N-CONSISTENT SEMIPARAMETRIC REGRESSION: UNIT ROOT TESTS WITH NONLINEARITIES

TED JUHL AND ZHIJIE XIAO

Abstract. We develop unit root tests using additional time series as suggested in Hansen (1995). However, we allow for the covariate to enter the model in a nonlinear fashion, so that our model is an extension of the semiparametric model analyzed in Robinson (1988). It is proven that the autoregressive parameter is estimated at rate N even though part of the model is estimated nonparametrically. The limiting distribution is a mixture of a standard normal and the Dickey-Fuller distribution. A Monte Carlo experiment is used to evaluate the performance of the tests for various linear and nonlinear specifications.

1. Introduction

Increasing power in unit root tests has become an important research topic in recent years. Elliot, Rothenberg, and Stock (1996) propose an estimation strategy which focuses on estimating potential trends under the alternative hypothesis in order to effectively reach the Gaussian power envelope for unit root tests. Another branch of the unit root literature focuses on using some other features of the time series data. For example, Lucas (1995) uses M-estimators to take advantage of non-Gaussian errors in unit root tests. His results show that power gains are possible, even if the M-estimator does not coincide with the true likelihood. Using rank based tests, Hasan and Koenker (1997) are also able to realize increased power under certain error distributions while experiencing a small loss in power if the errors are actually Gaussian. Seo (1999) simultaneously estimates GARCH effects along with the autoregressive coefficients to increase power. Shin and So (1999) and Beelders (1999) use adaptive estimation to nonparametrically estimate the density of errors, and again obtain large

Ted Juhl, Department of Economics, University of Kansas and Zhijie Xiao, Department of Economics, University of Illinois at Urbana Champaign.

power gains, particularly if the error terms are heavy-tailed. Hansen (1995) shows that inclusion of stationary covariates can generate more precise estimates of the autoregressive parameter, translating into higher power for unit root tests.

The extension of these methods to the multivariate case in models of cointegration has been explored as well. The multivariate treatment of estimating trends is provided in Xiao and Phillips (1998). Lucas (1997) and Boswijk and Lucas (1999) use likelihood-ratio type tests and adaptive estimation to test for the number of cointegrating vectors in a multivariate system. Seo (1998) extends Hansen's (1995) result to incorporate additional stationary time series in a multivariate system. Phillips (1995) proposes estimating cointegrating relationships using least absolute deviations or M-estimation and illustrates the improved performance of the estimators.

In all of the papers listed above, some additional information is used to improve "efficiency" in the estimation of autoregressive parameters. In this way, the goal is to improve power against linear alternatives. To treat potential nonlinearities, Phillips and Park (1999) propose an even more general framework in which they allow for a nonlinear autoregressive structure. The convergence of the nonparametric estimator of the autoregressive function in the unit root case is at rate $N^{1/4}$.

In this paper, we allow for an unknown nonlinear function of covariates to influence the time series while still retaining a partially linear model. In allowing such a general structure, we hope to further increase the power gains from using covariates, particularly if there is a nonlinear relationship with the chosen covariate. Since the form of the nonlinearity is unknown, we estimate this part of the model nonparametrically while retaining the linear specification for the autoregressive component.

There are several findings in this paper. First, by using the compromise of a partial linear model, the convergence for the autoregressive component remains at rate N. This is an important extension of Robinson's (1988) result to the nonstationary case. In addition, the limiting distribution of the unit root test is identical to the distribution found in Hansen (1995) where covariates are used in a linear fashion.

This implies that, asymptotically, there is no loss from our general framework using an unknown nonlinear component rather than assuming a linear structure and using OLS. Finally, in the course of proving our theorem, we show that nonparametrically regressing an I(1) series on an I(0) series is asymptotcially equivalent to an OLS regression of the I(1) series on a constant.

The outline of the paper is as follows. In Section 2, we develop the model and provide a brief description of the estimation procedure. Section 3 provides the assumptions and asymptotic distribution of the test. A small Monte Carlo experiment is given in Section 4 and Section 5 concludes.

Notation is standard with weak convergence denoted by $\Rightarrow$ and convergence in probability by $\xrightarrow{p}$.

2. The Model

We begin with a simple time series model with deterministic component d_i

$$y_i = d_i + s_i$$

$$\Delta s_i = \delta s_{i-1} + v_i$$

where the error term v_i has mean zero. However, there are additional stationary covariates which help explain v_i, so that

$$v_i = g(x_i) + \epsilon_i$$

with ϵ_i and x_i iid random variables. Let $g : \mathbb{R}^q \to \mathbb{R}$ and $E(g(x_i)) = \mu_g$. Following Hansen (1995), we define $\sigma^2_{v\epsilon} = E(v_i\epsilon_i)$, $\sigma^2_v = E(v_i^2)$, $\sigma^2_\epsilon = E(\epsilon_i^2)$, and

$$\rho^2 = \frac{\sigma^2_{v\epsilon}}{\sigma^2_v\sigma^2_\epsilon}.$$

First, consider the case where $d_i = \mu$ so that the model becomes

$$\Delta y_i = \mu^* + \delta y_{i-1} + g(x_i) + \epsilon_i, \tag{2.1}$$

where $\mu^* = -\delta\mu - \mu_g$. Following Robinson (1988), we obtain an estimate of δ in two steps. First, we regress Δy_i and y_{i-1} nonparametrically on x_i. The nonparametric estimation uses a Nadaraya-Watson kernel estimator which we illustrate below. Let $k(u)$ be the univariate kernel and we denote $K(u) = \prod_{p=1}^{q} k(u_p)$ if u is q dimensional. In addition, let

$$K_{ij} = K\left(\frac{x_i - x_j}{a}\right)$$

where a is a bandwidth parameter. Then we have

$$\hat{f}_i = (Na^q)^{-1}\sum_{j=1}^{N} K_{ij} \qquad \hat{y}_i = \frac{\sum_{j=1}^{N} K_{ij} y_j}{\sum_{j=1}^{N} K_{ij}} \qquad \hat{\Delta y}_i = \frac{\sum_{j=1}^{N} K_{ij}\Delta y_j}{\sum_{j=1}^{N} K_{ij}}.$$

