(\eta,x)-\tau\sum_{t=1}^{[n\eta]}\phi_\epsilon(x_{t-s,n})\right|=0.$$

In addition, using arguments similar to those used in the proof of Theorem 1 we get

$$\lim_{\epsilon\to0}\lim_{n\to\infty}\sup_{0\leq\eta\leq1}\mathbf{E}\left|L_n(\eta)-L_{n,\epsilon}(\eta)\right|=0.$$

■

7 Appendix B: Proofs of the Main Results

Proof of Theorem 1. Set,

$$L_n(\eta)=\frac{c_n}{n}\sum_{t=1}^{[nr]}\int_{-\infty}^{\infty}f\left(d_n x_{t-1,n}\right)g\left[c_n\left(x_{t,n}-\frac{x}{d_n}\right)\right]\phi(z)dz.$$

$$L_{n,\epsilon}(\eta)=\frac{c_n}{n}\sum_{t=1}^{[nr]}\int_{-\infty}^{\infty}f\left[d_n\left(x_{t-1,n}+\epsilon z\right)\right]g\left[c_n\left(x_{t,n}-\frac{x}{d_n}+\epsilon z\right)\right]\phi(z)dz.$$

Then

$$\lim_{\epsilon\to0}\lim_{n\to\infty}\sup_{0\leq r\leq1}\mathbf{E}\left|L_n(\eta)-L_{n,\epsilon}(\eta)\right|=0, \tag{30}$$

and the stated results follow as in WP. We proceed to prove (30). In what follows, we use A as a generic constant whose value may change in each location.

Set

$$Y_{t,n}(z) = f\left(d_n x_{t-1,n}\right) g\left[c_n\left(x_{t,n} - \frac{x}{d_n}\right)\right] - f\left[\sqrt{n}\left(x_{t-1,n} + \epsilon z\right)\right] g\left[c_n\left(x_{t,n} - \frac{x}{d_n} + \epsilon z\right)\right]$$

Notice that

$$\sup_{0 \le r \le 1} \mathbf{E}\left|L_n(\eta) - L_{n,\epsilon}(\eta)\right| \le \frac{c_n}{n} \int_{-\infty}^{\infty} \sum_{t=1}^{[n\eta]} \mathbf{E}\left|Y_{t,n}(z)\right| \phi(z) dz.$$

Next, we have

$$\begin{aligned}
& c_n \mathbf{E}\left|Y_{k,n}(z)\right| \\
= \ & c_n \mathbf{E}\left| f\left(d_n x_{k-1,n}\right) g\left[c_n\left(x_{k,n} - \frac{x}{d_n}\right)\right] - f\left[\sqrt{n}\left(x_{k-1,n} + \epsilon z\right)\right] g\left[c_n\left(x_{k,n} - \frac{x}{d_n} + \epsilon z\right)\right]\right| \\
\le \ & \int_s \int_v \left| f\left(d_n d_{k-1,0,n} s - v + x\right) g\left(c_n d_{k-1,0,n} s\right) - \right. \\
& \left. f\left(\sqrt{n}\left(d_{k-1,0,n} s + \epsilon z\right) - v + x\right) g\left[c_n\left(d_{k-1,0,n} s + \epsilon z\right)\right]\right| p(v)\, dv ds \\
\le \ & \frac{2A}{d_{k-1,0,n}} \int_s \int_v \left| f\left(\frac{d_n}{c_n} s - v + x\right) g(s)\right| p(v)\, dv ds \\
\le \ & \frac{2A}{d_{k-1,0,n}} \int_s \int_v f_0(s, v, x) \left|g(s)\right| p(v) dv ds,
\end{aligned}$$

for n large enough. In view of this condition (a) of Theorem 1 and (9) we get

$$\frac{c_n}{n} \sup_{0 \le \eta \le 1} \mathbf{E}\left|\sum_{k=1}^{[n\eta]} Y_{k,n}(z)\right| \le A_1 \frac{1}{n} \sum_{k=1}^{n} \left(d_{k-1,0,n}\right)^{-1} < \infty.$$

Set

$$\Lambda_n(\epsilon) \equiv \left(\frac{c_n}{n}\right)^2 \sup_{0 \le r \le 1} \mathbf{E}\left(\sum_{k=1}^{[nr]} Y_{k,n}(z)\right)^2.$$

In view of the above and dominated convergence, it would suffice to show that for each z

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \Lambda_n(\epsilon) = 0,$$

which is what we now set out to do. Notice that

$$\begin{aligned}
\Lambda_n(\epsilon) \ & \le \ \left(\frac{c_n}{n}\right)^2 \mathbf{E} \sum_{k=1}^{n} Y_{k,n}^2(z) + \frac{2c_n^2}{n^2} \sum_{k=1}^{n} \left|\mathbf{E} Y_{k,n}(z) Y_{k+1,n}(z)\right| + \frac{2c_n^2}{n^2} \sum_{k=1}^{n} \sum_{l=k+2}^{n} \left|\mathbf{E} Y_{k,n}(z) Y_{l,n}(z)\right| \\
& : \ = \Lambda_{1n}(\epsilon) + \Lambda_{2n}(\epsilon) + \Lambda_{3n}(\epsilon).
\end{aligned}$$

