Latent Variable Nonparametric Cointegrating Regression*

Qiying Wang[†], Peter C. B. Phillips[‡], and Ioannis Kasparis[§]

August 14, 2019

Abstract

This paper studies the asymptotic properties of empirical nonparametric regressions that partially misspecify the relationships between nonstationary variables. In particular, we analyze nonparametric kernel regressions in which a potential nonlinear cointegrating regression is misspecified through the use of a proxy regressor in place of the true regressor. Such models occur in linear and nonlinear regressions where the regressor suffers from measurement error or where the true regressor is a latent or filtered variable as in mixed-data-sampling. The treatment allows for endogenous regressors as the latent variable and proxy variables that cointegrate asymptotically with the true latent variable, including correctly specified as well as misspecified systems, and is therefore intermediate between nonlinear nonparametric cointegrating regression and completely spurious nonparametric nonstationary regressions. The results relate to recent work on dynamic misspecification in nonparametric nonstationary systems and the limit theory accommodates regressor variables with autoregressive roots that are local to unity and whose errors are driven by long memory and short memory innovations, thereby encompassing applications with a wide range of economic and financial time series. Some implications for forecasting under misspecification are also examined.

Keywords: Cointegrating regression, Kernel regression, Latent variable, Local time, MIDAS regression, Misspecification, Nonlinear nonparametric nonstationary regression

JEL classification: C23

---

*The authors thank the CoEditor, Rob Taylor, and two referees for helpful comments on the original manuscript, which have led to many improvements. Wang acknowledges research support from the Australian Research Council and Phillips acknowledges research support from the Kelly Fund, University of Auckland.

[†]School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia, email: qiying@maths.usyd.edu.au

[‡]University of Auckland, Auckland, New Zealand; Yale University, University of Southampton, Singapore Management University, email: peter.phillips@yale.edu.

[§]University of Cyprus, Nicosia, Cyprus, email: kasparis@ucy.ac.cy.

1 Introduction

Kernel regression methods are commonly used in empirical research when theory suggests no obvious model specification or when there is uncertainty about a given parametric specification and tests of that specification against a general nonparametric alternative may be desired. Time series models typically involve additional uncertainties about temporal dependence, nonstationarity, or memory properties of the variables in the regression. Such properties may be assessed by prior estimation or tests, but with the additional consequence of pre-test implications for inference. It is therefore desirable to have econometric methods of estimation and testing that accommodate a wide range of temporal dependence characteristics in the data. Recent research has shown that standard methods of nonparametric kernel regression may be conducted when the regressor has nonstationary characteristics of unknown and unspecified form, including autoregressive unit root, local unit root, or fractional unit root properties (Wang and Phillips, 2009a, 2009b, 2011, 2016 – hereafter WP; Duffy, 2014; Gao and Dong, 2018). In nonstationary cases an important aspect of this work is that the results apply even when the regressor is endogenous, thereby including nonparametric cointegrating regressions.

The present paper extends these results to include nonparametric cointegrating regressions in which the true regressor is a latent variable and a proxy variable is used in the empirical regression in place of the latent variable. Such regressions arise naturally when the true regressor is measured with error and/or when the proxy variable cointegrates asymptotically with the true latent variable. A typical example arises when the true regressor appears in filtered form as in mixed-data-sampling (MIDAS) regressions (Ghysels et al., 2004, 2005) or HP filtering. In this framework, the nonparametric nonstationary model suffers simultaneously from endogeneity of the latent regressor variable and measurement error in the observed proxy regressor. Such a framework is intermediate between correctly specified nonlinear nonparametric cointegrating regressions of the type studied in WP (2009a, 2009b) and completely spurious nonparametric regressions such as those in Phillips (2009). Important special cases of the present framework include nonparametric nonstationary systems in which the regressor is dynamically misspecified, as in the work by Kasparis and Phillips (2012), or similar nonstationary systems in which the true variable is measured with a stationary error, as in Duffy (2014).

The asymptotic results reveal the effects of misspecification, including the asymptotic bias, in nonparametric nonstationary regression. In certain cases such as when the true regression function is convex, the direction of the bias may be determined.[1] In general, when linkages between the observed regressor and the latent variable are

[1] In recent work on linear dynamic systems with nonstationary regressors Duffy and Hendry (2018) analyzed the effects of measurement error and were able to sign these effects in certain special cases.

‘close’, in a sense that will be made precise, an empirical nonparametric regression has a clear interpretation in terms of its pseudo-true value limit as a local average of the true cointegrating regression function. The findings of the paper therefore contribute in several ways to our present understanding of nonparametric cointegrating regression theory. They are particularly helpful in appreciating the combined impact of endogeneity and measurement error in such regressions.

The results of the paper also complement a large literature of recent microeconometric work on nonparametric estimation in the setting of a single endogenous regressor and independently identically distributed (iid) data. In such models, instrumental variable methods and regularization techniques are used to overcome the inconsistency of standard nonparametric estimation by kernel or sieve methods (e.g., Hall and Horowitz, 2005; Horowitz, 2011; Chen and Reiss, 2011). When the explanatory variable suffers from measurement error, these methods are typically inconsistent even in the iid setting. Schennach (2004) studied such problems of nonparametric regression in the presence of measurement error, but without addressing endogeneity of the regressor. Most recently in this literature, Ausumilli and Otsu (2018) have developed wavelet basis methods for dealing simultaneously with an endogenous regressor that is measured with error, showing that the impact of measurement error is to reduce the (already slow) convergence rate of nonparametric IV estimation. The results of the present paper show that in the nonstationary time series setting under endogeneity and measurement error, the standard nonparametric kernel estimator is convergent at the usual rate but to a local average of the nonparametric cointegrating regression function. Like most of the above cited work in nonparametric IV regression in microeconometrics this paper deals with a nonparametric regression of a single endogenous regressor. This limitation is somewhat mitigated in the current time series setting by other features of generality in the model, including the nonstationarity of the observed variables and the temporal dependence properties of the equation errors.

The paper is organized as follows. Section 2 describes the latent variable nonparametric model of cointegration studied here. This model involves dual sources of endogeneity that arise from (i) the use of a proxy variable in the empirical regression, leading to measurement error, as well as (ii) inherent endogeneity in the regressor. Section 3 provides assumptions under which the asymptotics are developed and gives the limiting stochastic processes that are involved in the limit theory. Section 4 provides a general result on the limit behavior of sample nonlinear functionals, which extends many existing results on weak convergence to local time. This result is applied to deliver asymptotic results for sample covariance functionals that appear in latent variable nonparametric cointegrating regressions with a proxy variable regressor. Section 5 gives a weak consistency result and asymptotic distribution theory for such regressions, extending the limit theory in WP (2009b, 2016) for correctly specified cointegrated models to latent and filtered variable cases. Section 6 concludes.

Proofs of the main results are given in Appendix A and supplementary results in Appendix B.

Throughout the paper, we make use of the following notation: for $x = (x_1, ..., x_d)$, $||x|| = \sum_{j=1}^{d} |x_j|$, $a_n \sim b_n$ stands for $0 < \liminf_{n\to\infty} a_n/b_n \leq \limsup_{n\to\infty} a_n/b_n < \infty$. We denote constants by $C, C_1, ...$, which may differ at each appearance.

2 Partially Misspecified Cointegrated Models

We suppose that two nonstationary variables (y_t, z_t) are linked according to the nonparametric regression model

$$y_t = g(z_t) + \eta_t, \ \ t = 1, ..., n \tag{2.1}$$

where g is an unknown function and η_t is a zero mean stationary disturbance whose properties are detailed below. In such models, it is natural to employ kernel regression methods to estimate the function g. When z_t is an integrated, near-integrated or fractionally integrated process, the model (2.1) is now commonly known as a nonlinear nonparametric cointegrating regression. The model may be estimated by kernel regression just as in stationary regression cases.

Importantly, and in contrast to stationary nonparametric regression, such kernel regression is consistent even when the regressor z_t is endogenous. The reason for this robustness to endogeneity in the regressor, as explained in WP (2009b), is that nonstationary regressors such as unit root processes have a wandering character that assists in tracing out the true regression function. In effect, a nonstationary regressor such as z_t serves as its own instrument in nonparametric kernel regresion by delineating the shape of a smooth curve g as y varies over the entire real line by virtue of the recurrence of the limit process corresponding to a standardized version of z_t. This advantage might suggest that such nonparametric regressions might also show some degree of immunity even to measurement error in the regressor. However, Kasparis and Phillips (2012) discovered that this is not so by demonstrating that dynamic misspecification in the timing of the regressor dependence produces inconsistency in nonlinear nonparametric regressions, a result that differs markedly from parametric linear cointegrated regression where the dynamic timing of the regressor has no asymptotic import. Our following analysis reveals the effects of misspecification of a nonlinear regression function in a wide range of nonstationary cases that include measurement error in the regressor.

To fix ideas, we suppose that the regressor z_t in the true model (2.1) is latent and unavailable to the econometrician, whereas another variable x_t is observed and is used in the regression in place of z_t. The fitted nonparametric regression then has the form

$$y_t = \hat{g}(x_t) + \hat{\eta}_t, \tag{2.2}$$

where

$$\hat{g}(x) = \frac{\sum_{t=1}^n y_t K[(x_t - x)/h]}{\sum_{t=1}^n K[(x_t - x)/h]} = \frac{\sum_{t=1}^n \eta_t K[(x_t - x)/h]}{\sum_{t=1}^n K[(x_t - x)/h]} + \frac{\sum_{t=1}^n g(z_t)\, K[(x_t - x)/h]}{\sum_{t=1}^n K[(x_t - x)/h]}$$
$$=: \frac{P_n}{\sum_{t=1}^n K[(x_t - x)/h]} + \frac{S_n}{\sum_{t=1}^n K[(x_t - x)/h]}, \tag{2.3}$$

where $K(x)$ is a non-negative kernel function and $0 < h \equiv h_n \to 0$ is a bandwidth. We study the asymptotic behavior of the fitted nonparametric regression function $\hat{g}(x)$ under certain regularity conditions that prescribe the relationship between x_t and the latent variable z_t and the generating mechanism of (x_t, z_t, η_t). Analysis of (2.3) requires consideration of two components. The asymptotic behavior of the first component follows in much the same way as for a correctly specified system, which is given in WP (2009b). The second component embodies the effects of the misspecification in the numerator S_n. Its asymptotic behavior involves the study of sample covariances of the form

$$S_n = \sum_{t=1}^n g(z_t)\, K[(x_t - x)/h], \tag{2.4}$$

where $h = h_n \to 0$ as the sample size $n \to \infty$. Importantly, (2.4) depends on two nonstationary time series (z_t, x_t), so that the limit theory for the sample function S_n depends on any linkages that are involved in the generating mechanism of these two series. This complicaation leads to considerable technical difficulties in the asymptotic development which are resolved in the paper. We now proceed to analyze sample functions involving such sample covariances of nonlinear functions of related nonstationary variables (z_t, x_t). To begin, we define the conditions on these variables, the regression function in (2.1), and the properties of the errors η_t.

3 Assumptions and Preliminaries

Let $\lambda_i = (\epsilon_i, e_i), i \in \mathbb{Z}$ be a sequence of iid random vectors with $\mathbb{E}\lambda_0 = 0$, $\mathbb{E}||\lambda_0||^2 < \infty$ and $\lim_{|t|\to\infty} |t|^a \, [|\mathbb{E}e^{it\epsilon_0}| + |\mathbb{E}e^{ite_0}|] < \infty$ for some $a > 0$. The variates λ_i form primitive innovations in linear processes that are described below. It should be mentioned that no restriction is imposed on the relation between ϵ_i and e_i. We make use of the following assumptions about the components of (2.1) and (2.2) for the development of the asymptotic theory in our main results.

A1. $x_k = \rho_n x_{k-1} + \xi_k$, where $x_0 = 0$, $\rho_n = 1 - \tau n^{-1}$ for some constant $\tau \geq 0$, and $\xi_k = \sum_{j=0}^{\infty} \phi_j \epsilon_{k-j}$. The coefficients $\phi_k, k \geq 0$, satisfy one of the following conditions:

LM: $\phi_k \sim k^{-\mu}\, l(k), 1/2 < \mu < 1$ and $l(k)$ is a function slowly varying at ∞;

SM: $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$.

A2. (i) $u_k = \sum_{j=0}^{\infty} \psi_j\, \lambda_{k-j}'$ with $\psi_j = (\psi_{1j}, \psi_{2j})$ satisfying $\sum_{j=0}^{\infty}(|\psi_{1j}|+|\psi_{2j}|) < \infty$;

(ii) $w_{nk} = w_n(\lambda_{k+d}, \lambda_{k+d-1}, ...)$, where $d \geq 0$ is an integer, is an arbitrary random array satisfying the bounding inequality $|w_{nk}| \leq \delta_n \sum_{j=-d}^{\infty} \big(|\varphi_{1j}\, \epsilon_{k-j}| + |\varphi_{2j}\, e_{k-j}|\big)$ where the coefficients $\varphi_j = (\varphi_{1j}, \varphi_{2j})$ themselves satisfy $\sum_{j=0}^{\infty}(|\varphi_{1j}| + |\varphi_{2j}|) < \infty$ and the scalar sequence $\delta_n \to 0$, as $n \to \infty$.

