Econometric Theory, **21**, 2005, 877–906. Printed in the United States of America.
DOI: 10.1017/S0266466605050450

PARTIALLY LINEAR MODELS WITH UNIT ROOTS

TED JUHL
University of Kansas
ZHIJIE XIAO
Boston College

This paper studies the asymptotic properties of a nonstationary partially linear regression model. In particular, we allow for covariates to enter the unit root (or near unit root) model in a nonparametric fashion, so that our model is an extension of the semiparametric model analyzed in Robinson (1988, *Econometrica* 56, 931–954). It is proved that the autoregressive parameter can be estimated at rate N even though part of the model is estimated nonparametrically. Unit root tests based on the semiparametric estimate of the autoregressive parameter have a limiting distribution that is a mixture of a standard normal and the Dickey–Fuller distribution. A Monte Carlo experiment is conducted to evaluate the performance of the tests for various linear and nonlinear specifications.

1. INTRODUCTION

In recent years, statistical models incorporating nonlinearity have received increased attention in econometrics. One type of these models is the following partially linear regression:

$$y_t = \gamma' z_t + g(x_t) + \epsilon_t, \qquad t = 1, \ldots, N, \tag{1.1}$$

where $g(\cdot)$ is an unknown real function and γ is the vector of unknown parameters that we want to estimate. This type of specification arises when the primary interest is in the parameter γ, whereas the relation of the mean response to additional variables x_t is not easily parameterized. Such a strategy provides an intermediate class of models that are more flexible than standard linear regression, with the potential for greater precision than purely nonparametric models.

There is a large literature in econometrics and statistics on the study of partially linear models. Wahba (1986), Engle, Granger, Rice, and Weiss (1986), and many others studied the penalized least squares method in partially linear

We thank Bruce Hansen, Roger Koenker, Helmut Lütkepohl, Peter Phillips, three referees, and participants of the 8th World Congress of the Econometric Society and the 10th Midwest Econometrics Group Meeting for helpful comments on an earlier version of this paper. This investigation was supported by the University of Kansas General Research Fund allocation 2301789-003. Address correspondence to Zhijie Xiao, Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467, USA; e-mail: xiaoz@bc.edu.

© 2005 Cambridge University Press 0266-4666/05 $12.00

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

regression estimation. Heckman (1986), Chen (1988), Rice (1986), and Speckman (1988) studied $\sqrt{N}$-consistency of γ under different assumptions and using various methods. Using higher order Nadaraya–Watson kernel estimates to eliminate the unknown function, Robinson (1988) introduced a feasible least squares estimator for γ. Under regularity and smoothness conditions, $\sqrt{N}$-consistency and asymptotic normality are obtained. Robinson also showed that when the errors are independently and identically distributed (i.i.d.) normal, this estimator achieves the semiparametric information bound. A higher order asymptotic analysis of the partially linear regression estimators is given by Linton (1995). Fan, Li, and Stengos (1995) extended the $\sqrt{N}$-consistency and asymptotic normality results for partially linear models with conditionally heteroskedastic disturbances. In time series, Fan and Li (1999) established these results for models with weakly dependent process x_t. For other work on partially linear regression models, see Chen (1988), Shiller (1984), and Schick (1986), among others. However, all of the previous studies have been focused on either i.i.d. or stationary cases, and, to our knowledge, there has been no study in the existing literature on nonstationary partially linear models. The current paper attempts to provide a first step toward investigation of such models. In particular, we consider a partially linear model with a unit root.

Unit root models have been an important subject in econometric analysis and have attracted a large amount of research effort in the last 15 years. Testing for the presence of unit roots is now a common practice in applied macroeconomics. Although the unit root hypothesis has been tested in hundreds of time series, it is well known that the discriminatory power of unit root tests is generally low. As a result, increasing power in unit root tests has become an important research topic. There has been a branch of unit root literature that uses various features of the time series data to increase power in recent years. For example, 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. 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. Elliot, Rothenberg, and Stock (1996) propose an estimation strategy that focuses on estimating potential trends under the local alternative hypothesis to effectively reach the Gaussian power envelope for unit root tests. 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 generalized autoregressive conditional heteroskedasticity (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 power gains, particularly if the error terms are heavy-tailed. In the current paper, we propose a partially linear unit root model to improve the power of unit root tests.

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

Although many observed time series seem to display nonstationary characteristics, nonlinearity also seems to be an important feature in a range of applications. In fact, a lot of economic models do contain nonlinear elements (see, e.g., among others, Tong, 1990; Granger and Terasvirta, 1993; Granger, 1995). For this and other reasons, one of the directions in which the subject is presently moving is the study of nonstationary models with nonlinearity. In particular, to treat potential nonlinearities, Phillips and Park (1998) studied nonlinear autoregressive models and showed that the nonparametric estimator of the autoregressive function converges at rate $N^{1/4}$ in the unit root case.

In this paper, we consider a partially linear autoregression with nonstationarity:

$$\Delta y_t = \delta y_{t-1} + \sum_{j=1}^{p} \pi_j \Delta y_{t-j} + g(x_t) + \epsilon_t,$$

where δ is close to zero and x_t is a vector of stationary covariates. When δ is exactly zero, y_t follows a unit root process. Otherwise, it is characterized as a near unit root process. In this model, our primary interest is still the estimation and test of the parameter δ, but we allow for an unknown nonlinear function of covariates, $g(x_t)$, to influence the time series. The model can be viewed as a semiparametric extension of the covariate augmented Dickey-Fuller (CADF) regression of Hansen (1995). In allowing such a general structure, we hope to improve the efficiency in estimating the autoregressive (AR) parameter and further increase the power gains of unit root tests from using covariates, particularly if there is a nonlinear relationship with the chosen covariate. Because the form of the nonlinearity is unknown, we estimate this part of the model nonparametrically while modeling the autoregressive component linearly. The motivation stems from the fact that we are adding variables to the model with the hope that they may explain some of the variation in y_t. In this exercise, we have little information about the influence of series x_t on y_t so it is natural to refrain from taking a stand on the functional form for $g(x_t)$. As we illustrate in the paper, the power loss associated with nonparametrically estimating the function $g(x_t)$ (relative to a correct specification) is small, yet the power gain from nonparametric estimation relative to an incorrect specification is quite large.

The proposed estimation strategy parallels that of Robinson (1988). However, the technical issues addressed here are different than those treated in the stationary case. In particular, we must bound the average of the difference between an integrated variable and its local average over the x values. Moreover, we must show that functions of the local average converge to known (nonstandard) distributions.

There are several findings in this paper. First, by using the compromise of a partially linear model, the convergence for the autoregressive component remains at rate N. This is an important extension of the result of Robinson (1988) to the nonstationary case. In addition, asymptotic distributions of partially linear estimates of δ and its modified t-statistic are derived. The limiting distribution of

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

the resulting unit root test is identical to the distribution found in Hansen (1995), where covariates are used in a linear fashion. A limited Monte Carlo experiment reveals that there is little loss in using our more general test statistic when covariates enter the model linearly or not at all, and the power gains from using our partially linear model when there are nonlinear effects are substantial. Finally, in the course of proving our theorem, we show that, in the density-weighted regression, nonparametrically regressing an I(1) series on an I(0) series is asymptotically equivalent to an ordinary least squares (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 estimator and the test statistic. Extensions to the case with weakly dependent covariates are discussed in Section 4. Section 5 reports some Monte Carlo results, and Section 6 applies the proposed tests to U.S. monthly macroeconomic time series. Notation is standard with weak convergence denoted by $\Rightarrow$ and convergence in probability by $\xrightarrow{p}$. Hadamard multiplication is indicated by the symbol $\odot$. Integrals with respect to Lebesgue measure such as $\int_0^1 W(s)\,ds$ are usually written as $\int_0^1 W$, or simply $\int W$ when there is no ambiguity over limits. All limits in the paper are taken as the sample size $N \to \infty$, except as otherwise noted.

2. THE MODEL AND THE ESTIMATOR

We begin with the following time series model:

$$\Delta y_t = \mu + \delta y_{t-1} + \sum_{j=1}^{p} \pi_j \Delta y_{t-j} + v_t \tag{2.1}$$

and denote the pth order polynomial $\Pi(L) = (1 - \pi_1 L - \cdots - \pi_p L^p)$. It is assumed that all roots of $\Pi(L)$ lie outside the unit circle. We are interested in the case where the largest autoregressive root of y_t is close to unity. Thus we focus our discussion on the case where δ is close to zero and assume for simplicity of exposition that $y_0 = 0$. Our primary interest is the estimation and tests on δ. We consider the model that contains a unit root under the null hypothesis $H_0: \delta = 0$ and allows for local departures from the hypothesis by setting $\delta = -c\Pi(1)/N$ as the alternative.

As argued in Hansen (1995), utilizing useful information contained in related time series can bring substantial power gain to the estimation and tests in unit root models. We assume that there are q additional stationary covariates, x_t, that help explain v_t, so that

$$v_t = [g(x_t) - \mu] + \epsilon_t, \tag{2.2}$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

where $g:\mathbb{R}^q \to \mathbb{R}$ and $E(g(x_t)) = \mu$. For identification reasons, we absorb the "intercept" term into the nonlinear function $g(x_t)$; see the further discussion on this issue at the end of this section. Combining (2.1) and (2.2) gives

$$\Delta y_t = \delta y_{t-1} + \sum_{j=1}^{p} \pi_j \Delta y_{t-j} + g(x_t) + \epsilon_t. \tag{2.3}$$

Regression (2.3) would be used to test the unit root hypothesis against the alternative of stationarity around a fixed mean.

