Dynamic Misspecification in Nonparametric Cointegrating Regression

Ioannis Kasparis
University of Cyprus

and

Peter C. B. Phillips*

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

November, 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 (2009a) 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, in general, 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-true 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. Moreover, we show that when dynamic misspecification is severe, convergence to some pseudo-true function may not hold. The estimator can be divergent or vanishing in this case. 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

1   Introduction

Arguably, any econometric model is an abstraction of reality rather than a true data generating mechanism. Hence, any econometric model is potentially misspecified. Even if observed economic data were generated by the econometric models used in practice, there is a myriad of ways one could depart from the true data generating mechanism. Therefore, it is important to know the limit properties of various estimators, when the underling model is misspecified. A series of papers in the econometric and statistics literature attempts to cast light on this problem. See for example Berk (1966, 1970), Huber (1967), White (1981, 1982), Domowitz and White (1982), Gourieroux, Monfort and Trognon (1984) *inter alia.* Some of the questions raised by the aforementioned papers are summarised by White (1982):

> "*If one does not assume that the probability model is correctly specified, it is natural to ask what happens to the properties of the [maximum likelihood] estimator. Does it still converge to some limit asymptotically, and does this limit have any meaning? If the estimator is somehow consistent, is it also asymptotically normal?*"

It is well known that, under certain conditions, parametric estimators of stationary misspecified models have a well defined limit referred to in the econometric literature as "pseudo-true value". The pseudo-true value can be different than the parameter of interest and is determined by the value that optimises certain limit criterion function. For instance, when NLS is the relevant estimation procedure, the pseudo-true value is the argument that minimises the asymptotic mean square error between the true and the fitted objective functions (see for example White, 1981; Bierens, 1984). For maximum likelihood estimation, the pseudo-true value is the argument that maximises the Kullback-Leibler Information Criterion (KLIC) (see for example Huber 1967; Akaike, 1973; White 1982). Further, it is known that estimators of misspecified stationary models are $\sqrt{n}$-convergent and have Gaussian limit distribution. The asymptotic analysis of misspecified models is not only of theoretical interest. To obtain asymptotic power rates for various specification tests e.g. Ramsey (1969), Bierens (1990) (tests without specific alternative) knowledge about the asymptotic behaviour of the estimator under misspecication is necessary. Moreover, to determine the limit distribution of certain model selection statistics under the null hypothesis, e.g. Cox (1961, 1962), Davidson and MacKinnon (1981) and Vuong (1989) (tests with specific alternative), the estimator's limit distribution about the pseudo-true value, is required.

The aforementioned literature focuses on stationary parametric misspecified models. Misspecified nonstationary models have received less attention. The analysis of nonstationary misspecified models became possible after recent theoretical developments. Although limit theory for linear models with unit roots was provided about twenty five years ago (e.g. Phillips, 1986, 1987; Chan and Wei, 1987) limit theory for nonlinear models with integrated variables is a relatively recent development. 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 subsequent work (Guerre, 2004; Karlsen, Mykelbust and Tjøstheim, 2007; Schienle, 2008; Wang and Phillips, 2009a,b,c;) has developed a limit theory for nonparametric cointegrating regression using Markov chain and local time

asymptotics. In a recent paper Kasparis (2009) provides asymptotic analysis of parametric misspecified models, utilising the Park and Phillips (2001) framework. It shown that the asymptotic behaviour of estimators relating to nonstationary misspecified models can be drastically different than that known for stationary misspecified models. In many cases estimators converge to boundary points of the parameter space. When the parameter space in unbounded, convergence to pseudo-true value does not always hold. Estimators can be divergent. Phillips (2009) provides limit theory for spurious non-parametric regression. In particular, Phillips (2009) considers the case where a nonstationary process is regressed on a possibly irrelevant integrated process by kernel methods. When the regression is spurious, the resultant estimator is $\sqrt{n}$-divergent. The current paper takes the Wang and Phillips (2009a; hereafter WP) framework and analyses 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.

The current work shows that the consequences of dynamic misspecification in a nonstationary framework, largely depend on the nature of the regression function and on the nature of the functions involved in the estimation procedure. As shown in Park and Phillips (1999, 2000, 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, misspecifying 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

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

where θ is an unknown parameter, f_o some regression function, 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):

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

If the regression function f_o 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

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

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_o is integrable then it follows directly from our limit theory (Theorem 1 below) that

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

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 $f(x) = \theta f_o(x)$ 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 (9) 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. We also consider the case of severe dynamic misspecification i.e. the lag differential between the true and the fitted models is large. For badly misspecified models, limit theory is substantially different. In this case, the NW estimator may be divergent, vanishing or converging to some stochastic integral.

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 (Section 2) 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. This limit theory is also useful for parametric models. For instance, the basic limit result we provide extends earlier work on parametric models (e.g. Park and Phillips, 1999; Marmer, 2007; Kasparis, 2009) to a dynamic framework. Moreover, the limit theory of Section 2 is utilised by Kasparis and Phillips (2009a) who provide robust inference in cointegrated systems, when exact integration properties of the covariates are unknown. 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. Before proceeding to the next section, we introduce some notation. For any two real numbers a and b, $a \vee b$ $(a \wedge b)$ denotes their maximum (minimum). As usual, $=_d$ stands for distributional equality.

2 First results

This section provides asymptotic theory for sample covariances of integrable transformations of nonstationary sequences and locally integrable transformation of time sequences. In particular we consider sample covariance terms of the form

$$S_n(\eta) := \frac{c_n}{n} \sum_{t=1}^{[n\eta]} f(x_{t-r}) g\left[c_n\left(\frac{x_{t-s} - x}{\sqrt{n}}\right)\right], \ 0 < \eta \leq 1, \tag{2}$$

where x_t is an I(1) or nearly integrated process, r and s are (possibly different) positive integers. Moreover, g is some integrable function whereas the function f is locally inte-

grable. Finally, c_n is some deterministic sequence of real numbers. The purpose of c_n is to allow for a bandwidth parameter. The bandwidth parameter, h say, can be obtained by setting $h_n = \sqrt{n}/c_n$. Therefore, (2) involves a sample covariance of some locally integrable transformation and some integrable transformation of time translated time series. The term under consideration relates to the recent work of WP. Setting $f = 1$ in (2), we obtain the sample sum analysed by WP.

The sample term of (2) is relevant to both parametric and non-parametric estimation. In non-parametric estimation, the integrable function g would typically play the role of some kernel smoother, whereas the locally integrable function f would correspond to some regression function. Our limit results for S_n are utilised in the next section, to analyse the properties of kernel regressions under dynamic misspecification. An application of our limit theory to correctly specified functional coefficient models with a unit root is also provided (see Example 5 below). In parametric estimation, g could correspond to some instrument. For instance, the covariance asymptotic results, provided in this section, are useful for the analysis of parametric distributed lag models, where the regression function is an integrable transformation of some integrated covariate (see Example 6 below). Limit theory for integrable regression functions, of trending variables, was first provided by Park and Phillips (1999, 2001). Marmer (2007) and Kasparis (2009) utilise this earlier work and develop specification tests for integrable regressions which are subsequently employed to test for stock returns predictability. The limit theory of this section can be readily utilised to extend the work of Park and Phillips (1999, 2001), Marmer (2007) and Kasparis (2009a) to dynamic models. In addition, our covariance asymptotics are exploited by Kasparis and Phillips (2009) who develop robust inference, free of nuisance parameters, when the integration properties of the regressor are unknown (see also Kasparis, Phillips and Magdalinos, 2008). The key idea in the aforementioned work is the utilisation of some integrable instrument. This approach provides standard inference when the regressor is I(0), I(1) or nearly integrated. In addition, inference is free of the nuisance local to unity parameter, when some covariate is a nearly integrated process.

We next specify the proporties of x_t, f and g in detail. The variable x_t is a nonstationary process 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

$$x_t = \rho_n x_{t-1} + v_t, \ x_0 = 0, \tag{3}$$

with $\rho_n = 1 - \frac{c_o}{n}$ for some constant c_o. 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).

Assumptions 2.1 and 2.2 below are largely based on WP. We start by introducing the following notation used in that work. First, c_n is a sequence of real numbers satisfying $c_n \rightarrow \infty$. It is convenient to standardise x_t as follows: $x_{t,n} = x_t/\sqrt{n}$. Then, $x_{t,n}$ $0 \leq t \leq n, n \geq 1$ is a triangular array and the standardisation ensures that $x_{t,n}$ has a limit distribution. We also introduce the sequence of real numbers $d_{l,k,n} = \sqrt{l-k}/\sqrt{n}$. Note that $(x_{l,n} - x_{k,n})/d_{l,k,n}$ has a limit distribution as $l-k \rightarrow \infty$. The sequence c_n is a secondary sequence which differs from $\sqrt{n}$ by a bandwidth factor, so that we usually have $c_n = \sqrt{n}/h_n$ for some bandwidth sequence $h_n \rightarrow 0$ arising in the kernel estimation. As in WP, it is convenient also to use

6

the set notation.

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

Assumption 2.1

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

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

(ii) for some $k_0 > 0$,

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

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

(b) 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 $x \in \mathbb{R}$.

Assumption 2.2

(a) The process $x_{[n\eta],n} := x_{[n\eta]}/\sqrt{n}$ on the Skorohod space $D[0,1]$, converges weakly to a Gaussian process $G(\eta)$ that has a continuous local time process $L_G(\eta, s)$.

(b) On a suitable probability space there exists a process $x_{t,n}^o$ such that $\left(x_{t,n}^o, \ 1 \le t \le n\right) =_d$ $(x_{t,n}, \ 1 \le t \le n)$ and $\sup_{0 \le \eta \le 1} \left|x_{[n\eta],n}^o - G(\eta)\right| = o_p(1)$.