Under condition (b) of Theorem 1 and using similar arguments as before it can be shown that

$$\Lambda_{1n}(\epsilon) \le \frac{c_n}{n^2} \sum_{k=1}^{n} \frac{2A}{d_{k-1,0,n}} \int_s \int_v f_0(s, v, x)^2 g(s)^2 p(v) dv ds \le A \frac{c_n}{n} \to 0.$$

Similarly, it can be shown that $\Lambda_{2n}(\epsilon) \to 0$. Next, we consider $\Lambda_{3n}(\epsilon)$. Recall that $x_{k,n}$ is adapted to $\mathcal{F}_{k-1,n}$ and conditional on $\mathcal{F}_{k-1,n}$, $(x_{l-1,n} - x_{k,n})/d_{l-1,k,n}$ has density $h_{l-1,k,n}(s)$ which is uniformly bounded. Write $\Omega_n = \Omega_n\left(\delta^{1/(2k_0)}\right)$. We have

$$
\begin{aligned}
& c_n d_{l-1,k,n} \left|\mathbf{E}_{k-1} Y_{l,n}(z)\right| \\
= & \left|\mathbf{E}_{k-1}\left\{f\left(d_n x_{l-1,n}\right) g\left[c_n\left(x_{l,n} - \frac{x}{d_n}\right)\right] - f\left[\sqrt{n}\left(x_{l-1,n} + \epsilon z\right)\right] g\left[c_n\left(x_{l,n} - \frac{x}{d_n} + \epsilon z\right)\right]\right\}\right| \\
= & \left|\mathbf{E}_{k-1} \int_v \left\{f\left(d_n x_{l-1,n}\right) g\left[c_n\left(x_{l-1,n} + \frac{v}{d_n} - \frac{x}{d_n}\right)\right]\right.\right. \\
& \left.\left. - f\left[d_n\left(x_{l-1,n} + \epsilon z\right)\right] g\left[c_n\left(x_{l-1,n} + \frac{v}{d_n} - \frac{x}{d_n} + \epsilon z\right)\right]\right\} p(v) dv\right| \\
= & \left|\mathbf{E}_{k-1} \int_v \left\{f\left[d_n\left(x_{k,n} + (x_{l-1,n} - x_{k,n})\right)\right] g\left[c_n\left(x_{k,n} + (x_{l-1,n} - x_{k,n}) + \frac{v}{d_n} - \frac{x}{d_n}\right)\right]\right.\right. \\
& \left.\left. - f\left[d_n\left(x_{k,n} + (x_{l-1,n} - x_{k,n}) + \epsilon z\right)\right] g\left[c_n\left(x_{k,n} + (x_{l-1,n} - x_{k,n}) + \frac{v}{d_n} - \frac{x}{d_n} + \epsilon z\right)\right]\right\} p(v) dv\right| \\
= & \left|\int_s \int_v \left\{f\left[d_n\left(x_{k,n} + d_{l,k-1,n} s\right)\right] g\left[c_n\left(x_{k,n} + d_{l,k-1,n} s - \frac{x}{d_n}\right)\right]\right.\right. \\
& \left.\left. - f\left[d_n\left(x_{k,n} + d_{l,k-1,n} s + \epsilon z\right)\right] g\left[c_n\left(x_{k,n} + d_{l,k-1,n} s - \frac{x}{d_n} + \epsilon z\right)\right]\right\} p(v) h_{l-1,k,n}(s) ds\right| \\
\leq & \int_y \int_v \left|f\left(\frac{d_n}{c_n} y - v\right)\right| |g(y)| \left|V\left(y, c_n x_{k,n}, v\right)\right| p(v) dv dy \\
\leq & \begin{cases} A, \text{ for } (l-1,k) \notin \Omega_n, \\ A \int_{|y| \geq \sqrt{c_n}} \int_v f_o(y,v,x) |g(y)| p(v) dv dy \\ + \int_{|y| < \sqrt{c_n}} \int_v f_0(y,v,x) |g(y)| \left|V\left(y, c_n x_{k,n}, v\right)\right| p(v) dv dy, \text{ for } (l-1,k) \in \Omega_n \end{cases}
\end{aligned}
$$

where

$$
V(y,r,v) = h_{l-1,k,n}\left(\frac{y - r + c_n \frac{x-v}{d_n}}{c_n d_{l-1,k,n}}\right) - h_{l-1,k,n}\left(\frac{y - r + c_n \frac{x-v}{d_n} - c_n z \epsilon}{c_n d_{l-1,k,n}}\right).
$$