Another important model in the unit root literature is the case where the time series y_t contains a linear time trend t.[1] In this case, we add a linear time trend in the regression and consider

$$\Delta y_t = \delta y_{t-1} + \theta t + \sum_{j=1}^{p} \pi_j \Delta y_{t-j} + g(x_t) + \epsilon_t, \tag{2.4}$$

where t is a simple time trend. In practice, (2.4) would be used to test if a series has a unit root with a drift versus (linear) trend stationarity.

In this paper, we consider unit root tests based on partial linear regressions (2.3) and (2.4). We introduce the following notation: let z_t be the vector of (linear) regressors and γ be the vector of unknown parameters associated with these regressors. Thus, corresponding to regression (2.3), $z_t = (y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p})'$, $\gamma = (\delta, \pi_1, \ldots, \pi_p)^\top$; and corresponding to regression (2.4), $z_t = (y_{t-1}, t, \Delta y_{t-1}, \ldots, \Delta y_{t-p})^\top$, $\gamma = (\delta, \theta, \pi_1, \ldots, \pi_p)^\top$. Consequently, our regression models can be rewritten in the following general format of a partially linear regression:

$$\Delta y_t = \gamma' z_t + g(x_t) + \epsilon_t. \tag{2.5}$$

Conditional on the covariates x_t, one gets

$$E(\Delta y_t | x_t) = \gamma' E(z_t | x_t) + g(x_t).$$

Taking the difference of the preceding two equations leads to

$$\Delta y_t - E(\Delta y_t | x_t) = \gamma' [z_t - E(z_t | x_t)] + \epsilon_t.$$

If the conditional expectations were known, regressing $\Delta y_t - E(\Delta y_t | x_t)$ on $z_t - E(z_t | x_t)$ would give us an estimate of γ. However, the quantities $E(\Delta y_t | x_t)$ and $E(z_t | x_t)$ are unknown. Thus, we consider the following procedure that estimates β, and thus δ, by two steps. First, regress Δy_t and z_t nonparametrically on x_t and denote the nonparametric regression residuals as $\hat{e}_{dt}$ and $\hat{e}_{zt}$, respectively. Next, regress the residuals $\hat{e}_{dt}$ on $\hat{e}_{zt}$ by least squares to get an estimate of γ.

The nonparametric estimation uses a Nadaraya–Watson kernel estimator that we illustrate subsequently. Let $k(u)$ be the univariate kernel and denote $K(u) = \prod_{r=1}^{q} k(u_r)$ if u is q-dimensional. In addition, let

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

$$K_{tj} = K\left(\frac{x_t - x_j}{a}\right),$$

where a is a bandwidth parameter. Let $f(x)$ be the density of x_t and denote $f(x_t)$ as f_t; then we have

$$\hat{f}_t = \frac{1}{Na^q}\sum_{j=1,j\neq t}^{N} K_{tj}; \qquad \hat{E}(\Delta y_t|x_t) = \frac{1}{Na^q}\frac{\sum_{j=1,j\neq t}^{N} K_{tj}\Delta y_j}{\hat{f}_t};$$

$$\hat{E}(z_t|x_t) = \frac{1}{Na^q}\frac{\sum_{j=1,j\neq t}^{N} K_{tj} z_j}{\hat{f}_t}$$

and the nonparametric regression residuals are

$$\hat{e}_{dt} = \Delta y_t - \hat{E}(\Delta y_t|x_t) \quad \text{and} \quad \hat{e}_{zt} = z_t - \hat{E}(z_t|x_t).$$

The vector and matrix of these residuals are denoted $\hat{e}_d$ and $\hat{e}_z$, respectively.

Next, we estimate γ by regressing $\hat{e}_d$ on $\hat{e}_z$. The preceding kernel regression necessarily involves a random denominator, a problem we circumvent by using a density-weighted estimate as in Powell, Stock, and Stoker (1989) and Fan and Li (1999). Thus, we regress $\hat{e}_d$ on $\hat{e}_z$ using OLS and incorporating the density weighting. Thus, we have

$$\hat{\gamma} = ((\hat{e}_z \odot \hat{f})^\top(\hat{e}_z \odot \hat{f}))^{-1}(\hat{e}_z \odot \hat{f})^\top(\hat{e}_d \odot \hat{f}), \tag{2.6}$$

where $\odot$ denotes the Hadamard product and $\hat{f}$ is the vector of the estimated density evaluated at each x_t.

Notice that by the nature of the semiparametric partially linear regression, an intercept term is not identified unless the model is further restricted. Consequently, the estimation of δ on the model with no constant term is the same as with a constant term. The apparent lack of identification arises because we have already implicitly estimated an intercept in the nonparametric regression, and no such effect remains. As argued by Robinson (1988), the fact that one cannot separate these cases is less a drawback than a consequence of the generality of the semiparametric model. Furthermore, in practice 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.[2]

3. MAIN RESULTS

We derive the asymptotic properties of the proposed partially linear regression estimation in this section. Our attention is focused on the case where the autoregressive root is close to unity, and we consider statistical tests for the null

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

hypothesis of a unit root. For purposes of determining asymptotic distributions, we use local to unity asymptotics (Phillips, 1987; Chan and Wei, 1987) so that $\delta = -c\Pi(1)/N$. Under the null hypothesis of a unit root, $c = 0$, whereas under $c \neq 0$, the alternative hypothesis becomes increasingly difficult to detect as the sample size increases. We assume that the system is initialized by setting $y_0 = 0$ (or, more generally, any random variable with finite variance). We follow convention and denote $W^c(r)$ as the solution to the stochastic differential equation

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

where $W(r)$ is a continuous stochastic process. When $W(r)$ is a Brownian motion, $W^c(r)$ is the conventional Ornstein–Uhlenbeck process.

Following Hansen (1995), we define

$$\sigma_{v\epsilon f} = \mathrm{Cov}(v_t, \epsilon_t f_t^2), \qquad \sigma_v^2 = \mathrm{Var}(v_t^2), \qquad \sigma_{\epsilon f}^2 = E(\epsilon_t^2 f_t^4),$$

$$\text{and} \quad \rho^2 = \frac{\sigma_{v\epsilon f}^2}{\sigma_v^2 \sigma_{\epsilon f}^2}.$$

We establish the N-consistency and derive the limiting distribution of the partially linear regression estimator $\hat{\delta}$ under conditions similar to those used in Robinson (1988) and Fan and Li (1999). We will use the definitions for the class of kernel functions of order l, $\mathcal{K}_l$, and the class of regression functions $\mathcal{G}_\mu^\alpha$ as defined in Robinson (1988, pp. 937, 939). In particular, $\mathcal{K}_l$ characterizes the class of lth order kernels and $\mathcal{G}_\mu^\alpha$ imposes moment and smoothness conditions for the nonlinear function and the density of the covariate. We also assume the following assumptions to facilitate the asymptotic analysis.

Assumptions. (A1) $\{x_t\}$ is i.i.d.; (A2) ϵ_t is i.i.d. with mean zero and is independent with x, $E|\epsilon|^4 < \infty$; (A3) x has probability density function (p.d.f.) $f \in \mathcal{G}_\lambda^\infty$, for some $\lambda > 0$; (A4) $g \in \mathcal{G}_\nu^4$, for some $\nu > 0$; (A5) as $N \to \infty$, $N^{-1}a^{-2q} \to 0$, $\sqrt{N}a^{2\min(\lambda+1,\nu)} \to 0$; (A6) $k \in \mathcal{K}_{l+n-1}$ for integers l and n such that $l - 1 < \lambda \leq l$, $n - 1 < \nu \leq n$; (A7) $\sigma_v^2 > 0$ and $\rho^2 > 0$; (A8) $\delta = -c\Pi(1)/N$.

Initially, we assume that $\{x_t\}$ and $\{\varepsilon_t\}$ are i.i.d. for simplicity. We discuss extensions to more general cases in the following section. The following lemma provides an important and interesting result regarding nonparametric regression of I(1) on I(0) processes. The proof of Lemma 3.1 also plays an important role in the proof of our main result.

LEMMA 3.1. *Under Assumptions A1–A8, for any* $t = 2, \ldots, N$,

$$\frac{1}{\sqrt{N}}(\hat{y}_{t-1} - \bar{y})\hat{f}_t = o_p(1),$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

where $\bar{y} = (1/N)\sum_{t=1}^{N} y_t$ *and*

$$\hat{y}_{t-1} = \frac{1}{Na^q} \frac{\sum_{j=1, j\neq t}^{N} K_{tj}\, y_{j-1}}{\hat{f}_t}.$$

The preceding result indicates that, in our density-weighted regression, if one nonparametrically regresses y_{t-1} on (stationary) x_t, the predicted value behaves asymptotically as if we used the sample mean. The asymptotic distribution of the semiparametric partially linear regression estimator $\hat{\delta}$ is summarized in the following theorem.

THEOREM 3.2. *Under Assumptions A1–A8,*

$$N(\hat{\delta} - \delta) \Rightarrow \Pi(1) \frac{\sigma_{\epsilon f}}{\sigma_v E(f^2)} \left(\int (\underline{W}_1^c)^2 \right)^{-1} \times \left(\rho \int \underline{W}_1^c \, dW_1 + \sqrt{1-\rho^2} \int \underline{W}_1^c \, dW_2 \right), \tag{3.1}$$

where W_2 *and* W_1 *are independent standard Brownian motions,* $\underline{W}_1^c = W_1^c(r) - \int W_1^c(s)\, ds$ *for model (2.3), and* $\underline{W}_1^c = W_1^c(r) + (6r - 4)\int W_1^c(s)\, ds - (12r - 6)\int W_1^c(s) s\, ds$ *for model (2.4) (with* $\theta = 0$*).*

The limiting distribution given in Theorem 3.2 is similar to Theorem 2 in Hansen (1995). To test the unit root hypothesis and apply the critical values in Hansen (1995), we construct a modified *t*-ratio. To this end, we denote $t^*(\delta)$ as

$$t^*(\hat{\delta}) = \frac{\hat{\delta}}{\hat{\sigma}_{\epsilon f}} \sqrt{\frac{\hat{E}(f^2)}{S^{11}}},$$

where S^{11} is the first entry of the matrix

$$((\hat{e}_z \odot \hat{f})^\top (\hat{e}_z \odot \hat{f}))^{-1}$$

and the terms $\hat{\sigma}_{\epsilon f}$ and $\hat{E}(f^2)$ are consistent estimators for $\sigma_{\epsilon f}$ and $E(f^2)$. The consistent estimators are presented later in this section. The following theorem provides the limiting distribution of our modified *t*-ratio.