In some cases it is more convenient to work with the Skorokhod copy $x_{t,n}^o$, instead of $x_{t,n}$. In the paper, we establish weak convergence of the NW estimator to some well defined deterministic limit (pseudo-true function), when x_t is the regression covariate. In addition, we provide limit distribution theory for the NW about the pseudo-true function. Therefore, for our purposes, there is no loss of generality if we assume that $\left(x_{t,n}^o, \ 1 \le t \le n\right) = (x_{t,n}, \ 1 \le t \le n)$ instead of $\left(x_{t,n}^o, \ 1 \le t \le n\right) =_d (x_{t,n}, \ 1 \le t \le n)$. Due to this convention, $\xrightarrow{p}$ convergence, for sample functionals of x_t, should be interpreted as $\xrightarrow{d}$ convergence, unless the limit is deterministic.

Assumption 2.3 Set $0 < \gamma \le 1$.

(a) $\lim_{n \to \infty} \sqrt{n}/c_n = 0$, where c_n satisfies $c_n \to \infty$;

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

(c) $\int_z |z| |g(z)| dz$ and $\int_v |f(x - v)|^q p_{r-s}(v) |v| dv < \infty$ for all x and some $q > 1$.

Assumption 2.3[*] Set $0 < \gamma \le 1$.

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

(b) *for n large enough,* $\left| f\left(\frac{\sqrt{n}}{c_n}z + x - v\right) - f\left(m_0 z + x - v\right) \right| \leq \left| \frac{\sqrt{n}}{c_n} - m_0 \right|^{\gamma} f_1(z, x, v)$
with $\int_v \int_z f_1(z, x, v) \, |g(z)| \, p(v) dz dv < \infty$, *for each* x.
    (c) $\int_s \int_v |f\left(m_0 z + x - v\right) g(z)| \, p_{r-s}(v)\left(|v| + |z|\right) dv dz < \infty$, *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 Assumption 2.1 (a) is the same as Assumption 2.3 of WP. Note that when x_t is given by (3), Assumption 2.2 is satisfied. In this case, $G(t)$ is either a Brownian Motion or an Ornstein-Uhlenbeck process. 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 $c_n = \sqrt{n}/h$, 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 local Lipschitz and integrability conditions on f, which are useful technical conditions.
    The following result provides a limit theory for the term S_n of (2). The result is 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 (2). The scale constant τ in the limit results (4) and (5), below, similarly involves the function f, whereas in WP, τ is the energy functional $\tau = \int_{-\infty}^{\infty} g(z) \, dz$ 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) $\left| f\left(\frac{\sqrt{n}}{c_n}z + x - v\right) \right| \leq f_0(z, x, v)$ *for n large enough, with* $\int_v \int_z f_0(z, x, v) \, |g(z)| \, p_{r-s}(v) dz dv <$
∞,
$\int_v \left\{ \int_z |f_0(z, x, v)| \, |g(z)| \, dz \right\}^2 p_{r-s}(v) dv < \infty$ *and* $\int_v \int_z f_0^2(z, x, v) g^2(z) p_{r-s}(v) dz dv < \infty$,
for and each $x \in \mathbb{R}$, *and* $r, s \in \mathbb{N}$;
    (b) *Assumption 2.3 holds and*

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

or
    (c) *Assumption 2.3* holds and*

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

---

[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 (similarly $x_{t-r} - x_{t-s} =_d - \sum_{j=s+1}^{r} v_j$, for $s < r$).

We have the following:

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

$$S_n(\eta) \xrightarrow{d} \tau L(\eta, 0). \tag{4}$$

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

$$\sup_{0 \le \eta \le 1} |S_n(\eta) - \tau L(\eta, 0)| \xrightarrow{p} 0. \tag{5}$$

When $f = 1$, S_n reduces to

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

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

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

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 Nadaraya-Watson kernel regression estimator in the case of dynamic misspecification. It is well known (e.g. White, 1981; White 1982; Domowitz and White, 1982;) that, under certain regularity conditions, parametric estimators of misspecified models convergence to some well defined limit, referred to in the econometric literature as "pseudo-true value". The pseudo-true value is typically different than the parameter of interest. In the current paper it is demonstrated that, when the fitted model suffers from dynamic misspecification, and under certain regularity conditions, the NW estimator has a well defined limit. When the dynamic misspecification is mild -i.e. the lag differential between the true models is finite-, the NW has a "pseudo-true function" limit. The pseudo-true function corresponds to the true regression function as long as the latter is linear. In general the pseudo-true function differs from the true function and is determined by some local average of the true regression function. If dynamic misspecification is severe -i.e. the lag differential between the true and fiitted models goes to infinity in large samples-, there is no pseudo-true function limit. In this case, the NW divereges, vanishes or converges to some random limit, depending on the properties of the true regression function.

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

$$y_t = f(x_{t-r}) + u_t, \text{ for some integer lag } r \ge 0. \tag{6}$$

where f is a locally integrable regresson function. The regressor x_t is given by (3), whereas the regression error u_t is a martingale difference sequence. Both x_t and u_t are defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

We concentrate on the case where a version of (6) 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

$$y_t = \hat{f}(x_{t-s}) + \hat{u}_t, \text{ for some fixed integer lag } s \geq 0, \, r \neq s, \tag{7}$$

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

$$\hat{f}(x) = \frac{\sum_{t=s}^{n} K\left(\frac{x_{t-s}-x}{h}\right) y_t}{\sum_{t=s}^{n} K\left(\frac{x_{t-s}-x}{h}\right)}, \tag{8}$$

for some kernel function K. In Subsection 3.1, the integer lags r and s, in (6), (7), are assumed to be fixed. In Subsection 3.2 we consider the case of severe dynamic misspecification i.e. we assume that the lag differential $|r - s| \to \infty$, as $n \to \infty$.

3.1 Mild dynamic misspecification

The limit results of Theorem 1 are utilised in this subsection, for the asymptotic analysis of $\hat{f}$. In particular, the function g of Theorem 1 will play the role of some kernel function relating to (8). We need to be more specific about the components of (6), (7) and (8). We start with the some regularity conditions on the regression function and the kernel. The subsequent conditions are similar to those used in WP.

Assumption 3.1 *Assume that the following hold:*
(a) *The regression function f of (6) satisfies Assumption 2.3.*
(b) *The kernel K of (8) equals the function g of Assumption 2.3 and $c_n := \sqrt{n}/h$. In addition, $\int_{-\infty}^{\infty} K(s)ds = 1$ 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, $|\mathbf{E}f\left(hy + x + \sum_{rs} v_i\right) - \mathbf{E}f\left(x + \sum_{rs} v_i\right)| \leq h^\gamma f_1(y, x)$ with $0 < \gamma \leq 1$, 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 such that $\mathbf{E}(u_t^2|\mathcal{F}_{n,t-1}) = \sigma_{u,t}^2 \to \sigma_u^2$, a.s. as $t \to \infty$.*

Assumption 3.4. *For any $m > 0$, $\sup_{1 \leq t \leq n} \mathbf{E}(u_t^{2+m}|\mathcal{F}_{n,t-1}) < \infty$ a.s.*

Assumption 3.2 is a technical condition that imposes smoothness on the limit of $\hat{f}(x)$. Assumption 3.3 is common in the literature of nonlinear models with integrated time series e.g. Park and Phillips (2000, 2001), Wang and Phillips (2009a). Note that Assumption 2.2(b) and Assumption 3.3 postulate that y_t is predetermined i.e. $\mathbf{E}(y_t|\mathcal{F}_{n,t-1}) = f(x_{t-r})$. Assumption 3.3 is important for our derivations as it allows the use of martingale convergence methods. Assumption 3.3 has been recently relaxed by Wang and Phillips (2009b) who consider structural nonparametric regressions with unit roots. Relaxation of the martingale difference assumption complicates asymptotic theory substantially. Wang and

Phillips (2009b) develop novel approximate martingale convergence methods. Their approach is significantly different than that followed in the current paper. 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 2.1, 2.2. and 3.1-3.3 hold.*
(b) *The bandwidth h satisfies $\sqrt{n}h \to \infty$ and $h \to 0$ as $n \to \infty$.*

Then, as $n \to \infty$,

$$\hat{f}(x) \xrightarrow{p} \mathbf{E}f\left(x + \sum_{rs} v_i\right). \tag{9}$$

In addition, suppose the following hold:
(c) *Assumption 3.4 holds.*
(d) *The component functions $\{f^2, f^4\}$ and the power kernel functions $\{K^2, K^4\}$ in the sample quantities $\frac{c_n}{n}\sum_{t=1}^n f^2\left(\sqrt{n}x_{t-r,n}\right)K^2\left[c_n\left(x_{t-s,n} - \frac{x}{\sqrt{n}}\right)\right]$ and
$\frac{c_n}{n}\sum_{t=1}^n f^4\left(\sqrt{n}x_{t-r,n}\right)K^4\left[c_n\left(x_{t-s,n} - \frac{x}{\sqrt{n}}\right)\right]$ both satisfy the conditions of Theorem 1.*
(e) *The bandwidth parameter h satisfies $\sqrt{n}h^{1+2\gamma} \to \infty$.*
Then, as $n \to \infty$,

$$\left(\sum_{t=1}^n K_h\left(x_{t-s} - x\right)\right)^{1/2}\left(\hat{f}(x) - \mathbf{E}f\left(x + \sum_{rs} v_i\right)\right) \xrightarrow{d} N\left(0, \sigma^2\right), \tag{10}$$

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

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

$$\mathbf{E}f\left(x + \sum_{rs} v_i\right) = \int f(x + w)\, p_{r-s}(w)\, dw, \tag{11}$$

where $\sum_{rs} v_i$ has density $p_{r-s}(w)$.[2] The limit (11) 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

$$\left(\sum_{t=1}^n K_h\left(x_{t-s} - x\right)\right)^{1/2}\left(\hat{f}(x) - f(x)\right) \xrightarrow{d} N\left(0, \sigma_u^2 \int_{-\infty}^{\infty} K(s)^2 ds\right), \tag{12}$$

---

[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 $p_{r-s}(w)$.

as in WP under suitable undersmoothing or choice of h in the regression. Both (12) and (10) may be adjusted to account for a bias term of $O\left(h^2\right)$ in the limit theory, as shown in Wang and Phillips (2009), but in view of the inconsistency already present in (10) there is little reason to provide that development in the case of misspecification.