We "trim" out small values of $\hat{f}_i$ which will appear in the denominator of our nonparamteric estimates. Hence, we define $I_i = I(|\hat{f}_i| > b)$ for some $b > 0$. The residuals from regressing Δy_i and y_{i-1} on x_i are denoted $\hat{e}_d$ and $\hat{e}_y$ respectively. Next, we regress $\hat{e}_d$ on $\hat{e}_y$ using OLS and incorporating the trimming to obtain

$$\hat{\delta} = \left(\sum_{i=1}^{N} \hat{e}_{yi}^2 I_i\right)^{-1} \sum_{i=1}^{N} \hat{e}_{yi}\hat{e}_{di} I_i.$$

Now consider the case where $d_i = \mu + \theta i$ so that the model becomes

$$\Delta y_i = \mu^* + \theta^* i + \delta y_{i-1} + g(x_i) + \epsilon_i, \tag{2.2}$$

where $\mu^* = \theta - \delta\mu - \mu_g$ and $\theta^* = -\delta\theta$. We introduce another term which accounts for the estimated trend,

$$\hat{i} = \frac{\sum_{j=1}^{N} K_{ij} j}{\sum_{j=1}^{N} K_{ij}}.$$

Let the residual from regressing the trend on x_i be denoted $\hat{e}_{ti} = i - \hat{i}$. In order to get an estimate of δ, we regress $\hat{e}_d$ on $\hat{e}_y$ and $\hat{e}_t$ using OLS and incorporating the trimming procedure. Then $\hat{\delta}$ is the estimated coefficient on $\hat{e}_y$.

3. Limiting Distribution

We derive the limiting distribution of our estimator in this section. For purposes of determining asymptotic distributions, we use local to unity asymptotics so that $\delta = -c/N$. Under the null hypothesis of a unit root, $c = 0$, while under $c \neq 0$, the alternative hypothesis becomes increasingly difficult to detect as the sample size increases. We follow convention and denote $W^c(r)$ the solution to the stochastic differential equation

$$dW^c(r) = -cW^c(r) + dW(r),$$

where $W(r)$ is a continuous stochastic process.

The following definitions are given in Robinson (1988).

Definition 1: $\mathcal{K}_l$, $l > 1$, is the class of even functions $k : \mathbb{R} \to \mathbb{R}$ satisfying

$$\int_{\mathbb{R}} u^i k(u)du = \delta_{i0} \qquad (i = 0, 1, \dots, l-1)$$
$$k(u) = O\left((1 + |u|^{l+1+\epsilon})^{-1}\right) \text{ for some } \epsilon > 0.$$

Definition 2: $\mathcal{G}_\mu^\alpha$, $\alpha > 0$, $\mu > 0$, is the class of functions $g : \mathbb{R}^q \to \mathbb{R}$ satisfying: g is $(m-1)$ times partially differentiable, for $(m-1) \le \mu \le m$; for some $\eta > 0$, $\sup_{y \in \phi_{z\eta}} |g(y) - g(z) - Q_g(y, z)|/|y-z|^\mu \le h_g(z)$ for all z, where $\phi_{z\eta} = \{y : |y-z| < \eta\}$; $Q_g = 0$ when $m = 1$, Q_g is an $(m-1)$th degree homogeneous polynomial in $y-z$ with coefficients the partial derivatives of g at z of orders 1 through $m-1$ when $m > 1$; and $g(z)$, its partial derivatives of order $m-1$ and less, and $h_g(z)$, have finite αth moments.

These definitions are used to put conditions on the number of zero moments of the kernel and to provide moment and smoothness conditions for the nonlinear function and the density of the covariate.

We begin by stating a Lemma which is used throughout the proof of the main theorem.

Lemma 3.1. *Let $f(x)$ be the density of x_i. If $\sup_x f(x) < \infty$, $E|g(X)| < \infty$, and $\sup_u |k(u)| + \int |k(u)|du < \infty$, then*

$$\frac{1}{\sqrt{N}}(\hat{y}_i - \bar{y})I_i = o_p(1). \tag{3.1}$$

The above result indicates that if one nonparametrically regresses y_i on x_i, the predicted value behaves asymptotically as if we used the sample mean. This is intuitive because we are attempting to explain a nonstationary series with a stationary series. Since such a regression is inconsistent, the sample mean is the default result. The lemma can be generalized to cases where y_i is generated independently of x_i rather than in the manner suggested in (2.1).

Theorem 3.2. *Suppose the following conditions hold: (i) x_i and ϵ_i are independent and identically distributed; (ii) $E|\epsilon|^4 < \infty$; (iii) x has pdf $f \in \mathcal{G}_\lambda^\infty$, for some $\lambda > 0$; (iv) $g \in \mathcal{G}_\nu^4$, for some $\nu > 0$; (v) as $N \to \infty$, $N^{-1}a^{-2q}b^{-4} \to 0$, $a^{2\min(\lambda+1,\nu)-q}b^{-4} \to 0$; (v) $k \in \mathcal{K}_{l+n-1}$ for integers l and n such that $l-1 < \lambda \leq l$,$n-1 < \nu \leq n$; (vi) $\sigma_v^2 > 0$ and $\rho^2 > 0$; (vii) $\mu_g = 0$. Then*