A3. $m(x, y)$ is a real function on $\mathbb{R}^2$ satisfying the following conditions:

(i) $\mathbb{E}\big|m(x, u_1)\big|$, where u_t is given in **A2**(i), is bounded and integrable;

(ii) there exist $\delta > 0$, integer $\beta \geq 1$ and a bounded and integrable function $T(x)$ such that, for all $x, y, t \in \mathbb{R}$,

$$|m(x, y) - m(x, t)| \leq T(x)\, |y - t|^{\delta}\, (1 + |y|^{\beta} + |t|^{\beta})\,. \tag{3.5}$$

Assumption **A1** allows for nearly integrated autoregressive transforms of short and long memory linear processes, a general linear process set up that has been widely used in the nonparametric nonstationary time series literature – see WP (2009a, b, 2011, 2016) and Wang (2014, 2015). The initialization $x_0 = 0$ is assumed for convenience but this is not necessary. The main results still hold if $x_0 = o_P(d_n)$, as is clear from analysis of the proofs. The process $x_t, t \geq 1$, is strictly speaking a triangular array due to the dependence of ρ_n on n. But the use of a double index on n will be omitted whenever there is no confusion. Assumption **A2** ensures that u_k is a stationary linear process and w_{nk} is an asymptotically negligible random component that plays a role in bounding one of the components of the misspecification error in our main result. As will be apparent in Section 4, the array component w_{nk} is useful in accommodating fitted coefficients in temporal averaging processes such as MIDAS. By virtue of the boundedness and integrability requirement of $\mathbb{E}|m(x, u_1)|$ given in **A3** (i) there exists a finite constant c_0 such that the function $m(x, c_0)$ is bounded and integrable. This fact will be used in the proofs that follow without further reference.

Assumption **A3** (ii) is a weak condition of uniform continuity for multi-argument functions, which is stronger than local Hölder continuity and weaker than uniform Hölder continuity. It is easy to verify in applications. To give an illustration, let $m(x, y) = K(x)\, f(x+y)$ or $m(x, y) = K(x)\, f(y)$. If $f(.)$ and $K(.)$ satisfy the following condition **A4**, then $m(x, y)$ satisfies (3.5).

A4 (i) There exist $\delta > 0$ and integer $\beta \geq 1$ such that, for all $y, t \in \mathbb{R}$,

$$|f(y) - f(t)| \leq C\, |y - t|^{\delta}\, (1 + |y|^{\beta} + |t|^{\beta}); \tag{3.6}$$

(ii) $\int_{-\infty}^{\infty} K(x)dx = 1$ and $(1+|x|^{1+\beta})K(x)$ is bounded and integrable, where β is given in (3.6).

It is readily seen that functions such as $f(y) = |y|^{\alpha}$ and $1/(1+|y|^{\alpha})$, where $\alpha > 0$, satisfy (3.6).

As will become apparent in what follows, the role of the function $m(x, y)$ in applications is to provide a linkage between the observable time series x_t and the dependent variable y_t in the model (2.1) and fitted regression (2.2). This linkage function allows for potential cointegrating links between y_t and x_t as well as measurement error. Examples of this linkage function are given in Section 4 and include functions of the type $m(x, y) = K(x)\, f(y)$ mentioned above. Corollary 4.1 and its proof employ such linkage functions and show how the limit theory of sample covariance functionals of two nonstationary time series that is given in our Theorem 4.1 can be applied to analyze misspecification components such as those arising from the second term in (2.3).

To complete this section, we define some stochastic processes that appear in the limit theory. In the following, let $d_n^2 = var(\sum_{j=1}^{n} \xi_j)$ and, for $t \geq 0$, define the continuous stochastic processes

$$
\begin{aligned}
Z_t &= W(t) + \tau \int_0^t e^{-\tau(t-s)} W(s) ds, \\
W(t) &= \begin{cases} B_{3/2-\mu}(t), & \text{under } \mathbf{LM}, \\ B_{1/2}(t), & \text{under } \mathbf{SM}, \end{cases}
\end{aligned}
$$

where $B_H(t)\}$ is fractional Brownian motion with Hurst exponent H. In this event, it is well known that Z_t is a fractional Ornstein-Uhlenbeck process, having continuous local time which we denote by $L_Z(t, x)$. We further have the following asymptotic orders

$$
(\mathbb{E}\epsilon_0^2)^{-1}\, d_n^2 \sim \begin{cases} c_\mu\, n^{3-2\mu}\, l^2(n), & \text{under } \mathbf{LM}, \\ \phi^2\, n, & \text{under } \mathbf{SM}, \end{cases}
$$

where $c_\mu = \frac{1}{(1-\mu)(3-2\mu)} \int_0^\infty x^{-\mu}(x+1)^{-\mu} dx$ (e.g., see Giraitis et al., 2012, or Wang et al., 2003). Finally, it is known (Jeganathan, 2008; WP, 2009a) that the functional law $Z_{n\lfloor nt \rfloor} \Rightarrow Z_t$ holds in the Skorohod space $D[0, 1]$ for the standardized element $Z_{nk} = x_k/d_n$ where $k = \lfloor nt \rfloor$ is the integer part of nt.

4 Nonlinear Functionals of Nonstationary Processes

Our first main result concerns the limit behavior of a standardized sample mean functional of a nonlinear function with multiple arguments that involve stationary and nonstationary processes. Such functionals turn out to be very useful in determining asymptotics for sample covariance functionals such as (2.4) that appear in misspecified fitted nonparametric regressions like (2.2). The following result is given in a general form to enhance its usefulness both in the present context and other applications.

Theorem 4.1. *In addition to* ***A1–A3****, suppose that* $\mathbb{E}||\lambda_0||^{2\beta+2} < \infty$ *where* β *is given as in (3.5) of* ***A3****. For any* $c_n \to \infty$ *with* $c_n/n \to 0$ *and* $z \in \mathbb{R}$*, we have*

$$\frac{c_n}{n}\sum_{t=1}^{n} m\big[c_n(Z_{nt}+c_n' z), u_t + w_{nt}\big] \to_D \int_{-\infty}^{\infty} m_1(t)dt\, \widetilde{L}_Z(1,z), \tag{4.7}$$

where $Z_{nk} = x_k/d_n$*,* $m_1(t) = \mathbb{E}m(t,u_1)$ *and*

$$\widetilde{L}_Z(r,z) = \begin{cases} L_Z(r,0), & \text{if } c_n' \to 0, \\ L_Z(r,-z), & \text{if } c_n' = 1. \end{cases}$$

Remark 1. An important element of the proof of Theorem 4.1 involves demonstrating that the sample mean functional can be asymptotically approximated so that the residual difference

$$R_n := \frac{c_n}{n}\sum_{t=1}^{n}\Big[m\big(c_n Z_{nt}, u_t + w_{nt}\big) - m_1(c_n\, Z_{nt})\Big] = o_P(1). \tag{4.8}$$

The rate at which R_n converges to zero heavily depends on the sequence δ_n given in **A2** (ii). Indeed, by letting $m(x,y) = K(x)\,y$, where $w_{nt} = \delta_n\,|\epsilon_t|$ and K satisfies certain smoothness conditions, as in Proposition 7.2 of WP (2016), it is readily seen that

$$R_n = \frac{c_n}{n}\sum_{t=1}^{n} K\big(c_n Z_{nt}\big)\left(u_t + \delta_n|\epsilon_t|\right) = O_P(\delta_n).$$

This rate cannot be improved. The representation (4.8) suggests that the existence of a limit distribution for the sample mean functional in (4.7), and hence that of a suitably standardized version of (2.4), relies on the validity of an asymptotic approximation of the form (4.8). Indeed, it seems unrealistic to consider the asymptotic distribution of sample functionals such as (4.7), at least in this framework, except in cases where the approximating residual element R_n has the property that $w_{nt} \to_P 0$ so that (4.8) holds with some convergence rate. In some specialized cases of the latter situation, the asymptotic behavior of R_n is known and has been considered in WP (2016) and Duffy (2014) for some particular functions $m(x,y)$, specifically $m(x,y) = K(x)y$. It would be interesting to consider more general extensions of this framework in which residual elements such as R_n do not necessarily have the property that $w_{nt} \to_P 0$ and where the approximating component is not necessarily of the form m_1. Such extensions would assist in analyzing spurious nonparametric regressions with nonstationary series, which have been considered in a special case by Phillips (2009), as discussed below. But such an extension seems beyond the scope of our present methods and is therefore left as a challenge for future work.

Remark 2. In spite of this limitation, Theorem 4.1 still provides a very general extension of existing results on convergence of sample functions to quantities that involve scaled local time, as is apparent from the form of (4.7). In previous research, WP (2009a, 2009b) [see also Jeganathan (2008) and Chapter 1 of Wang (2015)] established similar results for the statistic $\frac{c_n}{n}\sum_{t=1}^{n} K\big[c_n(Z_{nt}+c_n' z)\big]\, m(\lambda_t, ..., \lambda_{t-m_0})$, where m_0 is a fixed constant. Moving toward a general formulation of $m(x, y)$, for x_t satisfying **A1** with $\tau = 0$ and the coefficients ϕ_k satisfying the *SM* condition, Duffy (2014) provided a result for $n^{-1/2}\sum_{t=1}^{n} m(x_k, u_k)$ under a strong smoothness condition on $m(x, y)$. Our Theorem 4.1 has the advantage that it allows for nearly integrated long memory as well as short memory linear processes, which are now widely used in the applied literature, in addition to processes that satisfy some general linkage relationships, as will be apparent in our applications below. Furthermore, our formulation of $m(x, y)$ enables easy implementation of Theorem 4.1 to several useful practical applications, as is indicated in the following corollaries.

Corollary 4.1. *Let* $a_k = \gamma_{nk}\, x_k + w_{nk} + u_k$, *where* $\max_{1\le k\le n} |\gamma_{nk} - \gamma_0| \to 0$ *as* $n \to \infty$, *for some* $\gamma_0 \in \mathbb{R}$. *If, in addition to* ***A1****,* ***A2*** *and* ***A4****, suppose that* $\mathbb{E}||\lambda_0||^{2\beta+2} < \infty$ *where* β *is given as in (3.6). Then, for any fixed* $x \in \mathbb{R}$,

$$\frac{d_n}{nh}\sum_{t=1}^{n} f(a_t)\, K\big[(x_t - x)/h\big] \to_D \mathbb{E}f\big(\gamma_0\, x + u_1\big)\, L_Z(1, 0), \tag{4.9}$$

whenever $h := h_n \to 0$ *and* $d_n/nh \to 0$. *When* $h = 1$, *we have*

$$\frac{d_n}{n}\sum_{t=1}^{n} f(a_t)\, K\big(x_t - x\big) \to_D \int_{-\infty}^{\infty} \mathbb{E}f\big(\gamma_0\, t + \gamma_0\, x + u_1\big)\, K(t)dt\, L_Z(1, 0). \tag{4.10}$$

Remark 3. Phillips (2009) gave the first investigation of asymptotics for sample covariance functionals of the form $\sum_{t=1}^{n} f(a_t)\, K\big[(x_t - x)/h\big]$ where both x_t and a_t are $I(1)$ processes. The argument in that work essentially imposed independence between the time series x_t and a_t so that there was no linkage at all between the variables[2], thereby extending the standard spurious linear regression framework (Phillips, 1986; Granger and Newbold, 1974) to nonparametric regression. The limit distribution in that spurious nonparametric regression framework for x_t and a_t differs from Corollary 4.1 in this paper where there is an explicit linkage between the variables. In particular, the situation considered here is that a_t is “close” to being linearly cointegrated with x_t with an asymptotically constant coefficient and a stationary shift subject to an

[2]It is unclear at the moment whether more general versions of a spurious regression result exist under the same setting as Phillips (2009) but without imposing independence between x_t and a_t.

asymptotically negligible error. This framework clearly specializes to include the standard case of a regressor that is measured with error where $\gamma_{nk} = \gamma_0 = 1$ and $w_{nk} = 0$. In the present general framework the latent variable a_t in the nonparametric regression is replaced with a proxy variable x_t whose long run properties relate to those of a_t but via an asymptotically cointegrating linkage in which measurement error is permitted. Moreover, the specification of the proxy variable x_t via the equation $a_k = \gamma_{nk}\, x_k + w_{nk} + u_k$ allows for sample size dependent coefficients γ_{nk}, which accommodates estimated coefficients. In particular, we mention that γ_{nk} can be taken to be an arbitrary random array satifying

$$\max_{1\le k\le n} |\gamma_{nk} - \gamma_0| = o_P(1), \tag{4.11}$$

as $n \to \infty$, for some $\gamma_0 \in \mathbb{R}$. This generalization only involves a minor modification in the proof of Corollary 4.1, the details of which are omitted. Finally, we note that the limit distribution given in (4.10) still involves the local time of the Gaussian process Z_t associated with the weak limit of the process $Z_{n\lfloor nt\rfloor}$ based on standardized versions of the sample observations $x_{\lfloor nt\rfloor}$.