THEOREM 3.3. *Under Assumptions A1–A7, the t-statistic based on* $\hat{\delta}$ *has limiting distribution*

$$t^*(\hat{\delta}) \Rightarrow -\frac{cE(f^2)\sigma_v}{\sigma_{\epsilon f}} \left(\int (\underline{W}_1^c)^2 \right)^{1/2} + \left(\int (\underline{W}_1^c)^2 \right)^{-1/2} \left(\rho \int \underline{W}_1^c \, dW_1 \right) + \sqrt{1-\rho^2} N(0,1), \tag{3.2}$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

where the $N(0,1)$ variable is independent of W_1 and $\underline{W}_1^c(r)$ is defined as in Theorem 3.2. In particular, under the null hypothesis of a unit root,

$$t^*(\hat{\delta}) \Rightarrow \left(\int (\underline{W}_1)^2 \right)^{-1/2} \left(\rho \int \underline{W}_1 \, dW_1 \right) + \sqrt{1-\rho^2} N(0,1).$$

The limiting distribution in Theorem 3.3 is identical to that in Hansen (1995) and very similar to other distributions appearing in various related unit root tests. In particular, nearly 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 was employed.

The distribution has the disadvantage that ρ is a remaining nuisance parameter. There have been various approaches for dealing with the nuisance parameter ρ, ranging from simulating critical values for each value of the parameter to using conservative critical values to cover the range of possible ρ. We must estimate ρ in addition to the other parameters used to construct the modified t-ratio. Consider the following consistent estimate of $E(f^2)$:

$$\hat{E}(f^2) = \frac{1}{N} \sum_{t=1}^{N} \hat{f}_t^2.$$

We estimate v_t by the residual from an OLS regression of Δy_t on z_t, and we estimate $\epsilon_t f_t$ by $\tilde{\epsilon}_t \hat{f}_t$ where

$$\tilde{\epsilon}_t = \hat{e}_{dt} - \hat{\beta}^\top \hat{e}_{zt}.$$

The estimates $\hat{v}_t$ and $\tilde{\epsilon}_t \hat{f}_t^2$ can then be used to obtain a consistent estimate of ρ as in Hansen (1995). Finally, with the estimated value of ρ in hand, we compare $t^*(\hat{\delta})$ to the critical values in Table 1 of Hansen (1995).

4. EXTENSIONS

In this section, we list several possible extensions to our simple model. In the previous section we assumed that the covariates x_t are i.i.d. This assumption substantially simplifies the proof. However, x_t is not i.i.d. in most empirical settings, and it is desirable to extend our results to allow for some dependence in x_t. In stationary partially linear models, Fan and Li (1999) show that one can obtain the usual $\sqrt{N}$ convergence rate when the data are absolutely regular. Using the approach of Fan and Li (1999) to deal with the correlation in x_t, our analysis can be extended to the case with weakly dependent covariates.

We modify Assumption A1 to A1′ to allow for weak dependence in x_t.

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

Assumption A1′. $\{x_t\}$ is a stationary absolutely regular process with coefficient β_j satisfying $\beta_j^{\delta/(1+\delta)} = O(j^{-2-\epsilon})$ for some $0 < \epsilon < 1$, $0 < \delta \leq \frac{1}{2}$.

Assumption A1′ assumes that $\{x_t\}$ is absolutely regular with some restrictions on the decay rate of the mixing coefficient so that summations such as $\sum_{j=1}^{\infty} j^2 \beta_j^{\delta/(1+\delta)} < \infty$. Many well-known time series, including stationary autoregressive moving average (ARMA) processes (Pham and Tran, 1985), are absolutely regular with geometric decay rates and thus satisfy Assumption A1′. See Fan and Li (1999) for more discussion on this type of assumption and related references. Assuming that the stationary covariate is β-mixing and using results of Yoshihara (1976) for β-mixing processes, we can extend the results in Section 3 to the case with weakly dependent covariates.

THEOREM 4.1. *Under Assumptions A1′ and A2–A7, the partial linear regression estimator $\hat{\delta}$ has the same limiting distributions as that of (3.1) in Theorem 3.2.*

The preceding extension requires careful treatment of the nonstationarity introduced to the partially linear model framework as we have presented it. The mathematical complexity of the proofs is greatly increased in the presence of nonstationarity and nonparametric characteristics. For a sketch of a proof for this extension, the readers are referred to Appendix D (available from the authors upon request).

If one allows for dependence in x_t, as would be expected in most applications, the estimation of ρ is based on a long-run variance estimator. That is, we use the analogue of equation (17) in Hansen (1995) given by

$$\begin{pmatrix} \hat{\sigma}_v^2 & \hat{\sigma}_{v\epsilon f} \\ \hat{\sigma}_{v\epsilon f} & \hat{\sigma}_{\epsilon f}^2 \end{pmatrix} = \sum_{j=-T+1}^{T+1} w\left(\frac{j}{b_T}\right) \sum_t \begin{pmatrix} \hat{v}_t \\ \tilde{\epsilon}_t \hat{f}_t^2 \end{pmatrix} (\hat{v}_{t-j} \quad \tilde{\epsilon}_{t-j} \hat{f}_{t-j}^2),$$

where $w(\cdot)$ is a kernel function and b_T is a bandwidth parameter. The parameter ρ is estimated using these long-run variance estimates.

The assumption about the errors may also be relaxed to the case that ϵ_t is a martingale difference sequence and allow for conditional heteroskedasticity as in Fan et al. (1995) and Fan and Li (1999). The condition proposed in Fan and Li (1999) is $E(\epsilon_t | x_t, y_{t-1}, \mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_{t-1}$ is the sigma algebra generated by all past y_{t-1} and x_{t-1}. This condition is much weaker than independence of x_t and ϵ_t, and it allows for heteroskedasticty in addition to GARCH-type effects. We conjecture that our results still hold in the presence of this condition, and we investigate these extensions in Monte Carlo experiments in the next section.

5. MONTE CARLO

In this section, a small simulation study is conducted to examine the finite-sample performance of the nonstationary partially linear estimation and the

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

associated unit root test. We consider several specifications of $g(x)$, both linear and nonlinear, to compare the standard Dickey–Fuller test, the CADF test of Hansen (1995), and the new test $t^*(\hat{\delta})$ using the partially linear model, which we denote as partially linear model unit root (PLMUR). The data generating process is

$$\Delta y_t = \delta y_{t-1} + g_j(x_t) + \epsilon_t, \qquad j = 1,\ldots,5.$$

The different functions are listed here:

$$g_1(x) = 0; \qquad g_2(x) = 2x; \qquad g_3(x) = 2x_1 x_2;$$

$$g_4(x) = x^2 - 1; \qquad g_5(x) = x^3 - x.$$

The x variables are all standard normal. 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 because the covariates enter linearly. We include $g_3(x)$ for the purpose of checking the ability of the PLMUR test to use multiple covariates. As shown in the Monte Carlo experiment of Robinson (1988), the nonparametric estimates $\widehat{\Delta y}$ and $\hat{y}$ are likely to worsen as the dimension of x increases. In addition, it is easy to check that the OLS coefficients on x_1 and x_2 will converge to zero so that the CADF test should have similar performance to the Dickey–Fuller test for $g_3(x)$. The other specifications are also nonlinear, so the PLMUR tests should be more powerful if the nonlinearity is estimated reasonably.

Given a density associated with x, smaller values of ρ are indicative of the effectiveness of covariates in explaining variation in $v_t = g(x_t) + \epsilon_t$. Therefore, we expect more powerful tests if ρ is small. Straightforward calculations show that

$$b^2\rho_1^2 = 1; \qquad b^2\rho_2^2 = \frac{1}{5}; \qquad b^2\rho_3^2 = \frac{1}{5}; \qquad b^2\rho_4^2 = \frac{1}{3}; \qquad b^2\rho_5^2 = \frac{1}{11},$$

where ρ_j^2 is associated with $g_j(x)$ and $b^2 = E(f^4)/(E(f^2))^2$.

For the PLMUR test, we need to select a kernel and a bandwidth. In our experiment, we chose a Gaussian kernel. The bandwidth is set to $N^{-1/5}$ for all experiments.

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 $\delta = -(c/N)$. For each specification, we generate samples of size 100 and compute 10,000 replications for models (2.3) and (2.4).[3] The numerical results for size appear in Table 1, and we provide graphs of the power functions in Figures 1–3. In particular, Figures 1–3 correspond to g_1, g_2, and g_5. The results for g_3 and g_4 are similar to g_5 and are omitted to conserve space.

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

TABLE 1. Size

		Model (2.3)			Model (2.4)		
		DF	CADF	PLMUR	DF	CADF	PLMUR
$c = 0$	g_1	0.05	0.05	0.06	0.06	0.06	0.05
	g_2	0.05	0.05	0.04	0.05	0.05	0.04
	g_3	0.05	0.05	0.05	0.05	0.05	0.04
	g_4	0.05	0.05	0.04	0.05	0.05	0.04
	g_5	0.05	0.05	0.04	0.05	0.05	0.03

For $c = 0$, we have a unit root, and we compare the size for each of the tests. All three tests have reasonable size for all of the specifications, with no test being severely oversized. The size result for the PLMUR test indicates that the asymptotic theory provides an accurate approximation for the distribution of the statistic. This is remarkable given the choice of the same bandwidth for all of the widely different choices of $g(x)$.