$$N(\hat{\delta} - \delta) \Rightarrow \left(\int (W_1^\tau)^2\right)^{-1} \left(\rho^2 \int W_1^\tau dW_1 + \sqrt{\rho^2 - \rho^4} \int W_1^\tau dW_2\right), \tag{3.2}$$

where W_2 and W_1 are independent standard Brownian motions, $W_1^\tau = W_1^c(r) - \int W_1^c(s)ds$ for model (2.1) and $W_1^\tau = W_1^c(r) + (6r-4)\int W_1^c(s)ds - (12r-6)\int W_1^c(s)s\ ds$ for model (2.2). The t-statistic based on $\hat{\delta}$ has limiting distrubution

$$t(\hat{\delta}) \Rightarrow -\frac{c}{\rho}\left(\int (W_1^\tau)^2\right)^{\frac{1}{2}} + \left(\int (W_1^\tau)^2\right)^{-\frac{1}{2}} \left(\rho \int W_1^\tau dW_1\right) + \sqrt{1-\rho^2}N(0,1). \tag{3.3}$$

Assumptions (i)-(v) are similar to Robinson (1988). The limiting distribution given in Theorem 3.2 is identical to Theorem 2 in Hansen (1995). However, unlike Hansen, we are unable to estimate a third model, one without a constant term. The reason is the well known fact that in this form of semiparametric estimation, the intercept term is not idenitifed. The apparent lack of identification arises because we have already implicitly estimated an intercept in the nonparametric regression, and no such effect remains. That we cannot estimate a model without an intercept is not a drawback in the nonstationary case since one would at least estimate an intercept even in the simplest unit root test and even if an intercept is not present under the null hypothesis.[1] As further evidence that this effect is indeed accounted for, notice the presence of demeaned Ornstein-Uhlenbeck processes in the limiting distribution of the proposed estimator, as expected from estimating an intercept.

The limiting distribution in Theorem 3.2 also appears in various other related unit root tests. In particular, similiar (or identical) limiting distributions arise in Hasan and Koenker (1997) for their unmodified statistic S_T based on ranks, in Lucas (1995) for unit root tests based on M-estimators, and in Seo (1999) for unit root tests allowing for GARCH effects. Beelders (1999) and Shin and So (1999) also obtained the same limiting distribution for unit root tests when adaptive estimation is used. The distribution has the disadvantage that ρ is a remaining nuisance parameter. There have been various approaches for dealing with the nuisance parameter, ranging from simulating critical values for each value of the parameter to using conservative critical values to cover the range of possible ρ. We use the simulated critical values from Hansen (1995) in a limited Monte Carlo experiment given in the next section.

4. Monte Carlo

We consider several specifications of $g(x)$, both linear and nonlinear to compare the standard Dickey-Fuller test, Hansen's (1995) CADF test, and the new tests using the partial linear model which we denote PLMUR. The data generating process is

$$\Delta y_i = \delta y_{i-1} + g_j(x_i) + \epsilon_i, \quad j = 1, \dots, 5.$$

[1]See Hamilton (1994), chapter 17 for a discussion on inclusion of deterministic terms in tests for unit roots. The case with an estimated intercept when no intercept is present corresponds to case 2 in chapter 17 of Hamilton.

The different functions are listed below.

$$g_1(x) = 0$$

$$g_2(x) = 2x$$

$$g_3(x) = \log(x)$$

$$g_4(x) = x^2 - 1$$

$$g_5(x) = x^3 - x$$

The x variables are all standard normal except in $g_3(x)$ where x is log-normally distributed. When $g_1(x)$ is used, we expect the Dickey-Fuller test to perform the best as there is no x effect to detect. The function $g_2(x)$ gives the CADF test of Hansen the advantage since the covariate enter linearly. The other specifications are nonlinear, so that the PLMUR tests should be more powerful if the nonlinearity is estimated reasonably.

Smaller values of ρ are indicative of the effectiveness of covariates in explaining variation in $v_i = g(x_i) + \epsilon_i$. Therefore, we expect more powerful tests if ρ is small. Straightforward calculations show that

$$\rho_1^2 = 1$$

$$\rho_2^2 = 0.20$$

$$\rho_3^2 = 0.50$$

$$\rho_4^2 = 0.25$$

$$\rho_5^2 = 0.10$$

where ρ_j^2 is associated with $g_j(x)$.

For the PLMUR test, we need to select a kernel and a bandwidth. In our experiment, we chose the Epanechnikov kernel given by

$$k(u) = \begin{cases} .75(1-u^2) & \text{if } u \in [-1,1], \\ 0 & \text{otherwise.} \end{cases}$$

The bandwidth was set to $N^{-\frac{1}{5}}$ for all specifications and sample sizes.

The PLMUR test and the CADF test both require estimates of ρ. We compute these using the residuals from each of the regressions and then use the resulting estimate to select a critical value from Table 1 in Hansen (1995). We explore size and power by changing the value of c in $\delta = -\frac{c}{N}$. For each specification, we generate sample sizes of 100 and 200 and compute 10,000 replications.[2] The results appear in Table 1.

For $c = 0$, we have a unit root and we compare the size for each of the tests. The DF test has size close to the nominal 5% for all choices of $g(x)$ but the CADF is actually undersized for the linear, log, and cubic cases. The PLMUR test is slightly oversized when there are no covariates in the data generating process, yet this distortion is mitigated by the increase in sample size. The size result for the PLMUR test indicates that the asymptotic theory provides an accurate approximation for the distribution of the statistic.