Remark 4. Kasparis and Phillips (2012) investigated the asymptotics of $S_n := \sum_{t=1}^n f(x_{t+d})\, K\big[(x_t - x)/h\big]$ under certain strict conditions on x_t, essentially requiring x_t to be a random walk with iid innovations. As a direct consequence of Theorem 4.1, we may establish similar results under far less restrictive conditions. To illustrate, for some $d \ge 1$, let

$$a_k = \sum_{j=-d}^{d} \gamma_{nk}(j) x_{k+j}, \quad \text{where } \max_{\substack{1\le k\le n\\ -d\le j\le d}} |\gamma_{nk}(j) - \gamma(j)| \to 0. \tag{4.12}$$

As discussed in Remark 5 below, this formulation allows for various fixed data filtering methods (such as MIDAS) that are commonly used in time series regressions. The sample size dependence of the coefficients $\gamma_{nk}(j)$ in (4.12) allows for an even wider range of filters that may adjust to sample size. Note that for any $j \ge 1$,

$$\begin{aligned} x_{k+j} &= \rho_n x_{k+j-1} + \xi_{t+j} = \ldots = \rho_n^j x_k + \sum_{i=1}^{j} \rho_n^{j-i} \xi_{k+i}, \\ x_{k-j} &= \rho_n^{-1} x_{k-j+1} - \rho_n^{-1} \xi_{k-j+1} = \ldots = \rho_n^{-j} x_k - \sum_{i=0}^{j-1} \rho_n^{-j+i} \xi_{k-i}. \end{aligned}$$

We may therefore write

$$\begin{aligned} a_k &= x_k \sum_{j=-d}^{d} \gamma_{nk}(j)\rho_n^j + \sum_{j=1}^{d} \gamma_{nk}(j) \sum_{i=1}^{j} \rho_n^{j-i} \xi_{k+i} - \sum_{j=1}^{d} \gamma_{nk}(-j) \sum_{i=0}^{j-1} \rho_n^{-j+i} \xi_{k-i} \\ &= x_k \sum_{j=-d}^{d} \gamma_{nk}(j)\rho_n^j + \sum_{i=1}^{d} \xi_{k+i} \sum_{j=i}^{d} \gamma_{nk}(j)\rho_n^{j-i} - \sum_{i=1}^{d} \xi_{k-i+1} \sum_{j=i}^{d} \gamma_{nk}(-j)\rho_n^{-j+i-1} \\ &= x_k\, \gamma_{nk}^* + w_{nk} + u_k, \end{aligned} \tag{4.13}$$

where, by letting $\delta_n = \max_{\substack{1\le k\le n\\ -d\le j\le d}} |\gamma_{nk}(j)-\gamma(j)| + \max_{-d\le j\le d} |\rho_n^j - 1|$, we have

$$\begin{aligned}
\gamma_{nk}^* &= \sum_{j=-d}^{d} \gamma_{nk}(j)\rho_n^j = \sum_{j=-d}^{d} \gamma(j) + O(\delta_n),\\
u_k &= \sum_{i=1}^{d} \xi_{k+i} \sum_{j=i}^{d} \gamma(j) - \sum_{i=1}^{d} \xi_{k-i+1} \sum_{j=i}^{d} \gamma(-j),
\end{aligned}$$

and

$$\begin{aligned}
|w_{nk}| &\le \sum_{i=1}^{d} |\xi_{k+i}| \sum_{j=i}^{d} \big|\gamma_{nk}(j)\rho_n^{j-i} - \gamma(j)\big|\\
&\quad + \sum_{i=1}^{d} |\xi_{k-i+1}| \sum_{j=i}^{d} \big|\gamma_{nk}(-j)\rho_n^{-j+i-1} - \gamma(-j)\big|\\
&\le C\delta_n \sum_{i=1}^{d} \big(|\xi_{k+i}| + |\xi_{k-i+1}|\big).
\end{aligned}$$

Now suppose that $\xi_t = \sum_{j=0}^{\infty} \phi_j \epsilon_{t-j}$ with $\sum_{j=0}^{\infty} |\phi_j| < \infty$ and $\phi = \sum_{j=0}^{\infty} \phi_j \neq 0$, so that x_t satisfies **A1** with *SM* memory. Since $\delta_n \to 0$, it is readily seen from (4.13) that w_{nt} and u_t given in (4.12) satisfy **A2**. As an immediate consequence of Corollary 4.1, we have the following result.

Corollary 4.2. *Suppose that **A1** with coefficients satisfying SM holds. Suppose also that **A4** holds and $\mathbb{E}||\lambda_0||^{2\beta+2} < \infty$, where β is given as in (3.6). Then, for any fixed $x \in \mathbb{R}$ and a_t defined in (4.12),*

$$\begin{aligned}
&\frac{1}{\sqrt{n}h} \sum_{t=1}^{n} f(a_t)\, K\big[(x_t - x)/h\big]\\
\to_D\quad & \phi^{-1}(\mathbb{E}\epsilon_0^2)^{-1/2}\, \mathbb{E} f\Big(x \sum_{j=-d}^{d} \gamma(j) + \sum_{i=1}^{d} \xi_i \sum_{j=i}^{d} \gamma(j) - \sum_{i=1}^{d} \xi_{-i+1} \sum_{j=i}^{d} \gamma(-j)\Big)\\
&\times L_Z(1,0), \qquad (4.14)
\end{aligned}$$

whenever $h \to 0$ and $n^2 h \to \infty$. In particular, we have

$$\frac{1}{\sqrt{n}h} \sum_{t=d}^{n} f(x_t) K\left(\frac{x_{t-d}-x}{h}\right) \to_D \frac{1}{\phi(\mathbb{E}\epsilon_0^2)^{1/2}}\, \mathbb{E}\left\{f\big(x + \sum_{j=1}^{d} \xi_j\big)\right\} L_Z(1,0),$$
$$\frac{1}{\sqrt{n}h} \sum_{t=d}^{n} f(x_{t-d}) K\left(\frac{x_{t-d}-x}{h}\right) \to_D \frac{1}{\phi(\mathbb{E}\epsilon_0^2)^{1/2}} \mathbb{E}\left\{f\big(x - \sum_{j=1}^{d} \xi_j\big)\right\} L_Z(1,0).$$

Using (4.10), results for $h = 1$ can be derived similarly. The details are omitted.

Remark 5. It is evident from Corollary 4.2 that the kernel weighted sample moment function fails to deliver a consistent kernel regression estimator for the function $f(x)$ but leads instead to a local average pseudo-true-value function (PTVF) about x. Apart from dynamically misspecified regression, the formulation of the observed variable (4.12) is relevant in many other situations that arise in applied econometric work. For instance, the inconsistency and PTVF depicted in Corollary 4.2 may occur when there is temporal aggregation of the data. If the covariate a_t is sampled more frequently than the dependent variable, Ghysels, Santa-Vlara and Valkanov (2004, 2006) propose the following mixed-data-sampling (MIDAS) form of weighted average specification $a_t = \sum_{j=0}^{d} \gamma(j) x_{t+j}$ where $0 \leq \gamma(j) \leq 1$ and $\sum_{j=0}^{d} \gamma(j) = 1$. In practical work, preliminary estimates $\hat{\gamma}_n(j)$ are typically required for the MIDAS coefficients $\gamma(j)$, leading to the empirically filtered variable $\hat{z}_t = \sum_{j=0}^{d} \hat{\gamma}_{j,n} x_{t+j}$ with $\max_{0 \leq j \leq d} |\hat{\gamma}_n(j) - \gamma(j)| = o_P(1)$. It follows that if use of a MIDAS structure misspecifies the generating mechanism, then the kernel regression estimator is inconsistent. The present formulation of (4.11) allows for this type of filtering with estimated coefficients $\hat{\gamma}_n(j)$. A similar linear pre-filtering of the regressors is considered by Bollerslev et al (2013) in a different context. In particular, that work proposes a predictive regression for stock returns of the form $y_t = f\left(\sum_{j=1}^{d} \gamma(j) x_{t-j}\right) + \eta_t$ involving the temporal aggregate $\sum_{j=1}^{d} \gamma(j) x_{t-j}$ as regressor. The purpose of the linear filtering in this case is to adjust any persistence in the predictor (x_t) to conform with the temporal properties of stock returns (y_t) thereby balancing the predictive regression equation. Again preliminary estimators for $\gamma(j)$ are typically required in practice and are covered by the present formulation. Further extensions are possible by allowing for two-sided temporal average specifications in which the parameter $d = d_n \to \infty$ as $n \to \infty$. These are particularly relevant in applied time series work that use trend detection filtering methods such as the Whittaker filter (Whittaker, 1923; Phillips, 2010; Phillips and Jin, 2015), the Hodrick Prescott filter (Leser, 1961; Hodrick and Prescott, 1997), and bandpass filters (Baxter and King, 1999). Conditions for validity in such cases will be considered in subsequent work.

5 Applications

This section develops a nonparametric regression application of our limit theory for sample covariance functionals of nonstationary time series. Except where mentioned explicitly, the notation used here is the same as in Sections 2-3.

Suppose that the time series (y_t, x_t) are observed but the real data generating process has the form

$$y_t = g(z_t) + \eta_t, \quad x_t = \alpha_{nt}\, z_t + w_{nt} + u_t, \tag{5.15}$$

where $g(x)$ is an unknown regression function, x_t, u_t and w_{nt} satisfy **A1** and **A2**, and η_t is an error process defined by

$$\eta_t = \sum_{j=0}^{\infty} \theta_j \, \lambda'_{t-j}$$

with $\theta_j = (\theta_{1j}, \theta_{2j})$ satisfying $\sum_{j=0}^{\infty} j^{1/4}(|\theta_{1j}| + |\theta_{2j}|) < \infty$. We further assume that

$$\alpha_n := \max_{1 \le k \le n} |\alpha_{nk}^{-1} - 1| \to 0, \quad \text{as } n \to \infty,$$

This formulation involves a nonparametric regression model with latent endogenous regressor variable z_t that is observed with error via a proxy variable x_t that is asymptotically linked through an approximate cointegrating relation to z_t. The resulting fitted regression is a partially misspecified nonlinear nonparametric cointegrating regression. In particular, the previous section showed that for certain choices of α_{nt}, w_{nt} and u_t the model is dynamically misspecified. Further, for $\epsilon_i \perp e_i$ and $\psi_{1j} = b_{1j} = \theta_{2j} = 0$ we have classical measurement error, i.e. the measurement error $u_t + w_{nt}$ is exogenous and independent of x_t.

Since data on only (y_t, x_t) is observed, standard kernel estimation of the function g leads to

$$\hat{g}(x) = \frac{\sum y_t K[(x_t - x)/h]}{\sum K[(x_t - x)/h]},$$

where K is a non-negative kernel function and the bandwidth $h := h_n \to 0$. The following result shows the limit behavior of $\hat{g}(x)$.

Theorem 5.1. *If, when $f(x)$ is replaced by $g(x)$, **A4** holds and $\mathbb{E}||\lambda_0||^{2\beta+2} < \infty$ where β is given in **A4**, then*

$$\hat{g}(x) \to_P g_1(x) := \mathbb{E}g\big(x - u_1\big), \tag{5.16}$$

for any fixed x and $h \to 0$ satisfying $d_n/nh \to 0$.

Remark 6. Since $g_1(x) \neq g(x)$ in general, the nonparametric estimate $\hat{g}(x)$ will usually produce an inconsistent estimate of $g(x)$. The nature of the asymptotic bias $g_1(x) - g(x)$ depends on the degree of nonlinearity of g in conjunction with the shape and support of the density of the measurement error. If $g(x)$ is convex (concave) then by Jensen's inequality $g_1(x) \geq g(x)$ ($g_1(x) \leq g(x)$), indicating a positive (negative) bias in this misspecified nonparametric regression. The actual deviation of $g_1(x)$ from $g(x)$ can be calculated by application of the mean value theorem for smooth functions[3]. If $g(x)$ is monotonic, then so too is $g_1(x)$. Figure 1 shows the effect of misspecification in the case of the cubic function $g(x) = x^3$ when $u_1 \sim \mathcal{N}(0,1)$. When g is linear, $g_1(x) = g(x)$ and $\hat{g}(x)$ is consistent, just as in linear cointegrating regression with stationary measurement error.