For $c \neq 0$, the departure from the unit root becomes apparent in the increased rejection frequencies. For $g_1(x)$ (Figure 1), the power of the CADF test is very

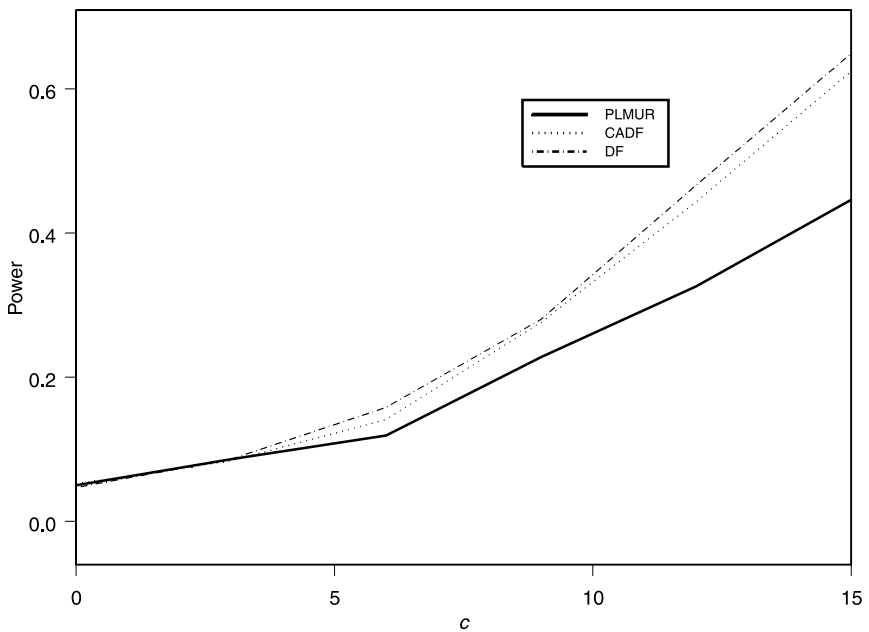


FIGURE 1. $g_1(x) = 0$.

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

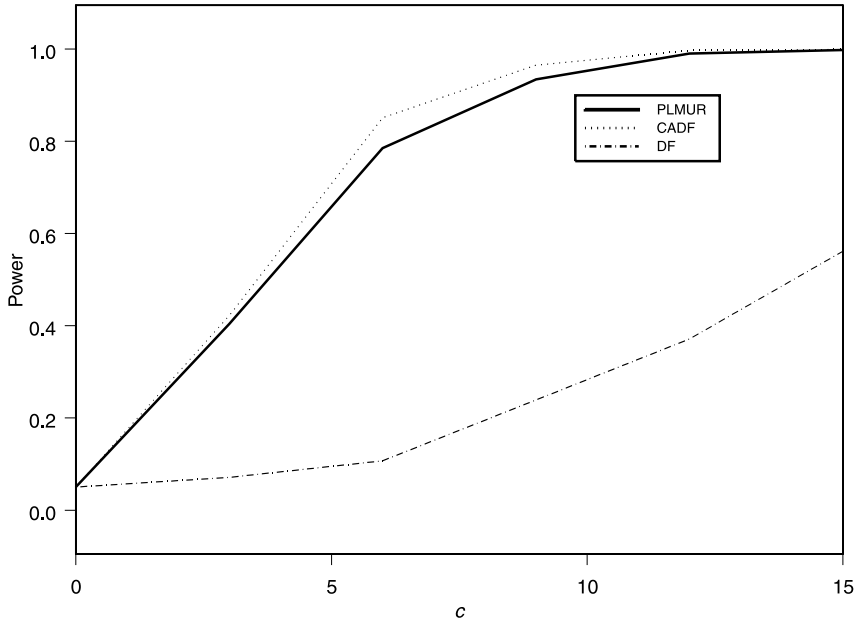


FIGURE 2. $g_2(x) = 2x$.

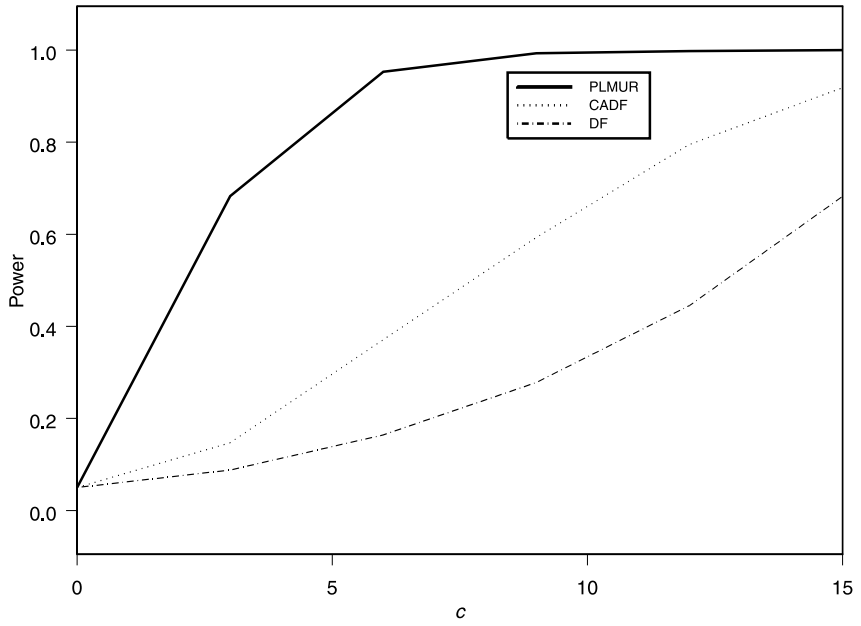


FIGURE 3. $g_5(x) = x^3 - x$.

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

close to the Dickey–Fuller test for the range of local alternatives considered. However, the PLMUR test is not as powerful as either the Dickey–Fuller or the CADF test when there is no covariate effect. For g_2 (Figure 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. The advantage of the PLMUR test becomes apparent when g_5 (Figure 3) is considered. As expected, the covariate is successfully used to reduce the variance of the estimator of δ. Using the cubic function, power is again much higher than the competing tests. The CADF test has more power than the Dickey–Fuller because using x linearly does help explain some of the variance of v_t so that δ is estimated with more precision. Simple calculations show that the estimated linear regression coefficient should converge to 2 so that the estimated value of ρ converges to $\frac{7}{11}$. 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.

The preceding experiments were all conducted using i.i.d. x_t and ϵ_t. We relax these conditions and explore the effects on the PLMUR test. In all additional cases, we explore the function g_4. First, we consider the effects of dependence in x_t by employing the data generating process

$$x_t = 0.7x_{t-1} + e_t, \tag{5.1}$$

where e_t is i.i.d. $N(0,1)$. In addition, we generate heteroskedasticity in ϵ_t according to the following process:

$$\epsilon_t = u_t|x_t|,$$

where u_t is i.i.d. $N(0,1)$ and x_t is generated according to (5.1). This implies that $E(\epsilon_t^2|x_t) = x_t^2$ so that there is heteroskedasticity in addition to dependence in x_t. Finally, we combine dependence in x_t (via (5.1)) with ARCH(1) errors given by

$$\epsilon_t = u_t(1 + 0.5\epsilon_{t-1}^2)^{1/2},$$

where u_t is again i.i.d. $N(0,1)$ so that $E(\epsilon_t^2|\epsilon_{t-1}) = 1 + 0.5\epsilon_{t-1}^2$.

For each of the final three experiments, we allow dependence in x_t. As mentioned in the previous section, dependence in x_t causes dependence in v_t so that we must use a long-run variance to estimate the parameters σ_v and $\sigma_{v\epsilon f}$. We employ a kernel-based estimator using a quadratic spectral kernel and the data-dependent bandwidth of Andrews (1991) applied to the series v_t. We present the output in Table 2. We denote the data generating processes as heteroskedasticity (HET), dependent (DEP), and autoregressive conditional heteroskedasticity (ARCH). We also list the size-adjusted power for each test in Table 2. We see that power is still higher for the PLMUR test in all cases. However, power for the PLMUR test varies slightly for the different data generating processes.

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

TABLE 2. Size and size-adjusted power for g_4

	HET			DEP			ARCH		
c	DF	CADF	PLMUR	DF	CADF	PLMUR	DF	CADF	PLMUR
0	0.02	0.03	0.03	0.03	0.03	0.04	0.02	0.03	0.04
3	0.02	0.05	0.96	0.02	0.05	0.84	0.03	0.05	0.68
6	0.02	0.08	1.00	0.02	0.08	0.99	0.03	0.09	0.94
9	0.04	0.16	1.00	0.03	0.14	0.99	0.05	0.16	0.99
12	0.07	0.24	1.00	0.06	0.21	1.00	0.09	0.23	1.00
15	0.13	0.35	1.00	0.10	0.30	1.00	0.15	0.34	1.00

The main reason this is true is that given the new data generating processes, the value of ρ varies. As noted before, the lower the value of ρ, the higher the power.

The results of the experiment indicate that power is indeed higher for the PLMUR test when there are nonlinear effects in the covariates. Moreover, the presence of heteroskedasticity, dependence in x_t, and ARCH effects do not alter the size of the tests.