For $c = 3$, the departure from the unit root becomes apparent in the increased rejection frequencies. All of the tests perform similarly when there is no covariate effect, indicating that little is lost when estimating a partial linear model even when it is not warranted. Moreover, for g_2, the linear effect, the PLMUR test competes favorably with the CADF test, suggesting that when there is a linear effect, the loss in using the more general PLMUR test is small as well. The advantage of the PLMUR test becomes apparent when the log specification is considered. Power is roughly

[2]The computations took around 10 days on a Pentium 600 computer. The programs were written in Ox 2.0, see Doornik (1998).

double the competing tests here as the covariate is successfully used to reduce the variance of the estimator of δ. For the quadratic function, the difference is more pronounced with the PLMUR test generating triple the power of the competing tests. Finally, using the cubic function, power is six times that of the other tests since the true value of ρ is .10 in this case. Similar results are obtained for the other local alternatives of $c = 6$ and $c = 9$ with all tests increasing in power as the alternative becomes more obvious. In all cases where covariates are correctly chosen, both the CADF and the PLMUR test dominate the standard Dickey-Fuller tests. In all cases where there is a nonlinear effect, the PLMUR test is the most powerful, with power increasing as ρ decreases.

5. Conclusions

We have proposed a unit root test based on estimating a partial linear model where a covariate enters the model nonlinearly. The relevant asymptotic theory was developed and we provided a limited Monte Carlo experiment to examine the performance of the test. The results indicate that the test effectively exploits the nonlinear effect to increase power substantially. In addition, it appears that in our simple case, estimating a partial linear model is benign even in cases where the covariates do not have a nonlinear effect.

There are several issues which remain. First, an extension to allow non iid settings is necessary. An extension to an augmented Dickey-Fuller test is straightforward since the coefficients of the lagged Δy_i terms will converge at a slower rate than the nonstationary components. A more ambitious project is to allow dependence in x_i. The complication arises because the kernels will be evaluated at values of $x_i - x_j$ which will depend on the difference $i - j$. As a practical matter, the issue of bandwidth selection needs to be treated carefully, with the development of some type of cross-validation procedure. Finally, an obvious extension to the multivariate case of cointegration is possible. This is currently being undertaken by the authors.

TABLE 1: Size and Power

		N=100			N=200		
		DF	CADF	PLMUR	DF	CADF	PLMUR
	g_1	0.0457	0.0478	0.0810	0.0484	0.0495	0.0670
	g_2	0.0487	0.0263	0.0540	0.0484	0.0213	0.0479
$c = 0$	g_3	0.0464	0.0339	0.0660	0.0482	0.0362	0.0594
	g_4	0.0498	0.0493	0.0568	0.0512	0.0510	0.0507
	g_5	0.0530	0.0391	0.0400	0.0531	0.0371	0.0410
	g_1	0.0940	0.0950	0.1260	0.0894	0.0916	0.1130
	g_2	0.0894	0.3639	0.4638	0.0856	0.3368	0.4661
$c = 3$	g_3	0.0872	0.1071	0.2245	0.0886	0.1021	0.2039
	g_4	0.0857	0.0878	0.3056	0.0812	0.0832	0.2978
	g_5	0.0702	0.1036	0.6382	0.0693	0.0949	0.6822
	g_1	0.1671	0.1707	0.1998	0.1505	0.1530	0.1734
	g_2	0.1589	0.7897	0.8460	0.1537	0.7789	0.8547
$c = 6$	g_3	0.1640	0.2594	0.4692	0.1525	0.2365	0.4424
	g_4	0.1620	0.1679	0.6457	0.1456	0.1517	0.6496
	g_5	0.1431	0.2777	0.9140	0.1365	0.2605	0.9480
	g_1	0.2956	0.2970	0.3366	0.2707	0.2720	0.2910
	g_2	0.3043	0.9569	0.9780	0.2760	0.9559	0.9747
$c = 9$	g_3	0.3019	0.4812	0.7052	0.2745	0.4532	0.7046
	g_4	0.2866	0.3039	0.8667	0.2725	0.2822	0.8714
	g_5	0.2673	0.5282	0.9800	0.2602	0.4907	0.9923

APPENDIX A.

To begin, use $y_i = \sum_{l=0}^{i}(1 - c/N)^{i-l}(g(x_l) + \epsilon_l)$. Without loss of generality, we find the convergence rates for $c = 0$, the case of a unit root. Notice that we also are assuming that the initial value is $y_0 = 0$.

Proof of Lemma 3.1: The proof is given in a technical appendix available from the authors upon request.

APPENDIX B.

We prove Theorem 3.2 for model (2.1) in this appendix. Note that $N(\hat{\delta} - \delta) = D^{-1}(E + F))$ where

$$D = \frac{1}{N^2} \sum_{i=1}^{N} (y_{i-1} - \hat{y}_{i-1})^2 I_i$$

$$E = \frac{1}{N} \sum_{i-1} (y_{i-1} - \hat{y}_{i-1}) (g(x_i) - \hat{g}(x_i) + \hat{\epsilon}_i) I_i$$

$$F = \frac{1}{N} \sum_{i-1} (y_{i-1} - \hat{y}_{i-1}) \epsilon_i I_i$$

The theorem holds if we can show that

$$D \Rightarrow \sigma_v^2 \int (W_1^\tau(s))^2 \, ds,$$

$$E \xrightarrow{p} 0,$$

$$F \Rightarrow \sigma_v \sigma_\epsilon \left(\rho \int W_1^\tau dW_1 + \sqrt{1-\rho^2} \int W_1^\tau dW_2 \right)$$

We prove these results in a series of six propositions.