[3] Taylor expansion yields $g(x+u_1) = g(x) + g'(x)u_1 + \frac{1}{2}g''(\xi)u_1^2$ for some ξ between x and $x+u_1$, and the claim follows from $g''(\xi) \geq 0$, $\mathbb{E}u_1 = 0$ and $\mathbb{E}u_1^2 > 0$.

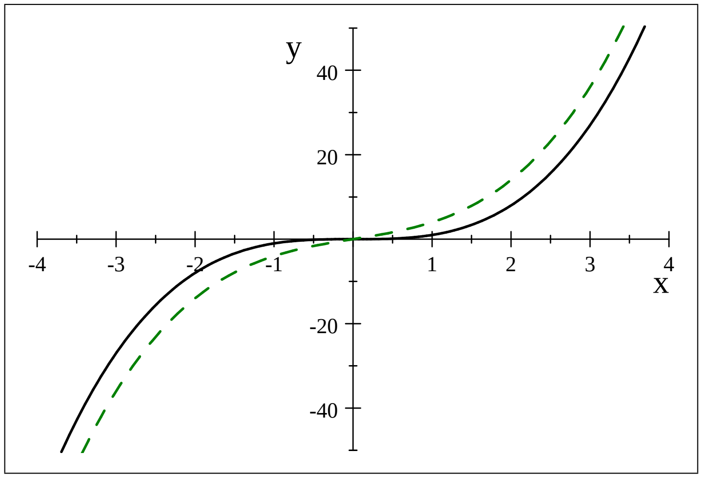

Figure 1: Nonparametric function $g(x) = x^3$ (solid black) and pseudo-true-value function (PTVF) $g_1(x) = \int (x-u)^3 \varphi(u)\, du = x^3 + 3x$ (dashed green) with standard Gaussian $\varphi(u)$.

Remark 7. The pseudo true regression function $g_1(x) = \mathbb{E}g\big(x - u_1\big)$ averages g around its local value at x with the weighting delivered by the density of the measurement error u in (5.15). By standard theory, a pseudo true value minimizes the distance (with respect to some asymptotic L_2-norm) between the fitted regression function and some (possibly) unknown true regression function (e.g. White, 1981). In parametric cases, the pseudo true value carries information about parameters of interest which is exploited by methods such as indirect inference which lead to consistent estimation of the true parameters under certain conditions. In the nonparametric case, information about the true function g can be obtained by deconvolution methods when the measurement error density is known or can be estimated. Such methods are used in nonparametric instrumental variables estimation of nonlinear nonparametric microeconometric models, as discussed earlier. Information about g can be obtained in other cases and is usefully employed in prediction. For example, suppose that $\{y_t, x_t\}_{t=1}^n$ are observed with y_t determined by the predictive regression $y_t = g(z_{t-1}) + \eta_t$ with z_t being a latent nonstationary variable with $z_t = x_t + u_t$, $x_t \perp u_t$ (giving a classical errors-in-variables formulation) and a martingale difference equation error, η_t, with respect to some filtration $\mathcal{F}_t$ to which (x_t, u_t) are adapted. Suppose that we are interested in obtaining the mean out-of-sample forecast $\mathbb{E}\left(y_{n+1}|\mathcal{F}_n\right) = g(z_n)$ utilizing $\{x_t\}$ as a proxy for the unobservable z_t via nonparametric estimation. As demonstrated in Theorem 5.1, under certain conditions the kernel regression estimator approximates $g_1(x) = \mathbb{E}g(x - u_1)$. Interestingly, the pseudo-true regression function $g_1(x)$ forms a better predictor for $\mathbb{E}\left(y_{n+1}|\mathcal{F}_n\right)$ than the true function $g(x)$, when the proxy variable $\{x_t\}$ is employed and used in prediction. In particular,

$$\mathbb{E}\left\{\left[\mathbb{E}\left(y_{n+1}|\mathcal{F}_n\right) - g_1(x_n)\right]\right\} = \mathbb{E}\left[g(z_n) - g_1(x_n)\right]$$

$$= \mathbb{E}\left\{\mathbb{E}\left[g(x_n + u_n) - g_1(x_n)|x_n\right]\right\} \overset{x_t \perp u_t}{=} \mathbb{E}\left[g_1(x_n) - g_1(x_n)\right] = 0.$$

Therefore, $g_1(x_n)$ is an unbiased estimator for the mean of $\mathbb{E}\left(y_{n+1}|\mathcal{F}_n\right)$. In fact, the deviation of the pseudo-true regression function $g_1(x)$ from the true function $g(x)$ on average compensates for fact that $\{z_t\}$ is only partially observable and values of the proxy variable $\{x_t\}$ are used instead in prediction. On the other hand if we were to utilize the true regression function $g(x)$ together with the proxy $\{x_t\}$, we would have

$$\mathbb{E}\left\{\left[\mathbb{E}\left(y_{n+1}|\mathcal{F}_n\right) - g(x_n)\right]\right\} = \mathbb{E}\left[g(x_n + u_n) - g(x_n)\right],$$

which is non zero in general.

Under somewhat stronger conditions on $g(x)$, u_t, α_{nk} and δ_n, it is possible to establish the asymptotic distribution of $R_n(x) := \hat{g}(x) - g_1(x)$. Theorem 5.2 below provides this limit theory under the following assumption.

A5. (i) For some integer $\beta \geq 1$, when $y, t \in \mathbb{R}$, we have

$$|g(y) - g(t)| \leq C\,|y - t|\,(1 + |y|^\beta + |t|^\beta);$$

(ii) $\int_{-\infty}^{\infty} K(x)dx = 1$ and $K(x)$ has a finite compact support;

(iii) x_t is defined as in **A1** and $u_k = \sum_{j=0}^{\infty} \psi_j\, \lambda'_{k-j}$, where the coefficents $\psi_j = (\psi_{1j}, \psi_{2j})$ satisfy that $\sum_{j=0}^{\infty} ||\psi_j|| < \infty$ and $n\sum_{j=\nu_n}^{\infty} ||\psi_j||^2 = o(1)$ with $\nu_n = (n/d_n)^\delta$ for some $\delta < 1/3$;

(iv) $nh/d_n \to \infty$, $nh^3/d_n \to 0$ and $\alpha_n + \delta_n = O(h)$;

(v) $E||\lambda_1||^{2(\beta+1)} < \infty$.

Theorem 5.2. *Under* ***A5****, for any fixed* x*, we have*

$$\Big(\sum_{k=1}^{n} K[(x_k - x)/h]\Big)^{1/2}\big[\hat{g}(x) - g_1(x)\big] \to_D \Lambda \times \mathcal{N}(0,1), \qquad (5.17)$$

where $\Lambda^2 = \mathbb{E}\big[\eta_1 + g(x - u_1) - g_1(x)\big]^2 \int_{-\infty}^{\infty} K^2(y)dy$.

Remark 8. WP (2016) established a version of (5.17) without investigating the effect of misspecification in the model, thereby imposing the conditions that $u_t = w_{nt} = 0$, and $\alpha_{nk} = 1$ on the present framework. Duffy (2014) allowed for $u_t \neq 0$, while still imposing $w_{nt} = 0$ and $\alpha_{nk} = 1$, but requiring $\tau = 0$ for the time series x_t defined in **A1** and requiring the coefficients ϕ_k to have *SM* memory, thereby restricting attention to $I\,(1)$ time series. Our Theorem 5.2 provides a general result for nonparametric regression under misspecification that allows for nearly integrated short and long memory latent variables that are observed with error. This result substantially extends the existing literature on nonparametric nonstationary regression to a latent variable framework that covers many potential time series applications in econometric work in which measurement error effects may be expected.

Remark 9. Endogeneity does not affect the probability limit or the limit distribution expressions of the kernel estimators when there are measurement errors (c.f. Theorems 5.1-5.2). This finding is analogous to that of Wang and Phillips (2009b) who consider structural regressions without measurement errors. Nonetheless, under endogeneity the limit variance of the kernel estimator may be larger. According to Theorem 5.2 the limit variance under measurement errors is of the form $\Lambda^2 = \mathbb{E}\left[\eta_1 + g(x - u_1) - g_1(x)\right]^2$. It is apparent that this expression will be larger if η_1 and $g(x - u_1)$ are positively correlated, which may be so for certain functions g when the measurement error is endogenous.

6 Conclusion

The present framework focuses on latent variable nonparametric cointegrating regressions which are partially misspecified through the presence of measurement error or the use of proxy variables in the regression. The limit theory reveals that such regressions lead to bias in estimation yet may be interpreted as estimating locally weighted averages of the true regression function and are amenable to inference. The latent variable framework does not include fully spurious nonparametric regresssion systems of the type studied in Phillips (2009). Extensions to such systems are of interest not only from the perspective of completing the limit theory for linear spurious regression (Phillips, 1986) to include nonlinear nonparametric regression but also because the present results seem close to the limit of what is possible for partially misspecified regressions arising from latent variable measurement error. It is therefore of interest to understand how gross misspecification, as distinct from partial misspecification due to measurement error, affects such regressions with randomly trending variables.

Nonparametric regressions offer empirical researchers considerably more flexibility than linear regressions in establishing 'empirical relationships'. Given the well-known tendency of trending variables to produce plausible regression findings in the absence of an underlying relationship between the variables, it is important to understand the implications of conducting nonparametric regressions with such variables when the linkages between the variables are no longer as 'close' as the partially misspecified linkages studied in the present paper. What the present paper does show is that when there are 'close' linkages between the observed regressor and the latent or filtered variables used in practice, an empirical nonparametric regression has a clear interpretation in terms of a local average relationship of the true regression function. In this sense, there is useful interpretable information that can be recovered from the pseudo-true value in nonparametric nonstationary regression with latent unobserved variables. In addition, when proxy variables are used in estimation and prediction, it is the property of the forecast given the proxy variable that is most relevant in practice and this forecast, as we have shown, is unbiased.

7 Appendix A: Proofs of the main results

We first introduce two lemmas that play a key role in the proofs of our main results. Notation is the same as in previous sections except where explicitly mentioned.

Lemma 7.1. *Let $p(s, s_1, ..., s_l)$ be a real function of its components and $t_1, ..., t_l \in \mathbb{Z}$, where $l \geq 0$. There exists an $m_0 > 0$ such that the following results hold.*

(i) For any $h > 0$ and $k \geq 2l + m_0$, we have

$$\mathbb{E}|\, p(x_k/h, \lambda_{t_1}, ..., \lambda_{t_l})| \leq \frac{C\, h}{d_k} \int_{-\infty}^{\infty} \mathbb{E}|p(t, \lambda_1, ..., \lambda_l)|dt. \tag{7.18}$$

(ii) For any $h > 0$, $k - j \geq 2l + m_0$ and $j + 1 \leq t_1, ..., t_l \leq k$, we have

$$\mathbb{E}\big[\big|p(x_k/h, \lambda_{t_1}, ..., \lambda_{t_l})\big| \mid \mathcal{F}_j\big] \leq \frac{C\, h}{d_{k-j}} \int_{-\infty}^{\infty} \mathbb{E}|p(t, \lambda_1, ..., \lambda_l)|dt. \tag{7.19}$$

If in addition $\mathbb{E}p(t, \lambda_1, ..., \lambda_l) = 0$ for any $t \in \mathbb{R}$, then

$$\begin{aligned} &\big|\mathbb{E}\left[p(x_k/h, \lambda_{t_1}, ..., \lambda_{t_l}) \mid \mathcal{F}_j\right]\big| \\ \leq\ & \frac{Ch \sum_{j=0}^{k-\min\{t_1,...,t_l\}} |\phi_j|}{d_{k-j}^2} \int_{-\infty}^{\infty} \mathbb{E}\Big\{|p(y,\ \lambda_1, ..., \lambda_l)| \sum_{j=1}^{l} |\epsilon_j|\Big\}\, dy. \end{aligned} \tag{7.20}$$

Proof. The proof of Lemma 7.1 is similar to Lemma 2.1 of Wang (2015). See, also, Lemma 8.1 of Wang and Phillips (2016). A proof of (7.20) is given in Appendix B for convenience. $\square$

For any $c_{nk} = (c_{n,1k}, c_{n,2k})$ satisfying $\sup_{n\geq 1} \sum_{k=0}^{\infty}(|c_{n,1k}| + |c_{n,2k}|) < \infty$, let $L_{nt} = \sum_{k=0}^{\infty} c_{nk}\lambda'_{t-k}$. Note that, for $\gamma \geq 1$,

$$\begin{aligned} |L_{nt}|^{\gamma} &\leq \Big(\sum_{k=0}^{\infty} |c_{nk}\lambda'_{t-k}|\Big)^{\gamma} \\ &\leq \Big[\sum_{k=0}^{\infty}(|c_{n,1k}| + |c_{n,2k}|)\Big]^{\gamma-1} \sum_{k=0}^{\infty}(|c_{n,1k}| + |c_{n,2k}|)\, ||\lambda_{t-k}||^{\gamma} \\ &\leq C \sum_{k=0}^{\infty}(|c_{n,1k}| + |c_{n,2k}|)\, ||\lambda_{t-k}||^{\gamma}, \end{aligned} \tag{7.21}$$

by Hölder's inequality. A simple application of (7.18) yields that, if $\mathbb{E}||\lambda_1||^\gamma < \infty$ and $p(x)$ is a positive bounded integrable function, then

$$\begin{aligned} &\sum_{t=1}^n \mathbb{E}\big[(1+|L_{nt}|^\gamma)p(x_t/h)\big] \\ \le\ & \sum_{t=1}^n \mathbb{E}p(x_t/h) + C\sum_{k=0}^\infty(|c_{n,1k}|+|c_{n,2k}|)\sum_{t=1}^n \mathbb{E}\big[||\lambda_{t-k}||^\gamma\, p(x_t/h)\big] \\ \le\ & C_1\, nh/d_n, \end{aligned} \tag{7.22}$$

for any $h>0$, i.e., $\sum_{t=1}^n(1+|L_{nt}|^\gamma)p(x_t/h) = O_P(nh/d_n)$. Similarly, if $\mathbb{E}||\lambda_1||^{\gamma+\gamma_1}<\infty$, then

$$\sum_{t=1}^n \mathbb{E}\big[||\lambda_{t-j}||^{\gamma_1}(1+|L_{nt}|^\gamma)p(x_t/h)\big] \ \le\ C_1\, nh/d_n, \tag{7.23}$$

for any $h>0$ and uniformly for $j\in\mathbb{Z}$. Results (7.22) and (7.23) will be directly used in the following proofs without further explanation.