The lag misspecification in the fitted nonparametric cointegrating relation (7) produces both inconsistency and a reduction in precision in the limit theory for the NW estimator. The limit distributions (10) and (12) differ in terms of both centering and variance. The centering is explained by the inconsistency (9) under mistiming $(r \neq s)$ of the lagged relationship. The additional variance in the limit distribution (10) occurs due to the term $\mathbf{Var} f\left(x + \sum_{rs} v_i\right)$, which is non zero whenever $r \neq s$. After scaling and centering the NW estimator of (8) on its probability limit we get

$$
\left(\sqrt{n} h\right)^{1/2}\left[\hat{f}(x) - p\lim \hat{f}(x)\right] = \left[\left(\sqrt{n} h\right)^{-1} \sum_{t=s}^{n} K_h\left(x_{t-s} - x\right)\right]^{-1}
$$

$$
\times \left(\sqrt{n} h\right)^{-1/2}\left[\sum_{t=s}^{n} K_h\left(x_{t-s} - x\right) u_t + \sum_{t=s}^{n}\left[f(x_{t-r}) - p\lim \hat{f}(x)\right] K_h\left(x_{t-s} - x\right)\right]. \text{(13)}
$$

When the fitted model is correctly specified, i.e. $r = s$, the second term on the last line of (13) is asymptotically negligible. In fact, the limit distribution of additional term is determined by the distribution of the increment $x_{t-r} - x_{t-s} = \sum_{rs} v_i$, which is of course degenerate, when $r = s$. Under dynamic misspecification however, the limit distribution of the increment process contributes in the limit variance. The extra component in the variance is $\mathbf{Var}\left\{f\left(x + \sum_{rs} v_i\right)\right\}$, which arises as in (4) 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. Therefore misspecification in the nonparametric framework necessarily results in larger limit variance.[3]

In the special case of linear cointegration with $f\left(x_t\right) = \theta x_t$, we have from (9)

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

so that kernel regression is consistent under lag misspecification, corresponding to the temporal invariance of linear cointegrating regression. In this case, (10) becomes

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

---

[3] Note misspecification in the parametric framework does not necessarily result in larger asymptotic variance. For instance denote by $\hat{\theta}$ the NLS estimator relating to some misspecified parametric model. Then the limit distribution of $\hat{\theta}$ about its probability limit (i.e. the pseudo-true value) is determined by a term of the form (e.g. White 1981; Kasparis 2009)

$$
d_n\left(\hat{\theta} - p\lim \hat{\theta}\right) = -H_n^{-1} \times \mathbf{s}_n,
$$

where d_n is some scaling sequence, H_n is the second derivative of the NLS objective function and $\mathbf{s}_n$ is the first derivative of the NLS objective function. Under misspecification, $\mathbf{s}_n$ involves additional components. This is analogous to the nonparametric estimator of (13). Nevertheless, the limit variance of $\hat{\theta}$ is not necessarily larger than that obtained under correct specification. This because, under misspecifcation, H_n involves additional components as well.

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 (6) is the true model and (7) 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

$$\hat{f}(x) \xrightarrow{p} \mathbf{E}f\left(x_{t-r}|x_{t-s} = x\right), \tag{14}$$

which is the analogue for the stationary time series x_t of the inconsistency shown in (9). 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 (14) 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 (9) 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 (2009b) 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

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

13

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

$$\hat{t}(x, \hat{\theta}) \xrightarrow{d} N(0, 1). \tag{15}$$

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

$$\hat{t}(x, \hat{\theta}) \sim (\sqrt{n}h)^{1/2} \left\{ \mathbf{E}f\left(x + \sum_{rs} v_i\right) + \frac{\kappa_f(\sqrt{n}) \int_{-\infty}^{\infty} s H_f(s) L_G(1, s) ds}{\sqrt{n} \int_{-\infty}^{\infty} s^2 L_G(1, s) ds} x \right\}, \tag{16}$$

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 $x f(x)$) we find that

$$\hat{t}(x, \hat{\theta}) \sim (\sqrt{n}h)^{1/2} \left\{ \mathbf{E}f\left(x + \sum_{rs} v_i\right) + \frac{\int_{-\infty}^{\infty} s f(s) ds L_G(1, 0)}{n^{3/2} \int_{-\infty}^{\infty} s^2 L_G(1, s) ds} x \right\}. \tag{17}$$

Results (16) and (17) show that the simple linearity test statistic $\hat{t}(x, \hat{\theta})$ in (15) has power against both homogeneous and integrable nonlinear functions and is robust to dynamic specification.

3.2 Severe dynamic misspecification

Next, we consider the consequences of severe dynamic misspecification. Suppose that the true model is given by (6) and the fitted is

$$y_t = \hat{f}(x_{t-s_n}) + \hat{u}_t, \tag{18}$$

where the sequence $s_n := [cn]$ with $0 < c < 1$. Consider the processes $(x_{[n\eta]-r}, \ x_{[n\eta]-s_n})/\sqrt{n}$ on $D[0, 1]$. Then under (3) and the assumptions of Section 2 we have the following weak convergence result on $D[0, 1]$

$$(x_{[n\eta]-r}, \ x_{[n\eta]-s_n})/\sqrt{n} \Rightarrow \left(G(\eta), \tilde{G}(\eta)\right),$$

where $\tilde{G}(\eta) = G(\eta - c)\mathbf{1}(1 \geq \eta \geq c)$.

The NW estimator of (18) does not have a pseudo-true function limit. In particular, $\hat{f}$ diverges when f is unbounded locally integrable. For bounded locally integrable f, the estimator $\hat{f}$ has a stochastic integral limit. Finally, if f is integrable, $\hat{f}$ vanishes. The limit properties of the NW estimator are demonstrated by the following result.

Theorem 4. *Suppose that:*

   (a) *The true model is given by (6) and the fitted model by (18).*
   (b) *Assumptions 2.1, 2.2. and 3.1-3.3 hold.*
   (c) *The bandwidth h satisfies $\sqrt{n}h \to \infty$ and $h \to 0$ as $n \to \infty$.*

(i) *Suppose that f is asymptotically homogeneous with homogeneous function H_f and asymptotic order κ_f i.e. $f(\lambda x) = \kappa_f(\lambda)H_f(x) + \nu(x,\lambda)$ with $\sup_x |\nu(x,\lambda)| = o(\kappa_f(\lambda))$ as $\lambda \to \infty$. In addition, $\kappa_f(\sqrt{n})\sqrt{h\sqrt{n}} \to \infty$. Then, as $n \to \infty$,*

$$\kappa_f\left(\sqrt{n}\right)^{-1}\hat{f}(x) \xrightarrow{d} \frac{1}{L_G(1-c,0)} \int_c^1 H_f\left(G(\eta)\right) dL_G(\eta - c, 0).$$

(ii) *Suppose that f is integrable with $|f(hp_1 + p_2)| \le f_0(p_1, p_2)$ for n large enough, with $\int_{p_1}\int_{p_1} f_0(p_1,p_2)K(p_1)dp_1 dp_2 < \infty$. Then, as $n \to \infty$,*

$$\sqrt{n^{1/2}h}\,\hat{f}(x) \xrightarrow{d} MN\left(0,\ \sigma_u^2 \int_{-\infty}^{\infty} K(s)^2\, ds L_G^{-1}(1-c,0)\right)$$

Remarks.

(a) For asymptotically homogeneous f the estimator can be divergent. In particular, if the regression is unbounded, $\hat{f}$ diverges. The divergence rate is determined by the asymptotic order function κ_f. For bounded asymptotically homogeneous functions, κ_f is fixed. In this case $\hat{f}$ has a stochastic integral limit.

(b) Unlike the mild misspecification case, $\hat{f}$ is inconsistent even if f is linear. In fact, $\hat{f}$ is $\sqrt{n}$-divergent, when the true regression function is a linear one. The consequences of dynamic misspecifation in this case are analogous to those of spurious nonparanetric regression (see Phillips, 2009).

(c) For integrable f, $\hat{f}$ converges in probability to zero at rate $\sqrt{n^{1/2}h}$. The limit distribution about the estimator's limit is mixed normal, just like the mild misspecification case. In addition, limit variance is larger in this case as well. Note that $L_G^{-1}(1-c,0) > L_G^{-1}(1,0)$, when $1 > c > 0$.

Examples.

(a) Suppose that the regression function f is $f(x) = e^x/(1+e^x)$ or similar to the Michaelis-Menten model: $f(x) = x(1+x)^{-1}\mathbf{1}(x \ge 0)$ (see Bates and Watts, 1998). In both cases, $H_f(x) = \mathbf{1}(x \ge 0)$ and $\kappa(\lambda) = 1$. Therefore, by virtue of Theorem 4 we get the stochastic integral limit $\hat{f}(x) \xrightarrow{d} \frac{1}{L_G(1-c,0)} \int_c^1 \mathbf{1}(G(\eta) > c)\, dL_G(\eta - c, 0)$.

(b) Suppose that $f(x) = \ln|x|$. Note that f is asymptotically homogenous with $H_f(x) = 1$ and $\kappa(\lambda) = \ln(\lambda)$. For this regression function the NW diverges. In particular, $\ln(\sqrt{n})\hat{f}(x) \xrightarrow{p} 1$.

