DYNAMIC MISSPECIFICATION IN NONPARAMETRIC COINTEGRATING REGRESSION

By

Ioannis Kasparis and Peter C.B. Phillips

COWLES FOUNDATION PAPER NO. 1374

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

2013

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

Journal of Econometrics 168 (2012) 270–284

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Dynamic misspecification in nonparametric cointegrating regression

Ioannis Kasparis [a,*], Peter C.B. Phillips [b,c,d,e]

[a] University of Cyprus, Cyprus
[b] Yale University, United States
[c] University of Auckland, New Zealand
[d] University of Southampton, United Kingdom
[e] Singapore Management University, Singapore

A R T I C L E   I N F O

Article history:
Received 26 March 2009
Received in revised form
17 January 2012
Accepted 19 January 2012
Available online 11 February 2012

JEL classification:
C22
C32

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

A B S T R A C T

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 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 ($\sqrt{h\sqrt{n}}$, h is a bandwidth term) but the variance of the limit distribution is larger. The practical import of the results for index models, functional regression models, temporal aggregation and specification testing are discussed. Two nonparametric linearity tests are considered. The proposed tests are robust to dynamic misspecification. Under the null hypothesis (linearity), the first test has a χ^2 limit distribution while the second test has limit distribution determined by the maximum of independently distributed χ^2 variates. Under the alternative hypothesis, the test statistics attain a $h\sqrt{n}$ divergence rate.

1. Introduction

Arguably, all econometric models abstract from reality and are potentially misspecified in uncertain ways. Even if the form of an econometric model were to accurately characterise reality, there is still a myriad of ways in which the generating mechanism for the observed data can depart from the posited model. Therefore, it is important to know the limit properties of various estimators when the underlying 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), Domowitz and White (1982), Gourieroux et al. (1984), Huber (1967), White (1981, 1982) *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 a *pseudo-true value*.[1] The asymptotic analysis of misspecified models is not only of theoretical interest. To obtain asymptotic power rates for various specification tests e.g. Bierens (1990), Ramsey (1969) (or tests without a specific alternative) knowledge about the asymptotic behaviour of the estimator under misspecification is necessary. Moreover, to determine the limit distribution of certain model selection statistics under the null hypothesis, e.g. Cox (1961, 1962), Davidson and McKinnon (1981) and Voung (1989) (tests

* Corresponding author. Tel.: +357 2289 2439.
E-mail address: kasparis@ucy.ac.cy (I. Kasparis).

[1] The pseudo-true value can be different than the parameter of interest and is determined by the value that optimises a certain limit criterion function (see for example Huber, 1967; Akaike, 1973; White, 1982; Bierens, 1984).

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2012.01.037

with a specific alternative), the estimator's limit distribution about the pseudo-true value, is required.

The current paper takes the Wang and Phillips (2009a, hereafter WP)[2] framework and analyses the effects of misspecification relating to the lag structure of a cointegrating model. Further, two nonparametric linearity tests that are robust to dynamic misspecification are proposed. This kind of dynamic misspecification is potentially relevant in a variety of contexts and is especially germane in situations where temporal aggregation issues arise. We show 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 (see Theorem 1 and Example 1). The current work also relates to Phillips (2009) and Kasparis (2011). Phillips (2009) analyses spurious nonparametric regression, while Kasparis (2011) considers the effects of functional form misspecification in the presence of stochastic trends.

One of the main results of the present paper is to show that the Nadaraya–Watson (NW) kernel estimator under dynamic misspecification exhibits inconsistency in nonstationary regression due to the use of integrable functions in the construction of the kernel regression function. It will be shown that, under certain regularity conditions, the effect of the lag misspecification is to induce a shift in the limit, based on a local average of the function around each regression point (i.e. the NW estimator has a pseudo-true function limit). This kind of behaviour is similar to the limit in the case of misspecified dynamics in a stationary time series setting. In this respect, we find analogous results for dynamically misspecified nonparametric models between stationary and nonstationary cases. On the other hand, there is a big difference between nonlinear cointegration models where dynamic mistiming induces inconsistency, as shown here, and linear cointegration models where consistency continues to hold under dynamic mistiming.

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 where the lag differential between the true and the fitted models is large. For badly misspecified models, the limit theory is substantially different. In this case, the NW estimator may be divergent, vanish or converge to a limit involving 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).

Moreover, two linearity tests that are robust to dynamic misspecification are proposed. The test statistics under consideration involve a comparison of the NW kernel estimator with parametric least squares. Asymptotic properties of the tests are derived. Under the null hypothesis of linearity, the first test has a χ^2 limit distribution while the second test has a limit distribution determined by the maximum of independently distributed χ^2 variates. The tests are consistent against integrable and locally integrable alternatives. The divergence rate is of order $h\sqrt{n}$.

[2] Wang and Phillips (2009a) provide a limit theory for nonparametric cointegrating regression. For related work see Guerre (2004), Karlsen et al. (2007), Schienle (2008) and Wang and Phillips (2009b, 2011).

The remainder of the paper is organised as follows. Section 2 provides limit theory for kernel regression under dynamic misspecification. Section 3 provides some applications in contexts of interest for applied work. Section 4 develops linearity tests that are robust to dynamic misspecification. The finite sample properties of the linearity tests are explored in a simulation experiment. Section 5 concludes. Technical results and proofs are given in the Appendices. Notation is fairly standard. For instance, we use $a \vee b$ ($a \wedge b$) to denote the maximum (minimum) of two real numbers a and b, and $=_d$ represents distributional equality. Throughout the paper summations such as $\sum_{t\geq 1}^{n}$ are interpreted as sums over $1 \leq r \vee s \vee l \leq n$ whenever there are integer parameters such as r, s, l governing the initialisation. Finally, I denotes the integrable family of functions, and LI denotes the locally integrable family of functions, that are not integrable.

2. Kernel regression under dynamic misspecification

This section develops a limit theory for the Nadaraya–Watson kernel regression estimator in the case of dynamic misspecification. It is well known (e.g. White, 1981, 1982; Domowitz and White, 1982) that, under certain regularity conditions, parametric estimators of misspecified models converge to some well defined pseudo-true value that 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 — that is, 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 in the sense that the lag differential between the true and fitted models goes to infinity in large samples there is no pseudo-true function limit. In this case, the NW diverges, vanishes or converges to some random limit, depending on the properties of the true regression function.

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

$$y_t = f(x_{t-r}) + u_t, \quad \text{for some integer lag } r \geq 0, \tag{1}$$

where f is a locally integrable regression function. 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

$$x_t = \rho_n x_{t-1} + v_t, \quad x_0 = 0, \tag{2}$$

where v_t is some error term whose properties are specified later (Assumptions 2.2 and 2.3 below) and $\rho_n = 1 - \frac{c_o}{n}$ 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 (2). The regression error u_t is a martingale difference sequence. Both x_t and u_t are defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The exact properties of f, x_t and u_t will be specified in detail later.

We concentrate on the case where a version of (1) 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

$$y_t = \hat{f}(x_{t-s}) + \hat{u}_t, \quad \text{for some fixed integer lag } s \geq 0,\ r \neq s, \tag{3}$$

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

$$\hat{f}(x) = \frac{\sum_{t=s}^{n} K\left(\frac{x_{t-s}-x}{h}\right) y_t}{\sum_{t=s}^{n} K\left(\frac{x_{t-s}-x}{h}\right)}, \quad (4)$$

where $K(\cdot)$ is some kernel function, h is some bandwidth term ($h \to 0$, as $n \to \infty$) and $K_h(\cdot) = K(\cdot/h)$. The main focus of the paper is on the case where the integer lags r and s, in (1), (3), are fixed (mild dynamic misspecification). Some discussion for the severe dynamic misspecification case where the lag differential $|r-s| \to \infty$, as $n \to \infty$ is provided at the end of this section.

For the subsequent analysis we introduce some technical conditions. Assumptions 2.1 and 2.2 below are largely based on WP, to which we refer readers for further discussion. Their notation is used here. First, it is convenient to standardise x_t in array form as $x_{t,n} = x_t/\sqrt{n}$ so that $x_{[nr],n}$ is compatible with a functional law as $n \to \infty$, where $[a]$ denotes the integer part of a. We introduce two companion sequences of real numbers c_n and $d_{l,k,n} = \sqrt{l-k}/\sqrt{n}$. 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 $h_n \to 0$ that arises in kernel estimation. We note that $\left(x_{l,n} - x_{k,n}\right)/d_{l,k,n}$ has a limit distribution as $l - k \to \infty$. As in WP, it is convenient to use the set notation.

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

Assumptions 2.1 and 2.2 deal with the density properties of x_t and their relation to the function f.

Assumption 2.1. For all $0 \le k < l \le n, n \ge 1$, there exist a sequence of σ-fields $\mathcal{F}_{k-1,n} \subseteq \mathcal{F}_{k,n}$ (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 $\mathcal{F}_{n,k-1}$, $\left(x_{l,n} - x_{k,n}\right)/d_{l,k,n}$ 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 $k_0 > 0$,
  $$\sup_{(l,k)\in\Omega_n\left(\delta^{1/(2k_0)}\right)} \sup_{|x|\le\delta} \left|h_{l,k,n}(x) - h_{l,k,n}(0)\right| = o_p(1),$$
  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 $x^o_{t,n}$ such that $\left(x^o_{t,n},\ 1 \le t \le n\right) =_d \left(x_{t,n},\ 1 \le t \le n\right)$ and $\sup_{0\le\eta\le1} \left|x^o_{[n\eta],n} - G(\eta)\right| = o_p(1)$.

Assumption 2.2 is standard in the nonstationary time series literature — e.g. Berkes and Horváth (2006), Park and Phillips (1999, 2000, 2001), Phillips (1991), Wang and Phillips (2009a). Assumption 2.1(a) is the same as Assumption 2.3 of WP. Note that if we set $\mathcal{F}_{k,n} = \sigma(v_0, \ldots, v_k)$, then Assumption 2.1(a) holds for nonstationary processes like the given by (2).[3] Assumption 2.1(b) 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. Finally, note that when x_t is given by (2), Assumption 2.2 is satisfied. In this case, $G(t)$ is either a Brownian Motion or an Ornstein–Uhlenbeck process.