Lemma 7.2. *If, in addition to* ***A5****,* $\mathbb{E}g(u_1)=0$*, then*

$$\begin{aligned} &\frac{d_n}{nh}\sum_{t=1}^n K\big[(x_t-x)/h\big],\ \Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n\big[\eta_t+g(u_t)\big]\,K\big[(x_t-x)/h\big] \\ \to_D\ & \big(L_Z(1,0),\ a_0 N\times L_Z(1,0)^{1/2}\big), \end{aligned}\tag{7.24}$$

where $a_0^2=\mathbb{E}\big[\eta_1+g(u_1)\big]^2\int_{-\infty}^\infty K^2(x)dx$ *and* $N\sim\mathcal{N}(0,1)$ *independent of* $L_Z(1,0)$*.*

Proof. For any $m>0$, let $u_{m,t}=\sum_{j=0}^m\psi_j\lambda_{t-j}'$ and $\eta_{m,t}=\sum_{j=0}^m\theta_j\lambda_{t-j}'$. By noting $\mathbb{E}g(u_t)=\mathbb{E}g(u_1)=0$, we may write

$$\begin{aligned} &\sum_{t=1}^n\big[\eta_t+g(u_t)\big]\,K\big[(x_t-x)/h\big] \\ =\ &\sum_{t=1}^n\big[\eta_{m,t}+g(u_{m,t})-\mathbb{E}g(u_{m,t})\big]\,K\big[(x_t-x)/h\big]+R_{n1}+R_{n2} \end{aligned}$$

where $R_{n1}=\sum_{t=1}^n\big(\eta_t-\eta_{m,t}\big)\,K\big[(x_t-x)/h\big]$ and

$$R_{n2}=\sum_{t=1}^n\big\{g(u_t)-g(u_{m,t})-\mathbb{E}\big[g(u_t)-g(u_{m,t})\big]\big\}\,K\big[(x_t-x)/h\big].$$

For any fixed $m>0$, it follows from Wang and Phillips (2009a, 2009b) that

$$\begin{aligned} &\Big(\frac{d_n}{nh}\sum_{t=1}^n K\big[(x_t-x)/h\big],\ \Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n\big[\eta_{m,t}+g(u_{m,t})-\mathbb{E}g(u_{m,t})\big]\,K\big[(x_t-x)/h\big]\Big) \\ \to_D\ &\big(L_Z(1,0),a_mN\times L_Z(1,0)^{1/2}\big), \end{aligned}\tag{7.25}$$

where $a_m^2 = \mathbb{E}\big[\eta_{m,1} + g(u_{m,1}) - \mathbb{E}g(u_{m,1})\big]^2 \int_{-\infty}^{\infty} K^2(x)dx$. Result (7.24) will follow if we prove

$$a_m^2 \to a_0^2, \tag{7.26}$$

as $m \to \infty$, and, for $i = 1$ and 2,

$$R_{ni} = o_p\big[(nh/d_n)^{1/2}\big], \tag{7.27}$$

as $n \to \infty$ first and then $m \to \infty$.

We only prove (7.27) for $i = 2$. The proof of (7.27) for $i = 1$ is similar [see, (8.13) of Wang and Phillips (2016), for instance]. Due to **A5** and $\mathbb{E}g(u_1) = 0$, the proof of (7.26) is trivial. We omit the details. To prove (7.27) for $i = 2$, we write

$$R_{n2} = R_{n3} + R_{n4}, \tag{7.28}$$

where, for ν_n given in **A5** (iii),

$$\begin{aligned} R_{n3} &= \sum_{t=1}^{n} \big\{g(u_t) - g(u_{\nu_n,t}) - \mathbb{E}\big[g(u_t) - g(u_{\nu_n,t})\big]\big\} K\big[(x_t - x)/h\big] \\ R_{n4} &= \sum_{t=1}^{n} \big\{g(u_{\nu_n,t}) - g(u_{m,t}) - \mathbb{E}\big[g(u_{\nu_n,t}) - g(u_{m,t})\big]\big\} K\big[(x_t - x)/h\big] \end{aligned}$$

Note that $u_t - u_{\nu_n,t} = \sum_{j=\nu_n+1}^{\infty} \psi_j \lambda'_{t-j}$. By using **A5** (i) and (7.22) with $\gamma = 2\beta$, it follows from the Hölder's inequality that

$$\begin{aligned} |R_{n3}| &\leq C \sum_{t=1}^{n} \big|u_t - u_{\nu_n,t}\big| \, (1 + |u_t|^\beta + |u_{\nu_n,t}|^\beta) K\big[(x_t - x)/h\big] \\ &\qquad + C \sum_{t=1}^{n} \mathbb{E}\big\{\big|u_t - u_{\nu_n,t}\big| \, (1 + |u_t|^\beta + |u_{\nu_n,t}|^\beta)\big\} K\big[(x_t - x)/h\big] \\ &\leq C \Big(\sum_{t=1}^{n} \big|u_t - u_{\nu_n,t}|^2\Big)^{1/2} \Big\{\sum_{t=1}^{n} (1 + |u_t|^{2\beta} + |u_{\nu_n,t}|^{2\beta}) K^2\big[(x_t - x)/h\big]\Big\}^{1/2} \\ &\qquad + C \, (\sum_{j=\nu_n}^{\infty} ||\psi_j||^2)^{1/2} \sum_{t=1}^{n} K\big[(x_t - x)/h\big] \\ &= O_P\big\{(n \sum_{j=\nu_n}^{\infty} ||\psi_j||^2)^{1/2} \, \big[(nh/d_n)^{1/2} + n^{-1/2}(nh/d_n)\big\} \\ &= o_P\big[(nh/d_n)^{1/2}\big]. \end{aligned}$$

Taking this into (7.28), to prove (7.27) for $i = 2$ it suffices to show that

$$R_{n4} = o_P\big[(nh/d_n)^{1/2}\big], \tag{7.29}$$

as $n \to \infty$ first and then $m \to \infty$.

To prove (7.29), let $g_t = g(u_{\nu_n,t}) - g(u_{m,t})$, $\widetilde{g}_t = g_t - \mathbb{E}g_t$ and $\nu_{1n} = 2\nu_n + m_0$, where m_0 is defined as in Lemma 7.1. By setting $\sum_{t=i}^{j} = 0$ if $i > j$, we may write

$$\begin{aligned}
\mathbb{E}R_{n4}^2 &= \sum_{s,t=1}^{n} \mathbb{E}\Big\{\widetilde{g}_s\,\widetilde{g}_t\,K\big[(x_s - x)/h\big]K\big[(x_t - x)/h\big]\Big\} \\
&\le 2\sum_{s=1}^{n}\Big(\sum_{t=s}^{s+m_0+1} + \sum_{t=s+m_0+2}^{s+\nu_{1n}} + \sum_{t=s+\nu_{1n}+1}^{n}\Big)\Big|\mathbb{E}\Big\{\cdots\Big\}\Big| \\
&=: \Delta_{1n} + \Delta_{2n} + \Delta_{3n}.
\end{aligned} \tag{7.30}$$

Hölder's inequality implies that

$$\begin{aligned}
&|\Delta_{1n}| \\
\le\ & \sum_{s=1}^{n}\sum_{t=s}^{s+m_0+1}\Big(\mathbb{E}\Big\{|\widetilde{g}_s|^2\,K^2\big[(x_s - x)/h\big]\Big\}\Big)^{1/2}\Big(\mathbb{E}\Big\{|\widetilde{g}_t|^2\,K^2\big[(x_t - x)/h\big]\Big\}\Big)^{1/2} \\
\le\ & C\,m_0\sum_{s=1}^{n}\mathbb{E}\Big\{|\widetilde{g}_s|^2\,K\big[(x_s - x)/h\big]\Big\}.
\end{aligned}$$

As in the estimation of R_{n3}, by using

$$\begin{aligned}
|g_t|^2 &\le C\Big[\sum_{j=m}^{\infty}|\psi_j\lambda'_{t-j}|\,(1 + |u_{\nu_n,t}|^{\beta} + |u_{m,t}|^{\beta})\Big]^2 \\
&\le C_1\sum_{j=m}^{\infty}(|\psi_{1j}| + |\psi_{2j}|)\,||\lambda_{t-j}||^2\,(1 + |u_{\nu_n,t}|^{2\beta} + |u_{m,t}|^{2\beta})
\end{aligned}$$

due to **A5**(i) and (7.21), it follows from (7.23) with $\gamma + \gamma_1 = 2\beta + 2$ that, whenever $\nu_n \ge m$,

$$\begin{aligned}
|\Delta_{1n}| &\le C\,m_0\sum_{t=\nu_{1n}+1}^{n}\mathbb{E}\Big\{\big[|g_t|^2 + \big(\mathbb{E}|g_t|\big)^2\big]\,K\big[(x_t - x)/h\big]\Big\} \\
&\le C\,m_0\sum_{j=m}^{\infty}(|\psi_{1j}| + |\psi_{2j}|) \\
&\qquad \sum_{t=\nu_{1n}+1}^{n}\mathbb{E}\big\{\big(1 + ||\lambda_{t-j}||^2\big)\,(1 + |u_{\nu_n,t}|^{2\beta} + |u_{m,t}|^{2\beta})K\big[(x_t - x)/h\big]\big\} \\
&\le C_1\sum_{j=m}^{\infty}(|\psi_{1j}| + |\psi_{2j}|)\,(nh/d_n) = o(nh/d_n),
\end{aligned} \tag{7.31}$$

as $n \to \infty$ first and then $m \to \infty$.

We next consider Δ_{3n}. As $\mathbb{E}\widetilde{g}_t = 0$ and $\widetilde{g}_t$ depends only on $\lambda_{t-\nu_n}, ..., \lambda_t$, it follows from (7.20) with $k = t, j = s$ and $l = \nu_n$ that, for $t-s \ge 2\nu_n+m_0$ (hence $t-\nu_n \ge s+1$),

$$\begin{aligned}
\Big|\mathbb{E}\big[\widetilde{g}_t\,K\big[(x_t - x)/h\big] \mid \mathcal{F}_s\big]\Big| &\le \frac{Ch\sum_{j=0}^{\nu_n}|\phi_j|}{d_{t-s}^2}\int K(y - x/h)\mathbb{E}\Big\{|\widetilde{g}_{\nu_n}|\sum_{j=1}^{\nu_n}|\epsilon_j|\Big\}dy \\
&\le \frac{C_1\nu_n\,h\sum_{j=0}^{\nu_n}|\phi_j|}{d_{t-s}^2},
\end{aligned}$$

where ϕ_j are given as in **A1**. Note that $\sum_{j=0}^{\nu_n} |\phi_j| \sum_{j=\nu_n}^{n} d_j^{-2} \le C(1+\log n)$ under both **LM** and **SM**, and as in the estimation of (7.31) [by using (7.23)],

$$\sum_{s=1}^{n} \mathbb{E}\Big\{|\widetilde{g}_s|\, K\big[(x_s-x)/h\big]\big(1+\,||\lambda_{s-j}||^{\beta+1}\big)\Big\} \le Cnd_n/h, \tag{7.32}$$

unifromaly for $j \ge 0$. As a consequence, conditional arguments yield that

$$\begin{aligned} |\Delta_{3n}| &\le 2\sum_{s=1}^{n}\sum_{t=s+\nu_{1n}+1}^{n} \mathbb{E}\Big\{|\widetilde{g}_s\, K\big[(x_s-x)/h\big]|\,\big|\mathbb{E}\big[\widetilde{g}_t\, K\big[(x_t-x)/h\big] \mid \mathcal{F}_s\big]\big|\Big\} \\ &\le C\sum_{s=1}^{n} \mathbb{E}\Big\{|\widetilde{g}_s|\, K\big[(x_s-x)/h\big]\Big\} \sum_{t=s+\nu_{1n}}^{n} \frac{\nu_n\, h \sum_{j=0}^{\nu_n}|\phi_j|}{d_{t-s}^2} \\ &\le C\,\nu_n\, h(1+\log n)\,(nd_n/h) = o(nd_n/h), \end{aligned} \tag{7.33}$$

due to $h\nu_n(1+\log n) = o(nh^3/d_n)^{\delta} = o(1)$.