6. EMPIRICAL ILLUSTRATION

In this section, we demonstrate the use of PLMUR tests using monthly data for the United States from January 1970 to January 2001. The series of interest are the unemployment rate and total capacity utilization.[4] As a preliminary analysis, we conduct a simple unit root test on the unemployment rate and include a constant and trend in the regression. We include four lags of Δy_t where the number of lags was determined by the modified Akaike information criterion (AIC) criterion of Ng and Perron (2001). The Dickey–Fuller t-statistic is -2.86, which is not significant at the 10% level. Hence, we fail to reject the unit root hypothesis. Using the preceding notation, the unemployment rate is denoted y_t, and we take the lag of the first difference of total capacity utilization as x_t. The model of interest is

$$\Delta y_t = \theta t + \delta y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \gamma_3 \Delta y_{t-3} + \gamma_4 \Delta y_{t-4} + g(x_t) + \epsilon_t.$$

For comparison purposes, we also calculate the CADF test of Hansen (1995), which assumes that $g(x_t)$ is a linear function. The estimated value of ρ^2 for the CADF is 0.71, so that the appropriate conservative critical values at the 10% and 5% levels are -2.97 and -3.27, respectively. The CADF t-statistic for unemployment using the lag of the first difference of total capacity utilization as a covariate is -2.39, so we again fail to reject the null hypothesis of a unit root. The PLMUR test uses a nonparametric estimate of $g(x_t)$, so the estimate of δ is semiparametric. We calculate the PLMUR test using three different

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

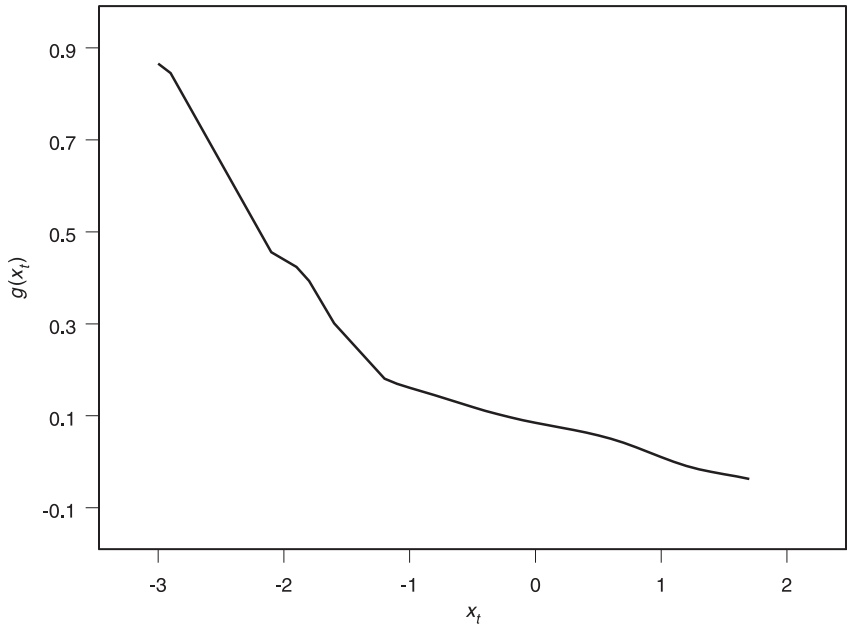


FIGURE 4. Estimated $g(x_t)$ for unemployment.

choices of bandwidth parameter a, with $a = dN^{-1/5}$ with $d = 1,2,3$. The tests and estimated ρ parameters associated with each bandwidth are then indexed by d. The corresponding estimates of ρ are $\hat{\rho}_1 = 0.23$, $\hat{\rho}_2 = 0.29$, and $\hat{\rho}_3 = 0.38$. These estimates suggest that the parameter δ will be estimated with increased precision, allowing for more powerful testing of the unit root hypothesis. The test statistics are $\text{PLMUR}_1 = -3.22$, $\text{PLMUR}_2 = -3.11$, and $\text{PLMUR}_3 = -3.09$. The 5% critical value associated with $\hat{\rho}_1$ and $\hat{\rho}_2$ is -2.73, and the 5% critical value for $\hat{\rho}_3$ is -2.87. We see that in all cases the null hypothesis of a unit root is rejected. A graph of the estimated nonlinear function $\hat{g}(x_t)$ using $a = 2 \times N^{-1/5}$ is shown in Figure 4. This example illustrates how using a partially linear model with covariates can lead to a rejection of the unit root hypothesis when standard tests fail to reject.

NOTES

1. See, e.g., Hansen (1995) for alternative representations and more discussions on unit root models with a deterministic component.

2. See Hamilton (1994, Ch. 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.

3. The programs were written in Ox 2.0; see Doornik (1998).

4. The series were extracted from the FRED II database at the St. Louis Federal Reserve.

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Beelders, O. (1999) Adaptive Estimation and Unit Root Tests. Working paper, Emory University.

Beveridge, S. & C.R. Nelson (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle." *Journal of Monetary Economics* 7, 151–174.

Chan, N.H. & C.Z. Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.

Chen, H. (1988) Convergence rates for parametric components. *Annals of Statistics* 16, 135–146.

Doornik, J.A. (1998): *Ox—An Object Oriented Matrix Programming Language*. Timberlake Consultants Press.

Elliot, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

Engle, R.F., C.W.J. Granger, J. Rice, & A. Weiss (1986) Semiparametric estimates for the relation between weather and electricity sales. *Journal of the American Statistical Association* 81, 310–320.

Fan, Y. & Q. Li (1999) Root-N-consistent estimation of partially linear time series models. *Journal of Nonparametric Statistics* 11, 251–269.

Fan, Y., Q. Li, & T. Stengos (1995) Root-N-consistent semiparametric regression with conditional heteroskedastic disturbances. *Journal of Quantitative Economics* 11, 229–240.

Granger, C.W.J. (1995) Modeling nonlinear relationships between extended-memory variables. *Econometrica* 63, 265–279.

Granger, C.W.J. & T. Terasvirta (1993) *Modelling Nonlinear Economic Relationships*. Oxford University Press.

Hamilton, J.D. (1994) *Time Series Analysis.* Princeton University Press.

Hansen, B.E. (1995) Rethinking the univariate approach to unit root testing. *Econometric Theory* 11, 1148–1171.

Hasan, M.N. & R.W. Koenker (1997) Robust rank tests of the unit root hypothesis. *Econometrica* 65, 133–161.

Heckman, N.E. (1986) Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society, Series B* 48, 244–248.

Juhl, T. & Z. Xiao (2000) Partially Linear Models with Unit Roots. Working paper, University of Illinois.

Linton, O.B. (1995) Second order approximation in the partially linear regression model. *Econometrica* 63, 1079–1112.

Lucas, A. (1995) Unit root tests based on M estimators. *Econometric Theory* 11, 331–346.

Ng, S. & P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.

Pham, D. & L. Tran (1985) Some mixing properties of time series models. *Stochastic Process and Applications* 19, 297–303.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.

Phillips, P.C.B. & J.Y. Park (1998) Nonstationary Density Estimation and Kernel Autoregression. Working paper, Yale University.

Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1403–1430.

Rice, J. (1986) Convergence rates for partially linear spline models. *Statistics and Probability Letters* 4, 203–208.

Robinson, P.M. (1988) Root-N-consistent semiparametric regression. *Econometrica* 56, 931–954.

Schick, A. (1986) On asymptotically efficient estimation in semiparametric models. *Annals of Statistics* 14, 1139–1151.

Seo, B. (1999) Distribution theory for unit root tests with conditional heteroskedasticity. *Journal of Econometrics* 91, 113–144.

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

Shiller, R.J. (1984) Smoothness priors and nonlinear regression. *Journal of the American Statistical Association* 72, 420–423.

Shin, D.W. & B.S. So (1999) Unit root tests based on adaptive maximum likelihood estimation. *Econometric Theory* 15, 1–23.

Speckman, P. (1988) Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B* 50, 413–436.

Tong, H. (1990) *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press.

Wahba, G. (1986) Partial and interaction spline models for the semiparametric estimation of functions of several variables. In T.J. Boardman (ed.), *Proceedings of Computer Science and Statistics: 18th Symposium on the Interface.* American Statistical Association.

Yoshihara, K. (1976) Limiting behavior of U-statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 35, 237–252.

APPENDIXES

We define $E_1(\cdot) = E(\cdot|x_1)$ and $E_{X_N}(\cdot) = E(\cdot|X_N)$ where $X_N = (x_1, x_2, \ldots, x_N)$. Without loss of generality, we also assume that the initial value is $y_0 = 0$ and demonstrate the proofs for the demeaned case with $c = 0$ (unit root), other cases being similar (for some details of proofs and a discussion of the case with a deterministic trend, also see an early version of this paper, Juhl and Xiao, 2000). We give some useful propositions in Appendix A. Appendix B proves the main result using the results of Appendix A. Proofs of Lemma 3.1 and Propositions 1–14 are given in Appendix C. Appendix D (available from the authors upon request) extends the analysis to the weakly dependent covariate x_t. We also define the following terms:

$$\widehat{\Delta y}_{t-s} = \frac{1}{Na^q} \frac{\sum_{j=1, j\neq t}^{N} K_{tj} \Delta y_{j-s}}{\hat{f}_t},$$

$$\hat{\epsilon}_t = \frac{1}{Na^q} \frac{\sum_{j=1, j\neq t}^{N} K_{tj} \epsilon_j}{\hat{f}_t},$$

$$\hat{g}(x_t) = \frac{1}{Na^q} \frac{\sum_{j=1, j\neq t}^{N} K_{tj} g(x_j)}{\hat{f}_t}.$$

In addition, let $\hat{e}_v$ be the vector of $(\epsilon_t - \hat{\epsilon}_t + g(x_t) - \hat{g}(x_t))$ and $\hat{e}_y$ and $\hat{e}_s$ be the vectors of residuals from nonparametrically regressing y_{t-1} and Δy_{t-s} on x_t, respectively. Let $B(L) = \Pi(L)^{-1}$; we have $B(L) = B(1) + B^*(L)(1 - L)$ and $y_t = y_t^* + v_t^* - v_0^*$ where $y_t^* = B(1)\sum_{l=0}^{t} v_l$, $v_t^* = B^*(L)v_t$. Finally, define $M = I - (\hat{e}_s \odot \hat{f}) \times [(\hat{e}_s \odot \hat{f})^\top (\hat{e}_s \odot \hat{f})]^{-1} (\hat{e}_s \odot \hat{f})^\top$.