In some cases it is more convenient to work with the Skorohod copy $x^o_{t,n}$, instead of $x_{t,n}$. In the paper we establish weak convergence of the NW estimator to some well defined deterministic limit (a 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 $\left(x^o_{t,n},\ 1 \le t \le n\right) = \left(x_{t,n},\ 1 \le t \le n\right)$ instead of $\left(x^o_{t,n},\ 1 \le t \le n\right) =_d \left(x_{t,n},\ 1 \le t \le n\right)$. With this convention $\xrightarrow{p}$ convergence, for sample functionals of x_t, should be interpreted as $\xrightarrow{d}$ convergence unless the limit is deterministic.

The following assumptions impose regularity conditions on the regression function f and the kernel K in relation to the regressor x_t.

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, $\left|f\left(\frac{\sqrt{n}}{c_n}z + x - v\right) - f(x-v)\right| \le \left(\sqrt{n}/c_n\right)^\gamma f_1(z,x,v)$ with $\int_v \int_z f_1(z,x,v)\,|K(z)|\,p(v)dzdv < \infty$, for each x.
(c) $\int_z |z|\,|K(z)|\,dz$ and $\int_v |f(x-v)|^q\, p_{r-s}(v)\,|v|\,dv < \infty$, for all x and some $q > 1$.

When $c_n = \sqrt{n}/h$, Assumption 2.3(a) requires that the bandwidth sequence $h \to 0$ as $n \to \infty$. The remaining parts of Assumption 2.3 impose local Lipschitz and integrability conditions on f, which are useful technical conditions.

We need to impose some additional conditions on f, K, x_t and u_t.

Assumption 2.4. Assume that $K(\lambda) \ge 0$, $\int_{-\infty}^{\infty} K(\lambda)d\lambda = 1$ and $\sup_\lambda K(\lambda) < \infty$.

Assumption 2.5. $(u_t, \mathcal{F}_{n,t})$ is a martingale difference sequence such that $\mathbf{E}(u_t^2 \mid \mathcal{F}_{n,t-1}) = \sigma^2_{u,t} \to \sigma^2_u$, a.s. as $t \to \infty$.

Assumption 2.6. For some $m > 0$, $\sup_{1\le t\le n} \mathbf{E}(u_t^{2+m} \mid \mathcal{F}_{n,t-1}) < \infty$ a.s.

In what follows it will be convenient to use the notation

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

Remarks. (a) 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 $s < r$).
(b) Moreover, 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 $p_{r-s}(w)$.

Assumption 2.7. For given x, there exists a real function $f_1(s,x)$ such that, when h is sufficiently small, $\left|\mathbf{E}f\left(hy + x + \sum_{rs} v_i\right) - \mathbf{E}f\left(x + \sum_{rs} v_i\right)\right| \le h^\gamma f_1(y,x)$ with $0 < \gamma \le 1$, for all $y \in \mathbb{R}$ and $\int_{-\infty}^{\infty} K(s)f_1(s,x)ds < \infty$. Further, $\mathbf{E}f\left(x + \sum_{rs} v_i\right)^2 < \infty$.

Assumption 2.4 imposes some standard regularity conditions on the kernel function. Assumption 2.5 together with Assumption 2.2(b) postulate that y_t is predetermined i.e. $\mathbf{E}(y_t \mid \mathcal{F}_{n,t-1}) = f(x_{t-r})$. This type of requirement is common in the literature

[3] WP demonstrate that Assumption 2.1(a) holds for fractionally integrated processes under standard moment conditions. Using arguments similar to those in the proof of Corollary 2.2 of WP, it possible to show that Assumption 2.1 also holds for near-unit root processes.

of nonlinear models with integrated time series — see Park and Phillips (1999, 2000, 2001), as well as WP. Assumption 2.5 is important for our derivations as it allows the use of martingale convergence methods. Assumption 2.5 has been recently relaxed by Wang and Phillips (2009b, WP_2), who consider structural nonparametric regressions with unit roots. Relaxation of the martingale difference assumption complicates the asymptotic theory substantially. WP_2 (2009b) develop novel approximate martingale convergence methods that cope with this extension and further address (finite order) dependence between the innovations v_t and u_t. Their approach is significantly different from that followed in the current paper. It seems likely that the inconsistency result under dynamic misspecification that is proved in the current paper extends to this more complex cointegrating structure, but we do not provide that extension here. Finally, Assumption 2.7 is a technical condition that imposes smoothness on the limit of $\hat{f}(x)$. This requirement holds for a variety of regression functions and v_i innovations.

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

Theorem 1. *Suppose that:*

(a) *Assumptions 2.1–2.5 hold.*
(b) *The bandwidth h satisfies $\sqrt{n}h \to \infty$ and $h \to 0$ as $n \to \infty$.*

Then, as $n \to \infty$,

$$\hat{f}(x) \xrightarrow{p} \mathbf{E}f\left(x + \sum_{rs} v_i\right). \tag{5}$$

Under some additional conditions, we have the following limit distribution result.

Theorem 2. *Suppose that:*

(a) *Assumptions 2.1–2.7 hold.*
(b) *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 Assumption 2.3.*
(c) *The bandwidth parameter h satisfies $\sqrt{n}h^{1+2\gamma} \to 0$.*
(d) *K has an integrable Fourier transform.*[4]

Then, as $n \to \infty$,

$$\left(\sum_{t=1}^n K_h\left(x_{t-s} - x\right)\right)^{1/2} \left(\hat{f}(x) - \mathbf{E}f\left(x + \sum_{rs} v_i\right)\right) \xrightarrow{d} N\left(0, \sigma^2(x) \int_{-\infty}^{\infty} K(s)^2 ds\right), \tag{6}$$

where $\sigma^2(x) := \sigma_u^2 + \mathbf{Var}\left\{f\left(x + \sum_{rs} v_i\right)\right\}$.

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

$$\mathbf{E}f\left(x + \sum_{rs} v_i\right) = \int f\left(x + w\right) p_{r-s}\left(w\right) dw, \tag{7}$$

where $\sum_{rs} v_i$ has density $p_{r-s}(w)$. The limit (7) is an average of f taken around the value at x with respect to this density. 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

$$\left(\sum_{t=1}^n K_h\left(x_{t-s} - x\right)\right)^{1/2} \left(\hat{f}(x) - f(x)\right) \xrightarrow{d} N\left(0, \sigma_u^2 \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda\right), \tag{8}$$

as in WP under suitable undersmoothing or choice of h in the regression. The condition $\sqrt{n}h^{1+2\gamma} \to 0$ in (c) is sufficient to ensure the required degree of undersmoothing and relies on the Lipschitz parameter γ as in WP. When $\gamma = 1$, the condition requires that $h \to 0$ faster than $n^{-1/6}$, in comparison to the usual rate of $n^{-1/5}$ in stationary kernel regression. As is clear from the proof in Appendix B, the condition is useful because the bias term depends on differentials that involve random perturbations of the points at which the function is evaluated, such as $f(x_{t-1}) - f(x - v_t)$, which need to be majorized by integrable functions to bound the bias. The condition (c) is sufficient for this purpose but does not appear to be necessary.

The lag misspecification in the fitted nonparametric cointegrating relation (3) produces both inconsistency and a reduction in precision in the limit theory for the NW estimator. The limit distributions (6) and (8) differ in terms of both centring and variance. The centring is explained by the inconsistency (5) under mistiming ($r \neq s$) of the lagged relationship. The additional variance in the limit distribution (6) occurs due to the term $\mathbf{Var}f\left(x + \sum_{rs} v_i\right)$, which is non zero whenever $r \neq s$. The extra component in the variance, which arises as in (5) 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, contrary to the parametric case (e.g. White, 1981; Kasparis, 2011) misspecification in the nonparametric framework necessarily results in larger limit variance.[5]

In the special case of linear cointegration with $f(x_t) = \theta x_t$, we have from (5)

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

so that kernel regression is consistent under lag misspecification, corresponding to the temporal invariance of linear cointegrating regression. In this case, (6) becomes

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

with

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

since $\mathbf{Var}\left\{\sum_{rs} v_i\right\} = |s - r|\,\sigma_v^2$. 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 (1) is the true model and (3) 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 analysed in the literature. However, it is readily shown

[4] Note that if condition (d) holds, then $K(x) \to 0$ as $|x| \to \infty$, by the Riemann–Lebesgue Lemma.

[5] See Kasparis and Phillips (2009) for further discussion.

by conventional methods for stationary nonparametric regression that under suitable regularity and mixing conditions

$$\hat{f}(x) \xrightarrow{p} \mathbf{E}f\left(x_{t-r} \mid x_{t-s} = x\right), \tag{9}$$

which is the analogue for the stationary time series x_t of the inconsistency shown in (5). 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 (9) may be written in the form

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

which corresponds precisely to the limit in (5) 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

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

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* $n \to \infty$,

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

Remarks (*Severe Dynamic Misspecification*).