Consider

$$\begin{aligned}
& \mathbf{E}\left|Y_{k,n}(z)\right|\left|V\left(y,c_n x_{k,n},v\right)\right| \\
= \quad & A\mathbf{E}\int_w \left| f\left[d_n\left(x_{k-1,n}+z\epsilon\right)\right] g\left[c_n\left(x_{k-1,n}+\frac{w}{d_n}-\frac{x}{d_n}+z\epsilon\right)\right]\right| \\
& \times\left|V\left(y,c_n x_{k-1,n}+\frac{c_n}{d_n}w,v\right)\right| p(w)dw \\
= \quad & A\int_s\int_w \left| f\left(d_n\left(d_{k-1,0,n}s+z\epsilon\right)\right) g\left[c_n\left(d_{k-1,0,n}s+\frac{w}{d_n}-\frac{x}{d_n}+z\epsilon\right)\right]\right| \\
& \times\left|V\left(y,c_n d_{k-1,0,n}s+\frac{c_n}{d_n}w\right)\right| p(w)h_{k-1,0,n}(s)dwds \\
= \quad & \frac{A}{c_n d_{k-1,0,n}}\int_l\int_w \left| f\left(\frac{d_n}{c_n}l-w+x\right) g\left(l\right)\right| \\
& \times\left\{\left|V\left(y,l+\frac{c_n}{d_n}x-c_n z\epsilon,v\right)\right|+\left|V\left(y,l+\frac{c_n}{d_n}x,v\right)\right|\right\} p(w)dwdl \\
\leq \quad & \frac{A}{c_n d_{k-1,0,n}}\left[\int_{l\geq\sqrt{c_n}}\int_w f_0\left(l,w,x\right)\left|g\left(l\right)\right| p(w)dwdl+\sup_{|r|\leq 2C[1+|z|+|v|]\epsilon^{1/2}}\left|h_{l-1,k,n}\left(r\right)-h_{l-1,k,n}\left(0\right)\right|\right],
\end{aligned}$$

where the last inequality holds for n large enough, and can be established using similar arguments to those in WP.

In view of the above, for $(l-1,k)\notin\Omega_n$

$$\left|\mathbf{E}Y_{k,n}(z)Y_{l,n}(z)\right| = \left|\mathbf{E}Y_{k,n}(z)\mathbf{E}_{k-1}Y_{l,n}(z)\right| \leq A\left(c_n d_{l-1,k,n}\right)^{-1}\mathbf{E}_{k-1}\left|Y_{l,n}(z)\right| \leq A_1\left(c_n^2 d_{l-1,k,n}d_{k-1,0,n}\right)^{-1}.$$

On the other hand, for $(l-1,k)\in\Omega_n$

$$\begin{aligned}
\left|\mathbf{E}Y_{k,n}(z)Y_{l,n}(z)\right| \quad = \quad & \left|\mathbf{E}Y_{k,n}(z)\mathbf{E}_{k-1}Y_{l,n}(z)\right| \\
\leq \quad & A\left(c_n d_{l-1,k,n}\right)^{-1}\mathbf{E}\left|Y_{k,n}(z)\right|\int_{|y|\geq\sqrt{c_n}}\int_v f_0\left(y,v,x\right)\left|g(y)\right| p(v)dvdy \\
& +A\left(c_n d_{l-1,k,n}\right)^{-1}\int_{|y|<\sqrt{c_n}}\int_v f_0\left(y,v,x\right)\left|g(y)\right|\mathbf{E}\left|Y_{k,n}(z)\right|\left|V\left(y,c_n x_{k,n},v\right)\right| p(v)dvdy \\
\leq \quad & A_1\left(c_n^2 d_{l-1,k,n}d_{k-1,0,n}\right)^{-1}\left\{\int_{|y|\geq\sqrt{c_n}}\int_v f_0\left(y,v,x\right)\left|g(y)\right| p(v)dvdy\right. \\
& \left.+\int_y\int_v \sup_{|r|\leq 2C[1+|z|+|v|]\epsilon^{1/2}}\left|h_{l-1,k,n}\left(r\right)-h_{l-1,k,n}\left(0\right)\right|\left|g(y)\right| f_0\left(y,v,x\right)p(v)dvdy\right\}.
\end{aligned}$$

Notice that the last term above converges to zero as $n\to\infty$, due to dominated convergence. In view of the above and (6)-(9) we have for $\eta=\epsilon^{1/2}/C$

$$
\begin{aligned}
\Lambda_{3n}(\epsilon) &\leq \frac{2c_n^2}{n^2}\left[\sum_{l-1>k,(l-1,k)\notin\Omega_n}^{n} + \sum_{(l-1,k)\in\Omega_n}^{n}\right]|\mathbf{E}Y_{k,n}(z)Y_{l,n}(z)| \\
&\leq \frac{A_1}{n^2}\sum_{k=(1-\eta)n}^{n}(d_{k,0,n})^{-1}\max_{1\leq k\leq n-2}\sum_{l=k+2}^{n}(d_{l-1,k,n})^{-1} \\
&\quad +\frac{A_2}{n^2}\sum_{k=\eta n}^{(1-\eta)n}(d_{k,0,n})^{-1}\max_{0\leq k\leq(1-\eta)n}\sum_{l=k+2}^{k+\eta n}(d_{l-1,k,n})^{-1} \\
&\quad +\frac{A_3}{n^2}\sum_{k=1}^{\eta n}(d_{k,0,n})^{-1}\max_{0\leq k\leq n-2}\sum_{l=k+2}^{n}(d_{l-1,k,n})^{-1} \\
&\quad +\frac{A_4}{n^2}\sum_{k=1}^{n}(d_{k,0,n})^{-1}\max_{0\leq k\leq n-2}\sum_{l=k+2}^{n}(d_{l-1,k,n})^{-1} \\
&\quad \times\left\{\int_{|y|\geq\sqrt{c_n}}\int_v f_0(y,v,x)\,|g(y)|\,p(v)dvdy\right. \\
&\quad \left.+\int_y\int_v\sup_{|r|\leq 2C[1+|z|+|v|]\epsilon^{1/2}}|h_{l-1,k,n}(r)-h_{l-1,k,n}(0)|\,|g(y)|\,f_0(y,v,x)\,p(v)dvdy\right\} \\
&\to 0,
\end{aligned}
$$

as required. ■

Proof of Theorem 2. We prove the result for one lag differential (i.e., $|s-r|=1$) and the result for the general case follows in the same way.