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

APPENDIX A: Useful Propositions

PROPOSITION 1. *For* $s = 1, \ldots, p$,

$$\sum_{t=1}^{N} (y_{t-1} - \hat{y}_{t-1})(\Delta y_{t-s} - \widehat{\Delta y}_{t-s})\hat{f}_t^2 = O_p(N).$$

PROPOSITION 2.

$$(\hat{e}_y \odot \hat{f})^\top (\hat{e}_s \odot \hat{f}) = O_p(N).$$

PROPOSITION 3.

$$\sum_{t=1}^{N} (\Delta y_{t-s} - \widehat{\Delta y}_{t-s})(\Delta y_{t-r} - \widehat{\Delta y}_{t-r})\hat{f}_t^2 = O_p(N).$$

PROPOSITION 4.

$$(\hat{e}_s \odot \hat{f})^\top (\hat{e}_s \odot \hat{f}) = O_p(N).$$

PROPOSITION 5.

$$(\hat{e}_y \odot \hat{f})^\top (\hat{e}_s \odot \hat{f})[(\hat{e}_s \odot \hat{f})^\top (\hat{e}_s \odot \hat{f})]^{-1} (\hat{e}_s \odot \hat{f})^\top (\hat{e}_y \odot \hat{f}) = o_p(N^2).$$

PROPOSITION 6.

$$(\hat{e}_s \odot \hat{f})^\top (\hat{e}_v \odot \hat{f}) = o_p(N).$$

PROPOSITION 7.

$$(\hat{e}_y \odot \hat{f})^\top (\hat{e}_s \odot \hat{f})[(\hat{e}_s \odot \hat{f})^\top (\hat{e}_s \odot \hat{f})]^{-1} (\hat{e}_s \odot \hat{f})^\top (\hat{e}_v \odot \hat{f}) = o_p(N).$$

PROPOSITION 8.

$$\frac{1}{N} \sum_{t=1}^{N} (y_{t-1}^* - \bar{y}^*)\epsilon_t \hat{f}_t^2 \Rightarrow \frac{\sigma_v \sigma_{ef}}{\Pi(1)} \left(\rho \int \underline{W}_1^c \, dW_1 + \sqrt{1 - \rho^2} \int \underline{W}_1^c \, dW_2 \right), \tag{A.1}$$

where $\bar{y}^* = N^{-1} \sum_{t=1}^{N} y_t^*$.

PROPOSITION 9.

$$\frac{1}{N} \sum_{t=1}^{N} (\bar{y}^* - \hat{y}_{t-1}^*)\epsilon_t \hat{f}_t^2 \xrightarrow{p} 0. \tag{A.2}$$

PROPOSITION 10.

$$\frac{1}{N^2} \sum_{t=1}^{N} (y_{t-1}^* - \bar{y}^*)(\bar{y}^* - \hat{y}_{t-1}^*)\hat{f}_t^2 \xrightarrow{p} 0.$$

PROPOSITION 11.

$$\frac{1}{N^2} \sum_{t=1}^{N} (y_{t-1}^* - \hat{y}_{t-1}^*)^2 \hat{f}_t^2 \Rightarrow \Pi(1)^{-2} E(f^2) \sigma_v^2 \int (W_1 - \overline{W}_1)^2, \tag{A.3}$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

where

$$\hat{y}^*_{t-1} = \frac{1}{Na^q}\left(\sum_{p\neq t}^{N} K_{tp}\, y^*_{p-1}\right)\Big/\hat{f}_t.$$

PROPOSITION 12.

$$\frac{1}{N}\sum_{t=1}^{N}(y^*_{t-1} - \hat{y}^*_{t-1})\hat{\epsilon}_t \hat{f}_t^2 \xrightarrow{p} 0.$$

PROPOSITION 13.

$$\frac{1}{N}\sum_{t=1}^{N}(y^*_{t-1} - \hat{y}^*_{t-1})(g(x_t) - \hat{g}(x_t))\hat{f}_t^2 \xrightarrow{p} 0.$$

APPENDIX B: Proof of Theorem 3.2

By definition

$$\hat{\delta} = ((\hat{e}_y \odot \hat{f})^\top M(\hat{e}_y \odot \hat{f}))^{-1}((\hat{e}_y \odot \hat{f})^\top M(\hat{e}_d \odot \hat{f}))$$

$$= \delta + ((\hat{e}_y \odot \hat{f})^\top M(\hat{e}_y \odot \hat{f}))^{-1}((\hat{e}_y \odot \hat{f})^\top M(\hat{e}_v \odot \hat{f})).$$

Notice that

$$(\hat{e}_y \odot \hat{f})^\top M(\hat{e}_y \odot \hat{f}) = (\hat{e}_y \odot \hat{f})^\top(\hat{e}_y \odot \hat{f}) - (\hat{e}_y \odot \hat{f})^\top(\hat{e}_s \odot \hat{f})[(\hat{e}_s \odot \hat{f})^\top(\hat{e}_s \odot \hat{f})]^{-1}(\hat{e}_s \odot \hat{f})^\top(\hat{e}_y \odot \hat{f});$$

we have, by Proposition 5,

$$(\hat{e}_y \odot \hat{f})^\top M(\hat{e}_y \odot \hat{f}) = (\hat{e}_y \odot \hat{f})^\top(\hat{e}_y \odot \hat{f}) + o_p(N^2).$$

Similarly, notice that

$$(\hat{e}_y \odot \hat{f})^\top M(\hat{e}_v \odot \hat{f}) = (\hat{e}_y \odot \hat{f})^\top(\hat{e}_v \odot \hat{f}) - (\hat{e}_y \odot \hat{f})^\top(\hat{e}_s \odot \hat{f})[(\hat{e}_s \odot \hat{f})^\top(\hat{e}_s \odot \hat{f})]^{-1}(\hat{e}_s \odot \hat{f})^\top(\hat{e}_v \odot \hat{f}).$$

We obtain, by Proposition 7,

$$(\hat{e}_y \odot \hat{f})^\top M(\hat{e}_v \odot \hat{f}) = (\hat{e}_y \odot \hat{f})^\top(\hat{e}_v \odot \hat{f}) + o_p(N).$$

Thus,

$$N(\hat{\delta} - \delta) = \left(\frac{1}{N^2}(\hat{e}_y \odot \hat{f})^\top(\hat{e}_y \odot \hat{f})\right)^{-1}\left(\frac{1}{N}(\hat{e}_y \odot \hat{f})^\top(\hat{e}_v \odot \hat{f})\right) + o_p(1).$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

Notice that $y_t = B(1)\sum_{l=0}^{t} v_l + v_i^* - v_0^* = y_t^* + v_t^* - v_0^*$; then

$$
\begin{aligned}
&(\hat{e}_y \odot \hat{f})^\top (\hat{e}_y \odot \hat{f}) \\
&\quad = \sum_{t=1}^{N} \left(y_{t-1}^* - \hat{y}_{t-1}^* + v_{t-1}^* - v_0^* - \frac{1}{Na^q} \left(\sum_{p \neq t}^{N} K_{tp}[v_{p-1}^* - v_0^*] \right) \Big/ \hat{f}_t \right)^2 \hat{f}_t^2 \\
&\quad = \sum_{t=1}^{N} (y_{t-1}^* - \hat{y}_{t-1}^*)^2 \hat{f}_t^2 + \sum_{t=1}^{N} \left((v_{t-1}^* - v_0^*) - \frac{1}{Na^q} \left(\sum_{p \neq t}^{N} K_{tp}[v_{p-1}^* - v_0^*] \right) \Big/ \hat{f}_t \right)^2 \hat{f}_t^2 \\
&\qquad + 2 \sum_{t=1}^{N} (y_{t-1}^* - \hat{y}_{t-1}^*) \left((v_{t-1}^* - v_0^*) - \frac{1}{Na^q} \left(\sum_{p \neq t}^{N} K_{tp}[v_{p-1}^* - v_0^*] \right) \Big/ \hat{f}_t \right) \hat{f}_t^2
\end{aligned}
$$

and

$$
\sum_{t=1}^{N} \left((v_{t-1}^* - v_0^*) - \frac{1}{Na^q} \left(\sum_{p \neq t}^{N} K_{tp}[v_{p-1}^* - v_0^*] \right) \Big/ \hat{f}_t \right)^2 \hat{f}_t^2 = O_p(N),
$$

$$
\sum_{t=1}^{N} (y_{t-1}^* - \hat{y}_{t-1}^*) \left((v_{t-1}^* - v_0^*) - \frac{1}{Na^q} \left(\sum_{p \neq t}^{N} K_{tp}[v_{p-1}^* - v_0^*] \right) \Big/ \hat{f}_t \right) \hat{f}_t^2 = O_p(N);
$$

thus

$$
(\hat{e}_y \odot \hat{f})^\top (\hat{e}_y \odot \hat{f}) = \sum_{t=1}^{N} (y_{t-1}^* - \hat{y}_{t-1}^*)^2 \hat{f}_t^2 + o_p(N^2).
$$