(a) Suppose the true model is still given by (1) with a fixed lag r, but the fitted regression is

$$y_t = \hat{f}(x_{t-s_n}) + \hat{u}_t, \tag{10}$$

where the time lag in the specification is long and expressed as a fraction of the sample as $s_n := [cn]$ with $0 < c < 1$. We might similarly specify the true model as one involving a long lag and the fitted model as one with a fixed lag, with similar implications. Such distortions in specification might arise when there is a long gestation lag in response time which is not captured in the empirical specification. The NW estimator of (10) does not have a pseudo-true function limit in this case. In particular, $\hat{f}$ diverges when f is unbounded and locally integrable. For bounded locally integrable f, the estimator $\hat{f}$ has a stochastic integral limit. Finally, if f is integrable, $\hat{f}$ vanishes. Under certain regularity conditions, we have the following limit behaviour (see Kasparis and Phillips, 2009, for further details):

(i) Suppose that f is locally integrable (LI) and asymptotically homogeneous of the form

$$f(\lambda x) = \kappa_f(\lambda)H_f(x) + \nu_f(x, \lambda), \tag{11}$$

where $H_f(x) \in LI$ and $\sup_x \left|\nu_f(x, \lambda)\right| = o(\kappa_f(\lambda))$ as $\lambda \to \infty$.[6] Then, by Theorem 1 of Phillips (2009) as $n \to \infty$,

$$\kappa_f\left(\sqrt{n}\right)^{-1}\hat{f}(x) \xrightarrow{d} \frac{1}{L_G(1-c, 0)}\int_c^1 H_f\left(G(\eta)\right) dL_G(\eta - c, 0). \tag{12}$$

(ii) Suppose that f is integrable (I). Then, as $n \to \infty$

$$\sqrt{n^{1/2}h}\hat{f}(x) \xrightarrow{d} MN\left(0, \sigma_u^2 \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda L_G^{-1}(1-c, 0)\right). \tag{13}$$

(b) 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. Further, for integrable f, the estimator vanishes. Therefore in general, when the lag differential $|r - s|$ increases, $\left|\hat{f}\right|$ either increases or vanishes. 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 misspecification in this case are analogous to those of spurious nonparametric regression (see Phillips, 2009).

(c) These results may be compared with the stationary, weakly dependent case where the joint density of $\left(X_t, X_{t-s_n}\right)$ factors asymptotically so that $p_{X_t, X_{t-s_n}}(X, Y) \to p_{X_t}(X)\, p_{X_t}(Y)$ as $s_n \to \infty$. In this event, we find that $\hat{f}(x) \xrightarrow{p} \mathbf{E}\left(f\left(X_t\right)\right) = \int f(y)\, p_{X_t}(y)\, dy$, so that the stationary misspecified case is inconsistent also, but has a different, well defined, nonrandom limit — the mean of the function $f(X_t)$.

3. Some specific examples

Example 1 (*Dynamic Misspecification in Parametric Regression*). Suppose that the true model is

$$y_t = \theta g(x_{t-r}) + u_t, \tag{14}$$

where θ is an unknown parameter and g some regression function. Moreover, x_t is an integrated process, and u_t is some martingale difference error that both satisfy the assumptions of the current paper. In place of (14), suppose that the following dynamically misspecified model is estimated by least squares (LS):

$$y_t = \hat{\theta}g(x_{t-s}) + \hat{u}_t. \tag{15}$$

If the regression function g is continuous with $g \in LI$, it can be shown (see, for example Kasparis, 2008, Lemma A1(b)) that the LS estimator in this case satisfies

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

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

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 g is integrable then it follows directly from our limit theory (Proposition A in the Appendix) that

[6] The functions $H_f(\cdot)$ and $\kappa_f(\cdot)$ are the asymptotic homogeneous function and asymptotic order of f respectively. Limit theory for parametric models of this form (asymptotically homogeneous) is provided in Park and Phillips (1999, 2001).

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

$$= \theta \frac{\mathbf{E} \int_{-\infty}^{\infty} g(\lambda) g\left(\lambda + \sum_{rs} v_i\right) d\lambda}{\int_{-\infty}^{\infty} g(\lambda)^2 d\lambda} + o_p(1),$$

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

Example 2 (*KPSS Test Under Dynamic Misspecification*)**.** The standard KPSS test cannot detect dynamic misspecification. The KPSS test is a residual based goodness of fit test aimed to detect abnormal fluctuation in the (parametric) residuals, when the fitted model is misspecified. As shown in Example 1 above, smooth locally integrable regression models are invariant under temporal translation. As a result, the KPSS has no power against dynamic misspecification. Suppose that the true and fitted models are given by (14) and (15) respectively. Further, assume that the regression function g, of (14), has a continuous derivative $\dot{g}$, and $g, \dot{g} \in LI$ are asymptotically homogeneous as in (11). Consider the KPSS statistic

$$KPSS_n = n^{-1} \sum_{j=1}^{n} \left(\sum_{t=1}^{j} \frac{\hat{u}_t}{\sqrt{n}} \right)^2 \Big/ \sum_{t=1}^{n} \frac{\hat{u}_t^2}{n}$$

based on the parametric residuals $\hat{u}_t = y_t - \hat{\theta} g(x_{t-s})$. Under the additional assumption: $\kappa_{\dot{g}}(\lambda) \to \infty$, as $\lambda \to \infty$, it can be shown (see Kasparis, 2008, Lemma A1) that

$$\sqrt{n} \frac{\kappa_g\left(\sqrt{n}\right)}{\kappa_{\dot{g}}\left(\sqrt{n}\right)} \left(\hat{\theta} - \theta\right)$$

$$\xrightarrow{d} \frac{\sqrt{|r-s|} \int_0^1 H_g\left(B_x(\lambda)\right) H_{\dot{g}}\left(B_x(\lambda)\right) dB_x(\lambda)}{\int_0^1 H_g\left(B_x(\lambda)\right)^2 d\lambda} := Z_*,$$

$$\frac{1}{\sqrt{n}\kappa_{\dot{g}}\left(\sqrt{n}\right)} \sum_{t=1}^{n} \hat{u}_t$$

$$\xrightarrow{d} \theta \sqrt{|r-s|} \int_0^1 H_g\left(B_x(\lambda)\right) H_{\dot{g}}\left(B_x(\lambda)\right) dB_x(\lambda)$$

$$- Z_* \int_0^1 H_g\left(B_x(\lambda)\right) d(\lambda) := \bar{U}(1),$$

where B_x and B_u are Brownian motions. Further,

$$\frac{1}{\kappa_{\dot{g}}\left(\sqrt{n}\right)^2} \hat{\sigma}^2 = \frac{1}{n \kappa_{\dot{g}}\left(\sqrt{n}\right)^2} \sum_{t=1}^{n} \hat{u}_t^2$$

$$\xrightarrow{p} \theta^2 \mathbf{E}\left[\sum_{rs} v_i\right]^2 \int_0^1 H_{\dot{g}}\left(B_x(\lambda)\right)^2 d\lambda := s_*^2.$$

The above suggest that the KPSS test statistic is bounded in probability, even if $r \neq s$. In particular, we have:

$$KPSS_n \xrightarrow{d} \frac{1}{s_*^2} \int_0^1 \bar{U}(v)^2 dv.$$

A similar limit result holds for $KPSS_n$ if instead of $\lim_{\lambda\to\infty} \kappa_{\dot{g}}(\lambda) = \infty$, we assume $\lim_{\lambda\to\infty} \kappa_{\dot{g}}(\lambda) = 1$ (note that for linear g we have $\kappa_{\dot{g}}(\lambda) = 1$).

Example 3 (*Single Index Model*)**.** Suppose that y_t is generated by the single index model:

$$y_t = f(\theta_1 x_t + \theta_2 x_{t-1}) + u_t, \theta_1, \theta_2 \in \mathbb{R}$$

where the regressor x_t satisfies Assumptions 2.1 and 2.2 and u_t is a martingale difference sequence satisfying Assumptions 2.5 and 2.6. 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. It can be easily demonstrated that

$$\hat{f}(x) \xrightarrow{p} \mathbf{E} f\left((\theta_1 + \theta_2) x - \theta_2 v_t\right),$$

as in Theorem 1. 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 4 (*Temporal Aggregation*)**.** When a regressor x_t is sampled (two times) more frequently than y_t, Ghysels et al. (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

$$y_t = \lambda f(x_t) + (1 - \lambda) f(x_{t-1}) + u_t, \quad 0 \leq \lambda \leq 1, \tag{16}$$

and where x_t and u_t are as in Example 1. If the fitted model ignores the temporal aggregation in (16) and is a simple nonparametric regression of the form

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

then Theorem 1 shows (see also Proposition A) 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. Marmer (2007) and Kasparis (2010) test for the predictability of stock returns in the context of nonlinear models with integrated time series. It is natural to expect that temporal aggregation issues arise in this area. For example, Ghysels et al. (2005) use MIDAS regression to show that stock market volatility, aggregated using high frequency (intraday) information, can be used to explain low frequency (monthly) stock returns. In a similar context, Ludvigson and Ng (2007) have a regression model where the return volatility is sampled more frequently than stock returns.

Example 5 (*Nonparametric Unit Root Autoregression*)**.** Suppose that the true model is given by the autoregression

$$x_t = f(x_{t-1}) + u_t, \tag{17}$$

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

$$x_t = \hat{f}(x_{t-2}) + \hat{u}_t. \tag{18}$$

Under the true model (17) 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

$$\left(\sum_{t=1}^{n} K_h(x_{t-2} - x) \right)^{1/2} \left(\hat{f}(x) - x \right)$$

$$\xrightarrow{d} N\left(0, \ 2\sigma_u^2 \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda \right).$$

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 (18) 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

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

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

Example 6 (*Misspecified Functional Coefficient Models*)**.** Cai et al. (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

$$y_t = \beta\left(z_t\right)' x_t + u_t, \quad t = 1, \ldots, n \tag{19}$$

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 $\beta(\cdot)$ 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 (2009a), CLP showed that for any fixed $z\hat{\beta}(z)$ is consistent with mixed normal distribution. If (19) is estimated when the true response function is $\beta\left(z_{t-1}\right)$, the methods of the present paper may be used to show that the nonparametric estimate $\hat{\beta}(z)$ is inconsistent and convergent to the averaged coefficient $\mathbf{E}\left\{\beta\left(z - \Delta z_t\right)\right\}$ with 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 $h \to 0$, $\nu_0 = \int K(\lambda)^2 d\lambda$ 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.