To estimate Δ_{2n}, recall that $|g(x)| \le C(1+|x|^{\beta+1})$ by **A5** (i). We have that

$$\begin{aligned} |\widetilde{g}_t| &\le |g_t| + E|g_t| \le C\left[1+\big(\sum_{j=0}^{\nu_n}|\psi_j\lambda_{t-j}'|\big)^{\beta+1}\right] + E|g_t| \\ &\le C\big(1+\sum_{i=0}^{\nu_n}(|\psi_{1i}|+|\psi_{2i}|)\,||\lambda_{t-i}||^{\beta+1}\big). \end{aligned}$$

Note that, for $t-s \ge m_0+2$,

$$\mathbb{E}\big[(1+||\lambda_{t-i}||^{\beta+1})\,K\big[(x_t-x)/h\big] \mid \mathcal{F}_s\big] \le \frac{Ch}{d_{t-s}}\big(1+\,||\lambda_{t-i}||^{\beta+1}\, I_{(i\ge t-s)}\big)$$

due to Lemma 7.1 (ii) with $l=1$, where $I_{(.)}$ denotes the indicator function. It is readily seen from the conditional arguments that

$$\begin{aligned} |\Delta_{2n}| &\le 2\sum_{s=1}^{n}\sum_{t=s+m_0+2}^{s+\nu_{1n}} \Big|\mathbb{E}\Big\{\widetilde{g}_s\,\widetilde{g}_t\, K\big[(x_s-x)/h\big]K\big[(x_t-x)/h\big]\Big\}\Big| \\ &\le C\sum_{s=1}^{n}\mathbb{E}\Big\{|\widetilde{g}_s|\,K\big[(x_s-x)/h\big] \\ &\qquad \sum_{t=s+m_0+2}^{s+\nu_{1n}} \frac{h}{d_{t-s}}\Big[1+\sum_{i=0}^{\nu_n}(|\psi_{1i}|+|\psi_{2i}|)\,\big(1+\,||\lambda_{t-i}||^{\beta+1}\, I_{(i\ge t-s)}\big)\Big]\Big\} \\ &\le C\sum_{t=m_0+2}^{\nu_{1n}}\frac{h}{d_t}\sum_{s=1}^{n}\mathbb{E}\Big\{|\widetilde{g}_s|\,K\big[(x_s-x)/h\big]\Big\} + C\sum_{t=m_0+2}^{\nu_{1n}}\frac{h}{d_t}\sum_{i=0}^{\nu_n}(|\psi_{1i}|+|\psi_{2i}|) \\ &\qquad \sum_{s=1}^{n}\mathbb{E}\Big\{|\widetilde{g}_s|\,K\big[(x_s-x)/h\big]\big(1+\,||\lambda_{t+s-i}||^{\beta+1}\, I_{(i\ge t)}\big)\Big\} \\ &\le C\sum_{t=1}^{\nu_{1n}}\frac{h}{d_t}\,\big[1+\sum_{s=m_0+1}^{n}\frac{h}{d_s}\big] \\ &\le C\,\frac{h\nu_{1n}}{d_{\nu_{1n}}}\,nh/d_n = o\big[nh/d_n\big], \end{aligned} \tag{7.34}$$

due to $h\nu_{1n} \le Ch\nu_n = o(nh^3/d_n) = o(1)$ and $d_{\nu_{1n}}^{-1} \le C$, where we have used (7.32).

Combining (7.30)-(7.34), $R_{n4} = o_P[(nh/d_n)^{1/2}]$ as $n \to \infty$ first and then $m \to \infty$, i.e., we have (7.29). The proof of Lemma 7.2 is now complete. □

Proof of Theorem 4.1. We start with some preliminaries. Let $A_l = \sum_{j=l+1}^{\infty} \big(|\psi_{1j}| + |\psi_{2j}|\big)$, where l is chosen so large that $A_l \le 1$. Due to

$$u_t = \Big(\sum_{j=0}^{l} + \sum_{j=l+1}^{\infty}\Big)\psi_j\, \lambda'_{t-j} = u_{l,t} + u_{l,1t}, \quad \text{say},$$

it follows from (3.5) that, for any $x \in \mathbb{R}$,

$$|m(x,u_t) - m(x,u_{l,t})| \le T(x)\, |u_{l,1t}|^{\delta}\, (1 + |u_t|^{\beta} + |u_{l,t}|^{\beta}).$$

As in (7.21) and by recalling $A_0 = \sum_{j=0}^{\infty} \big(|\psi_{1j}| + |\psi_{2j}|\big) < \infty$, we have

$$\begin{aligned} |u_{l,1t}|^2 &\le A_l \sum_{j=0}^{\infty} \big(|\psi_{1j}| + |\psi_{2j}|\big)\, ||\lambda_{t-j}||^2, \\ 1 + |u_t|^{\beta} + |u_{l,t}|^{\beta} &\le 1 + C \sum_{j=0}^{\infty} \big(|\psi_{1j}| + |\psi_{2j}|\, ||\lambda_{t-j}||^{\beta}\big). \end{aligned}$$

Hence, by letting $\alpha = \max\{\beta, 2\}$, it follows that

$$\begin{aligned} & |m(x,u_t) - m(x,u_{l,t})| \\ \le\;& T(x)\, \big(|u_{l,t}|^2\big)^{\delta/2}\, (1 + |u_{l,t}|^{\beta} + |u_{l,1t}|^{\beta}) \\ \le\;& A_l^{\delta}\, T(x) \Big(1 + C\, \big[\sum_{j=0}^{\infty} \big(|\psi_{1j}| + |\psi_{2j}|\big) ||\lambda_{t-j}||^{\alpha}\big]\Big)^{1+\delta/2} \\ \le\;& C A_l^{\delta}\, T(x) \Big(1 + \sum_{j=0}^{\infty} \big(|\psi_{1j}| + |\psi_{2j}|\big) ||\lambda_{t-j}||^{3\alpha/2}\Big) \end{aligned} \tag{7.35}$$

Similarly, for any $x, x_0 \in \mathbb{R}$, we have

$$\begin{aligned} |m(x,u_{l,t}) - m(x,x_0)| &\le T(x)\, |u_{l,t} - x_0|^{\delta} (1 + |x_0|^{\beta} + |u_{l,t}|^{\beta}) \\ &\le C\, T(x)\, \big[1 + \sum_{j=0}^{l} \big(|\psi_{1j}| + |\psi_{2j}|\big)\, ||\lambda_{t-j}||^{\beta+1}\big]. \end{aligned} \tag{7.36}$$

We are now ready to prove Theorem 4.1. First assume $w_{nk} = 0$. This restriction will be removed later. For convenience of notation, we further assume $z = 0$. The removal of the restriction $z = 0$ is standard and so details are omitted.

Let $m_{1l}(x) = \mathbb{E} m(x, u_{l,1})$. We may write

$$\frac{c_n}{n} \sum_{t=1}^{n} m\big(c_n Z_{nt}, u_t\big) =: S_n + S_{1n} + S_{2n},$$

where $S_n = \frac{c_n}{n} \sum_{t=1}^n m_{1l}(c_n Z_{nt})$,

$$
\begin{aligned}
S_{1n} &= \frac{c_n}{n} \sum_{t=1}^n \Big[m(c_n Z_{nt},\, u_t) - m(c_n Z_{nt},\, u_{l,t})\Big], \\
S_{2n} &= \frac{c_n}{n} \sum_{t=1}^n \Big[m(c_n Z_{nt},\, u_{l,t}) - m_{1l}(c_n Z_{nt})\Big].
\end{aligned}
$$

It follows from Corollary 2.3 (i) of Wang (2015) that, for any $l \geq 1$,

$$
S_n \to_d \int_{-\infty}^{\infty} m_{1l}(x)\, dx\, L_Z(1,0).
$$

Since, by (7.35),

$$
\begin{aligned}
& \mathbb{E}|m(x,u_1) - m(x,u_{x,1})| \\
\leq\ & C\, T(x)\, A_l^{\delta/2}\, \mathbb{E}\Big(1 + \sum_{j=0}^{\infty} \big(|\psi_{1j}| + |\psi_{2j}|\big) ||\lambda_{t-j}||^{\max\{3\beta/2,3\}}\Big) \\
\leq\ & C_1\, T(x)\, A_l^{\delta/2},
\end{aligned}
$$

we have $\int_{-\infty}^{\infty} |m_{1l}(x) - m_1(x)|dx \leq C A_l^{\delta/2} \int_{-\infty}^{\infty} T(x)dx \to 0$, i.e.,

$$
\int_{-\infty}^{\infty} m_{1l}(x)dx \to \int_{-\infty}^{\infty} m_1(x)dx, \quad \text{as} \quad l \to \infty.
$$

Hence, to prove (4.7) with $\nu_{nk} = 0$, it suffices to show that

$$
S_{in} = o_P(1), \qquad i = 1,\ 2, \tag{7.37}
$$

as $n \to \infty$ first and then $l \to \infty$.

The proof of (7.37) for $i = 1$ is simple. Indeed, due to (7.35) and $\mathbb{E}||\lambda_0||^{\max\{3\beta/2,3\}} < \infty$, it follows from Lemma 7.1 (i) with $h = d_n/c_n$ and $m = 0$ that

$$
\begin{aligned}
\mathbb{E}|S_{1n}| &\leq \frac{c_n}{n} \sum_{t=1}^n \mathbb{E}\big|m(c_n Z_{nt},\, u_t) - m(c_n Z_{nt},\, u_{l,t})\big| \\
&\leq C A_l^{\delta/2} \frac{c_n}{n} \sum_{t=1}^n \mathbb{E}\Big(1 + \sum_{j=0}^{\infty} \big(|\psi_{1j}| + |\psi_{2j}|\big) ||\lambda_{t-j}||^{\max\{3\beta/2,3\}}\Big)\, T\big(c_n Z_{nt}\big) \\
&\leq C A_l^{\delta/2} \sup_{j\geq 0} \frac{c_n}{n} \sum_{t=1}^n \mathbb{E}\big(1 + ||\lambda_{t-i}||^{\max\{3\beta/2,3\}}\big)\, T\big(c_n Z_{nt}\big) \\
&\leq C A_l^{\delta/2} \big(1 + \frac{d_n}{n} \sum_{t=m_0}^n d_k^{-1}\big) \leq C A_l^{\delta/2} \to 0,
\end{aligned}
$$

as $n \to \infty$ first and then $l \to \infty$, which implies (7.37) with $i = 1$.