Proposition 1.

$$\text{(B.1)} \qquad \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i) \epsilon_i I_i \Rightarrow \sigma_v \sigma_\epsilon \left(\rho \int W_1^\tau dW_1 + \sqrt{1-\rho^2} \int W_1^\tau dW_2 \right)$$

Proof of Proposition 1: The proof follows if

$$\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - \bar{y}) \epsilon_i I_i \xrightarrow{p} 0,$$

which we show by

$$E \left(\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - \bar{y}) \epsilon_i I_i \right)^2 \to 0.$$

$$E\left(\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i-\bar{y})\epsilon_i I_i\right)^2 = E\left(\underbrace{\frac{1}{N^2}\sum_{i=1}(\hat{y}_i-\bar{y})^2\epsilon_i^2 I_i}_{A}+\underbrace{\frac{1}{N^2}\sum\sum_{i\neq j}(\hat{y}_i-\bar{y})(\hat{y}_j-\bar{y})\epsilon_i\epsilon_j I_i I_j}_{B}\right)$$

We have $E\frac{1}{N}(\hat{y}_i-\bar{y})^2 I_i = O(N^{-1}a^{-q}b^{-2})$ so that $E(A)=O(N^{-1}a^{-q}b^{-2})$.

For part B, we find the order of the N^2 terms of the form

$$\begin{aligned}&E(\hat{y}_i-\bar{y})(\hat{y}_j-\bar{y})\epsilon_i\epsilon_j\\&\quad = E\left(\frac{1}{N}\sum_{k=1}^{N}\left(\frac{NK_{ik}-\sum_{l=1}^{N}K_{il}}{\sum_{l=1}^{N}K_{il}}\right)y_k\right)\left(\frac{1}{N}\sum_{m=1}^{N}\left(\frac{NK_{jm}-\sum_{l=1}^{N}K_{jl}}{\sum_{l=1}^{N}K_{jl}}\right)y_m\right)\epsilon_i\epsilon_j\end{aligned}$$

Conditioning on $X_N=(x_0,\dots,x_N)$ and taking expectations, we find that the order is the same as the order of

$$E\frac{1}{N}\sum_{k=1}^{N}\left(\frac{NK_{ik}-\sum_{l=1}^{N}K_{il}}{\sum_{l=1}^{N}K_{il}}\right)\frac{1}{N}\sum_{m=1}^{N}\left(\frac{NK_{jm}-\sum_{l=1}^{N}K_{jl}}{\sum_{l=1}^{N}K_{jl}}\right). \tag{B.2}$$

By identity of distribution, (B.2) has N^2 terms of the form

$$\frac{1}{N^2}E\left(\frac{-2NK_{ik}K_{jm}-2NK_{ik}K_{jk}}{\sum_{l=1}^{N}K_{il}\sum_{l=1}^{N}K_{jl}}\right).$$

Taking the expected value of the absolute value the above term is $O(N^{-3}a^{-q}b^{-2})$, implying that $E(B)=O(N^{-1}a^{-q}b^{-2})$. □

Proposition 2.

$$\frac{1}{N^2}\sum_{i=1}^{N}(y_i-\bar{y})(\bar{y}-\hat{y}_i)I_i \xrightarrow{p} 0.$$

Proof of Proposition 2:

$$\frac{1}{N^4}E\left(\sum_{i=1}^{N}(y_i-\bar{y})(\bar{y}-\hat{y}_i)I_i\right)^2 \leq \frac{1}{N^2}\sum_{i=1}^{N}E(y_i-\bar{y})^2E\frac{1}{N}(\bar{y}-\hat{y}_i)^2I_i \tag{B.3}$$

by Loève's c_r inequality and Cauchy-Schwartz. Since $E\frac{1}{N}(\bar{y}-\hat{y}_i)^2I_i = O(N^{-1}a^{-q}b^{-2})$ and $E(y_i-\bar{y})^2=O(N)$, then (B.3) is $O(N^{-1}a^{-q}b^{-2})$. □

Proposition 3.

$$\frac{1}{N^2}\sum_{i=1}^{N}(\bar{y}-\hat{y}_i)^2 I_i \xrightarrow{p} 0$$

Proof of Proposition 3: The proof is given in a technical appendix available from the authors upon request.

Proposition 4.

$$\frac{1}{N^2}\sum_{i=1}^{N}(y_{i-1}-\hat{y}_{i-1})^2 I_i \Rightarrow \int (W_1-\bar{W}_1)^2$$

Proof of Proposition 4:

$$\frac{1}{N^2}\sum_{i=1}^{N}(y_{i-1}-\hat{y}_{i-1})^2 I_i = \frac{1}{N^2}\sum_{i=1}^{N}(y_{i-1}-\bar{y})^2 + 2\frac{1}{N^2}\sum_{i=1}^{N}(y_{i-1}-\bar{y})(\bar{y}-\hat{y}_i) I_i + \frac{1}{N^2}\sum_{i=1}(\bar{y}-\hat{y}_i)^2 I_i$$

The second and third terms on the right hand side converge to zero by Propositions 2 and 3. □

Proposition 5.

$$\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)\hat{\epsilon}_i I_i \xrightarrow{p} 0.$$

Proof of Proposition 5: The proof appears in a technical appendix available from the authors upon request.

Proposition 6.

$$\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)(g(x_i)-\hat{g}(x_i))I_i \xrightarrow{p} 0.$$

Proof of Proposition 6: The proof appears in a technical appendix available from the authors upon request.

Appendix C.