First, we consider the case $r>s$. Set $\tau:=\mathbf{E}f(x-v_t)$. We have

$$
\begin{aligned}
\hat{f}(x)-\mathbf{E}f(x-v_t) &= \frac{\overbrace{\left(\frac{d_n}{nh}\right)^{1/2}\sum\mathbf{E}_{t-2}\left\{\left[f(x_{t-1})-\mathbf{E}f(x-v_t)\right]K_h(x_t-x)\right\}}^{:=R_n}}{\left(\frac{d_n}{nh}\right)^{1/2}\sum K_h(x_t-x)} \\
&\quad +\frac{\sum\overbrace{f(x_{t-1})K_h(x_t-x)-\mathbf{E}_{t-2}f(x_{t-1})K_h(x_t-x)}^{:=\alpha_t}}{\sum K_h(x_t-x)} \\
&\quad +\frac{\sum\overbrace{\tau\mathbf{E}_{t-2}K_h(x_t-x)-\tau K_h(x_t-x)}^{:=\beta_t}}{\sum K_h(x_t-x)}+\frac{\sum\overbrace{K_h(x_t-x)u_t}^{:=\gamma_t}}{\sum K_h(x_t-x)} \\
&:= \frac{R_n+M_n}{\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}K_h(x_t-x)}.
\end{aligned}
$$

Notice that

$$
\begin{aligned}
& \mathbf{E}\left|R_n\right| \\
= & \mathbf{E}\left|\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left[\left(f\left(x_{t-1}\right)-\mathbf{E}f(x-v_t)\right)K\left(\frac{x_t}{h}-\frac{x}{h}\right)\right]\right| \\
= & \mathbf{E}\left|\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}\int_u\left[f\left(x_{t-1}\right)-\mathbf{E}f(x-v_t)\right]K\left(\frac{d_n x_{t-1,n}}{h}+\frac{v}{h}-\frac{x}{h}\right)p_1(v)dv\right| \\
\leq & \left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}\int_y\left|\int_u\left[f\left(d_n d_{t-1,0,n}y\right)-\mathbf{E}f(x-v_t)\right]K\left(\frac{d_n d_{t-1,0,n}y}{h}+\frac{v}{h}-\frac{x}{h}\right)p_1(v)dv\right| \\
& \times h_{t-1,0,n}(y)dy \\
= & \left(\frac{d_n}{nh}\right)^{1/2}\frac{h}{d_n}\sum_{t=1}^{n}(d_{t-1,0,n})^{-1}\int_z\left|\left[\mathbf{E}f(hz+x-v_t)-\mathbf{E}f(x-v_t)\right]K\left(z\right)\right|h_{t-1,0,n}\left(\frac{hz+x-v_t}{d_n d_{t-1,0,n}}\right)dz \\
\leq & \left(\frac{n}{hd_n}\right)^{1/2}\frac{h^{1+\gamma}}{n}\sum_{t=1}^{n}(d_{t-1,0,n})^{-1}\int_z f_1(z,x)K\left(z\right)dz\to 0.
\end{aligned}
$$

Next, $\{M_n,\mathcal{F}_{n-1}\}$ is a martingale sequence. We shall establish a martingale CLT for this term. Consider

$$
\begin{aligned}
& \frac{d_n}{nh}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left(\alpha_t+\beta_t+\gamma_t\right)^2 \\
= & \frac{d_n}{nh}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left\{\left[f(x_{t-1})K_h(x_t-x)-\mathbf{E}_{t-2}f(x_{t-1})K_h(x_t-x)\right]\right. \\
& \left.+\tau\left[\mathbf{E}_{t-2}K_h(x_t-x)-K_h(x_t-x)\right]+K_h(x_t-x)u_t\right\}^2 \\
= & \frac{d_n}{nh}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left[f^2(x_{t-1})K_h^2(x_t-x)\right]-\frac{d_n}{nh}\sum_{t=1}^{n}\left[\mathbf{E}_{t-2}f(x_{t-1})K_h(x_t-x)\right]^2 \\
& +\frac{d_n}{nh}\tau^2\sum_{t=1}^{n}\mathbf{E}_{t-2}K_h^2(x_t-x)-\frac{d_n}{nh}\tau^2\sum_{t=1}^{n}\left[\mathbf{E}_{t-2}K_h(x_t-x)\right]^2 \\
& +\frac{d_n}{nh}\sum_{t=1}^{n}\mathbf{E}_{t-2}K_h^2(x_t-x)u_t^2 \\
& -\frac{d_n}{nh}2\tau\sum_{t=1}^{n}f(x_{t-1})\mathbf{E}_{t-2}K_h^2(x_t-x)-\frac{d_n}{nh}2\tau\sum_{t=1}^{n}f(x_{t-1})\left[\mathbf{E}_{t-2}K_h(x_t-x)\right]^2
\end{aligned}
$$

$$
\begin{aligned}
= & \frac{d_n}{nh}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left[f^2(x_{t-1})K_h^2(x_t-x)\right]+\frac{d_n}{nh}\tau^2\sum_{t=1}^{n}\mathbf{E}_{t-2}K_h^2(x_t-x) \\
& +\frac{d_n}{nh}\sum_{t=1}^{n}K_h^2(x_t-x)\sigma_u^2-\frac{d_n}{nh}2\tau\sum_{t=1}^{n}f(x_{t-1})\mathbf{E}_{t-2}K_h^2(x_t-x)+o_p(1)\text{ (by Lemma 3)} \\
: & =T_n+o_p(1).
\end{aligned}
$$