4. Linearity tests robust to dynamic misspecification

This section develops two linearity tests that are robust to dynamic misspecification. Under the null hypothesis the regression function is linear, i.e., $H_0 : f(x) = \theta_0 + \theta_1 x$, for some unknown vector of parameters $\theta' = (\theta_0, \theta_1)$. Consider the following variance estimator for σ_u^2

$$\hat{\sigma}_w^2 = \sum_{t=1}^n \hat{u}_t^2 w\left(x_{t-s}\right) \Big/ \sum_{t=1}^n w\left(x_{t-s}\right), \tag{20}$$

where $\hat{u}_t$ are the least squares residuals under the null, viz., $\hat{u}_t = y_t - \hat{\theta}_0 - \hat{\theta}_1 x_{t-s}$, and $w(x)$ is a non-negative weight function such that $w(x), f(x)^2 w(x), x^2 w(x) \in I$. The proposed tests are based on the following nonparametric t-statistic

$$\hat{t}(x,\theta) := \left(\frac{\sum_{t=1}^n K_h\left(x_{t-s} - x\right)}{\hat{\sigma}_w^2 \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda}\right)^{1/2} \left(\hat{f}(x) - \theta_0 - \theta_1 x\right).$$

It can be shown that under linearity the t-statistic $\hat{t}(x,\theta) \xrightarrow{d} N(0,1)$ under both correct and incorrect dynamic specification. The $\hat{t}(x,\theta)$ statistic forms the basis of a linearity test that is robust to dynamic misspecification because: (a) for linear f, the NW estimator $\hat{f}$ is consistent even if the fitted model is misspecified in terms of dynamic structure; and (b) it can be shown that $\hat{\sigma}_w^2$ provides a consistent estimator for the limit variance of $\hat{f}$ under correct or incorrect dynamic specification, when f is linear. In particular, for $f(x) = \theta_0 + \theta_1 x$ we have

$$\hat{\sigma}_w^2 \xrightarrow{p} \sigma_u^2 + \mathbf{Var} f\left(x + \sum_{rs} v_i\right) = \sigma_u^2 + \theta_1^2 \mathbf{E}\left(\sum_{rs} v_i\right)^2.$$

Therefore, $\hat{\sigma}_w^2$ provides a valid standardisation even if the model is misspecified in terms of dynamic structure.

Under linearity, it transpires that all the following variance estimators

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

have the same limit $\sigma_u^2 + \theta_1^2 \mathbf{E}\left(\sum_{rs} v_i\right)^2$. Numerical results indicate that best test performance is achieved when $\hat{\sigma}_w^2$ is utilised. The nonparametric estimator, $\hat{\sigma}^2(x)$ results in oversized[7] tests. The parametric estimator $\tilde{\sigma}^2$ results in good test size, but the linearity tests become inconsistent when $\tilde{\sigma}^2$ is employed because under the alternative hypothesis (nonlinearity) it can be shown that $\tilde{\sigma}^2$ diverges faster than the numerator of $\hat{t}(x,\theta)$, and so $\hat{t}(x,\theta)$ vanishes as $n \to \infty$. The estimator $\hat{\sigma}_w^2$ is based on weighted parametric residuals. The integrable weight functions in (20) control the estimator's divergence rate under H_1. As a result, our linearity tests are consistent, when $\hat{\sigma}_w^2$ is used for standardisation in $\hat{t}(x,\theta)$.

The parameter vector θ is generally unknown and can be consistently estimated by the least squares estimator $\hat{\theta}$ under the null. Since $\hat{\theta}$ is $O_p\left(diag\left(\sqrt{n}, n\right)\right)$ consistent for θ under linearity we have

$$\hat{t}(x,\hat{\theta}) \xrightarrow{d} N(0,1). \tag{21}$$

Therefore, the feasible test statistic $\hat{t}(x,\hat{\theta})$ involves a comparison of the nonparametric estimator $\hat{f}(x)$ with the parametric estimator $\hat{\theta}_1 + \hat{\theta}_2 x$. The test statistics are based on making this comparison over a set of points. In particular, let X_k be a set of isolated points $X_k = \{\bar{x}_1, \ldots, \bar{x}_k\}$ in $\mathbb{R}$ for some $k \in \mathbb{N}$. The test statistics are the sum and sup statistics over this set, viz.,

$$\hat{F}_{\text{sum}} := \sum_{x \in X_k}\left[\hat{t}(x,\hat{\theta})\right]^2 \quad \text{and} \quad \hat{F}_{\max} := \max_{x \in X_k}\left[\hat{t}(x,\hat{\theta})\right]^2.$$

We consider the linearity hypothesis

$$H_0 : f(x) = \theta_0 + \theta_1 x, \quad \text{for some } \theta = (\theta_0, \theta_1) \in \mathbb{R}^2 \text{ and all } x \in X_k.$$

The limit properties of the test statistics under the null hypothesis are demonstrated by the following result.

[7] A preliminary simulation experiment has shown that when $\hat{\sigma}^2(x)$ is employed the linearity test under consideration exhibits severe size distortions. In some cases size is three times nominal size.

Theorem 4. *Suppose that the conditions of Theorem 2 hold. Then, under H_0, as $n \to \infty$,*

$$\hat{F}_{sum} \xrightarrow{d} \chi^2_k \quad \text{and} \quad \hat{F}_{\max} \xrightarrow{d} Y,$$

where the random variable Y has distribution function $F_Y(y) = \mathbf{P}(X \le y)^k$ with $X \sim \chi^2_1$.

The component statistics $\hat{t}(\bar{x}_1, \hat{\theta}), \dots, \hat{t}(\bar{x}_k, \hat{\theta})$ are asymptotically independent.[8] As a result, $\hat{F}_{\text{sum}}$ has a χ^2_k limit distribution. Similarly, the limit distribution of $\hat{F}_{\max}$ is determined as the maximum of independently distributed χ^2_1 variates.

The alternative hypothesis is

$$H_1 : f(x) \ne \theta_0 + \theta_1 x, \quad \text{for all } \theta = (\theta_0, \theta_1) \in \mathbb{R}^2 \text{ and some } x \in X_k \tag{22}$$

and the following result gives the asymptotic behaviour of $\hat{F}_{\text{sum}}$ and $\hat{F}_{\max}$ under this alternative. We consider nonlinear alternatives for f that are either integrable (I) or locally integrable (LI), as in (11). Suppose that (22) holds, and let $X'_k \subseteq X_k$ be the set of regression points that satisfy H_1. Further, define the vector $(\theta^*_0, \ \theta^*_1)$ as follows

$$\begin{bmatrix}\theta^*_0 \\ \theta^*_1\end{bmatrix} := \begin{bmatrix} 1 & \int_{-\infty}^{\infty} \lambda L_G(1,\lambda) d\lambda \\ \int_{-\infty}^{\infty} \lambda L_G(1,\lambda) d\lambda & \int_{-\infty}^{\infty} \lambda^2 L_G(1,\lambda) d\lambda \end{bmatrix}^{-1} \times \begin{bmatrix} \int_{-\infty}^{\infty} H_f(\lambda) L_G(1,\lambda) d\lambda \\ \int_{-\infty}^{\infty} \lambda H_f(\lambda) L_G(1,\lambda) d\lambda \end{bmatrix},$$

where $H_f(\lambda)$ is the limit homogeneous function of f, for $f \in LI$ e.g. (11). We have the following limit result.

Theorem 5. *Suppose that the conditions of Theorem 2 hold. Then, under H_1, as $n \to \infty$*

$$\frac{1}{h\sqrt{n}} \hat{F}_{sum} = \frac{L_G(1,0) \int_{-\infty}^{\infty} K(\lambda) d\lambda}{\sigma^2_* \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda} \sum_{x \in X'_k} D(x)^2 + o_p(1),$$

and

$$\frac{1}{h\sqrt{n}} \hat{F}_{\max} = \frac{L_G(1,0) \int_{-\infty}^{\infty} K(\lambda) d\lambda}{\sigma^2_* \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda} \max_{x \in X'_k} D(x)^2 + o_p(1),$$

where:

(i) *for $f \in LI$ with $\kappa_f(\sqrt{n}) = 1$,*

$$D(x) = \mathbf{E} f\left(x + \sum_{rs} v_i\right) - \theta^*_0,$$

and $$\sigma^2_* = \frac{\int_{-\infty}^{\infty} \left\{ \sigma^2_u + \left[\mathbf{E} f\left(\lambda + \sum_{rs} v_i\right) - \theta^*_0 \right]^2 \right\} w(\lambda) d\lambda}{\int_{-\infty}^{\infty} w(\lambda) d\lambda};$$

(ii) *for $f \in LI$ with $\kappa_f(\sqrt{n}) \to \infty$,*

$$D(x) = \theta^*_0 \quad \text{and} \quad \sigma^2_* = (\theta^*_0)^2;$$

(iii) *for $f \in I$ or $f \in LI$ with $\kappa_f(\sqrt{n}) \to 0$,*

$$D(x) = \mathbf{E} f\left(x + \sum_{rs} v_i\right) \quad \text{and}$$

$$\sigma^2_* = \frac{\int_{-\infty}^{\infty} \left\{ \sigma^2_u + \mathbf{E} f\left(\lambda + \sum_{rs} v_i\right)^2 \right\} w(\lambda) d\lambda}{\int_{-\infty}^{\infty} w(\lambda) d\lambda}.$$

Theorem 5 demonstrates that the tests have asymptotic power against both I and LI nonlinear functions. The term $D(x)$ shown above involves parametric and/or nonparametric pseudo-true values. Under H_1, the divergence rate is $h\sqrt{n}$. In particular, under H_1, we have $\hat{t}(x, \hat{\theta})^2 \sim h\sqrt{n} D(x)$. Part (i) of Theorem 5 considers the case where $f \in LI$ with fixed asymptotic order. For instance, for $f(x) = 1\{x \ge 0\}$ or $f(x) = 1/(1 + \exp(-x))$ the asymptotic order is $\kappa_f(\sqrt{n}) = 1$. In this case the leading terms of the test statistics are given by the parametric and the nonparametric estimators. Note that the term $D(x)$ shown above involves both the parametric and the nonparametric pseudo-true values. The focus in part (ii) is locally integrable alternatives with increasing asymptotic order. Here the leading terms of the test statistics are given by the parametric estimator only. Finally, part (iii) considers integrable alternatives and locally integrable alternatives with decreasing asymptotic order. In this case the power rate is driven by the nonparametric pseudo-true value only.