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

$$y_t = \lambda f(x_t) + (1 - \lambda)f(x_{t-1}) + u_t, \ 0 \le \lambda \le 1, \tag{19}$$

and where x_t and u_t are as in Example 1. If the fitted model ignores the temporal aggregation in (19) 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

$$x_t = f(x_{t-1}) + u_t, \tag{20}$$

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

$$x_t = \hat{f}(x_{t-2}) + \hat{u}_t. \tag{21}$$

Under the true model (20) Assumption 2.2 holds with $x_{[n\eta],n} = \frac{1}{\sqrt{n}} \sum_{t=3}^{[n\eta]-2} u_t \xrightarrow{d} G(\eta)$, where $G(\eta)$ is Brownian motion. In view of Theorem 2 we get

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

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 (21) 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

$$n \left(\hat{\rho} - 1 \right) \xrightarrow{d} 2 \left[\int_0^1 W^2 \right]^{-1} \int_0^1 W dW,$$

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

Example 4. (Misspecified functional coefficient 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

$$y_t = \beta \left(z_t \right)' x_t + u_t, \quad t = 1, ..., n \tag{22}$$

where y_t and z_t are scalar, z_t is an I(1) process, x_t is stationary, and u_t is a martingale difference sequence with constant conditional variance σ^2 and finite fourth moments. The functional coefficient $\beta(.)$ 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 $\hat{\beta}(z)$ is consistent with mixed normal distribution. If (22) 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

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

where $h \rightarrow 0$, $\nu_0 = \int K(s)^2 ds$ and $L_{W_z}(1,0)$ is the local time of some Brownian motion process. Misspecification of functional regression therefore leads to inconsistency and an increase in limiting variance. The extra component in the variance term is $\mathbf{Var} \left[\beta \left(z - \Delta z_t \right) \right]$. 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. (Functional coefficient model with a unit root) Consider the model

$$y_t = \beta \left(x_t \right) x_t + u_t, \quad t = 1, ..., n,$$

where x_t is an integrated process. Unlike to the functional coefficient model of CLP, the specification shown above does not involve any stationary regressor. Consider the local linear estimator

$$\begin{bmatrix} \widehat{\beta}(x) \\ \widehat{\beta}^{(1)}(x) \end{bmatrix} = \arg\min_{\theta_0,\,\theta_1} \sum_{t=1}^{n} [y_t - \theta_0 x_t - \theta_1 x_t(x_t - x)]^2 K_h(x_t - x),$$

where $\beta^{(1)}(x)$ is the first derivative of $\beta(x)$. The asymptotic analysis of the local linear estimator involves sample covariance terms like those that appear in (2), with $r = s$. Set $D_n = diag(1, h)$. Under certain regularity conditions, a repeated application of Theorem 1 together with the martingale CLT give

$$\sqrt{n^{1/2} h} D_n \left\{ \begin{bmatrix} \widehat{\beta}(x) \\ \widehat{\beta}^{(1)}(x) \end{bmatrix} - \begin{bmatrix} \beta(x) \\ \beta^{(1)}(x) \end{bmatrix} \right\} \xrightarrow{d} MN\left(0,\ \sigma_u^2 \Phi \left[L_G(1,0)x^2\right]^{-1}\right),$$

where

$$\Phi = \left\{ \int_{-\infty}^{\infty} \begin{bmatrix} 1 & s \\ s & s^2 \end{bmatrix} K(s)ds \right\}^{-1} \left\{ \int_{-\infty}^{\infty} \begin{bmatrix} 1 & s \\ s & s^2 \end{bmatrix} K(s)^2 ds \right\} \left\{ \int_{-\infty}^{\infty} \begin{bmatrix} 1 & s \\ s & s^2 \end{bmatrix} K(s)ds \right\}^{-1}.$$

Example 6. (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

$$y_t = \theta_1 f_1(x_t) + \theta_2 f_2(x_{t-1}) + u_t, \tag{23}$$

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 (23). 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 (23) 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:

$$\sqrt[4]{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} \sigma_u L_G(1,0)^{-1/2} V^{-1/2} Z, \tag{24}$$

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 (24).

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 (1992) 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 five lemmas extend the WP 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 $\phi_\epsilon(x) = (2\pi\epsilon^2)^{-1/2}\exp\left(-x^2/2\epsilon^2\right)$ and $\phi(x) = \phi_1(x)$. We also often write the density $p_1(v)$ as $p(v)$. Moreover, we introduce the following notation for conditional expectation and conditional probability respectively: $\mathbf{E}_t(.) = \mathbf{E}(.\mid\mathcal{F}_{n,t})$ and $\mathbf{P}_t(.) = \mathbf{P}(.\mid\mathcal{F}_{n,t})$. In the following proofs, we use A as a generic constant whose value may change in each location.

Lemma 1. *Suppose that*
  (a) *Assumption 2.1 holds.*
  (b) $\left| f\left(\frac{\sqrt{n}}{c_n}z + x - v\right) \right| \leq f_0(z,x,v)$, *for n large enough and*
(i) $\int_v \int_z f_0(z,x,v)\left|g(z)\right| p_{r-s}(v) dz dv < \infty$,
(ii) $\int_v \left\{\int_z \left|f_0(z,x,v)\right|\left|g(z)\right| dz\right\}^2 p_{r-s}(v) dv < \infty$ *and*
(iii) $\int_v \int_z f_0^2(z,x,v)g^2(z)p_{r-s}(v) dz dv < \infty$,
for $r,s \in \mathbb{N}$ and $x \in \mathbb{R}$.

Let

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

Then

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

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

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

$$= \frac{c_n}{n} \sum_{t=1}^{[n\eta]} \mathbf{E}_{t-2} z_t + \frac{c_n}{n} \sum_{t=1}^{[n\eta]} \left(z_t - \mathbf{E}_{t-2} z_t\right). \tag{25}$$

We show that the second term in (25) is $o_p(1)$. Notice that $\left\{(z_t - \mathbf{E}_{t-2} z_t), \mathcal{F}_{n,t-1}\right\}$ is a martingale difference sequence. Hence,

$$\mathbf{E}\mathbf{E}_{t-2}\left(\frac{c_n}{n} \sum_{t=1}^{[n\eta]} \left(z_t - \mathbf{E}_{t-2} z_t\right)\right)^2 = \left(\frac{c_n}{n}\right)^2 \mathbf{E}\left\{\sum_{t=1}^{[n\eta]} \mathbf{E}_{t-2} z_t^2 - \sum_{t=1}^{[n\eta]} \left(\mathbf{E}_{t-2} z_t\right)^2\right\}. \tag{26}$$

The first term on right hand side of (26) equals

$$\left(\frac{c_n}{n}\right)^2 \sum_{t=1}^{[n\eta]} \mathbf{E}\mathbf{E}_{t-2}\left\{\int_{-\infty}^{\infty} f\left(\sqrt{n}\left(x_{t-1,n} + z\epsilon\right)\right) g\left(c_n\left(x_{t,n} - \frac{x}{\sqrt{n}} + z\epsilon\right)\right) \phi(z)\, dz\right\}^2$$

$$= \left(\frac{c_n}{n}\right)^2 \sum_{t=1}^{[n\eta]} \mathbf{E}\int_v\left\{\int_l f\left(\sqrt{n}l + x - v\right) g\left(c_n l\right) \phi_\epsilon\left(l - x_{t-1,n} - \frac{v}{\sqrt{n}} + \frac{x}{\sqrt{n}}\right) dl\right\}^2 p(v)\, dv$$

$$\leq \phi_\epsilon^2(0) \frac{c_n^2}{n} \int_v\left\{\int_l f\left(\sqrt{n}l + x - v\right) g\left(c_n l\right) dl\right\}^2 p(v)\, dv$$

$$\leq \frac{Ac_n}{n} \int_v\left\{\int_m\left|f\left(\frac{\sqrt{n}}{c_n}m + x - v\right)\right| |g(m)|\, dm\right\}^2 p(v)\, dv$$

$$\leq \frac{Ac_n}{n} \int_v\left\{\int_m |f_0(m, x, v)|\, |g(m)|\, dm\right\}^2 p(v)\, dv \to 0,$$

where the last inequality holds for n large enough. The second term on the R.H.S of (26) equals

$$\left(\frac{c_n}{n}\right)^2 \mathbf{E}\sum_{t=1}^{[n\eta]}\left\{\mathbf{E}_{t-2}\int_{-\infty}^{\infty} f\left[\sqrt{n}\left(x_{t-1,n} + z\epsilon\right)\right] g\left(c_n\left(x_{t,n} - \frac{x}{\sqrt{n}} + z\epsilon\right)\right) \phi(z)\, dz\right\}^2$$

$$= \left(\frac{c_n}{n}\right)^2 \sum_{t=1}^{[n\eta]} \mathbf{E}\left\{\int_v\int_z f\left[\sqrt{n}\left(x_{t-1,n} + z\epsilon\right)\right] g\left[c_n\left(x_{t-1,n} + \frac{v}{\sqrt{n}} - \frac{x}{\sqrt{n}} + z\epsilon\right)\right] \phi(z) p(v)\, dzdv\right\}^2$$

$$\leq \frac{\phi_\epsilon^2(0)}{n}\left\{\int_v\int_l\left|f\left(\frac{\sqrt{n}}{c_n}l + x - v\right) g(l)\right| p(v)\, dvdl\right\}^2 \leq \frac{A}{n}\left\{\int_v\int_l f_0(l, x, v)\, |g(l)|\, p(v)\, dvdl\right\}^2 \to 0,$$

as required. ∎

Lemma 2. *Suppose that:*
  (a) *Assumption 2.1 holds.*
  (b) *Assumption 2.3 or Assumption 2.3* holds.*
Set