We next prove (7.37) with $i = 2$. Write $p(x,y) = m(x,y) - m_{1l}(x)$. First note that, by the boundedness and integrability of $\mathbb{E}|m(x,u_1)|$, there exists a finite constant x_0

such that $m(x, x_0)$ is bounded and integrable. For this x_0, it follows from (7.36) that, for any $l \geq 1$ and $x \in \mathbb{R}$,

$$\begin{aligned} |p(x, u_{l,t})| &\leq |m(x, u_{l,t}) - m(x, x_0)| + E|m(x, u_{l,t}) - m(x, x_0)| \\ &\leq C\, T(x) \Big[1 + \sum_{j=0}^{l} \big(|\psi_{1j}| + |\psi_{2j}|\big)\, ||\lambda_{t-j}||^{\beta+1}\Big]. \end{aligned} \tag{7.38}$$

Therefore, for any $l \geq 0$ and $x \in \mathbb{R}$, we have

$$\begin{aligned} & \mathbb{E}\big(|p(x, u_{l,t})| + |p(x, u_{l,t})|^2\big) \\ \leq\ & CT(x) + C\,T^2(x)\,\mathbb{E}\Big[1 + \sum_{j=0}^{l} \big(|\psi_{1j}| + |\psi_{2j}|\big)\, ||\lambda_{t-j}||^{\beta+1}\Big]^2 \\ \leq\ & C_1\, T(x)\Big[1 + \sum_{j=0}^{l} \big(|\psi_{1j}| + |\psi_{2j}|\big)\, \mathbb{E}||\lambda_{t-j}||^{2\beta+2}\Big] \\ \leq\ & C_2\, T(x). \end{aligned} \tag{7.39}$$

Similarly, for any $l \geq 0$ and $x \in \mathbb{R}$, it follows that

$$\begin{aligned} & \mathbb{E}|p(x, u_{l,t})|\big(1 + \sum_{j=1}^{l} |\epsilon_j|\big) \\ \leq\ & C\,T(x)\big[1 + \sum_{j=0}^{l} \big(|\psi_{1j}| + |\psi_{2j}|\big)\, \mathbb{E}\{||\lambda_{t-j}||^{\beta+1}\big(1 + \sum_{i=0}^{l} |\epsilon_i|\big)\} \\ \leq\ & C_1\,(1+l)\,T(x). \end{aligned} \tag{7.40}$$

Since $T(x)$ is bounded and integrable, (7.38) implies that

$$\mathbb{E}p^2(c_n\, Z_{nk},\ u_{l,k}) \leq C \tag{7.41}$$

for any $k \geq 1$; it follows from (7.39) and Lemma 7.1 (i) with $h = d_n/c_n$ that

$$\begin{aligned} & \mathbb{E}\big(|p(c_n Z_{nk}, u_{l,k})| + |p(c_n Z_{nk}, u_{l,k})|^2\big) \\ \leq\ & C\,\frac{d_n}{c_n\, d_k} \int_{-\infty}^{\infty} \mathbb{E}\big(|p(x, u_{l,l})| + |p(x, u_{l,l})|^2\big)\,dx \\ \leq\ & \frac{C_1\, d_n}{c_n\, d_k}, \end{aligned} \tag{7.42}$$

for any $k \geq 2l + m_0$; and by (7.40) and Lemma 7.1 (iii) with $h = d_n/c_n$, we have

$$\begin{aligned} & \big|\mathbb{E}\big[p(c_n Z_{nk}, u_{l,1k}) \mid \mathcal{F}_j\big]\big| \\ \leq\ & \frac{C\, d_n \sum_{j=0}^{l} |\phi_j|}{c_n\, d_{k-j}^2} \int_{-\infty}^{\infty} \mathbb{E}|p(x, u_{l,1l})|\big(1 + \sum_{j=1}^{l} |\epsilon_j|\big)\,dx \\ \leq\ & \frac{C\, d_n\, l\, \sum_{j=0}^{l} |\phi_j|}{c_n\, d_{k-j}^2}, \end{aligned} \tag{7.43}$$

for any $k - j \geq 2l + m_0$.

Result (7.37) with $i = 2$ can now be proved by using standard conditional arguments as those of Lemma 2.2 (ii) in Wang (2015). A outline is given as follows.

For each $l \geq 1$, we have

$$\begin{aligned}
\mathbb{E}S_{2n}^2 &= \big(\frac{c_n}{n}\big)^2 \sum_{s,t=1}^{n} \mathbb{E}\Big\{p(c_n\, Z_{ns},\, u_{l,s})\, p(c_n\, Z_{nt},\, u_{l,t})\Big\} \\
&= \big(\frac{c_n}{n}\big)^2 \Big(\sum_{|t-s|\leq 2l+m_0} + \sum_{|t-s|>2l+m_0}\Big) E\Big\{p(c_n\, Z_{ns},\, u_{l,s})\, p(c_n\, Z_{nt},\, u_{l,t})\Big\} \\
:= & \quad \Delta_{1n} + \Delta_{2n}.
\end{aligned}$$

Using (7.41) and (7.42), we have

$$\begin{aligned}
|\Delta_{1n}| &\leq \big(\frac{c_n}{n}\big)^2 \sum_{|t-s|\leq 2l+m_0} \mathbb{E}\Big\{p^2(c_n\, Z_{ns},\, u_{l,s}) + p^2(c_n\, Z_{nt},\, u_{l,t})\Big\} \\
&\leq C\,\Big[(l+m_0)^2\big(\frac{c_n}{n}\big)^2 + \frac{c_n d_n}{n^2} \sum_{|t-s|\leq 2l+m_0} (d_s^{-1} + d_t^{-1})\Big] \\
&\leq C_1\,\big[\frac{(l+m_0)c_n}{n} + \big(\frac{(l+m_0)c_n}{n}\big)^2\big].
\end{aligned}$$

Using (7.42)-(7.43) and conditional arguments, it follows that

$$\begin{aligned}
|\Delta_{2n}| &\leq 2\,\big(\frac{c_n}{n}\big)^2 \sum_{t-s>2l+m_0} \mathbb{E}\Big\{|p(c_n\, Z_{ns},\, u_{l,s})|\,\big|\mathbb{E}\big[p(c_n\, Z_{nt},\, u_{l,t}) \mid \mathcal{F}_s\big]\big|\Big\} \\
&\leq C\, l \sum_{j=0}^{l} |\phi_j|\, \frac{d_n^2}{n^2} \sum_{s=1}^{n} \frac{1}{d_s} \sum_{t=s+2l}^{n} \frac{1}{d_{t-s}^2} \\
&\leq C\, l\, \frac{d_n\, \log n}{n}.
\end{aligned}$$

Combining all these estimates, it follows that

$$\begin{aligned}
\mathbb{E}S_{2n}^2 &\leq \Delta_{1n} + \Delta_{2n} \\
&\leq C\, l\, \frac{d_n\, \log n}{n} + C_1\,\big[\frac{(l+m_0)c_n}{n} + \big(\frac{(l+m_0)c_n}{n}\big)^2\big] \to 0,
\end{aligned}$$

as $n \to \infty$ first and then $l \to \infty$, yielding (7.37) with $i = 2$.

The proof of (4.7) with $w_{nk} = 0$ is now complete. We next remove the restriction $w_{nk} = 0$. In fact, by (3.5), we have

$$\begin{aligned}
&\frac{c_n}{n} \sum_{t=1}^{n} m\big(c_n\, Z_{nt}, u_t + w_{nt}\big) \\
&= \frac{c_n}{n} \sum_{t=1}^{n} m\big(c_n\, Z_{nt}, u_t\big) + O(\delta_n^{\delta})\, \frac{c_n}{n} \sum_{t=1}^{n} (1 + |w_t|^{\beta+1} + |u_t|^{\beta})\, T\big(c_n\, Z_{nt}\big),
\end{aligned}$$

where $w_t = \sum_{j=-d}^{\infty} \big(|\varphi_{1j}\, \epsilon_{k-j}| + |\varphi_{2j}\, \eta_{k-j}| \big)$. Note that

$$(1 + |w_t|^{\beta+1} + |u_t|^{\beta}) \le (1 + \widetilde{u}_t)^{\beta+1} \le C \Big[1 + \sum_{j=0}^{\infty} \big(\theta_{1j} + \theta_{2j}\big)\, ||\lambda_{t-j}||^{\beta+1} \Big], \qquad (7.44)$$

where $\widetilde{u}_t = \sum_{j=-d}^{\infty} \big(\theta_{1j} + \theta_{2j}\big)\, ||\lambda_{t-j}||$ with (letting $\psi_{1j}, \psi_{2j} = 0$ if $j < 0$)

$$\theta_{1j} = |\varphi_{1j}| + |\psi_{1j}| \text{ and } \theta_{2j} = |\varphi_{2j}| + |\psi_{2j}|.$$

As $\delta_n \to 0$, result (4.7) follows from that of the first part with $w_{nk} = 0$ and (7.22) with $h = d_n/c_n$.

The proof of Theorem 4.1 is now complete. □

Proof of Corollary 4.1. We first prove (4.9). For any given x, write $\kappa_{nt} = \gamma_0 x + u_t + w_{nt}$, $\widetilde{\kappa}_{nt} = (\gamma_{nt} - \gamma_0)x + \kappa_{nt}$ and $\beta_n = \max_{1\le k\le n} |\gamma_{nk} - \gamma_0|$. Since $a_t = h\gamma_{nt}(x_t - x)/h + \widetilde{\kappa}_{nt}$, it follows from (3.6) that

$$\begin{aligned} |f(a_t) - f(\kappa_{nt})| &\le |f(a_t) - f(\widetilde{\kappa}_{nt})| + |f(\widetilde{\kappa}_{nt}) - f(\kappa_{nt})| \\ &\le C \left| h\gamma_{nt}(x_t - x)/h \right|^{\delta} \big(1 + |a_t|^{\beta} + |\widetilde{\kappa}_{nt}|^{\beta}\big) \\ &\qquad + C \left| (\gamma_{nt} - \gamma_0) x \right|^{\delta} \big(1 + |\widetilde{\kappa}_{nt}|^{\beta} + |\kappa_{nt}|^{\beta}\big) \\ &\le C\,(h^{\delta} + \beta_n^{\delta}) \Big[1 + |u_t|^{\beta} + |w_{nt}|^{\beta} + \big(|x_t - x|/h\big)^{\beta+1}\Big] \\ &\le C\,(h^{\delta} + \beta_n^{\delta}) \Big[1 + |\widetilde{u}_t|^{\beta} + \big(|x_t - x|/h\big)^{\beta+1}\Big], \end{aligned} \qquad (7.45)$$

where $\widetilde{u}_t$ is defined as in (7.44).

Let $K_1(s) = (1 + |s|^{\beta+1})K(s)$. It follows from (7.45) and (7.22) with $p(s) = K_1(s)$ that

$$\begin{aligned} &\frac{d_n}{nh} \sum_{t=1}^{n} f(a_t)\, K\big[(x_t - x)/h\big] \\ = \ &\frac{d_n}{nh} \sum_{t=1}^{n} f(\kappa_{nt})\, K\big[(x_t - x)/h\big] + O(h^{\delta} + \beta_n^{\delta})\, \frac{d_n}{nh} \sum_{t=1}^{n} (1 + |\widetilde{u}_t|^{\beta})\, K_1\big[(x_t - x)/h\big] \\ = \ &\frac{d_n}{nh} \sum_{t=1}^{n} f(\kappa_{nt})\, K\big[(x_t - x)/h\big] + o_P(1), \end{aligned}$$

due to $h \to 0$ and $\beta_n \to 0$. Result (4.9) now follows from (4.7) with

$$m(t, y) = K(t)\, f(\gamma_0 x + y), \quad c_n = d_n/h, \quad c_n' = 1/d_n, \quad z = -x.$$

The proof of (4.10) is similar. Indeed, in this case we may write

$$a_t = \beta_n(\gamma_{nt} - \gamma_0)x_t/\beta_n + \gamma_0 x_t + u_t + w_{nt},$$

and the similar arguments to the proof of (7.45) yield that, for any fixed x,

$$\begin{aligned} & |f(a_t) - f(\gamma_0 x_t + u_t + w_{nt})| \\ \leq\ & C\beta_n^{\delta}\, |x_t|^{\delta}\, \big(1 + |a_t|^{\beta} + |\gamma_0 x_t + u_t + w_{nt}|^{\beta}\big) \\ \leq\ & C\,\beta_n^{\delta}\,\big(1 + |x_t - x|^{1+\beta}\big)\,(1 + |\widetilde{u}_t|^{\beta}), \end{aligned}$$

implying that (letting $K_2(s) = (1 + |s|^{1+\beta})K(s)$)

$$\begin{aligned} & \frac{d_n}{nh}\sum_{t=1}^{n} f(a_t)\,K\big(x_t - x\big) \\ =\ & \frac{d_n}{nh}\sum_{t=1}^{n} f(\gamma_0 x_t + u_t + w_{nt})\,K\big(x_t - x\big) + O(\beta_n^{\delta})\,\frac{d_n}{nh}\sum_{t=1}^{n}(1 + |\widetilde{u}_t|^{\beta})\,K_2\big[(x_t - x)/h\big] \\ =\ & \frac{d_n}{nh}\sum_{t=1}^{n} f(\gamma_0 x_t + u_t + w_{nt})\,K\big(x_t - x\big) + o_P(1). \end{aligned}$$

Result (4.10) follows from (4.7) with

$$m(t, y) = K(t)\, f(\gamma_0 x + \gamma_0 t + y), \quad c_n = d_n, \quad c_n' = 1/d_n, \quad z = -x.$$

The proof of Corollary 4.1 is now complete. □

Proof of Theorem 5.1. We may write

$$\begin{aligned} \hat{g}(x) &= \frac{\sum \eta_t K[(x_t - x)/h]}{\sum K[(x_t - x)/h]} + \frac{\sum g(z_t)\,K[(x_t - x)/h]}{\sum K[(x_t - x)/h]} \\ &=: R_{1n} + R_{2n}. \end{aligned} \tag{7.46}$$