Remark (*Tests Under Severe Dynamic Misspecification*). Theorems 4 and 5 above suggest that the tests are robust to "mild" dynamic misspecification. Note however that when the lag differential $|r - s|$ is large the performance of the tests is likely to deteriorate. Suppose that $r = 0$ and $s = s_n = [cn]$. Then the effects of severe dynamic misspecification on the tests can be explained by the earlier limit results (12) and (13):

(i) Under severe misspecification the tests are likely to be oversized. When H_0 holds, the statistic $\hat{t}(x, \hat{\theta})$ is of order $O_p\left(\sqrt{h\sqrt{n}}\right)$. To keep the presentation simple suppose that the parametric model does not involve an intercept.[9] Then, using (12) and the limit theory of Phillips (2009) we find

$$\frac{1}{\sqrt{h\sqrt{n}}} \hat{t}(x, \hat{\theta}) \xrightarrow{d} \frac{\int_0^1 G(\eta) dL_{\tilde{G}}(\eta, 0)}{\left(L_{\tilde{G}}(1,0) S^*_w \int_{-\infty}^{\infty} K^2(\lambda) d\lambda\right)^{1/2}},$$

where $\tilde{G}(\eta) = G(\eta - c)\mathbf{1}(1 \ge \eta \ge c)$,

$$S^*_w = \frac{\int_{\lambda=-\infty}^{\infty} w(\lambda) \int_{\eta=c}^{1} \left(G(\eta) - \theta_* \tilde{G}(\eta)\right)^2 dL_G(\eta - c, \lambda) d\lambda}{\int_{-\infty}^{\infty} w(\lambda) d\lambda L_{\tilde{G}}(1,0)},$$

and $$\theta_* = \frac{\int_0^1 G(\eta) \tilde{G}(\eta) d\eta}{\int_0^1 \tilde{G}(\eta)^2 d\eta}.$$

(ii) Further, when misspecification is severe, the tests are expected to have poor asymptotic power against integrable alternatives. When H_1 holds, the statistic $\hat{t}(x, \hat{\theta})$ is convergent. In this case, using (13), Phillips (2009) and Chang et al. (2001) we find

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

[8] See Lemma 3(ii) and (iii) and the proof of Theorem 4 in the Appendix.

[9] A similar result holds when an intercept is included, but not reported here because of the length of the expressions involved.

Table 1
($r = 0$): Size and power of linearity tests F_{sum} and F_{max} under the null hypothesis of linearity (f_1) against alternatives (f_2–f_5) in (23).

F_{sum}	f_1	f_2	f_3	f_4	f_5	F_{max}	f_1	f_2	f_3	f_4	f_5
$n = 250$											
$b = 0.10$	0.125	1.000	0.943	0.890	0.936	$b = 0.10$	0.091	1.000	0.945	0.895	0.944
$b = 0.20$	0.021	1.000	0.912	0.832	0.917	$b = 0.20$	0.034	0.998	0.931	0.867	0.939
$b = 0.30$	0.018	0.993	0.854	0.732	0.856	$b = 0.30$	0.034	0.985	0.873	0.796	0.895
$b = 0.40$	0.016	0.920	0.844	0.676	0.799	$b = 0.40$	0.036	0.872	0.727	0.757	0.873
$b = 0.45$	0.017	0.863	0.767	0.618	0.709	$b = 0.45$	0.038	0.731	0.550	0.719	0.818
$n = 500$											
$b = 0.10$	0.115	1.000	0.992	0.972	0.958	$b = 0.10$	0.090	1.000	0.990	0.969	0.962
$b = 0.20$	0.032	1.000	0.976	0.934	0.941	$b = 0.20$	0.040	0.997	0.981	0.946	0.957
$b = 0.30$	0.025	0.992	0.951	0.871	0.914	$b = 0.30$	0.041	0.982	0.955	0.898	0.942
$b = 0.40$	0.025	0.947	0.940	0.824	0.862	$b = 0.40$	0.043	0.886	0.882	0.863	0.913
$b = 0.45$	0.023	0.870	0.888	0.748	0.768	$b = 0.45$	0.043	0.776	0.735	0.805	0.847
$n = 1000$											
$b = 0.10$	0.107	1.000	1.000	0.993	0.970	$b = 0.10$	0.088	0.999	0.999	0.991	0.976
$b = 0.20$	0.035	0.999	0.995	0.976	0.959	$b = 0.20$	0.042	0.996	0.996	0.979	0.972
$b = 0.30$	0.031	0.986	0.990	0.958	0.955	$b = 0.30$	0.042	0.969	0.981	0.961	0.972
$b = 0.40$	0.033	0.920	0.961	0.899	0.888	$b = 0.40$	0.045	0.854	0.919	0.905	0.923
$b = 0.45$	0.027	0.839	0.903	0.818	0.799	$b = 0.45$	0.043	0.737	0.801	0.841	0.863

Table 2
Size of linearity tests F_{sum} and F_{max}.

	F_{sum}						F_{max}					
	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
$n = 250$												
$b = 0.10$	0.125	0.177	0.282	0.372	0.438	0.481	0.091	0.109	0.174	0.245	0.307	0.351
$b = 0.20$	0.021	0.035	0.099	0.172	0.242	0.293	0.034	0.043	0.079	0.131	0.174	0.223
$b = 0.30$	0.018	0.022	0.047	0.078	0.110	0.142	0.034	0.036	0.056	0.081	0.104	0.133
$b = 0.40$	0.016	0.019	0.035	0.051	0.059	0.077	0.036	0.039	0.051	0.066	0.079	0.090
$b = 0.45$	0.017	0.016	0.031	0.040	0.048	0.057	0.038	0.039	0.046	0.059	0.067	0.076
$n = 500$												
$b = 0.10$	0.115	0.194	0.335	0.443	0.518	0.575	0.090	0.131	0.221	0.308	0.384	0.442
$b = 0.20$	0.032	0.044	0.109	0.194	0.278	0.349	0.040	0.048	0.094	0.158	0.205	0.253
$b = 0.30$	0.025	0.030	0.058	0.093	0.136	0.170	0.041	0.044	0.065	0.091	0.115	0.142
$b = 0.40$	0.025	0.027	0.044	0.056	0.069	0.091	0.043	0.041	0.054	0.070	0.077	0.088
$b = 0.45$	0.023	0.027	0.034	0.040	0.047	0.065	0.043	0.045	0.053	0.057	0.065	0.069
$n = 1000$												
$b = 0.10$	0.107	0.196	0.351	0.479	0.564	0.630	0.088	0.133	0.236	0.340	0.421	0.487
$b = 0.20$	0.035	0.050	0.127	0.216	0.296	0.372	0.042	0.054	0.102	0.161	0.215	0.258
$b = 0.30$	0.031	0.036	0.072	0.102	0.141	0.181	0.042	0.047	0.067	0.098	0.118	0.142
$b = 0.40$	0.033	0.035	0.048	0.061	0.067	0.084	0.045	0.047	0.056	0.066	0.076	0.084
$b = 0.45$	0.027	0.028	0.037	0.044	0.047	0.057	0.043	0.049	0.052	0.058	0.062	0.068

We report briefly some simulation findings concerning the finite sample properties of these linearity tests. The simulation is based on 5000 replications of the model

$$y_t = f(x_{t-r}) + u_t, \qquad x_t = x_{t-1} + v_t,$$

with $(v_t, u_t) \sim i.i.d.N(\mathbf{0}, I)$, and the following functions

$$f_1(x) = x, \qquad f_2(x) = \text{sign}(x)\,|x|^{\frac{3}{2}}, \qquad f_3(x) = 2\text{sign}(x)\,|x|^{\frac{1}{2}}$$

$$f_4(x) = \ln(0.1 + |x|), \qquad f_5(x) = 5e^{-x^2}. \tag{23}$$

We test the null of linearity ($f = f_1$) against the nonlinear alternatives $f = f_2, \ldots, f_5$. The functions f_2, f_3, f_4 are locally integrable, and f_5, is integrable. The nonparametric fitted regression from a model which is possibly misspecified in terms of its dynamic structure is $\hat{f}(x) = \sum_{t=1}^{n} K_h(x_t - x)\,y_t / \sum_{t=1}^{n} K_h(x_t - x)$ and the fitted parametric model is $y_t = \hat{\theta}_0 + \hat{\theta}_1 x_t + \hat{u}_t$.

We use the normal kernel for $K(x)$, sample sizes $n = 250$, 500, 1000 and explore robustness of the tests against a range of possible lags: $r = 0, 1, 2, 3, 4, 5$. The bandwidth is $h = n^{-b}$ with settings $b = 0.1, 0.2, 0.3, 0.4, 0.45$. The weight function w for the construction of the variance estimator $\hat{\sigma}_w^2$ is $w(x) = \exp(-|x|^{1/2}/2)$. The "grid" X_k is given by $\{-5, \ldots, -1, 0, 1, \ldots, 5\}$. Nominal size is set at 5%.

The findings reveal good size control for all sample sizes provided the bandwidth $h \leq n^{-0.2}$ (Table 1). The tests also show good power (Table 1). The F_{sum} test appears to have superior power against polynomial alternatives, while the F_{max} test is superior against logarithmic and integrable alternatives. Size becomes conservative as the bandwidth decreases. The F_{sum} test is more conservative than the F_{max} test. Size is also robust against lag misspecification, with greater resilience for smaller bandwidths (Table 2). Hence, if dynamic misspecification is suspected, then smaller bandwidth choice in testing is recommended for better size control.