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

Then

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

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

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

$$\frac{c_n}{n}\sum_{t=1}^{[n\eta]}\mathbf{E}_{t-2}\int_{-\infty}^{\infty} f\left(\sqrt{n}\left\{x_{t-1,n} + \epsilon z\right\}\right) g\left(c_n\left\{x_{t-1,n} + \frac{v_t}{\sqrt{n}} - \frac{x}{\sqrt{n}} + \epsilon z\right\}\right)\phi(z)dz$$

$$= \frac{c_n}{n}\sum_{t=1}^{[nr]}\int_v\int_z f\left(\sqrt{n}\left\{x_{t-1,n} + \epsilon z\right\}\right) g\left(c_n\left\{x_{t-1,n} + \frac{v}{\sqrt{n}} - \frac{x}{\sqrt{n}} + \epsilon z\right\}\right)\phi_\epsilon(z)p(v)dzdv$$

$$= \frac{1}{n}\sum_{t=1}^{[n\eta]}\int_v\int_z f\left(\sqrt{n}\left\{\frac{z}{c_n} + \frac{x}{\sqrt{n}} - \frac{v}{\sqrt{n}}\right\}\right) g(z)\,\phi_\epsilon\left(\frac{z}{c_n} - x_{t-1,n} + \frac{x}{\sqrt{n}} - \frac{v}{\sqrt{n}}\right)p(v)dzdu$$

$$\vdots \quad = T_n(\eta)$$

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

$$\left|f\left(\frac{\sqrt{n}}{c_n}z + x - v\right)\phi_\epsilon\left(\frac{z}{c_n} - x_{t-1,n} + \frac{v}{\sqrt{n}} - \frac{x}{\sqrt{n}}\right) - f(x-v)\phi_\epsilon(x_{t-1,n})\right|$$

$$\le |\phi_\epsilon(0)|\left|f\left(\frac{\sqrt{n}}{c_n}z + x - v\right) - f(x-v)\right| + |f(x-v)|\left|\phi_\epsilon\left(\frac{z}{c_n} - x_{t-1,n} + \frac{v}{\sqrt{n}} - \frac{x}{\sqrt{n}}\right) - \phi_\epsilon(x_{t-1,n})\right|$$

$$\le |\phi_\epsilon(0)|\left(\frac{\sqrt{n}}{c_n}\right)^\gamma f_0(z,v,x) + |f(x-v)|\,C\left|\frac{z}{c_n} + \frac{v}{\sqrt{n}} - \frac{x}{\sqrt{n}}\right|,$$

where C is a Lipschitz constant. Therefore,

$$\left|T_n(\eta) - \frac{1}{n}\sum_{t=1}^{[n\eta]}\int_v\int_z f(x-v)\,\phi_\epsilon(x_{t-1,n})\,g(z)p(v)dzdv\right| \le \left(\frac{\sqrt{n}}{c_n}\right)^\gamma|\phi(0)|\int_u\int_z f_0(z,v,x)\,|g(z)|\,p(v)dzdv$$

21

$$+C \int_v \int_z |f(x-v)| p(v) \left| \frac{z}{c_n} + \frac{v}{\sqrt{n}} - \frac{x}{\sqrt{n}} \right| |g(z)| \, dz dv \to 0$$

as required.

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

$$\left| f\left(\frac{\sqrt{n}}{c_n} z + x - v \right) \phi\left(\frac{z}{c_n} - x_{t-1,n} + \frac{v}{\sqrt{n}} - \frac{x}{\sqrt{n}} \right) - f(m_o z + x - v) \phi_\epsilon(x_{t-1,n}) \right|$$

$$\leq |\phi_\epsilon(0)| \left| \frac{\sqrt{n}}{c_n} - m_o \right|^\gamma f_0(z,v,x) + |f(m_o z + x - v)| C \left| \frac{z}{c_n} + \frac{v}{\sqrt{n}} - \frac{x}{\sqrt{n}} \right| \to 0,$$

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 2.1 holds.*
(b) $\left| f\left(\frac{\sqrt{n}}{c_n} z + x - v \right) \right| \leq f_0(z,x,v)$ *for n large enough with* $\int_v \int_z f_0(z,x,v) |g(z)| p_{r-s}(v) dz dv < \infty$, *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

$$\lim_{n \to \infty} \sup_{0 \leq \eta \leq 1} \frac{c_n}{n} \mathbf{E} \left| \underbrace{\sum_{t=1}^{[n\eta]} f\left(\sqrt{n} x_{t-r,n} \right) \left\{ \mathbf{E}_{t-r-1} g\left[c_n \left(x_{t-s,n} - \frac{x}{\sqrt{n}} \right) \right] \right\}}_{M_n(\eta)} \right|^q = 0.$$

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

$$\mathbf{E}|M_n(\eta)| = \frac{c_n}{n} \int_s \left| \sum_{t=1}^{[n\eta]} \int_{l_1} \cdots \int_{l_q} f\left(\sqrt{n} d_{t-1,0,n} l \right) g\left[c_n \left(d_{t-1,0,n} l + \frac{l_1}{\sqrt{n}} \right) \right] \cdots g\left[c_n \left(d_{t-1,0,n} l + \frac{l_q}{\sqrt{n}} \right) \right] \right.$$

$$\times p(l_1)p(l_q) dl_1 ... dl_q \Big| h_{t-1,0,n}(l) \, dl$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} \frac{1}{d_{t-1,0,n}} \int_m \int_{l_1} \cdots \int_{l_q} \left| f\left(\frac{\sqrt{n}}{c_n} m - l_1 \right) \right| |g(m)| \left| \prod_{i=2}^{q} g\left(m + \frac{c_n}{\sqrt{n}} (l_i - l_1) \right) \right|$$

$$\times p(l_1)...p(l_q) dl_1 ... dl_q dm$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} \frac{1}{d_{t-1,0,n}} \int_m \int_{l_1} \cdots \int_{l_q} f_0(m, l_1) |g(m)| \left| \prod_{i=2}^{q} g\left(m + \frac{c_n}{\sqrt{n}} (l_i - l_1) \right) \right| p(l_1)...p(l_q) dl_1 ... dl_q dm$$

$$\leq A \int_m \int_{l_1} \cdots \int_{l_q} f_0(m, l_1) |g(m)| \left| \prod_{i=2}^{q} g\left(m + \frac{c_n}{\sqrt{n}} (l_i - l_1) \right) \right| p(l_1)...p(l_q) dl_1 ... dl_q dm \to 0,$$

as $n \to \infty$. Note that the second inequality above holds for n large enough. Further, the limit above holds by dominated convergence since $g\left(m + \frac{c_n}{\sqrt{n}} (l_{i+1} - l_1) \right) \to 0$ almost everywhere with respect to the Lebesgue measure, $\int_{l_1} \int_m f_0(m, l_1) |g(m)| p(l_1) dl_1 dm < \infty$, and $\sup_s |g(s)| < \infty$. ∎

Lemma 4. *Suppose that*

    (a) *Assumption 2.2 holds.*

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

Set

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

Then

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

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

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

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*

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

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

Proof of Lemma 5: Set

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

and

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

It can be shown along the lines of Lemma 2 that

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

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

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \sup_{0 \leq \eta \leq 1} \mathbf{E} \left| L_n(\eta) - L_{n,\epsilon}(\eta) \right| = 0.$$

∎

7   Appendix B: Proofs of the Main Results

Proof of Theorem 1. Set,

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

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

Then

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \sup_{0 \leq \eta \leq 1} \mathbf{E} \left| L_n(\eta) - L_{n,\epsilon}(\eta) \right| = 0, \tag{27}$$

and the stated results follow as in WP. We proceed with the proof of (27).

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

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

Next, we have

$$c_n \mathbf{E} \left| Y_{k,n}(z) \right|$$

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

$$\leq \int_s \int_v \left| f\left(d_n d_{k-1,0,n} s - v + x\right) g\left(c_n d_{k-1,0,n} s\right) - \right.$$

$$\left. f\left(\sqrt{n}\left(d_{k-1,0,n} s + \epsilon z\right) - v + x\right) g\left[c_n\left(d_{k-1,0,n} s + \epsilon z\right)\right] \right| p(v) \, dv ds$$

$$\leq \frac{2A}{d_{k-1,0,n}} \int_s \int_v \left| f\left(\frac{d_n}{c_n} s - v + x\right) g(s) \right| p(v) \, dv ds$$

$$\leq \frac{2A}{d_{k-1,0,n}} \int_s \int_v f_0(s, v, x) \left| g(s) \right| p(v) dv ds,$$

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

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

24

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

$$\Lambda_n(\epsilon) \le \left(\frac{c_n}{n}\right)^2 \mathbf{E}\sum_{k=1}^n Y_{k,n}^2(z) + \frac{2c_n^2}{n^2}\sum_{k=1}^n |\mathbf{E}Y_{k,n}(z)Y_{k+1,n}(z)| + \frac{2c_n^2}{n^2}\sum_{k=1}^n \sum_{l=k+2}^n |\mathbf{E}Y_{k,n}(z)Y_{l,n}(z)|$$

$$: \quad = \Lambda_{1n}(\epsilon) + \Lambda_{2n}(\epsilon) + \Lambda_{3n}(\epsilon).$$

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

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

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

$$c_n d_{l-1,k,n}|\mathbf{E}_{k-1}Y_{l,n}(z)|$$

$$= \left|\mathbf{E}_{k-1}\left\{f(d_n x_{l-1,n})g\left[c_n\left(x_{l,n} - \frac{x}{d_n}\right)\right] - f[\sqrt{n}(x_{l-1,n} + \epsilon z)]g\left[c_n\left(x_{l,n} - \frac{x}{d_n} + \epsilon z\right)\right]\right\}\right|$$

$$= \left|\mathbf{E}_{k-1}\int_v \left\{f(d_n x_{l-1,n})g\left[c_n\left(x_{l-1,n} + \frac{v}{d_n} - \frac{x}{d_n}\right)\right]\right.\right.$$
$$\left.\left. - f[d_n(x_{l-1,n} + \epsilon z)]g\left[c_n\left(x_{l-1,n} + \frac{v}{d_n} - \frac{x}{d_n} + \epsilon z\right)\right]\right\}p(v)dv\right|$$

$$= \left|\mathbf{E}_{k-1}\int_v \left\{f[d_n(x_{k,n} + (x_{l-1,n} - x_{k,n}))]g\left[c_n\left(x_{k,n} + (x_{l-1,n} - x_{k,n}) + \frac{v}{d_n} - \frac{x}{d_n}\right)\right]\right.\right.$$
$$\left.\left. - f[d_n(x_{k,n} + (x_{l-1,n} - x_{k,n}) + \epsilon z)]g\left[c_n\left(x_{k,n} + (x_{l-1,n} - x_{k,n}) + \frac{v}{d_n} - \frac{x}{d_n} + \epsilon z\right)\right]\right\}p(v)dv\right|$$

$$= \left|\int_s \int_v \left\{f[d_n(x_{k,n} + d_{l,k-1,n}s)]g\left[c_n\left(x_{k,n} + d_{l,k-1,n}s - \frac{x}{d_n}\right)\right]\right.\right.$$
$$\left.\left. - f[d_n(x_{k,n} + d_{l,k-1,n}s + \epsilon z)]g\left[c_n\left(x_{k,n} + d_{l,k-1,n}s - \frac{x}{d_n} + \epsilon z\right)\right]\right\}p(v)h_{l-1,k,n}(s)ds\right|$$

$$\le \int_y \int_v \left|f\left(\frac{d_n}{c_n}y - v\right)\right||g(y)|\,|V(y,c_n x_{k,n},v)|\,p(v)dvdy$$

$$\le \begin{cases} A, \text{ for } (l-1,k) \notin \Omega_n, \\ A\int_{|y|\ge \sqrt{c_n}}\int_v f_o(y,v,x)|g(y)|\,p(v)dvdy \\ \quad + \int_{|y|<\sqrt{c_n}}\int_v f_0(y,v,x)|g(y)|\,|V(y,c_n x_{k,n},v)|\,p(v)dvdy, \text{ for } (l-1,k) \in \Omega_n \end{cases}$$