In addition, by Lemma 1 and Theorem 1, we get

$$
\begin{aligned}
T_n &\xrightarrow{p} \left(\mathbf{E}f^2(x-v_t)+\tau^2+\sigma_u^2-2\tau^2\right) L_G(1,0)\int_{-\infty}^{\infty} K^2(s)ds \\
&= \left(\mathbf{Var}f(x-v_t)+\sigma_u^2\right) L_G(1,0)\int_{-\infty}^{\infty} K^2(s)ds.
\end{aligned}
$$

Fix $\delta>0$ and consider

$$
\begin{aligned}
&\frac{d_n}{nh}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left(\alpha_t+\beta_t+\gamma_t\right)^2 1\left\{\left(\frac{d_n}{nh}\right)^{1/2}|\alpha_t+\beta_t+\gamma_t|>\delta\right\} \\
\le\ & 3\frac{d_n}{nh}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left(\alpha_t^2+\beta_t^2+\gamma_t^2\right) 1\left\{\left(\frac{d_n}{nh}\right)^{1/2}|\alpha_t+\beta_t+\gamma_t|>\delta\right\} \\
\le\ & 3\frac{d_n}{nh}\sum_{t=1}^{n}\left\{\mathbf{E}_{t-2}\left(\alpha_t^2+\beta_t^2+\gamma_t^2\right)^2\right\}^{1/2}\left\{\mathbf{P}_{t-2}\left[\left(\frac{d_n}{nh}\right)^{1/2}|\alpha_t+\beta_t+\gamma_t|>\delta\right]\right\}^{1/2} \\
\le\ & 3\frac{d_n}{nh}\sum_{t=1}^{n}\left\{\mathbf{E}_{t-2}\left(\alpha_t^2+\beta_t^2+\gamma_t^2\right)^2\right\}^{1/2}\left\{\frac{1}{\delta^4}\left(\frac{d_n}{nh}\right)^2\mathbf{E}_{t-2}\left[(\alpha_t+\beta_t+\gamma_t)^4\right]\right\}^{1/2} \\
\le\ & 9\sqrt{3}\frac{d_n}{nh}\sum_{t=1}^{n}\left\{\mathbf{E}_{t-2}\left(\alpha_t^4+\beta_t^4+\gamma_t^4\right)\right\}^{1/2}\left\{\frac{1}{\delta^4}\left(\frac{d_n}{nh}\right)^2\mathbf{E}_{t-2}\left[(\alpha_t+\beta_t+\gamma_t)^4\right]\right\}^{1/2} \\
&= 9\sqrt{3}\frac{1}{\delta^2}\frac{d_n}{nh}\left(\frac{d_n}{nh}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left(\alpha_t^4+\beta_t^4+\gamma_t^4\right)\right).
\end{aligned}
$$

We shall show that $\left(\frac{d_n}{nh}\right)^2\sum_{t=1}^n \mathbf{E}_{t-2}\alpha_t^4=o_p(1)$. We have

$$
\left(\frac{d_n}{nh}\right)^2\sum_{t=1}^{n}\mathbf{E}_{t-2}\alpha_t^4 \le 9\left(\frac{d_n}{nh}\right)^2\sum_{t=1}^{n}\left[\mathbf{E}_{t-2}f^4(x_{t-1})K_h^4(x_t-x)+f^4(x_{t-1})\left\{\mathbf{E}_{t-2}K_h(x_t-x)\right\}^4\right].
$$

By Lemma 1 and Theorem 1

$$
\left(\frac{d_n}{nh}\right)^2\sum_{t=1}^{n}\mathbf{E}_{t-2}f^4(x_{t-1})K_h^4(x_t-x)=O_p\left(\frac{d_n}{nh}\right),
$$

and by Lemma 3

$$
\left(\frac{d_n}{nh}\right)^2\sum_{t=1}^{n}f^4(x_{t-1})\left\{\mathbf{E}_{t-2}K_h(x_t-x)\right\}^4=o_p\left(\frac{d_n}{nh}\right).
$$