Finally, we provide some discussion about the role of the weight function that is utilised for the construction of the variance estimator $\hat{\sigma}_w^2$. The simulation results show that in some cases there is a power drop against the polynomial alternative f_2 when the sample size increases for the weight function $w(x)$ (Table 1). As mentioned, the purpose of the integrable weight function is to prevent the variance estimator diverging under H_1. Without a weight function the tests are inconsistent (the test statistics vanish as $n \to \infty$). On the other hand the weight function results in information loss because data away from the origin are down-weighted. In general, "thick tailed" weighting functions involve smaller information loss. However, there can be a drop in power when the term $f(\cdot)^2 w(\cdot)$ is large, as for example when $f(x)$ is a higher order polynomial. In general, this term assumes large values when f is a higher order polynomial. Therefore, against higher order polynomial alternatives, "thin tailed" weight functions are recommended.[10]

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 regression. In consequence, correct dynamic specification takes on new significance in nonlinear cointegrating systems. Specification tests for nonlinear cointegration 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.

Acknowledgements

The authors thank the CoEditor and three referees for helpful comments and suggestions on this paper. They are also grateful to Timos Papadopoulos for the simulation results. The second author acknowledges partial support from NSF Grants #SES 06-47086 and SES 09-56687.

Appendix A. Supporting results

The following six limit results extend the WP framework as needed to accommodate sample covariances of convolution integrable functions (f) and integrable kernels (K) involving x_t. It will be convenient to use notation $\phi_\epsilon(x) = \left(2\pi\epsilon^2\right)^{-1/2}\exp\left(-x^2/2\epsilon^2\right)$ and $\phi(x) = \phi_1(x)$. We also often write the density $p_1(v)$ as $p(v)$, and use the following (standard) notation for conditional expectation and conditional probability: $\mathbf{E}_t(\cdot) = \mathbf{E}(\cdot \mid \mathcal{F}_{n,t})$ and $\mathbf{P}_t(\cdot) = \mathbf{P}(\cdot \mid \mathcal{F}_{n,t})$. In the following proofs, we use A as a generic constant whose value may change in each location.

Set

$$S_n(\eta) := \frac{c_n}{n}\sum_{t=1}^{[n\eta]} f(x_{t-r}) K\left[c_n\left(\frac{x_{t-s}-x}{\sqrt{n}}\right)\right], \quad 0 < \eta \le 1.$$

[10] In more extensive simulations (not reported in detail here), a thicker tailed (than $w(x)$) weight function $w^*(x) = \exp(-|x|/2)$ was also used. For $w^*(x)$ all the tests have monotonic power against the polynomial alternative f_2. Nevertheless, the thinner tailed $w(x)$ results in overall better performance.

Proposition A below provides limit theory for $S_n(\eta)$, and is a partial generalisation of the main limit result of WP, for sample covariances of integrable transformations of nonstationary sequences and time translated sequences. Note that if we set $f = 1$ we obtain the sample average term of WP (Theorem 2.1). Proposition A is fundamental for the proof of Theorems 1–4 of the paper. It is also of independent interest (see the Remark below).

Proposition A. *Suppose that Assumptions 2.1–2.3 and the following conditions hold:* $\left|f\left(\frac{\sqrt{n}}{c_n}z + x - v\right)\right| \le f_0(z, x, v)$ *for n large enough, with* $\int_v\int_z f_0(z, x, v)\,|K(z)|\,p_{r-s}(v)dzdv < \infty$,

$$\int_v\left\{\int_z |f_0(z, x, v)|\,|K(z)|\,dz\right\}^2 p_{r-s}(v)dv < \infty \quad \text{and}$$
$$\int_v\int_z f_0^2(z, x, v)K^2(z)p_{r-s}(v)dzdv < \infty,$$

for and each $x \in \mathbb{R}$, *and* $r, s \in \mathbb{N}$;

Then, we have the following:

(i) *If Assumption 2.2 (a) holds and* $n \to \infty$

$$S_n(\eta) \xrightarrow{d} \mathbf{E}f\left(x + \sum_{rs} v_i\right)\int_{-\infty}^{\infty} K(\lambda)\,d\lambda L_G(\eta, 0).$$

(ii) *If Assumption 2.2 (b) holds and* $n \to \infty$

$$S_n(\eta) \xrightarrow{p} \mathbf{E}f\left(x + \sum_{rs} v_i\right)\int_{-\infty}^{\infty} K(\lambda)\,d\lambda L_G(\eta, 0),$$

uniformly in $\eta \in (0, 1]$.

Remark. Replace Assumption 2.3(a) by the requirement $\lim_{n\to\infty}\sqrt{n}/c_n = m_0 > 0$. Then, we have the following modification of Proposition A, which is relevant to parametric estimation (e.g. Example 1):

$$S_n(\eta) \xrightarrow{d} \int_{-\infty}^{\infty}\mathbf{E}f\left(x + \lambda m_0 + \sum_{rs} v_i\right)K(\lambda)\,d\lambda L_G(\eta, 0).$$

Lemma 1. *Suppose that*

(a) *Assumption 2.1 holds.*
(b) $\left|f\left(\frac{\sqrt{n}}{c_n}z + x - v\right)\right| \le f_0(z, x, v)$, *for n large enough and*
  (i) $\int_v\int_z f_0(z, x, v)\,|K(z)|\,p_{r-s}(v)dzdv < \infty$,
  (ii) $\int_v\left\{\int_z |f_0(z, x, v)|\,|K(z)|\,dz\right\}^2 p_{r-s}(v)dv < \infty$ *and*
  (iii) $\int_v\int_z f_0^2(z, x, v)K^2(z)p_{r-s}(v)dzdv < \infty$,

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

Let

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

Then

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

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

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

$$= \frac{c_n}{n}\sum_{t=1}^{[n\eta]}\mathbf{E}_{t-2}z_t + \frac{c_n}{n}\sum_{t=1}^{[n\eta]}\left(z_t - \mathbf{E}_{t-2}z_t\right). \quad (24)$$

We show that the second term in (24) 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,

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

The first term on right hand side of (25) equals

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

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

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

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

where the last inequality holds for n large enough. The second term on the R.H.S. of (25) equals

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

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

$$\le \frac{\phi_\epsilon^2(0)}{n}\left\{\int_v\int_l\left|f\left(\frac{\sqrt{n}}{c_n}l + x - v\right)K(l)\right|p(v)dvdl\right\}^2$$

$$\le \frac{A}{n}\left\{\int_v\int_l f_0(l,x,v)\left|K(l)\right|p(v)dvdl\right\}^2 \to 0,$$

as required. □

Lemma 2. *Suppose that Assumptions 2.1 and 2.3 hold. Set*

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

Then

$$\lim_{n\to\infty}\sup_{0\le\eta\le1}\left|L^*_{n,\epsilon}(\eta) - \frac{\tau}{n}\sum_{t=1}^{[n\eta]}\phi_\epsilon\left(x_{t-(r\vee s),n}\right)\right| = 0,$$

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

Proof of Lemma 2. In view of Assumption 2.3 and the Lipschitz continuity of ϕ_ϵ the result follows from standard arguments (see Kasparis and Phillips, 2009, for more details). □

Lemma 3. *Suppose that:*

(a) *Assumptions 2.1* and 2.4 *hold.*
(b) $\left|f\left(\frac{\sqrt{n}}{c_n}z + x - v\right)\right| \le f_0(z,x,v)$ *for n large enough with* $\int_v\int_z f_0(z,x,v)K(z)p_{r-s}(v)dzdv < \infty$, *for each* $x \in \mathbb{R}$, *and* $r > s \in \mathbb{N}$.
(c) *K has an integrable Fourier transform.*

Let $q \in \mathbb{N}$ *with* $q \ge 2$. *Then:*

(i)

$$\lim_{n\to\infty}\sup_{0\le\eta\le1}\frac{c_n}{n}\mathbf{E}\left|\sum_{t=1}^{[n\eta]}f\left(\sqrt{n}x_{t-r,n}\right) \times \left\{\mathbf{E}_{t-r-1}K\left[c_n\left(x_{t-s,n}-\frac{x}{\sqrt{n}}\right)\right]\right\}^q\right| = 0,$$

(ii)

$$\lim_{n\to\infty}\sup_{0\le\eta\le1}\frac{c_n}{n} \times \mathbf{E}\underbrace{\left|\sum_{t=1}^{[n\eta]}f\left(\sqrt{n}x_{t-r,n}\right)\prod_{j=1}^{q}\mathbf{E}_{t-r-1}K\left[c_n\left(x_{t-s,n}-\frac{\bar{x}_j}{\sqrt{n}}\right)\right]\right|}_{D_n(\eta)} = 0,$$

(iii)

$$\lim_{n\to\infty}\sup_{0\le\eta\le1}\frac{c_n}{n}\mathbf{E}\left|\sum_{t=1}^{[n\eta]}f\left(\sqrt{n}x_{t-r,n}\right)\prod_{j=1}^{q} \times K\left[c_n\left(x_{t-s,n}-\frac{\bar{x}_j}{\sqrt{n}}\right)\right]\right| = 0.$$

Proof of Lemma 3. We shall prove part (ii). Without loss of generality, assume that $r = 1, s = 0$.

$$\mathbf{E}\left|D_n(\eta)\right| = \frac{c_n}{n}\int_l\left|\sum_{t=1}^{[n\eta]}\prod_{j=1}^{q}\int_{l_j} f\left(\sqrt{n}d_{t-1,0,n}l\right) \times K\left[c_n\left(d_{t-1,0,n}l + \frac{l_j - \bar{x}_j}{\sqrt{n}}\right)\right]p(l_j)dl_j\right|h_{t-1,0,n}(l)\,dl$$

$$\le \frac{1}{n}\sum_{t=1}^{n}\frac{1}{d_{t-1,0,n}}\int_m\prod_{j=2}^{q}\int_{l_1}\int_{l_j} f_0(m,l_1) \times K(m)\,K\left(m + \frac{c_n}{\sqrt{n}}\left(l_j - l_1 + \bar{x}_j - \bar{x}_1\right)\right)p(l_j)dl_jdm$$

$$\leq A\int_m\int_{l_1}\prod_{j=2}^{q}\int_{l_j} f_0(m,l_1)K(m)$$
$$\times K\left(m+\frac{c_n}{\sqrt{n}}\left(l_j-l_1+\bar{x}_j-\bar{x}_1\right)\right)p(l_j)dl_jdm\to 0,$$

as $n\to\infty$. Note that the second inequality above holds for n large enough. Further, the limit above holds by dominated convergence since $K\left(m+\frac{c_n}{\sqrt{n}}(l_j-l_1+\bar{x}_j-\bar{x}_1)\right)\to 0$ almost everywhere with respect to the Lebesgue measure, $\int_{l_1}\int_m f_0(m,l_1)K(m)p(l_1)dl_1dm<\infty$, and $\sup_s K(s)<\infty$.