25

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

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

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 $(l-1, k) \notin \Omega_n$

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

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

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

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

$$\Lambda_{3n}(\epsilon) \leq \frac{2c_n^2}{n^2}\left[\sum_{l-1>k,(l-1,k)\notin\Omega_n}^{n} + \sum_{(l-1,k)\in\Omega_n}^{n}\right]|\mathbf{E}Y_{k,n}(z)Y_{l,n}(z)|$$

$$\leq \frac{A_1}{n^2}\sum_{k=(1-\eta)n}^{n}(d_{k,0,n})^{-1}\max_{1\leq k\leq n-2}\sum_{l=k+2}^{n}(d_{l-1,k,n})^{-1}$$

$$+\frac{A_2}{n^2}\sum_{k=\eta n}^{(1-\eta)n}(d_{k,0,n})^{-1}\max_{0\leq k\leq(1-\eta)n}\sum_{l=k+2}^{k+\eta n}(d_{l-1,k,n})^{-1}$$

$$+\frac{A_3}{n^2}\sum_{k=1}^{\eta n}(d_{k,0,n})^{-1}\max_{0\leq k\leq n-2}\sum_{l=k+2}^{n}(d_{l-1,k,n})^{-1}$$

$$+\frac{A_4}{n^2}\sum_{k=1}^{n}(d_{k,0,n})^{-1}\max_{0\leq k\leq n-2}\sum_{l=k+2}^{n}(d_{l-1,k,n})^{-1}$$

$$\times\left\{\int_{|y|\geq\sqrt{c_n}}\int_{v}f_0(y,v,x)|g(y)|p(v)dvdy\right.$$

$$+\int_{y}\int_{v}\sup_{|r|\leq 2C[1+|z|+|v|]\epsilon^{1/2}}|h_{l-1,k,n}(r)-h_{l-1,k,n}(0)||g(y)|f_0(y,v,x)p(v)dvdy\Bigg\}$$

$$\rightarrow 0,$$

and the result follows. ∎

Proof of Theorem 2. Equation (9) of Theorem 2 follows easily from Theorem 1. We shall prove (10). 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 $\tau:=\mathbf{E}f(x-v_t)$ and $K_h(.)=K(./h)$. We have

$$\sqrt{\sqrt{nh}}\left[\hat{f}(x)-\mathbf{E}f(x-v_t)\right]=\frac{\overbrace{\frac{1}{\sqrt{\sqrt{nh}}}\sum\mathbf{E}_{t-2}\{[f(x_{t-1})-\mathbf{E}f(x-v_t)]K_h(x_t-x)\}}^{:=R_n}}{\frac{1}{\sqrt{nh}}\sum K_h(x_t-x)}$$

$$+\frac{\frac{1}{\sqrt{\sqrt{nh}}}\sum\overbrace{f(x_{t-1})K_h(x_t-x)-\mathbf{E}_{t-2}f(x_{t-1})K_h(x_t-x)}^{:=\alpha_t}}{\frac{1}{\sqrt{nh}}\sum K_h(x_t-x)}+\frac{\frac{1}{\sqrt{\sqrt{nh}}}\sum\overbrace{\tau\mathbf{E}_{t-2}K_h(x_t-x)-\tau K_h(x_t-x)}^{:=\beta_t}}{\frac{1}{\sqrt{nh}}\sum K_h(x_t-x)}$$

$$+\frac{\frac{1}{\sqrt{\sqrt{nh}}}\sum\overbrace{K_h(x_t-x)u_t}^{:=\gamma_t}}{\frac{1}{\sqrt{nh}}\sum K_h(x_t-x)}:=\frac{R_n+M_n}{\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}K_h(x_t-x)}.$$

Notice that

$$\mathbf{E}|R_n|=\mathbf{E}\left|\left(\frac{1}{\sqrt{nh}}\right)^{1/2}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left[(f(x_{t-1})-\mathbf{E}f(x-v_t))K\left(\frac{x_t}{h}-\frac{x}{h}\right)\right]\right|$$

$$= \mathbf{E} \left| \left(\frac{1}{\sqrt{nh}} \right)^{1/2} \sum_{t=1}^{n} \int_{u} [f(x_{t-1}) - \mathbf{E}f(x - v_t)] K \left(\frac{\sqrt{n}x_{t-1,n}}{h} + \frac{v}{h} - \frac{x}{h} \right) p_1(v) dv \right|$$

$$\leq \left(\frac{1}{\sqrt{nh}} \right)^{1/2} \sum_{t=1}^{n} \int_{y} \left| \int_{u} [f(\sqrt{n}d_{t-1,0,n}y) - \mathbf{E}f(x - v_t)] K \left(\frac{\sqrt{n}d_{t-1,0,n}y}{h} + \frac{v}{h} - \frac{x}{h} \right) p_1(v) dv \right| h_{t-1,0,n}(y) dy$$

$$= \left(\frac{1}{\sqrt{nh}} \right)^{1/2} \frac{h}{\sqrt{n}} \sum_{t=1}^{n} (d_{t-1,0,n})^{-1} \int_{z} \left| [\mathbf{E}f(hz + x - v_t) - \mathbf{E}f(x - v_t)] K(z) \right| h_{t-1,0,n} \left(\frac{hz + x - v_t}{\sqrt{n}d_{t-1,0,n}} \right) dz$$

$$\leq \left(\sqrt{n}h^{1+2\gamma} \right)^{1/2} \frac{1}{n} \sum_{t=1}^{n} (d_{t-1,0,n})^{-1} \int_{z} f_1(z, x) K(z) dz \to 0.$$

Hence,

$$\sqrt{\sqrt{nh}} \left[\hat{f}(x) - \mathbf{E}f(x - v_t) \right] = \frac{M_n}{\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} K_h(x_t - x)} + o_p(1). \tag{28}$$

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

$$T_{1,n} := \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} \mathbf{E}_{t-2} (\alpha_t + \beta_t + \gamma_t)^2.$$

First, we shall show that

$$T_{1,n} \xrightarrow{p} \left(\mathbf{Var}f(x - v_t) + \sigma_u^2 \right) L_G(1, 0) \int_{-\infty}^{\infty} K^2(s) ds. \tag{29}$$

We have

$$T_{1,n} = \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} \mathbf{E}_{t-2} \{ [f(x_{t-1})K_h(x_t - x) - \mathbf{E}_{t-2}f(x_{t-1})K_h(x_t - x)]$$

$$+ \tau [\mathbf{E}_{t-2}K_h(x_t - x) - K_h(x_t - x)] + K_h(x_t - x)u_t \}^2$$

$$= \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} \mathbf{E}_{t-2} \left[f^2(x_{t-1})K_h^2(x_t - x) \right] - \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} [\mathbf{E}_{t-2}f(x_{t-1})K_h(x_t - x)]^2$$

$$+ \frac{1}{\sqrt{nh}} \tau^2 \sum_{t=1}^{n} \mathbf{E}_{t-2} K_h^2(x_t - x) - \frac{1}{\sqrt{nh}} \tau^2 \sum_{t=1}^{n} [\mathbf{E}_{t-2}K_h(x_t - x)]^2 + \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} \mathbf{E}_{t-2}K_h^2(x_t - x)u_t^2$$

$$- \frac{1}{\sqrt{nh}} 2\tau \sum_{t=1}^{n} f(x_{t-1})\mathbf{E}_{t-2}K_h^2(x_t - x) - \frac{1}{\sqrt{nh}} 2\tau \sum_{t=1}^{n} f(x_{t-1}) [\mathbf{E}_{t-2}K_h(x_t - x)]^2$$

$$= \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} \mathbf{E}_{t-2} \left[f^2(x_{t-1})K_h^2(x_t - x) \right] + \frac{1}{\sqrt{nh}} \tau^2 \sum_{t=1}^{n} \mathbf{E}_{t-2}K_h^2(x_t - x)$$

$$+ \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} K_h^2(x_t - x)\sigma_{t,u}^2 - \frac{1}{\sqrt{nh}} 2\tau \sum_{t=1}^{n} f(x_{t-1})\mathbf{E}_{t-2}K_h^2(x_t - x) + o_p(1) \text{ (by Lemma 3)}$$

$$: \quad = T_{2,n} + o_p(1).$$

Next, note that

$$\frac{\mathbf{E}}{\sqrt{nh}}\sum_{t=1}^{n}K_h^2(x_t-x)\left|\sigma_{t,u}^2-\sigma_u^2\right|\leq\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}\left\{\mathbf{E}K_h^{2q}(x_t-x)\right\}^{1/q}\left\{\mathbf{E}\left|\sigma_{t,u}^2-\sigma_u^2\right|^p\right\}^{1/p}\to0,$$
(30)

as $n\to\infty$ for some $p,q>1$ and $1/p+1/q=1$. Equation (30) holds by the Toeplitz lemma. To see this, notice that by Assumption 3.3 $\left|\sigma_{t,u}^2-\sigma_u^2\right|=o_{a.s.}(1)$. Hence, $\mathbf{E}\left|\sigma_{t,u}^2-\sigma_u^2\right|^p\to0$ by dominated convergence, for $\mathbf{E}\left|\sigma_{t,u}^2-\sigma_u^2\right|^p\leq2^{p-1}\left(\sup_t\mathbf{E}\sigma_{t,u}^{2p}+\sigma_u^{2p}\right)<\infty$, due to Assumption 3.4. Moreover, using arguments similar to those used in the proof of Lemma 3, we get $\mathbf{E}K_h^{2q}(x_t-x)\to0$ as $n\to\infty$ and $\sup_n(\sqrt{nh})^{-1}\sum_{t=1}^{n}\left\{\mathbf{E}K_h^{2q}(x_t-x)\right\}^{1/q}<\infty$. Therefore, (30) holds (e.g. Hall and Heyde, 1980 p.31).