Using similar arguments to those used above we have $\left(\frac{d_n}{nh}\right)^2\sum_{t=1}^n \mathbf{E}_{t-2}\left(\beta_t^4+\gamma_t^4\right)=o_p(1)$. Therefore, by Hall and Heyde (1980, Theorem 3.2)

$$
M_n \xrightarrow{d} \left\{\left(\mathbf{Var}f(x+v_t)+\sigma_u^2\right) L_G(1,0)\int_{-\infty}^{\infty} K^2(s)ds\right\}^{1/2} W := M,
$$

where W is a standard normal variate. Next, the quadratic variation of M_n, is $[M_n] = \frac{d_n}{nh}\sum_{t=1}^{n}(\alpha_t+\beta_t+\gamma_t)^2$. The following condition (see Jacod and Shiryaev, 1986)

$$\sup_n \left(\frac{d_n}{nh}\right)^{1/2} \max_{0\le t\le n} \mathbf{E}\left|\alpha_t+\beta_t+\gamma_t\right| < \infty$$

is sufficient for

$$([M_n], M_n) \xrightarrow{d} ([M], M).$$

For some $\gamma > 2$ we have

$$\begin{aligned}
& \left(\frac{d_n}{nh}\right)^{1/2} \max_{0\le t\le n} \mathbf{E}\left|\alpha_t+\beta_t+\gamma_t\right| \\
\le\ & \left(\frac{d_n}{nh}\right)^{1/2} \max_{0\le t\le n} \left\{\mathbf{E}\left|\alpha_t+\beta_t+\gamma_t\right|^\gamma\right\}^{\frac{1}{\gamma}} \le A\left(\frac{d_n}{nh}\right)^{1/2} \max_{0\le t\le n}\left\{\mathbf{E}\left(|\alpha_t|^\gamma+|\beta_t|^\gamma+|\gamma_t|^\gamma\right)\right\}^{\frac{1}{\gamma}} \\
=\ & A\left(\frac{d_n}{nh}\right)^{1/2}\left\{\mathbf{E}\max_{0\le t\le n}\left(|\alpha_t|^\gamma+|\beta_t|^\gamma+|\gamma_t|^\gamma\right)\right\}^{\frac{1}{\gamma}} \le A\left(\frac{d_n}{nh}\right)^{1/2}\left\{\mathbf{E}\sum_{t=1}^{n}\left(|\alpha_t|^\gamma+|\beta_t|^\gamma+|\gamma_t|^\gamma\right)\right\}^{\frac{1}{\gamma}}.
\end{aligned}$$

Consider the first summand. We have

$$\begin{aligned}
\left(\frac{d_n}{nh}\right)^{\gamma/2}\mathbf{E}\sum_{t=1}^{n}|\alpha_t|^\gamma &= \left(\frac{d_n}{nh}\right)^{\gamma/2}\mathbf{E}\sum_{t=1}^{n}\left|f(x_{t-1})K_h(x_t-x)-\mathbf{E}_{t-2}f(x_{t-1})K_h(x_t-x)\right|^\gamma \\
&\le A\left(\frac{d_n}{nh}\right)^{\gamma/2}\mathbf{E}\sum_{t=1}^{n}|f(x_{t-1})|^\gamma\,\mathbf{E}_{t-2}\left|K_h(x_t-x)\right|^\gamma = O\left(\left(\frac{d_n}{nh}\right)^{(\gamma-2)/2}\right),
\end{aligned}$$

where the last equality is due to Lemma 4. Dealing with the other terms in a similar way, we get

$$\left(\frac{d_n}{nh}\right)^{1/2}\max_{0\le t\le n}\mathbf{E}\left|\alpha_t+\beta_t+\gamma_t\right| = O\left(\left(\frac{d_n}{nh}\right)^{(\gamma-2)/2}\right) = o(1).$$

Next consider the predictable quadratic variation of M_n, $\langle M_n\rangle := \frac{d_n}{nh}\sum_{t=1}^{n}\mathbf{E}_{t-2}(\alpha_t+\beta_t+\gamma_t)^2$. We shall show that $[M_n] = \langle M_n\rangle + o_p(1)$. This will be sufficient for the joint convergence

$$(\langle M_n\rangle, M_n) \xrightarrow{d} ([M], M).$$

We have

$$\begin{aligned}
& \mathbf{E}\left|[M_n]-\langle M_n\rangle\right| \\
\le\ & \frac{d_n}{nh}\mathbf{E}\left|\sum_{t=1}^{n}\overbrace{\left[(\alpha_t+\beta_t+\gamma_t)^2-\mathbf{E}_{t-2}(\alpha_t+\beta_t+\gamma_t)^2\right]}^{z_t}\right| \\
\le\ & \left\{\left(\frac{d_n}{nh}\right)^2\mathbf{E}\left(\sum_{t=1}^{n}z_t\right)^2\right\}^{1/2} = \left\{\left(\frac{d_n}{nh}\right)^2\mathbf{E}\sum_{t=1}^{n}z_t^2\right\}^{1/2} \\
=\ & \left\{\left(\frac{d_n}{nh}\right)^2\mathbf{E}\sum_{t=1}^{n}\left[(\alpha_t+\beta_t+\gamma_t)^2-\mathbf{E}_{t-2}(\alpha_t+\beta_t+\gamma_t)^2\right]^2\right\}^{1/2} \\
\le\ & \left\{A\left(\frac{d_n}{nh}\right)^2\mathbf{E}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left(\alpha_t^4+\beta_t^4+\gamma_t^4\right)\right\}^{1/2} = o(1),
\end{aligned}$$

where the last equality can be established using similar arguments as those above (Lemma 4).