Parts (i) and (iii) follow using similar arguments to those used above. □

Lemma 4. *Suppose that:*
(a) *Assumptions 2.2 and 2.4 hold.*
(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* $\int_v\int_z f_0(z,x,v)K(z)p_{r-s}(v)dzdv<\infty$, *for each* $x\in\mathbb{R}$ *and* $r>s\in\mathbb{N}$.
Set

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

Then

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

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

$$\mathbf{E}\left|D_n(\eta)\right|=\left(\frac{c_n}{n}\right)\mathbf{E}\sum_{t=1}^{[n\eta]}\int_v\left|f\left(\sqrt{n}x_{t-1,n}\right)\right.$$
$$\left.\times K\left[c_n\left(x_{t-1,n}+\frac{v}{\sqrt{n}}-\frac{x}{\sqrt{n}}\right)\right]\right|p(v)dv$$
$$=\left(\frac{c_n}{n}\right)\mathbf{E}\sum_{t=1}^{[n\eta]}\int_s\int_v\left|f\left(\sqrt{n}d_{t-1,0,n}s\right)\right.$$
$$\left.\times K\left[c_n\left(d_{t-1,0,n}s+\frac{v}{\sqrt{n}}-\frac{x}{\sqrt{n}}\right)\right]\right|p(v)h_{t-1,0,n}(s)\,dvds$$
$$\leq A\int_m\int_v f_0(m,v,x)K(m)p(v)dvdm<\infty,$$

as required. □

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

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

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

Proof of Lemma 5. Set

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

and

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

It can be shown along the lines of Lemma 2 that

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

where τ is defined in Lemma 2. In addition, using arguments similar to those used in the proof of Theorem 1 in Kasparis and Phillips (2009) we get

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

Proof of Proposition A. We shall prove part (ii) of Proposition A. The proof for part (i) is similar. Set,

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

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

(note that $L_{n,\epsilon}(\eta)$ is as in Lemma 1). By Lemmas 1 and 2 we have

$$\lim_{n\to\infty}\sup_{0\leq\eta\leq 1}\left|L_{n,\epsilon}(\eta,x)-\mathbf{E}f\left(x+\sum_{rs}v_i\right)\right.$$
$$\left.\times\int_{-\infty}^{\infty}K(z)\,dz\frac{1}{n}\sum_{t=1}^{[n\eta]}\phi_\epsilon(x_{t-s,n})\right|=0.\tag{26}$$

Further, it can be shown that

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

The asymptotic result of (27) is an extension of Theorem 2.1 of WP to sample covariances. For a detailed proof of (27) see Kasparis and Phillips (2009). The subsequent arguments are similar to those in WP. Consider the term (Eq. (26))

$$\begin{aligned}\lim_{\epsilon\downarrow 0}\frac{1}{n}\sum_{t=1}^{[n\eta]}\phi_\epsilon(x_{t-s,n})&=\lim_{\epsilon\downarrow 0}\int_0^\eta\phi_\epsilon(G(\lambda))d\lambda+o_p(1)\\&=\lim_{\epsilon\downarrow 0}\int_{-\infty}^{\infty}\phi(x)L(\eta,\epsilon x)\,dx+o_p(1)\\&=L(\eta,0)\int_{-\infty}^{\infty}\phi(x)dx+o_{a.s.}(1)\\&=L(\eta,0)+o_{a.s.}(1),\end{aligned}\tag{28}$$

where the first equality above is a consequence of the strong approximation (Assumption 2.2), and holds uniformly in $0\leq\eta\leq 1$. The second equality, given above, follows from the occupation times formula (e.g. WP). In view of (26)–(28), the result follows. □

Appendix B. Proofs of the main results

Proof of Theorem 1. Write

$$\hat{f}(x)-\mathbf{E}f\left(x+\sum_{rs}v_i\right)$$
$$=\frac{\sum\left\{\left[f(x_{t-r})-\mathbf{E}f\left(x+\sum_{rs}v_i\right)\right]K_h(x_{t-s}-x)\right\}}{\sum K_h(x_{t-s}-x)}$$
$$+\frac{\sum K_h(x_{t-s}-x)u_t}{\sum K_h(x_{t-s}-x)}.$$

Then it follows directly from Proposition A that the first term on right hand side above is $o_p(1)$. Further, an application of the martingale CLT[11] together with Proposition A show that the second term is $O_p\left(1/\sqrt{h\sqrt{n}}\right)$, and the result of Eq. (5) follows. □

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

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

$$\sqrt{\sqrt{n}h}\left[\hat{f}(x)-\mathbf{E}f(x-v_t)\right]$$

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

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

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

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

Notice that

$$\mathbf{E}\left|R_n\right| = \mathbf{E}\left|\left(\frac{1}{\sqrt{n}h}\right)^{1/2}\sum_{t=1}^{n}\mathbf{E}_{t-2}\right.$$

$$\left.\times\left[\left(f(x_{t-1})\,\mathbf{E}f(x-v_t)\right)K\left(\frac{x_t}{h}-\frac{x}{h}\right)\right]\right|$$

$$\leq \left(\frac{1}{\sqrt{n}h}\right)^{1/2}\sum_{t=1}^{n}\int_y\left|\int_u\left[f\left(\sqrt{n}d_{t-1,0,n}y\right)-\mathbf{E}f(x-v_t)\right]\right.$$

$$\left.\times K\left(\frac{\sqrt{n}d_{t-1,0,n}y}{h}+\frac{v}{h}-\frac{x}{h}\right)p_1(v)dv\right|h_{t-1,0,n}(y)dy$$

$$= \left(\frac{1}{\sqrt{n}h}\right)^{1/2}\frac{h}{\sqrt{n}}\sum_{t=1}^{n}(d_{t-1,0,n})^{-1}$$

$$\times\int_z\left|\left[\mathbf{E}f(hz+x-v_t)-\mathbf{E}f(x-v_t)\right]K(z)\right|$$

$$\times h_{t-1,0,n}\left(\frac{hz+x-v_t}{\sqrt{n}d_{t-1,0,n}}\right)dz$$

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

Hence,

$$\sqrt{\sqrt{n}h}\left[\hat{f}(x)-\mathbf{E}f(x-v_t)\right]$$

$$= \frac{M_n(x)}{\frac{1}{\sqrt{n}h}\sum_{t=1}^{n}K_h(x_t-x)}+o_p(1). \quad (29)$$

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

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

First, we show that

$$T_{1,n}\xrightarrow{p}\left(\mathbf{Var}f(x-v_t)+\sigma_u^2\right)L_G(1,0)\int_{-\infty}^{\infty}K^2(s)ds. \quad (30)$$

By Lemma 3 we get

$$\begin{aligned}
T_{1,n} =\ & \frac{1}{\sqrt{n}h}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left[f^2(x_{t-1})K_h^2(x_t-x)\right] \\
& -\frac{1}{\sqrt{n}h}\sum_{t=1}^{n}\left[\mathbf{E}_{t-2}f(x_{t-1})K_h(x_t-x)\right]^2 \\
& +\frac{1}{\sqrt{n}h}\tau^2\sum_{t=1}^{n}\mathbf{E}_{t-2}K_h^2(x_t-x) \\
& -\frac{1}{\sqrt{n}h}\tau^2\sum_{t=1}^{n}\left[\mathbf{E}_{t-2}K_h(x_t-x)\right]^2 \\
& +\frac{1}{\sqrt{n}h}\sum_{t=1}^{n}\mathbf{E}_{t-2}K_h^2(x_t-x)u_t^2 \\
& -\frac{1}{\sqrt{n}h}2\tau\sum_{t=1}^{n}f(x_{t-1})\mathbf{E}_{t-2}K_h^2(x_t-x) \\
& +\frac{1}{\sqrt{n}h}2\tau\sum_{t=1}^{n}f(x_{t-1})\left[\mathbf{E}_{t-2}K_h(x_t-x)\right]^2 \\
=\ & \frac{1}{\sqrt{n}h}\sum_{t=1}^{n}\mathbf{E}_{t-2}\left[f^2(x_{t-1})K_h^2(x_t-x)\right] \\
& +\frac{1}{\sqrt{n}h}\tau^2\sum_{t=1}^{n}\mathbf{E}_{t-2}K_h^2(x_t-x) \\
& +\frac{1}{\sqrt{n}h}\sum_{t=1}^{n}K_h^2(x_t-x)\sigma_{t,u}^2 \\
& -\frac{1}{\sqrt{n}h}2\tau\sum_{t=1}^{n}f(x_{t-1})\mathbf{E}_{t-2}K_h^2(x_t-x)+o_p(1) \\
=:\ & T_{2,n}+o_p(1).
\end{aligned}$$

Next, note that

$$\frac{\mathbf{E}}{\sqrt{n}h}\sum_{t=1}^{n}K_h^2(x_t-x)\left|\sigma_{t,u}^2-\sigma_u^2\right|$$

$$\leq \frac{1}{\sqrt{n}h}\sum_{t=1}^{n}\left\{\mathbf{E}K_h^{2q}(x_t-x)\right\}^{1/q}\left\{\mathbf{E}\left|\sigma_{t,u}^2-\sigma_u^2\right|^p\right\}^{1/p}\to 0, \quad (31)$$

as $n\to\infty$ for some $p,q>1$ and $1/p+1/q=1$. Eq. (31) holds by the Toeplitz lemma. To see this, notice that by Assumption 2.3 $\left|\sigma_{t,u}^2-\sigma_u^2\right| = o_{a.s.}(1)$. Hence, $\mathbf{E}\left|\sigma_{t,u}^2-\sigma_u^2\right|^p\to 0$ by dominated convergence, for $\mathbf{E}\left|\sigma_{t,u}^2-\sigma_u^2\right|^p \leq 2^{p-1}\left(\sup_t\mathbf{E}\sigma_{t,u}^{2p}+\sigma_u^{2p}\right)<\infty$,

[11] An explicit martingale CLT for this term is obtained in the proof of Theorem 2.

due to Assumption 2.6. Moreover, using arguments similar to those used in the proof of Lemma 3, we get $\mathbf{E}K_h^{2q}(x_t - x) \to 0$ as $n \to \infty$ and $\sup_n \left(\sqrt{n}h\right)^{-1} \sum_{t=1}^{n} \left\{\mathbf{E}K_h^{2q}(x_t - x)\right\}^{1/q} < \infty$. Therefore, (31) holds (e.g. Hall and Heyde, 1980, p.31).