Hence, by (30), Lemma 1 and Theorem 1, we get

$$T_{2,n}\overset{p}{\to}\left(\mathbf{E}f^2(x-v_t)+\tau^2+\sigma_u^2-2\tau^2\right)L_G(1,0)\int_{-\infty}^{\infty}K^2(s)ds$$

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

Next, fix $\delta>0$ and $\zeta>0$ and consider

$$T_{3,n}:=\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left(\alpha_t+\beta_t+\gamma_t\right)^2\mathbf{1}\left\{\left(\frac{1}{\sqrt{nh}}\right)^{1/2}\left|\alpha_t+\beta_t+\gamma_t\right|>\delta\right\}.$$

We shall show that

$$T_{3,n}=o_p(1).$$
(31)

We have

$$T_{3,n}\leq\frac{A}{\sqrt{nh}}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left(\alpha_t^2+\beta_t^2+\gamma_t^2\right)\mathbf{1}\left\{\left(\frac{1}{\sqrt{nh}}\right)^{1/2}\left|\alpha_t+\beta_t+\gamma_t\right|>\delta\right\}$$

$$\leq\frac{A}{\sqrt{nh}}\sum_{t=1}^{n}\left\{\mathbf{E}_{t-2}\left(\alpha_t^2+\beta_t^2+\gamma_t^2\right)^{1+\zeta}\right\}^{1/(1+\zeta)}\left\{\mathbf{P}_{t-2}\left[\left(\frac{1}{\sqrt{nh}}\right)^{1/2}\left|\alpha_t+\beta_t+\gamma_t\right|>\delta\right]\right\}^{\zeta/(1+\zeta)}$$

$$\leq\frac{A}{\sqrt{nh}}\sum_{t=1}^{n}\left\{\mathbf{E}_{t-2}\left(\alpha_t^2+\beta_t^2+\gamma_t^2\right)^{1+\zeta}\right\}^{1/(1+\zeta)}\left\{\left(\frac{\delta^{-2}}{\sqrt{nh}}\right)^{1+\zeta}\mathbf{E}_{t-2}\left[\left(\alpha_t+\beta_t+\gamma_t\right)^{2(1+\zeta)}\right]\right\}^{\zeta/(1+\zeta)}$$

$$\leq\frac{A}{\sqrt{nh}}\sum_{t=1}^{n}\left\{\mathbf{E}_{t-2}\left(\alpha_t^{2(1+\zeta)}+\beta_t^{2(1+\zeta)}+\gamma_t^{2(1+\zeta)}\right)\right\}^{1/(1+\zeta)}\left\{\left(\frac{1}{\sqrt{nh}}\right)^{1+\zeta}\mathbf{E}_{t-2}\left[\left(\alpha_t+\beta_t+\gamma_t\right)^{2(1+\zeta)}\right]\right\}^{\zeta/(1+\zeta)}$$

$$\leq A\left(\frac{1}{\sqrt{nh}}\right)^{1+\zeta}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left(\alpha_t^{2(1+\zeta)}+\beta_t^{2(1+\zeta)}+\gamma_t^{2(1+\zeta)}\right).$$

Next, we shall show that $\left(\frac{1}{\sqrt{nh}}\right)^{1+\zeta}\sum_{t=1}^{n}\mathbf{E}_{t-2}\alpha_t^{2(1+\zeta)}=o_p(1)$. Notice that

$$\left(\frac{1}{\sqrt{nh}}\right)^{1+\zeta}\sum_{t=1}^{n}\mathbf{E}_{t-2}\alpha_t^{2(1+\zeta)}$$

$$\leq A\left(\frac{1}{\sqrt{nh}}\right)^{1+\zeta}\sum_{t=1}^{n}\left[\mathbf{E}_{t-2}f^{2(1+\zeta)}(x_{t-1})K_h^{2(1+\zeta)}(x_t-x)+f^{2(1+\zeta)}(x_{t-1})\left\{\mathbf{E}_{t-2}K_h(x_t-x)\right\}^{2(1+\zeta)}\right].$$

By Lemma 1 and Theorem 1,

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

and by Lemma 3(i),

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

In view of Assumption 3.4 and using similar arguments to those used above, we also have $\left(\frac{1}{\sqrt{nh}}\right)^{1+\zeta}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left(\beta_t^{2(1+\zeta)}+\gamma_t^{2(1+\zeta)}\right)=o_p(1)$. Therefore, (31) holds. Finally, in view of Hall and Heyde (1980, Theorem 3.2), (29) and (31) give

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

where W is a standard normal variate independent of $L_G(1,0)$.

Next, the quadratic variation of M_n, is $[M_n]:=\left(\sqrt{nh}\right)^{-1}\sum_{t=1}^{n}\left(\alpha_t+\beta_t+\gamma_t\right)^2$. The following condition (see Jacod and Shiryaev, 1986)

$$\sup_n\left(\sqrt{nh}\right)^{-1/2}\max_{0\le t\le n}\mathbf{E}\left|\alpha_t+\beta_t+\gamma_t\right|<\infty \tag{32}$$

is sufficient for

$$\left([M_n],M_n\right)\overset{d}{\to}\left([M],M\right). \tag{33}$$

We shall demonstrate that (32) holds. For any $\gamma>2$ we have

$$\left(\sqrt{nh}\right)^{-1/2}\max_{0\le t\le n}\mathbf{E}\left|\alpha_t+\beta_t+\gamma_t\right|$$

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

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

Consider the first summand. We have

$$\left(\sqrt{nh}\right)^{-\gamma/2}\mathbf{E}\sum_{t=1}^{n}|\alpha_t|^\gamma=\left(\sqrt{nh}\right)^{-\gamma/2}\mathbf{E}\sum_{t=1}^{n}\left|f(x_{t-1})K_h(x_t-x)-\mathbf{E}_{t-2}f(x_{t-1})K_h(x_t-x)\right|^\gamma$$

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

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

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

Next, consider the predictable quadratic variation of M_n, $\langle M_n \rangle := \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} \mathbf{E}_{t-2} \left(\alpha_t + \beta_t + \gamma_t \right)^2$. We shall show that

$$\lim_{n \to \infty} \mathbf{E} \left| [M_n] - \langle M_n \rangle \right| = 0. \tag{34}$$

In view of (33), (34) implies

$$\left(\langle M_n \rangle , M_n \right) \xrightarrow{d} \left([M] , M \right). \tag{35}$$

According to Hall and Heyde (1980, Theorem 2.23), (31) and tightness of $\langle M_n \rangle$ are sufficient for (34). Let $\lambda > 0$ and notice that

$$\lim_{\lambda \to \infty} \sup_n \mathbf{P} \left(\langle M_n \rangle > \lambda \right) \leq \lim_{\lambda \to \infty} \frac{1}{\lambda} \sup_n \mathbf{E} \left(\langle M_n \rangle \right) = 0,$$

for $\sup_n \mathbf{E} \left(\langle M_n \rangle \right) < \infty$ due to Lemma 3(ii) and Lemma 4. Therefore, the sequence $\langle M_n \rangle$ is tight.

Finally, it follows from (28) that the NW estimator

$$\left(\sum_{t=1}^{n} K_h(x_t - x) \right)^{1/2} \left[\hat{f}(x) - \mathbf{E} f(x + v_t) \right] = \frac{M_n}{\left((\sqrt{nh})^{-1} \sum_{t=1}^{n} K_h(x_t - x) \right)^{1/2}} + o_p(1)$$

$$= \frac{\langle M_n \rangle^{1/2}}{\left((\sqrt{nh})^{-1} \sum_{t=1}^{n} K_h(x_t - x) \right)^{1/2}} \frac{M_n}{\langle M_n \rangle^{1/2}} := A_n B_n.$$

Now by Theorem 1, it can be easily seen that

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

In addition, (35) implies that $B_n \xrightarrow{d} W$, and the result for $r > s$ follows. The proof for $r < s$ follows from Lemma 5 and arguments similar to those used in the previous part. ■

Proof of Theorem 3.

Write

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

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

It follows directly from Theorem 1 and Theorem 2 that

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

31

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

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

This shows the first part of Thorem 3.

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

Proof of Theorem 4. (i) For n large enough $s_n > r$ and

$$\left| \kappa\left(\sqrt{n}\right)^{-1} \hat{f}(x) - \frac{\sum_{t=s_n}^{n} K_h\left(x_{t-s_n} - x\right) H_f\left(x_{t-r,n}\right)}{\sum_{t=s_n}^{n} K_h\left(x_{t-s_n} - x\right)} \right|$$

$$\leq \frac{\left| \sum_{t=s_n}^{n} K_h\left(x_{t-s_n} - x\right) u_t \right| + \sup_{z \in \mathbb{R}} |\nu\left(z,n\right)| \sum_{t=s_n}^{n} K_h\left(x_{t-s_n} - x\right)}{\kappa\left(\sqrt{n}\right) \sum_{t=s_n}^{n} K_h\left(x_{t-s_n} - x\right)} = o_p(1).$$

Further, by Theorem 1 of Phillips (2009) we have the following limit result

$$\frac{\sum_{t=s_n}^{n} K_h\left(x_{t-s_n} - x\right) H_f\left(x_{t-r,n}\right)}{\sum_{t=s_n}^{n} K_h\left(x_{t-s_n} - x\right)} \xrightarrow{d} \frac{1}{L_{\tilde{G}}(1,0)} \int_0^1 H_f\left(G(\eta)\right) dL_{\tilde{G}}(\eta,0)$$

$$= \frac{1}{L_G(1-c,0)} \int_c^1 H_f\left(G(\eta)\right) dL_G(\eta - c,0),$$

where the last equality follows from the fact that $L_{\tilde{G}}(\eta,0) = L_G(\eta - c,0)\mathbf{1}\left(1 \geq \eta \geq c\right)$.