Hence, the NW estimator has the following form

$$\begin{aligned}
&\left(\sum_{t=1}^{n} K_h(x_t - x)\right)^{1/2} \left[\hat{f}(x) - \mathbf{E}f(x+v_t)\right] \\
= \quad & \frac{M_n}{\left(\frac{d_n}{nh}\sum_{t=1}^{n} K_h(x_t - x)\right)^{1/2}} = \frac{\langle M_n\rangle^{1/2}}{\left(\frac{d_n}{nh}\sum_{t=1}^{n} K_h(x_t - x)\right)^{1/2}} \frac{M_n}{\langle M_n\rangle^{1/2}} := A_n B_n.
\end{aligned}$$

Now by Theorem 1, it can be easily seen that

$$A_n \xrightarrow{p} \frac{\left(\left(\sigma_u^2 + \mathbf{Var}f(x-v_t)\right) L_G(1,0)\int_{-\infty}^{\infty} K^2(s)ds\right)^{1/2}}{\left(L_G(1,0)\int_{-\infty}^{\infty} K(s)ds\right)^{1/2}} = \left(\left(\sigma_u^2 + \mathbf{Var}f(x-v_t)\right)\int_{-\infty}^{\infty} K^2(s)ds\right)^{1/2}.$$

In addition, $B_n \xrightarrow{d} W$, and the result for $r > s$ follows.

Next, suppose that $r < s$. Set $\tau := \mathbf{E}f(x+v_t)$. We have

$$\begin{aligned}
\hat{f}(x) - \mathbf{E}f(x+v_t) \quad = \quad & \frac{\overbrace{\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}\left[\mathbf{E}_{t-2}f(x_t) - \mathbf{E}f(x+v_t)\right] K_h(x_{t-1}-x)}^{:=R_n}}{\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n} K_h(x_{t-1}-x)} \\
& + \frac{\sum_{t=1}^{n}\overbrace{\left[f(x_t) - \mathbf{E}_{t-2}f(x_t)\right] K_h(x_{t-1}-x)}^{\alpha_t}}{\sum_{t=1}^{n} K_h(x_{t-1}-x)} \\
& + \frac{\sum_{t=1}^{n}\overbrace{K_h(x_{t-1}-x)u_t}^{\beta_t}}{\sum_{t=1}^{n} K_h(x_{t-1}-x)} \\
: \quad = \quad & \frac{R_n + M_n}{\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n} K_h(x_t - x)}.
\end{aligned}$$

Notice that

$$
\begin{aligned}
& \mathbf{E}\left|R_n\right| \\
= & \mathbf{E}\left|\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}\left[\mathbf{E}_{t-2}f(x_t)-\mathbf{E}f(x+v_t)\right]K\left(\frac{x_{t-1}}{h}-\frac{x}{h}\right)\right| \\
= & \mathbf{E}\left|\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}\int_v\left[f\left(x_{t-1}+v\right)-\mathbf{E}f(x+v_t)\right]K\left(\frac{d_n x_{t-1,n}}{h}-\frac{x}{h}\right)p_1(v)dv\right| \\
\leq & \left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}\int_y\left|\int_v\left[f\left(d_n d_{t-1,0,n}y+v\right)-\mathbf{E}f(x+v_t)\right]K\left(\frac{d_n d_{t-1,0,n}y}{h}-\frac{x}{h}\right)p_1(v)dv\right| \\
& \times h_{t-1,0,n}(y)dy \\
= & \left(\frac{d_n}{nh}\right)^{1/2}\frac{h}{d_n}\sum_{t=1}^{n}(d_{t-1,0,n})^{-1}\int_z\left|\int_v\left[\mathbf{E}f(hz+x+v)-\mathbf{E}f(x+v_t)\right]K\left(z\right)p_1(v)dv\right| \\
& \times h_{t-1,0,n}\left(\frac{zh+x}{d_n d_{t-1,0,n}}\right)dy \\
= & \left(\frac{d_n}{nh}\right)^{1/2}\frac{h}{d_n}\sum_{t=1}^{n}(d_{t-1,0,n})^{-1}\int_z\left|\left[\mathbf{E}f(hz+x+v_t)-\mathbf{E}f(x+v_t)\right]K\left(y\right)\right|h_{t-1,0,n}\left(\frac{zh+x}{d_n d_{t-1,0,n}}\right)dz \\
\leq & \left(\frac{n}{hd_n}\right)^{1/2}\frac{h^{1+\gamma}}{n}\sum_{t=1}^{n}(d_{t-1,0,n})^{-1}\int_z f_1(z,x)K\left(z\right)dz\rightarrow 0.
\end{aligned}
$$

Next, notice that $\{M_n,\mathcal{F}_{n-1}\}$ is a martingale sequence and satisfies a martingale CLT. The proof is similar to that provided in the previous part. Consider

$$
\begin{aligned}
& \frac{d_n}{nh}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left(\alpha_t+\beta_t\right)^2 \\
= & \frac{d_n}{nh}\sum_{t=1}^{n}K_h^2(x_{t-1}-x)\mathbf{E}_{t-2}f^2(x_t)-\frac{d_n}{nh}\sum_{t=1}^{n}K_h(x_{t-1}-x)\left[\mathbf{E}_{t-2}f(x_t)\right]^2 \\
& +\frac{d_n}{nh}\sum_{t=1}^{n}K_h^2(x_{t-1}-x)\mathbf{E}_{t-2}u_t^2 \\
& \stackrel{p}{\rightarrow}\left(\mathbf{E}f^2(x+v_t)-\tau^2+\sigma_u^2\right)L_G(1,0)\int_{-\infty}^{\infty}K^2(s)ds \\
= & \left(\mathbf{Var}f(x+v_t)+\sigma_u^2\right)L_G(1,0)\int_{-\infty}^{\infty}K^2(s)ds,
\end{aligned}
$$

where the last limit is due to Lemma 1, Lemma 5 and Theorem 1. The remaining proof follows from similar arguments to those used in the previous part. ■

Proof of Theorem 3.