Hence, by (31), Lemma 1 and Proposition A, we get

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

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

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

Using Lemmas 1, 3(i) and Proposition A we can show that (see Kasparis and Phillips, 2009)

$$T_{3,n} = o_p(1). \tag{32}$$

Finally, in view of Hall and Heyde (1980, Theorem 3.2), (30) and (32) give

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

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

Next, the quadratic variation of $M_n(x)$, is $[M_n] := \left(\sqrt{n}h\right)^{-1} \sum_{t=1}^{n} (\alpha_t + \beta_t + \gamma_t)^2$. The following condition (see Theorem 6.1 and Corollary 6.7 of Jacod and Shiryaev (1986))

$$\sup_n \left(\sqrt{n}h\right)^{-1/2} \max_{0\le t\le n} \mathbf{E} |\alpha_t + \beta_t + \gamma_t| < \infty \tag{33}$$

is sufficient for

$$\left([M_n], \, M_n(x)\right) \xrightarrow{d} \left([M], \, M\right). \tag{34}$$

Using Lemma 4, it can be shown that (33) holds (see Kasparis and Phillips, 2009). Next, consider the predictable quadratic variation of $M_n(x)$, $\langle M_n \rangle := \frac{1}{\sqrt{n}h} \sum_{t=1}^{n} \mathbf{E}_{t-2} (\alpha_t + \beta_t + \gamma_t)^2$. We shall show that

$$\lim_{n\to\infty} \mathbf{E} \left|[M_n] - \langle M_n \rangle\right| = 0. \tag{35}$$

In view of (34), (35) implies

$$\left(\langle M_n \rangle, M_n(x)\right) \xrightarrow{d} \left([M], M\right). \tag{36}$$

According to Hall and Heyde (1980, Theorem 2.23), (32) and tightness of $\langle M_n \rangle$ are sufficient for (35). Let $\lambda > 0$ and notice that

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

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

Finally, it follows from (29) that the NW estimator

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

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

Now by Theorem 1, it can be easily seen that

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

In addition, (36) 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

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

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

It follows directly from Proposition A and Theorem 2 that

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

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

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

as required. □

Proof of Theorem 4. Consider $\hat{t}(x, \hat{\theta})$ for $x \in X_k$. In view of the $O_p\left(diag\left(\sqrt{n}, n\right)\right)$ consistency of $\hat{\theta}$, and using arguments similar to those in the proof of Theorem 2 we have

$$\hat{t}(x, \hat{\theta}) = \left(\frac{\sum_{t=1}^{n} K_h\left(x_{t-s} - x\right)}{\hat{\sigma}_w^2 \int_{-\infty}^{\infty} K(s)^2 ds}\right)^{1/2} \left(\hat{f}(x) - \theta_0 - \theta_1 x\right) + o_p(1)$$

$$= \left(\frac{\sum_{t=1}^{n} K_h\left(x_{t-s} - x\right)}{\hat{\sigma}_w^2 \int_{-\infty}^{\infty} K(s)^2 ds}\right)^{1/2} \frac{M_n(x)}{\sum_{t=1}^{n} K_h\left(x_{t-s} - x\right)} + o_p(1),$$

where $M_n(x)$ is as in the proof of Theorem 2. Let λ be a $(k \times 1)$ non zero vector. By Lemma 3(ii) and (iii), and using arguments similar to those used in the proof of Theorem 2 we get

$$\lambda' \left(M_n(\bar{x}_1), \ldots, M_n(\bar{x}_k)\right)' \xrightarrow{d} \left\{ L_G(1,0) \int_{-\infty}^{\infty} K^2(\lambda) d\lambda \sum_{j=1}^{k} \lambda_j^2 \sigma(\bar{x}_j)^2 \right\}^{1/2} N(0,1),$$

where $\sigma(\cdot)^2$ is given by (6), and L_G is independent of $N(0,1)$. Therefore, in view of the above, the Cramér–Wold device gives:

$$\left(M_n(\bar{x}_1), \ldots, M_n(\bar{x}_k)\right)' \xrightarrow{d} \left\{ L_G(1,0) \int_{-\infty}^{\infty} K^2(\lambda) d\lambda \right\}^{1/2} N(0, \Omega), \tag{37}$$

where $\Omega = diag\left(\sigma(\bar{x}_1)^2, \ldots, \sigma(\bar{x}_k)^2\right)$. Finally, by (37), Theorem 3 and in view the results in Jacod and Shiryaev (1986) on the joint convergence of a martingale (vector) and its quadratic variation, we get

$$\left(\hat{t}(\bar{x}_1, \hat{\theta}), \ldots, \hat{t}(\bar{x}_k, \hat{\theta})\right)' \xrightarrow{d} N(0, I_k).$$

The result of Theorem 4 now follows easily. □

Proof of Theorem 5. The result follows from Proposition A, Theorems 1 and 2, and the limit theory of Park and Phillips (1999, 2001) for locally integrable functions. □

References

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

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.

Berkes, I., Horváth, L., 2006. Convergence of integral functionals of stochastic processes. Econometric Theory 22, 304–322.

Bierens, H.J., 1984. Model specification testing of time series regressions. Journal of Econometrics 26, 323–353.

Bierens, H.J., 1990. A consistent test of functional form. Econometrica 58, 1443–1458.

Cai, Z., Li, Q., Park, J.Y., 2009. Functional coefficient models for nonstationary time series data. Journal of Econometrics 149, 101–113.

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

Cox, D.R., 1961. Tests for separate families of hypotheses. Journal of the Royal Statistical Society, Series B 32, 323–344.

Cox, D.R., 1962. Further results on tests for separate families of Hypotheses. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol 1. University of California Press, Berkeley.

Davidson, R., McKinnon, J.G., 1981. Several tests for model specification in the presence of alternative hypotheses. Econometrica 49, 781–793.

Domowitz, I., White, H., 1982. Misspecified models with dependent observations. Journal of Econometrics 20, 35–58.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2004. The Midas Touch: Mixed Data Sampling Regression Models. In: Scientific Series, Cirano, Montreal.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a risk–return trade-off after all. Journal of Financial Economics 76, 509–548.

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

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: theory. Econometrica 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., Heyde, C.C., 1980. Martingale Limit Theory and Its Application. Academic Press.

Hu, L., Phillips, P.C.B., 2004. Dynamics of the federal funds target rate: a nonstationary discrete choice approach. Journal of Applied Econometrics 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. University of California Press, Berkeley.

Jacod, J., Shiryaev, A.N., 1986. Limit theorems for stochastic processes. Springer-Verlag.

Karlsen, H.A., Mykelbust, T., Tjøstheim, D., 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., 2010. The Bierens test for certain nonstationary models. Journal of Econometrics 158, 221–230.

Kasparis, I., 2011. Functional form misspecification in regressions with a unit root. Econometric Theory 27, 285–311.

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

Ludvigson, S.C., Ng, S., 2007. The empirical risk-return relation: a factor analysis approach. Journal of Financial Economics 83, 171–222.

Li, Q., Racine, J.S., 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., Phillips, P.C.B., 1999. Asymptotics for nonlinear transformations of integrated time series. Econometric Theory 15, 269–298.

Park, J.Y., Phillips, P.C.B., 2000. Nonstationary binary choice. Econometrica 68, 1249–1280.

Park, J.Y., Phillips, P.C.B., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 117–161.

Phillips, P.C.B., 1991. Spectral regression for cointegrated time series. In: Barnett, W., Powell, J., Tauchen, G. (Eds.), Non-Parametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press, Cambridge.

Phillips, P.C.B., 2009. Local limit theory for spurious nonparametric regression. Econometric Theory 25, 1466–1497.

Ramsey, J.B., 1969. Classical model selection through specification error tests. Journal of the Royal Statistical Society, Series B 32, 350–371.

Schienle, M., 2008. Nonparametric nonstationary regression, Ph.D Thesis, University of Mannheim.

Voung, Q., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.

Wang, Q., Phillips, P.C.B., 2009a. Asymptotic theory for local time density estimation and nonparametric cointegrating regression. Econometric Theory 25, 710–738.

Wang, Q., Phillips, P.C.B., 2009b. Structural nonparametric cointegrating regression. Econometrica 77, 1901–1948.

Wang, Q., Phillips, P.C.B., 2011. Asymptotic theory for zero energy density estimation with nonparametric regression applications. Econometric Theory 27, 235–259.

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 of misspecified models. Econometrica 50, 1–25.

Xiao, Z., 2009. Functional-coefficient cointegration models. Journal of Econometrics 152, 81–92.