(ii) Note that the term

$$\mathbf{E}\left| \left(\frac{1}{\sqrt{n}h}\right)^{1/2} \sum_{t=s_n}^{n} \left[f\left(x_{t-r}\right) K_h\left(x_{t-s_n} - x\right)\right] \right|$$

$$\leq \mathbf{E}\left(\frac{1}{\sqrt{n}h}\right)^{1/2} \sum_{t=s_n}^{n} \mathbf{E}_{t-s_n} \left|f\left(x_{t-s_n} + x_{t-r} - x_{t-s_n}\right) K_h\left(x_{t-s_n} - x\right)\right|$$

$$\leq \mathbf{E}\left(\frac{1}{\sqrt{n}h}\right)^{1/2} \sum_{t=s_n}^{n} \mathbf{E}_{t-s_n} \left|f\left[\sqrt{n}d_{t-s_n,0,n}x_{t-s_n,n} + \sqrt{n}d_{t-r,t-s_n,n}\left(x_{t-r,n} - x_{t-s_n}\right)\right] K_h\left(\sqrt{n}d_{t-s_n,0,n}x_{t-s_n} - x\right)\right|$$

$$= \left(\frac{1}{\sqrt{n}h}\right)^{1/2} \sum_{t=s_n}^{n} \int_{l_1} \int_{l_2} \left|f\left[\sqrt{n}d_{t-s_n,0,n}l_1 + \sqrt{n}d_{t-r,t-s_n,n}l_2\right]\right.$$

$$\times K_h\left(\sqrt{n}d_{t-s_n,0,n}l_1 - x\right)\Big| h_{t-r,t-s_n,n}\left(l_2\right) h_{t-s_n,0,n}\left(l_1\right) dl_2 dl_1$$

$$\leq A\left(\sqrt{n}h\right)^{-1/2} \int_{p_1} \int_{p_2} f\left(h\left[p_1 + x\right] + p_2\right) K\left(p_1\right) dp_2 dp_1 \frac{h}{n} \sum_{t=s_n}^{n} \left(d_{t-s_n,0,n}d_{t-r,t-s_n,n}\right)^{-1}$$

$$\leq A\left(h/\sqrt{n}\right)^{1/2} \int_{p_1} \int_{p_1} f_0(p_1,p_2)K(p_1)dp_1dp_2 = o(1).$$

The last inequality above follows from the fact that $n^{-1}\sum_{t=1}^{n}\left(d_{t-s_n,0,n}d_{t-r,t-s_n,n}\right)^{-1}$ is bounded. In vew of this, the result follows from arguments similar to those used in the proof of Theorem 2. ∎

Proof of Example 5. Consider the local linear estimator

$$\hat{b}(x) = \begin{bmatrix} \widehat{\beta}(x) \\ \hat{\beta}^{(1)}(x) \end{bmatrix} = \arg\min_{\theta_0,\,\theta_1} \sum_{t=1}^{n} [y_t - \theta_0 x_t - \theta_1 x_t(x_t - x)]^2 K_h(x_t - x)$$

Further, assume that β has a continuous third dericative, the kernel function $K(.)$ compact support and $nh^{10} = o(1)$. Then

$$\hat{b}(x) =$$

$$\overbrace{\left\{ \sum_{t=1}^{n} \begin{bmatrix} x_t^2 & x_t^2(x_t - x) \\ x_t^2(x_t - x) & x_t^2(x_t - x)^2 \end{bmatrix} K_h(x_t - x) \right\}}^{H_n}{}^{-1} \sum_{t=1}^{n} y_t \begin{bmatrix} x_t \\ x_t(x_t - x) \end{bmatrix} K_h(x_t - x)$$

$$= \overbrace{\begin{bmatrix} \beta(x) \\ \beta^{(1)}(x) \end{bmatrix}}^{b(x)} + H_n^{-1} \left\{ \sum_{t=1}^{n} y_t \begin{bmatrix} x_t \\ x_t(x_t - x) \end{bmatrix} K_h(x_t - x) - H_n b(x) \right\}$$

$$= b(x) + H_n^{-1} \left\{ \sum_{t=1}^{n} (\beta(x_t) x_t + u_t) \begin{bmatrix} x_t \\ x_t(x_t - x) \end{bmatrix} K_h(x_t - x) - H_n b(x) \right\}$$

Using arguments similar to those of Wang and Phillips (2009c) we get

$$|R_n| := n^{-1/2} \left| \sum_{t=1}^{n} \beta(x_t) x_t \begin{bmatrix} x_t \\ x_t(x_t - x) \end{bmatrix} K_h(x_t - x) - H_n b(x) \right| =$$

$$h^{-1} n^{-1/2} \left| \sum_{t=1}^{n} K_h(x_t - x) \begin{bmatrix} x_t \\ x_t(x_t - x) \end{bmatrix} \left\{ \beta(x_t) x_t - \begin{bmatrix} x_t \\ x_t(x_t - x) \end{bmatrix}' b(x) \right\} \right| =$$

$$h^{-1} n^{-1/2} \left| \sum_{t=1}^{n} K_h(x_t - x) \begin{bmatrix} x_t \\ x_t(x_t - x) \end{bmatrix} x_t \left\{ \beta(x_t) - \beta(x) - \beta^{(1)}(x)(x_t - x) \right\} \right| =$$

$$h^{-1} n^{-1/2} \left| \sum_{t=1}^{n} K_h(x_t - x) \begin{bmatrix} x_t \\ x_t(x_t - x) \end{bmatrix} x_t \beta^{(2)}(\bar{x}_t)(x_t - x)^2 \right| \leq$$

$$A h^{-1} n^{-1/2} D_n D_n^{-1} \left| \sum_{t=1}^{n} K_h(x_t - x) \begin{bmatrix} x_t \\ x_t(x_t - x) \end{bmatrix} x_t(x_t - x)^2 \right|$$

$$\sim A h^2 D_n L_G(1,0) x^2 \int_{-\infty}^{\infty} \begin{bmatrix} s^2 \\ s^3 \end{bmatrix} K(s) ds, \tag{36}$$

where the last approximation above is due to Theorem 1 and $D_n = diag(1,h)$. Therefore, $D_n \sqrt{h} \sqrt{n} |R_n| = o_p(1)$ because we have assumed $nh^{10} = o(1)$. Further, by Theorem 1

$$h^{-1} n^{-1/2} D_n^{-1} H_n D_n^{-1} \xrightarrow{p} L_G(1,0) x^2 \int_{-\infty}^{\infty} \begin{bmatrix} 1 & s \\ s & s^2 \end{bmatrix} K(s) ds \tag{37}$$

In addition, by Theorem 1 and the martingale CLT

$$h^{-1/2} n^{-1/4} D_n^{-1} \sum_{t=1}^{n} u_t \begin{bmatrix} x_t \\ x_t(x_t - x) \end{bmatrix} K_h(x_t - x) \xrightarrow{d} MN\left(\sigma_u^2 L_G(1,0) x^2 \int_{-\infty}^{\infty} \begin{bmatrix} 1 & s \\ s & s^2 \end{bmatrix} K(s)^2 ds \right).$$

$$\tag{38}$$

The result follows from (36)-(38). ∎

8　References

Akaike, H. (1973)Information theory and an extension of the likelihood principle, in Proceedings of the Second International Symposium of Information Theory, ed. B. N. Petrov and F. Csaki. Budapest: Akademiai Kiado.

Berk, R. H. (1966) Limiting behavior of posterior distributions when the model is incorrect, *Annals of Mathematical Statistics*, 37, 51-58.

Berk, R. H. (1970) Consistency a posteriori, *Annals of Mathematical Statistics*, 41, 894-906.

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.

Domowitz, I. and H. White (1982) Misspecified models with dependent observations. *Journal of Econometrics*, 20, 35-58.

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.

Gourieroux, C., Monfort, A. and A. Trognon (1984). Pseudo maximum likelihood methods: Theory, *Econometrica*, vol. 52(3), 681-700.

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

Huber, P. J. (1967) The behavior of maximum likelihood estimates under nonstandard conditions, in Proceedings of the fifth Berkeley Symposium in Mathematical Statistics and Probability. Berkeley: University of California Press.

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. (2009a) Functional form misspecification in regressions with a unit root. *Econometric Theory*, accepted.

Kasparis, I. (2009b) The Bierens test for certain nonstationary models. *Journal of Econometrics*, accepted.

Kasparis, I. and P.C.B. Phillips (2009a) Robust and nuisance-parameter-free IV estimation when the integration order of the regression covariates is unknown. (In preparation)

Kasparis, I. and P.C.B. Phillips (2009b) Dynamic misspecification in nonparametric cointegrating regression. Cowles Foundation Discussion Paper 1700.

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.

Marmer, V., (2007). Nonlinearity, nonstationarity and spurious forecasts. *Journal of Econometrics*, 142. 1-27.

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* 25, 1466-1497.

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 (2009a) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, 25, 710-738.

Wang, Q. and P.C.B. Phillips (2009b) Structural nonparametric cointegrating regression. *Econometrica* (in press).

Wang, Q. and P.C.B. Phillips (2009c) Asymptotic Theory for Zero Energy Density Estimation with Nonparametric Regression Applications. Cowles Foundation Discussion Paper 1687.

White, H. (1981) Consequences and detection of misspecified nonlinear regression models, *Journal of the American Statistical Association*, 76, 419-433.

White, H. (1982) Maximum likelihood estimation if misspecified models. *Econometrica*, Vol. 50, 1-25.

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

Notation

$\mathbf{E}_t\left(.\right) = \mathbf{E}(. \mid \mathcal{F}_{n,t})$ $\qquad$ $a \vee b = \max(a,b)$

$\mathbf{P}_t\left(.\right) = \mathbf{P}(. \mid \mathcal{F}_{n,t})$ $\qquad$ $a \wedge b = \min(a,b)$

$K_h(.) = K\left(\frac{.}{h}\right)$ $\qquad$ $=_d$ distributional equality

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