Write

$$
\left\{\frac{d_n}{nh}\sum_{t=1}^{n}K_h(x_{t-s}-x)\right\}\hat{\sigma}^2=
$$

$$= \frac{d_n}{nh} \sum_{t=1}^{n} \left[f(x_{t-r}) - \hat{f}(x) \right]^2 K_h(x_{t-s} - x) + \frac{d_n}{nh} \sum_{t=1}^{n} u_t^2 K_h(x_{t-s} - x)$$
$$+ \frac{2d_n}{nh} \sum_{t=1}^{n} u_t \left[f(x_{t-r}) - \hat{f}(x) \right] K_h(x_{t-s} - x) := \alpha_n + \beta_n + \gamma_n.$$

It follows directly from Theorem 1 and Theorem 2 that

$$\alpha_n \xrightarrow{p} \mathbf{Var} \left\{ f \left(x + \sum_{rs} v_i \right) \right\} \int_{-\infty}^{\infty} K(s) ds.$$

In addition, manipulations similar to those used in the proof of Theorem 2 give

$$\beta_n + \gamma_n = \sigma_u^2 \int_{-\infty}^{\infty} K(s) ds + O_p \left(\left(\frac{d_n}{nh} \right)^{1/2} \right).$$

This shows the first part of Thorem 3.

In view of the above and Theorem 2, it can be easily seen that $\hat{t}(x, \theta) \xrightarrow{d} N(0, 1)$. ■

8 References

Cai, Z., Q. Li, and J. Y. Park (2009). "Functional coefficient models for nonstationary time series data", *Journal of Econometrics*, 149, 101-113.

Chang, Y., Park J.Y. and P.C.B. Phillips (2001) Nonlinear econometric models with cointegrated and deterministically trending regressors. *Econometrics Journal* 4, 1-36.

Ghysels, E., P. Santa-Clara and R Valkanov (2004). The Midas touch: mixed data sampling regression models, Scientific Series, Cirano, Montreal.

Ghysels, E., P. Santa-Clara and R Valkanov (2006). Predicting volatility: getting the most out of return data sampled at different frequencies, *Journal of Econometrics*, 131, 59-95.

Guerre, E. (2004) Design-adaptive pointwise non-parametric regression estimation for recurrent Markov time series. CREST Working Paper No. 2004-22

Hall, P. and C.C. Heyde (1980) *Martingale limit theory and its application.* Academic Press.

Hu, L. and P. C. B. Phillips (2004): "Dynamics of the Federal Funds Target Rate: A Nonstationary Discrete Choice Approach," *Journal of Applied Econometric*s, 19, 851-867

Jacod, J. and A.N. Shiryaev (1986) *Limit theorems for stochastic processe*s, Springer-Verlag.

Karlsen, H. A., Mykelbust, T. and D. Tjøstheim (2007). Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics*, 35, 252-299.

Kasparis, I. (2008) Detection of functional form misspecification in cointegrating relations. *Econometric Theory*, 24, 1373-1403.

Kasparis, I., Phillips P.C.B. and T. Magdalinos (2008) Instrumental variables estimation of integrable nonstationary models. unpublished manuscript.

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

Park, J.Y. and P.C.B. Phillips (1999) Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 15, 269-298.

Park, J.Y. and P. C. B. Phillips (2000). Nonstationary binary choice, *Econometrica*. 68, 1249-1280.

Park, J.Y. and P. C. B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69, 117-161.

Phillips, P. C. B. (2001): Descriptive Econometrics for Nonstationary Time Series with Empirical Illustrations. *Journal of Applied Econometrics,* 16, 389-413.

Phillips, P.C.B. (2009) Local limit theory for spurious nonparametric regression. *Econometric Theory*, forthcoming.

Schienle, M. (2008) Nonparametric nonstationary regression, PhD thesis, University of Mannheim.

Teräsvirta, T. and A. Eliasson (2001) Nonlinear error correction model and the UK demand for broad money. *Journal of Applied Econometrics* 16, 277-288.

Wang, Q. and P.C.B. Phillips (2009) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, forthcoming.

Wang, Q. and P.C.B. Phillips (2008) Structural nonparametric cointegrating regression. *Econometrica* (in press).

Xiao, Z. (2009). "Functional coefficient cointegrating regression". *Journal of Econometrics* (in press)

Notation

- $\mathbf{E}_t(.) = \mathbf{E}(. \mid \mathcal{F}_t)$
- $\mathbf{P}_t(.) = \mathbf{P}(. \mid \mathcal{F}_t)$
- $K_h(\cdot) = K\left(\frac{\cdot}{h}\right)$
- $a \vee b = \max(a, b)$

- $a \wedge b = \min(a, b)$
- $\sum_{rs} v_i = \mathbf{1}\,(r > s) \sum_{i=s+1}^{r} v_i - \mathbf{1}\,(s > r) \sum_{i=r+1}^{s} v_i$