We prove Theorem 3.2 for model (2.2). We define

$$w_i = \theta i + y_i$$

so that the relationship between the previous propositions and those given in this appendix is more transparent. The joint estimators for (δ, β) are given by

$$\begin{pmatrix} \hat{\delta} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \hat{e}_w^\top \hat{e}_w & \hat{e}_w^\top \hat{e}_t \\ \hat{e}_t^\top \hat{e}_w & \hat{e}_t^\top \hat{e}_t \end{pmatrix}^{-1} \begin{pmatrix} \hat{e}_w^\top \hat{e}_{\Delta w} \\ \hat{e}_i^\top \hat{e}_{\Delta w} \end{pmatrix},$$

where $\hat{e}_w$, $\hat{e}_{\Delta w}$, and $\hat{e}_t$ are the residuals from nonparametrically regressing w, Δw, and a trend on x_i. A rotation of the type in Fuller (1976) is applied to facilitate the proof.

$$\begin{pmatrix} N & 0 \\ -\theta N^{\frac{3}{2}} & N^{\frac{3}{2}} \end{pmatrix} \begin{pmatrix} \hat{\delta} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \frac{(\hat{e}_w - \theta \hat{e}_t)^\top (\hat{e}_w - \theta \hat{e}_t)}{N^2} & \frac{(\hat{e}_w - \theta \hat{e}_t)^\top \hat{e}_t}{N^{\frac{5}{2}}} \\ \frac{\hat{e}_t^\top (\hat{e}_w - \theta \hat{e}_t)}{N^{\frac{5}{2}}} & \frac{\hat{e}_t^\top \hat{e}_t}{N^3} \end{pmatrix} \begin{pmatrix} \frac{(\hat{e}_w - \theta \hat{e}_t)^\top \hat{e}_{\Delta w}}{N} \\ \frac{\hat{e}_t^\top \hat{e}_{\Delta w}}{N^{\frac{3}{2}}} \end{pmatrix}$$

We show the (joint) convergence of the above terms in a series of propositions. Notice that using the rotation above, we generate

$$\begin{aligned} \hat{e}_{wi} - \theta \hat{e}_{ti} &= w_i - \hat{w}_i - \theta (i - \hat{i}) \\ &= \theta i + y_i - \theta \hat{i} - \hat{y}_i - \theta i + \theta \hat{i} \\ &= y_i - \hat{y}_i. \end{aligned} \tag{C.1}$$

This allows us to use some of the results in Appendix B in the remaining proofs.

Proposition 7.

$$\frac{1}{N^2} \sum_{i=1}^{N} (\hat{e}_{wi} - \theta \hat{e}_{ti})^2 I_i \xrightarrow{d} \int \left(W_1 - \bar{W}_1 \right)^2 \tag{C.2}$$

Proof of Proposition 7 The result follows by Proposition 4.

Proposition 8.

$$E\left(\frac{\hat{i}}{N}-\frac{1}{2}\right)^2 I_i = O_p(N^{-1}a^{-q}b^{-2}) \tag{C.3}$$

Proof of Proposition 8

$$E\left(\frac{\sum_j K_{ij}\left(\frac{j}{N}\right)}{\sum_j K_{ij}}-\frac{1}{2}\right)^2 I_i \le (Na^q b)^{-2} E\left(\sum_j K_{ij}(\frac{j}{N}-\frac{1}{2})\right)^2.$$

Then

$$E\left(\sum_j K_{ij}(\frac{j}{N}-\frac{1}{2})\right)^2 = \underbrace{\sum_j EK_{ij}^2\left(\frac{j}{N}-\frac{1}{2}\right)^2}_{A} + \underbrace{\sum\sum_{j\neq l} EK_{ij}K_{il}\left(\frac{j}{N}-\frac{1}{2}\right)\left(\frac{l}{N}-\frac{1}{2}\right)}_{B}$$

Part A is $O(Na^q)$. Notice that we can pull out $EK_{ij}K_{il}$ by identity of distribution and note

$$\begin{aligned}\sum\sum_{j\neq l}\left(\frac{j}{N}-\frac{1}{2}\right)\left(\frac{l}{N}-\frac{1}{2}\right) &= \left(\sum_j \frac{j}{N}-\frac{1}{2}\right)^2 - \sum_j\left(\frac{j}{N}-\frac{1}{2}\right)^2 \\ &= \frac{1}{4N^2}-\left(\frac{N}{3}-\frac{N+1}{2}+\frac{N}{4}\right) \\ &= O(N),\end{aligned}$$

so that part B is $O(Na^q)$. $\square$

Proposition 9.

$$N^{-\frac{5}{2}}\sum_{i=1}^{N}(\hat{e}_{wi}-\theta\hat{e}_{ti})\hat{e}_{ti}I_i \xrightarrow{d} \int\left(W_1-\bar{W}_1\right)sds \tag{C.4}$$

Proof of Proposition 9

$$N^{-\frac{5}{2}}\sum_{i=1}^{N}(\hat{e}_{yi}-\theta\hat{e}_{ti})\hat{e}_{ti}I_i = N^{-\frac{5}{2}}\sum_{i=1}^{N}(y_i-\hat{y}_i)iI_i - N^{-\frac{5}{2}}\sum_{i=1}^{N}(y_i-\hat{y}_i)\hat{i}I_i.$$

We first write

$$N^{-\frac{5}{2}}\sum_{i=1}^{N}(y_i-\hat{y}_i)i = N^{-\frac{5}{2}}\sum_{i=1}^{N}(y_i-\bar{y})i + N^{-\frac{5}{2}}\sum_{i=1}^{N}(\bar{y}-\hat{y}_i)i.$$

Then

$$E\left(N^{-\frac{5}{2}}\sum_{i=1}^{N}(\bar{y}-\hat{y}_i)i\right)^2 = N^{-5}E\sum_{i=1}^{N}(\bar{y}-\hat{y}_i)^2 i^2 I_i + N^{-5}E\sum\sum_{i\neq j}(\bar{y}-\hat{y}_i)(\bar{y}-\hat{y}_j)ijI_iI_j$$