As in Wang and Phillips (2016) (see, also, Lemma 7.2), it is easy to show that $R_{1n} = O_P\big[(nh^2)^{1/4}\big]$. On the other hand, a simple application of Corollary 4.1 yields that

$$\begin{aligned} & \Big\{\frac{d_n}{nh}\sum_{t=1}^{n} K\big[(x_t - x)/h\big],\ \frac{d_n}{nh}\sum_{t=1}^{n} g(z_t)\,K\big[(x_t - x)/h\big]\Big\} \\ \to_D\ & \Big(L_Z(1, 0),\ \mathbb{E}g\big(x - u_1\big) \times L_Z(1, 0)\Big). \end{aligned} \tag{7.47}$$

The result (5.16) follows from the continuous mapping theorem. □

Proof of Theorem 5.2. We may write

$$\sum g(z_t)\,K[(x_t - x)/h] =: y_{n1} + y_{n2},$$

where

$$\begin{aligned} y_{n1} &= \sum g(x - u_t)\,K[(x_t - x)/h] \\ y_{n2} &= \sum \big[g\big[\alpha_{nt}^{-1}(x_t - w_{nt} - u_t)\big] - g(x - u_t)\big]\,K[(x_t - x)/h] \end{aligned}$$

Due to **A5** (i), (iv) and the fact that $K(x)$ has a compact support, for any fixed x, we have

$$
\begin{aligned}
|y_{n2}| &\leq C \sum \big[|\alpha_{nt}^{-1}(||x_t - x| + |w_{nt}|) + |\alpha_{nt}^{-1} - 1|\, |x - u_t|\big) \\
&\qquad\qquad \big(1 + |x - u_t|^{\beta} + |\alpha_{nt}^{-1}(x - u_t - w_{nt})|^{\beta}\big) K[(x_t - x)/h] \\
&\leq C_x \, (h + \alpha_n + \delta_n) \sum_{k=1}^{n} \big(1 + |\tilde{u}_t|^{\beta+1}\big)\, K[(x_t - x)/h] \\
&= O_P\big[(h + \alpha_n + \delta_n) \frac{nh}{d_n}\big] = o_P\left[\Big(\frac{nh}{d_n}\Big)^{1/2}\right],
\end{aligned}
$$

due to $\alpha_n + \delta_n = O(h)$ and $nh^3/d_n \to 0$, where $\widetilde{u}_t$ is defined as in (7.44) and we have used Hölder's inequality and (7.22). Taking these facts into (7.46), simple calculations show that (5.17) will follow if we prove

$$
\begin{aligned}
& \frac{d_n}{nh} \sum_{t=1}^{n} K\big[(x_t - x)/h\big], \ \big(\frac{d_n}{nh}\big)^{1/2} \sum_{t=1}^{n} A_t \, K\big[(x_t - x)/h\big] \\
\to_D \quad & \big(L_Z(1,0), \Lambda N \times L_Z(1,0)^{1/2}\big), \qquad (7.48)
\end{aligned}
$$

where $A_t = \eta_t + g(x - u_t) - Eg(x - u_t)$ and $N \sim \mathcal{N}(0,1)$ independent of $L_Z(1,0)$. This follows from a simple application of Lemma 7.2, since $\tilde{g}(y) = g(x - y) - Eg(x - u_1)$ still satisfies **A5**(i) with $E\tilde{g}(u_1) = 0$ for any fixed $x \in \mathbb{R}$. $\Box$

8 Appendix B: Proof of (7.20)

Note that

$$
\begin{aligned}
x_k - x_j &= \sum_{i=j+1}^{k} \rho_n^{k-i}\, \xi_i + \sum_{i=1}^{j} (\rho_n^{k-i} - \rho_n^{j-i})\, \xi_i \\
&= \sum_{i=j+1}^{k} \rho_n^{k-i}\, \Big(\sum_{u=j+1}^{i} + \sum_{u=-\infty}^{j} \Big) \epsilon_u \phi_{i-u} + \sum_{i=1}^{j} (\rho_n^{k-i} - \rho_n^{j-i})\, \xi_i .
\end{aligned}
$$

We may have

$$
x_k = x_{1jk} + x_{2jk}, \tag{8.49}
$$

where $x_{1jk} = \sum_{i=j+1}^{k} \epsilon_i\, a_{k,i}$ with

$$
a_{k,i} = \sum_{u=i}^{k} \rho_n^{k-u} \phi_{u-i} = a_{k-i},
$$

and x_{2jk} depends only on $\epsilon_j, \epsilon_{j-1}, ...$

Let $\Lambda_m = \sum_{j=1}^{m} \epsilon_{t_j} a_{k-t_j}$ and $y_{jk}^* = x_{1jk} - \Lambda_m$. It is readily seen that there exists an $m_0 > 0$ such that, whenever $k - j \ge 2m + m_0$, $0 < a_1 \le E(y_{jk}^*)^2 / d_{k-j}^2 \le a_2 < \infty$, where a_1 and a_2 are constants. As a consequence, similar arguments as in the proof of Theorem 2.18 of Wang (2015) [In particular, see part (ii), the fact **F,** and (2.66)] yields that, whenever $k - j \ge 2m + m_0$, y_{jk}^*/d_{k-j} has a density function $f_{jk}(x)$, which is uniformly bounded over x by a constant C and

$$
\sup_x |f_{jk}(x+u) - f_{jk}(x)| \le C \min\{|u|,\ 1\}. \tag{8.50}
$$

This, together with (8.49) and the independence of ϵ_i, implies that

$$
\begin{aligned}
&\mathbb{E}\big\{ p(x_k/h, \lambda_{t_1}, ..., \lambda_{t_m}) \mid \mathcal{F}_j \big\} = \mathbb{E}\big\{ p\big[(x_{2jk} + \Lambda_m + y_{jk}^*)/h,\ \lambda_{t_1}, ..., \lambda_{t_m} \big] \ \mid \mathcal{F}_j \big\} \\
&= \mathbb{E}\Big\{ \int_{-\infty}^{\infty} p\big[(x_{2jk} + \Lambda_m + d_{k-j} y)/h,\ \lambda_{t_1}, ..., \lambda_{t_m} \big]\, f_{jk}(y)\, dy \mid \mathcal{F}_j \Big\} \\
&= \frac{h}{d_{k-j}} \int_{-\infty}^{\infty} \mathbb{E}\Big\{ p(y,\ \lambda_{t_1}, ..., \lambda_{t_m})\, f_{jk}\big(\frac{-x_{2jk} - \Lambda_m + hy}{d_{k-j}} \big) \,\big|\mathcal{F}_j \Big\} dy. \qquad (8.51)
\end{aligned}
$$

As x_{2jk} depends only on $\epsilon_j, \epsilon_{j-1}, ...$, and $j+1 \le t_1, ..., t_m \le k$, we have

$$
\begin{aligned}
&\mathbb{E}\Big\{ p(y,\ \lambda_{t_1}, ..., \lambda_{t_m})\, f_{jk}\big(\frac{-x_{2jk} + hy}{d_{k-j}} \big) \,\big|\mathcal{F}_j \Big\} \\
&= f_{jk}\big(\frac{-x_{2jk} + hy}{d_{k-j}} \big)\, \mathbb{E}\big\{ p(y,\ \lambda_{t_1}, ..., \lambda_{t_m}) \,\big\} = 0.
\end{aligned}
$$

Taking this fact into (8.51) and using (8.50), we have

$$\begin{aligned}
& \left|\mathbb{E}\left\{p(x_k/h, \lambda_{t_1}, ..., \lambda_{t_m}) \mid \mathcal{F}_j\right\}\right| \\
\le\ & \frac{h}{d_{k-j}} \int_{-\infty}^{\infty} \mathbb{E}\Big\{p(y,\ \lambda_{t_1}, ..., \lambda_{t_m}) \\
& \qquad \Big| f_{jk}\big(\frac{-x_{2jk} - \Lambda_m + hy}{d_{k-j}}\big) - f_{jk}\big(\frac{-x_{2jk} + hy}{d_{k-j}}\big)\Big| \,\big|\mathcal{F}_j\Big\}\, dy \\
\le\ & \frac{Ch}{d_{k-j}} \int_{-\infty}^{\infty} \mathbb{E}\Big\{|p(y,\ \lambda_{t_1}, ..., \lambda_{t_m})| \ \min\{|\Lambda_m|/d_{k-j},\ 1\}\Big\}\, dy \\
\le\ & \frac{Ch \sum_{j=0}^{k-\min\{t_1,...,t_m\}} |\phi_j|}{d_{k-j}^2} \int_{-\infty}^{\infty} \mathbb{E}\Big\{|p(y,\ \lambda_1, ..., \lambda_m)| \sum_{j=1}^{m} |\epsilon_j|\Big\}\, dy,
\end{aligned}$$

as required. □

REFERENCES

Adusumilli, K. and T. Otsu (2018). Nonparametric instrumental regression with errors in variables. *Econometric Theory,* 34, 1256-1280.

Baxter, M., and R. G. King (1999): Measuring business cycles: approximate band-pass filters for economic time series, *Review of Economics and Statistics*, 81(4), 575 593.

Bollerslev, T., D. Osterrieder, N. Sizova and G. Tauchen (2013). Risk and return: Long-run relations, fractional cointegration and return predictability. *Journal of Financial Economics* 108, 409–424.

Chen, X. and Reiss, M. (2011). On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 27, 497-521.

Duffy, J. A. (2014). Three essays on the nonparametric estimation of nonlinear cointegrating regression. Doctoral Dissertation, Yale University.

Duffy, J. A. and D. F. Hendry (2017). The impact of integrated measurement errors on modelling long-run macroeconomic time series. *Econometric Reviews,* 36, 568-587.

Gao, J. and C. Dong (2018). Specification testing driven by orthogonal series for nonlinear cointegration with endogeneity, *Econometric Theory,* 34, 754-789.

Giraitis, L., H. L. Houl and D. Surgailis (2012). *Large Sample Inference for Long Memory Processes.* London: Imperial College Press.

Granger, C. W. J. and P. Newbold (1974). Spurious Regressions in Econometrics, *Journal of Econometrics*, 74, 111–120.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The MIDAS touch: MIxed DAta Sampling regression models", mimeo, Chapel Hill, N.C.

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

Hall, P. and J. L. Horowitz (2005). Nonparametric methods for inference in the presence of instrumental variables, *Annals of Statistics*, 33, 2904-2929.

Hodrick, R. J., and E. C. Prescott (1997). Postwar US business cycles: an empirical investigation, *Journal of Money, Credit, and Banking*, pp. 1-16.

Horowitz, J. L. (2011). Applied nonparametric instrumental variables estimation, *Econometrica*, 79, 347-394.

Kasparis, I. and Phillips, P. C. B. (2012). Dynamic misspecification in nonparametric cointegrating regression. *Journal of Econometrics,* 168, 270-284.

Jeganathan, P. (2008). Limit theorems for functional sums that converge to fractional Brownian and stable motions. Cowles Foundation Discussion Paper No. 1649, Cowles Foundation for Research in Economics, Yale University.

Leser, C. E. V., 1961. A simple method of trend construction. *Journal of the Royal Statistical Society*, B, 23, 91-107.

Phillips, P. C. B. (1986). "Understanding spurious regressions in econometrics," *Journal of Econometrics* 33, 311–340.

Phillips, P. C. B (2009). Local limit theory and spurious nonparametric regression. *Econometric Theory,* **25,** 1466-1497.

Phillips, P. C. B. (2010). Two New Zealand pioneer econometricians, *New Zealand Economic Papers*, 44, 1-26.

Phillips, P. C. B. and S. Jin (2015). Business Cycles, Trend Elimination, and the HP Filter, *Cowles Foundation Discussion Paper No. 2005,* Yale University.

Schennach, S. M. (2004). Nonparametric regression in the presence of measurement error, *Econometric Theory*, 20, 1046-1093.

Wang, Q. (2014). Martingale limit theorem revisited and nonlinear cointegrating regression, *Econometric Theory*, **30** , 509–535.

Wang, Q. (2015). *Limit theorems for nonlinear cointegrating regression.* Singapore: World Scientific.

Wang, Q., Lin, Y. X. and Gulati, C. M. (2003). Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory*, **19**, 143–164.

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

Wang, Q. and Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression, *Econometrica* **77**, 1901-1948.

Wang, Q. and Phillips, P. C. B. (2011). Asymptotic theory for zero energy functionals with nonparametric regression applications. *Econometric Theory*, **27**, 235–259.

Wang, Q. and Phillips, P. C. B. (2016). Nonparametric cointegrating regression with endogeneity and long memory, *Econometric Theory*, 32, 359-401.

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

Whittaker, E. T. (1923). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Association*, 78, 81-89.