Similarly,

$$
\begin{aligned}
&\frac{1}{N} (\hat{e}_y \odot \hat{f})^\top (\hat{e}_v \odot \hat{f}) \\
&\quad = \frac{1}{N} \sum_{t=1}^{N} (y_{t-1} - \hat{y}_{t-1}) \hat{f}_t^2 (\epsilon_t - \hat{\epsilon}_t + g(x_t) - \hat{g}(x_t)) \\
&\quad = \frac{1}{N} \sum_{i=1}^{N} \left(y_{t-1}^* - \hat{y}_{t-1}^* + v_{t-1}^* - v_0^* - \frac{1}{Na^q} \left(\sum_{p \neq t}^{N} K_{tp}[v_{p-1}^* - v_0^*] \right) \Big/ \hat{f}_t \right) \\
&\qquad \times (\epsilon_t - \hat{\epsilon}_t + g(x_t) - \hat{g}(x_t)) \hat{f}_t^2 \\
&\quad = \frac{1}{N} \sum_{i=1}^{N} (y_{t-1}^* - \hat{y}_{t-1}^*)(\epsilon_t - \hat{\epsilon}_t + g(x_t) - \hat{g}(x_t)) \hat{f}_t^2 \\
&\qquad + \frac{1}{N} \sum_{t=1}^{N} \left(v_{t-1}^* - v_0^* - \frac{1}{Na^q} \left(\sum_{p \neq t}^{N} K_{tp}[v_{p-1}^* - v_0^*] \right) \Big/ \hat{f}_t \right) (\epsilon_t - \hat{\epsilon}_t + g(x_t) - \hat{g}(x_t)) \hat{f}_t^2 \\
&\quad = \frac{1}{N} \sum_{t=1}^{N} (y_{t-1}^* - \hat{y}_{t-1}^*)(\epsilon_t - \hat{\epsilon}_t + g(x_t) - \hat{g}(x_t)) \hat{f}_t^2 + o_p(1).
\end{aligned}
$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

The theorem holds because

$$\frac{1}{N^2}\sum_{t=1}^{N}(y^*_{t-1}-\hat{y}^*_{t-1})^2\hat{f}_t^2 \Rightarrow E(f^2)\sigma_v^2\int(W_1-\overline{W}_1)^2/\Pi(1)^2$$

by Proposition 11, and

$$\frac{1}{N}\sum_{t=1}^{N}(y^*_{t-1}-\hat{y}^*_{t-1})(\hat{\epsilon}_t+g(x_t)-\hat{g}(x_t))\hat{f}_t^2 \xrightarrow{p} 0,$$

by Propositions 12 and 13, and

$$\frac{1}{N}\sum_{t=1}^{N}(y^*_{t-1}-\hat{y}^*_{t-1})\epsilon_t\hat{f}_t^2 \Rightarrow \frac{\sigma_v\sigma_{\epsilon f}}{\Pi(1)}\left(\rho\int\underline{W}_1^c\,dW_1+\sqrt{1-\rho^2}\int\underline{W}_1^c\,dW_2\right)$$

by Proposition 8. ■

APPENDIX C: Proofs of Lemmas and Propositions

Proof of Lemma 3.1. Using a BN (Beveridge and Nelson, 1981) decomposition for $B(L)=\Pi(L)^{-1}$ we have $B(L)=B(1)+B^*(L)(1-L)$, where $B^*(L)$ has all roots outside the unit root circle; then

$$y_t=\sum_{l=0}^{t}B(L)v_l=B(1)\sum_{l=0}^{t}v_l+v_t^*-v_0^*$$

and

$$\frac{1}{\sqrt{N}}(\hat{y}_{t-1}-\bar{y})\hat{f}_t=\frac{1}{\sqrt{N}}(\hat{y}^*_{t-1}-\bar{y}^*)\hat{f}_t+o_p(1).$$

We next show that

$$\frac{1}{\sqrt{N}}(\hat{y}^*_{t-1}-\bar{y}^*)\hat{f}_t=o_p(1).$$

Let

$$A=E\left(\frac{1}{N}(\hat{y}^*_{t-1}-\bar{y}^*)^2\hat{f}_t^2\right)=B(1)^2E\left(\frac{1}{N}(\bar{d}+\bar{h}-\hat{d}_{t-1}-\hat{h}_{t-1})^2\hat{f}_t^2\right),$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

where $d_t = \sum_{l=0}^{t} \epsilon_l$ and $h_t = \sum_{l=0}^{t} g(x_l)$, and we verify the order of magnitude of A. First,

$$E\left(\frac{1}{N}(\bar{d} - \hat{d}_{t-1})^2 \hat{f}_i^2\right)$$

$$= N^{-3}a^{-2q}E\sum_{p\neq t}^{N} K_{tp}^2(\bar{d} - d_{p-1})^2$$

$$+ N^{-3}a^{-2q}E\sum\sum_{p\neq p'\neq t} K_{tp}K_{tp'}(\bar{d} - d_{p-1})(\bar{d} - d_{p'-1})$$

$$= A_1 + A_2.$$

Using Cauchy–Schwarz, $A_1 = O(N^{-1}a^{-q})$ because $E(\bar{d} - d_{p-1})^2 = O(N)$ and $E(K_{tp})^2 = O(a^q)$ by a direct verification of moment. For A_2, condition on $X_N = (x_0, \ldots x_N)$ so that

$$A_2 = N^{-3}a^{-2q}E\left(K_{tp}K_{tp'}\sum\sum_{p\neq p'\neq t} E_{X_N}(\bar{d} - d_{p-1})(\bar{d} - d_{p'-1})\right).$$

From the identity $E(\sum_t^N (d_t - \bar{d}))^2 \equiv 0$ we find that $\sum\sum_{j\neq m\neq t} E(\bar{d} - d_{j-1})(\bar{d} - d_{m-1}) = O(N^2)$, so $A_2 = O(N^{-1})$ because $E(K_{tp}K_{tp'}) = O(a^{2q})$.

Next,

$$E\left(\frac{1}{N}(\bar{h} - \hat{h}_{t-1})^2 \hat{f}_t^2\right)$$

$$= N^{-3}a^{-2q}E\sum_{p\neq i}^{N} K_{ip}^2(\bar{h} - h_{p-1})^2$$

$$+ N^{-3}a^{-2q}E\sum\sum_{p\neq p'\neq t} K_{tp}K_{tp'}(\bar{h} - h_{p-1})(\bar{h} - h_{p'-1})$$

$$= A_3 + A_4.$$

Here $A_3 = O(N^{-1}a^{-q})$ by the same argument used for A_1. Now we consider the summands in A_4. First,

$$E(K_{tp}K_{tp'}\bar{h}^2) = \left(\frac{1}{N^2}\sum_{r=0}^{N} r^2\right)E(K_{tp}K_{tp'}g(x_s)^2) + \frac{(N-i)^2}{N^2}E(K_{tp}K_{tp'}g(x_t)^2)$$

$$+ \frac{(N-p)^2}{N^2}E(K_{tp}K_{tp'}g(x_p)^2) + \frac{(N-p')^2}{N^2}E(K_{tp}K_{tp'}g(x_{p'})^2).$$

If $p' < p$ and $t < p$,

$$E(K_{tp}K_{tp'}\bar{h}h_{p-1}) = \left(\frac{1}{N}\sum_{r=0}^{p-1}(N-r)\right)E(K_{tp}K_{tp'}g(x_s)^2) + \frac{(N-p')}{N}E(K_{tp}K_{tp'}g(x_{p'})^2)$$

$$+ \frac{(N-t)}{N}E(K_{tp}K_{tp'}g(x_t)^2).$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

If $p' > p$, then the second term is not present in the preceding equation, and if $t > p$, the third term is zero. Similarly, if $t < \min(p-1, p'-1)$,

$$E(K_{tp}K_{tp'}h_{p-1}h_{p'-1}) = \min(p-1,p'-1)E(K_{tp}K_{tp'}g(x_s)^2) + E(K_{tp}K_{tp'}g(x_t)^2).$$

Therefore,

$$|A_4| \leq N^{-3}a^{-2q}E(|K_{tp}K_{tp'}|g(x_s)^2)|A_{41}| + C(Na^q)^{-1}E(|K_{tp}K_{tp'}|g(x_p)^2),$$

where

$$A_{41} = \sum\sum_{p\neq p'\neq i}\left(\frac{1}{N^2}\sum_{r=0}^{N} r^2 - \frac{1}{N}\sum_{r=0}^{p-1}(N-r) - \frac{1}{N}\sum_{r=0}^{p'-1}(N-r) + \min(p-1,p'-1)\right).$$

However, A_{41} is equivalent to the expectation we find in $E\sum\sum_{p\neq p'\neq t}(\bar{h} - h_{p-1}) \times (\bar{h} - h_{p'-1})$, which is $O(N^2)$. The expression $E(|K_{tp}K_{tp'}|g(x_s)^2) = O(a^{2q})$ and $E(|K_{tp}K_{tp'}|g(x_p)^2) = O(a^q)$. Thus, we have $|A_4| = O(N^{-1}) + O(N^{-1}a^{-q})$. Using Cauchy–Schwarz, we can show that

$$E\left(\frac{1}{N}(\bar{d} - \hat{d}_{t-1})(\bar{h} - \bar{h}_{t-1})\hat{f}_t^2\right) = O(N^{-1}a^{-q}),$$

so that $A = O(N^{-1}a^{-q})$ and $N^{-1/2}(\bar{y} - \hat{y}_{t-1})\hat{f}_t = O_p(N^{-1/2}a^{-q/2})$. ■