Using the Cauchy-Schwartz inequality and the fact that $E(\bar{y}-\hat{y}_i)^2 I_i = O(a^{-q}b^{-2})$, the above term is $O(N^{-1}a^{-q}b^{-2})$. It is well known that

$$N^{-\frac{5}{2}}\sum_{i=1}^{N}(y_i-\bar{y})i \Rightarrow \int(W_1-\bar{W}_1)sds,$$

so it remains to show that

$$N^{-\frac{5}{2}}\sum_{i=1}^{N}(y_i-\hat{y}_i)\hat{i}I_i \xrightarrow{p} 0. \tag{C.5}$$

It is easy to see that

$$N^{-\frac{3}{2}}\sum_{i=1}^{N}(y_i-\hat{y}_i)\frac{1}{2}I_i \xrightarrow{p} 0$$

using Lemma 3.1 so it is enough to show that

$$N^{-\frac{3}{2}}\sum_{i=1}^{N}(y_i-\hat{y}_i)(\frac{\hat{i}}{N}-\frac{1}{2})I_i \xrightarrow{p} 0.$$

$$N^{-\frac{3}{2}}\sum_{i=1}^{N}(y_i-\hat{y}_i)(\frac{\hat{i}}{N}-\frac{1}{2})I_i = \underbrace{N^{-\frac{3}{2}}\sum_{i=1}^{N}(y_i-\bar{y})(\frac{\hat{i}}{N}-\frac{1}{2})I_i}_{A} + \underbrace{N^{-\frac{3}{2}}\sum_{i=1}^{N}(\bar{y}-\hat{y}_i)(\frac{\hat{i}}{N}-\frac{1}{2})I_i}_{B}$$

The expectation of part A squared is

$$N^{-3}\sum_{i=1}^{N}E(y_i-\bar{y})^2(\frac{\hat{i}}{N}-\frac{1}{2})^2 I_i + N^{-3}\sum\sum_{i\neq j}E(y_i-\bar{y})(y_j-\bar{y})(\frac{\hat{i}}{N}-\frac{1}{2})(\frac{\hat{j}}{N}-\frac{1}{2})I_iI_j$$

This term is clearly $O(N^{-1}a^{-q}b^{-2})$ from Proposition 8. The expectation of part B squared is

$$N^{-3}\sum_{i=1}^{N}E(\bar{y}-\hat{y}_i)^2(\frac{\hat{i}}{N}-\frac{1}{2})^2 I_i + N^{-3}\sum\sum_{i\neq j}E(\bar{y}-\hat{y}_i)(\bar{y}-\hat{y}_j)(\frac{\hat{i}}{N}-\frac{1}{2})(\frac{\hat{j}}{N}-\frac{1}{2})I_iI_j$$

which is seen to be $O(N^{-2}a^{-2q}b^{-4})$. □

Proposition 10.

$$N^{-3}\sum_{i=1}^{N}\hat{e}_{ti}^2 I_i \xrightarrow{p} \frac{1}{12}$$

Proof of Proposition 10

$$N^{-3}\sum_{i=1}^{N}\hat{e}_{ti}^2 = N^{-2}\sum_{i=1}^{N}\left(\frac{i}{N}-\frac{1}{2}+\frac{1}{2}-\frac{\hat{i}}{N}\right)^2$$

which can be written as

$$N^{-2}\sum_{i=1}^{N}\left(\left(\frac{i}{N}-\frac{1}{2}\right)^2+\left(\left(\frac{i}{N}-\frac{1}{2}\right)\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)\right)+\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)^2\right)$$

For the first term,

$$N^{-2}\sum_{i=1}^{N}\left(\frac{i}{N}-\frac{1}{2}\right)^2 \to \frac{1}{12}$$

since $N^{-2}\sum_{i}^{N} i \to 1/2$ and $N^{-3}\sum_{i=1}^{N} i^2 \to 1/3$. Next,

$$E\left(N^{-2}\sum_{i=1}^{N}\left(\frac{i}{N}-\frac{1}{2}\right)\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)\right)^2 = N^{-4}E\sum_{i=1}^{N}\left(\frac{i}{N}-\frac{1}{2}\right)^2\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)^2$$
$$+N^{-4}E\sum\sum_{i\neq j}\left(\frac{i}{N}-\frac{1}{2}\right)\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)\left(\frac{j}{N}-\frac{1}{2}\right)\left(\frac{1}{2}-\frac{\hat{j}}{N}\right)$$

which is $O(N^{-1}a^{-q}b^{-2})$ by Proposition 8 and Cauchy-Schwartz. Then

$$E\left(\sum_{i=1}^{N}\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)^2\right)^2 = \underbrace{N^{-4}E\sum_{i=1}^{N}\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)^4}_{A}+\underbrace{N^{-4}E\sum\sum_{i\neq l}\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)^2\left(\frac{1}{2}-\frac{\hat{l}}{N}\right)^2}_{B}$$

For part A, consider

$$\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)^4 \leq (Na^q b)^{-4}E\left(\sum_{j=1}^{N}K_{ij}\left(\frac{j}{N}-\frac{1}{2}\right)\right)^4.$$

Using Loève's c_r inequality, we have

$$E\left(\sum_{j=1}^{N} K_{ij}(\frac{i}{N}-\frac{1}{2})\right)^4 \leq N^3 \sum_{j=1}^{N} K_{ij}^4(\frac{j}{N}-\frac{1}{2})^4,$$

so that part A is $O(N^{-3}a^{-3q}b^{-4})$. By Cauchy-Schwartz and Proposition 8, part B is $O(N^{-4}a^{-2q}b^{-4})$. □

Proposition 11.

$$N^{-\frac{3}{2}}\sum_{i=1}^{N}(i-\hat{i})\epsilon_i I_i \xrightarrow{d} \int(s-\frac{1}{2})dW_2(s)$$

Proof of Proposition 11: The proof is completed by showing that

$$N^{-\frac{1}{2}}\sum_{i=1}^{N}\left(\frac{\hat{i}}{N}-\frac{1}{2}\right)\epsilon_i \xrightarrow{p} 0.$$

Conditioning on X_N and using the independence of ϵ_i gives

$$E\left(N^{-\frac{1}{2}}\sum_{i=1}^{N}\left(\frac{\hat{i}}{N}-\frac{1}{2}\right)\epsilon_i\right)^2 = N^{-1}E\sum_{i=1}^{N}\left(\frac{\hat{i}}{N}-\frac{1}{2}\right)^2\epsilon_i^2$$

which is $O(N^{-1}a^{-q}b^{-2})$. □

Proposition 12.

$$N^{-\frac{3}{2}}\sum_{i=1}^{N}(i-\hat{i})\hat{\epsilon}_i I_i \xrightarrow{p} 0$$

Proof of Proposition 12: Again, we break this term into two parts:

$$N^{-\frac{3}{2}}\sum_{i=1}^{N}(i-\hat{i})\hat{\epsilon}_i I_i = \underbrace{N^{-\frac{1}{2}}\sum_{i=1}^{N}(\frac{i}{N}-\frac{1}{2})\hat{\epsilon}_i I_i}_{A} + \underbrace{N^{-\frac{1}{2}}\sum_{i=1}^{N}(\frac{1}{2}-\frac{\hat{i}}{N})\hat{\epsilon}_i I_i}_{B}.$$

For the part A,

$$E\left(N^{-\frac{1}{2}}\sum_{i=1}^{N}(\frac{i}{N}-\frac{1}{2})\hat{\epsilon}_i I_i\right)^2 = N^{-1}E\sum_{i=1}^{N}(\frac{i}{N}-\frac{1}{2})^2\hat{\epsilon}_i^2 I_i$$
$$+ N^{-1}E\sum\sum_{i\neq l}(\frac{i}{N}-\frac{1}{2})(\frac{i}{N}-\frac{1}{2})\hat{\epsilon}_i\hat{\epsilon}_l I_i I_l$$

The first part of the above term is $O(N^{-1}a^{-q}b^{-2})$. Using the fact that $E\hat{\epsilon}_i\hat{\epsilon}_l I_i I_l$ is identical for $i \neq l$ and the fact that $\sum\sum_{i\neq l}(i/N - 1/2)(l/N - 1/2) = O(N)$, part A is $O_p(N^{-1/2}a^{-q/2}b^{-1})$.

For part B,

$$E\left(N^{-\frac{1}{2}}\sum_{i=1}^{N}(\frac{1}{2}-\frac{\hat{i}}{N})\hat{\epsilon}_i I_i\right)^2 = N^{-1}E\sum_{i=1}^{N}(\frac{1}{2}-\frac{\hat{i}}{N})^2\hat{\epsilon}_i^2 I_i$$
$$+ N^{-1}E\sum\sum_{i\neq l}(\frac{1}{2}-\frac{\hat{i}}{N})(\frac{1}{2}-\frac{\hat{l}}{N})\hat{\epsilon}_i\hat{\epsilon}_l I_i I_l$$

Using Cauchy-Schwartz, (**??**), and Proposition 8, part B is $O_p(N^{-1/2}a^{-q}b^{-2})$. □

Proposition 13.

$$N^{-\frac{3}{2}}\sum_{i=1}^{N}(i-\hat{i})(g(x_i)-\hat{g}(x_i))I_i \xrightarrow{p} 0$$

Proof of Proposition 13: The proof is identical to the proof of Proposition 12 with the exception that Proposition 1 of Robinson is applied to $g(x_i) - \hat{g}(x_i)$ as (**??**) was applied to $\hat{\epsilon}_i$. The order is $O_p(N^{-1/2}a^{-q/2}b^{-1} + a^{\zeta}b^{-1})$ where $\zeta = \min(\lambda + 1, \nu)$ as in Proposition 6. □

REFERENCES

BEELDERS, O. (1999): "Adaptive Estimation and Unit Root Tests," Emory University.

BOSWIJK, H. P., AND A. LUCAS (1997): "Semi-nonparametric Cointegration Testing," Vrije Universiteit, Amsterdam.

DOORNIK, J. A. (1998): *Ox-An Object Oriented Matrix Programming Language*. London: Timberlake Consultants Press.

ELLIOT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

HANSEN, B. E. (1995): "Rethinking the Univariate Approach to Unit Root Testing," *Econometric Theory*, 11, 1148–1171.

HASAN, M. N., AND R. W. KOENKER (1997): "Robust Rank Tests of the Unit Root Hyptothesis," *Econometrica*, 65, 133–161.

LUCAS, A. (1995): "Unit Root Tests Based on M Estimators," *Econometric Theory*, 11, 331–346.

——— (1997): "Cointegration Testing Using Pseudolikelihood Ratio Tests," *Econometric Theory*, 13, 149–169.

PHILLIPS, P. C. B. (1995): "Robust Nonstationary Regression," *Econometric Theory*, 11, 912–951.

PHILLIPS, P. C. B., AND J. Y. PARK (1998): "Nonstationary Density Estimation and Kernel Autoregression," Yale University.

ROBINSON, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

SEO, B. (1998): "Statistical Inference on Cointegrating Rank in Error Correction Models with Stationary Covariates," *Journal of Econometrics*, 85, 339–385.

——— (1999): "Distribution Theory for Unit Root Tests with Conditional Heteroskedasticity," *Journal of Econometrics*, 91, 113–144.

SHIN, D. W., AND B. S. SO (1999): "Unit Root Tests Based on Adaptive Maximum Likelihood Estimation," *Econometric Theory*, 15, 1–23.

XIAO, Z., AND P. C. B. PHILLIPS (1999): "Efficient Detrending in Cointegrating Regression," *Econometric Theory*, 15, 519–548.