Proof of Proposition 1. Notice that

$$\begin{aligned}
&\sum_{t=1}^{N}(y_{t-1} - \hat{y}_{t-1})(\Delta y_{t-s} - \widehat{\Delta y}_{t-s})\hat{f}_t^2 \\
&\quad= \sum_{t=1}^{N}(Na^q)^{-2}\left(\sum_{p\neq t}^{N}K_{tp}\left(\sum_{i=1}^{t-1}B(L)[g(x_i)+\epsilon_i] - \sum_{i=1}^{p-1}B(L)[g(x_i)+\epsilon_i]\right)\right) \\
&\qquad\times\left(\sum_{r\neq t}^{N}K_{tr}(B(L)[g(x_{t-s})+\epsilon_{t-s}] - B(L)[g(x_{r-s})+\epsilon_{r-s}])\right) \\
&\quad= \sum_{t=1}^{N}(Na^q)^{-2}\left(\sum_{p\neq t}^{N}K_{tp}\left(\sum_{i=1}^{t-1}\sum_{l=0}^{\infty}b_l[g(x_{i-l})+\epsilon_{i-l}] - \sum_{i=1}^{p-1}\sum_{l=0}^{\infty}b_l[g(x_{i-l})+\epsilon_{i-l}]\right)\right) \\
&\qquad\times\left(\sum_{r\neq t}^{N}K_{tr}\left(\sum_{\nu=0}^{\infty}b_\nu[g(x_{t-s-\nu})+\epsilon_{t-s-\nu}] - \sum_{\nu=0}^{\infty}b_\nu[g(x_{r-s-\nu})+\epsilon_{r-s-\nu}]\right)\right).
\end{aligned}$$

For the term

$$\begin{aligned}
&\sum_{t=1}^{N}(Na^q)^{-2}\left(\sum_{p\neq t}^{N}K_{tp}\left(\sum_{i=1}^{t-1}\sum_{l=0}^{\infty}b_l g(x_{i-l})\right)\right)\left(\sum_{r\neq t}^{N}K_{tr}\left(\sum_{\nu=0}^{\infty}b_\nu g(x_{t-s-\nu})\right)\right) \\
&\quad= \sum_{t=1}^{N}(Na^q)^{-2}\sum_{p\neq t}^{N}\sum_{r\neq t}^{N}\sum_{l=0}^{\infty}\sum_{\nu=0}^{\infty}\sum_{i=1}^{t-1}K_{tp}K_{tr}b_l b_\nu g(x_{i-l})g(x_{t-s-\nu}), \qquad \textbf{(C.1)}
\end{aligned}$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

the second moment of (C.1) is

$$\left(\frac{1}{Na^q}\right)^4 \sum_{t=1}^{N}\sum_{p\neq t}^{N}\sum_{r\neq t}^{N}\sum_{l=0}^{\infty}\sum_{\nu=0}^{\infty}\sum_{i=1}^{t-1}\sum_{\alpha=1}^{N}\sum_{j\neq\alpha}^{N}\sum_{k\neq\alpha}^{N}\sum_{\beta=0}^{\infty}\sum_{\mu=0}^{\infty}\sum_{\tau=1}^{\alpha-1}$$

$$K_{tp}K_{tr}K_{\alpha j}K_{\alpha k}b_l b_\nu b_\beta b_\mu g(x_{i-l})g(x_{t-s-\nu})g(x_{\tau-\beta})g(x_{\alpha-s-\mu}),$$

whose leading term (when $i = t - s + l - \nu$ and $\tau = \alpha - s + \beta - \mu$)

$$\left(\frac{1}{Na^q}\right)^4 \sum_{t=1}^{N}\sum_{p\neq t}^{N}\sum_{r\neq t}^{N}\sum_{l=0}^{\infty}\sum_{\nu=0}^{\infty}\sum_{\alpha=1}^{N}\sum_{j\neq\alpha}^{N}\sum_{k\neq\alpha}^{N}\sum_{\beta=0}^{\infty}\sum_{\mu=0}^{\infty}$$

$$b_l b_\nu b_\beta b_\mu K_{tp}K_{tr}K_{\alpha j}K_{\alpha k}g(x_{t-s-\nu})^2 g(x_{\alpha-s-\mu})^2$$

is of order $O(N^2)$ because $\sum_{l=0}^{\infty}\sum_{\nu=0}^{\infty}\sum_{\beta=0}^{\infty}\sum_{\mu=0}^{\infty} b_l b_\nu b_\beta b_\mu = O(1)$ and

$$E[K_{tp}K_{tr}K_{\alpha j}K_{\alpha k}g(x_{t-s-\nu})^2 g(x_{\alpha-s-\mu})^2] = O(a^{4q}).$$

Thus, (C.1) is of order $O_p(N)$. By similar methods, we can verify that the remaining terms are $O_p(N)$. Consequently, $\sum_{t=1}^{N}(y_{t-1} - \hat{y}_{t-1})(\Delta y_{t-s} - \widehat{\Delta y}_{t-s})\hat{f}_t^2$ is $O_p(N)$. ■

Proof of Proposition 2. This is a direct result from Proposition 1. ■

Proof of Proposition 3. First we use the BN decomposition to write $\sum_{i=1}^{N}(\Delta y_{t-s} - \widehat{\Delta y}_{t-s})(\Delta y_{t-r} - \widehat{\Delta y}_{t-r})\hat{f}_t^2$ into summation of the following terms:

$$P_1 = \frac{1}{N^2a^{2q}}\sum_{t=1}^{N}\sum_{l=0}^{\infty} b_l \sum_{\nu=0}^{\infty} b_\nu \sum_{p\neq t}^{N}\sum_{r'\neq t}^{N} K_{tp}K_{tr'}[g(x_{t-s-l}) - g(x_{p-s-l})]$$

$$\times [g(x_{t-r-\nu}) - g(x_{r'-r-\nu})],$$

$$P_2 = \frac{1}{N^2a^{2q}}\sum_{t=1}^{N}\sum_{l=0}^{\infty} b_l \sum_{\nu=0}^{\infty} b_\nu \sum_{p\neq t}^{N}\sum_{r'\neq t}^{N} K_{tp}K_{tr'}[g(x_{t-s-l}) - g(x_{p-s-l})][\epsilon_{t-r-\nu} - \epsilon_{r'-r-\nu}],$$

$$P_3 = \frac{1}{N^2a^{2q}}\sum_{t=1}^{N}\sum_{l=0}^{\infty} b_l \sum_{\nu=0}^{\infty} b_\nu \sum_{p\neq t}^{N}\sum_{r'\neq t}^{N} K_{tp}K_{tr'}[\epsilon_{t-s-l} - \epsilon_{p-s-l}][g(x_{t-r-\nu}) - g(x_{r'-r-\nu})],$$

$$P_4 = \frac{1}{N^2a^{2q}}\sum_{t=1}^{N}\sum_{l=0}^{\infty} b_l \sum_{\nu=0}^{\infty} b_\nu \sum_{p\neq t}^{N}\sum_{r'\neq t}^{N} K_{tp}K_{tr'}[\epsilon_{t-s-l} - \epsilon_{p-s-l}][\epsilon_{t-r-\nu} - \epsilon_{r'-r-\nu}].$$

Then, similarly to the proof of Proposition 1, we verify that all these terms are $O_p(N)$. ■

Proof of Proposition 4. This is a direct result from Proposition 3. ■

Proof of Proposition 5. This can be obtained by using the results of Propositions 2 and 4. ■

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

Proof of Proposition 6. Similarly to the proofs of Proposition 3, we write

$$\sum_{t=1}^{N}(\Delta y_{t-s}-\widehat{\Delta y}_{t-s})(\epsilon_t-\hat{\epsilon}_t+g(x_t)-\hat{g}(x_t))\hat{f}_t^2=\sum_{i=1}^{6}Q_i,$$

where $Q_1=\sum_{t=1}^{N}\hat{f}_t^2B(L)v_{t-s}\epsilon_t$, $Q_2=\sum_{t=1}^{N}\hat{f}_t(1/Na^q\sum_{p\neq t}^{N}K_{tp}B(L)v_{p-s})\epsilon_t$, $Q_3=\sum_{t=1}^{N}\hat{f}_t^2B(L)v_{t-s}\hat{\epsilon}_t$, $Q_4=\sum_{t=1}^{N}\hat{f}_t(1/Na^q\sum_{p\neq t}^{N}K_{tp}B(L)v_{p-s})\hat{\epsilon}_t$, $Q_5=\sum_{t=1}^{N}\hat{f}_t^2B(L)\times v_{t-s}(g(x_t)-\hat{g}(x_t))$, $Q_6=\sum_{t=1}^{N}\hat{f}_t(1/Na^q\sum_{p\neq t}^{N}K_{tp}B(L)v_{p-s})(g(x_t)-\hat{g}(x_t))$. By straightforward but tedious moment verification, we can show that all of the preceding terms are of order $o_p(N)$. ■

Proof of Proposition 7. The results can be obtained by the results of Propositions 2, 4, and 6. ■

Proof of Proposition 8. We write $N^{-1}\sum_{t=1}^{N}(y_{t-1}^*-\bar{y}^*)\epsilon_t\hat{f}_t^2$ into the summation of $C_1+C_2+C_3$, where $C_1=1/N\sum_{t=1}^{N}(y_{t-1}^*-\bar{y}^*)\epsilon_tf_t^2$, $C_2=1/N\sum_{t=1}^{N}(y_{t-1}^*-\bar{y}^*)\times\epsilon_t(\hat{f}_t-f_t)f_t$, $C_3=1/N\sum_{i=1}^{N}(y_{t-1}^*-\bar{y}^*)\epsilon_t(\hat{f}_t-f_t)^2$. Notice that C_1 converges to the expression given in (A.1) by Lemma 3 of Hansen (1995). We show that the remaining parts converge to zero in mean square. First,

$$\begin{aligned}EC_2^2=&\frac{1}{N^2}E\sum_{t=1}^{N}(y_{t-1}^*-\bar{y}^*)^2\epsilon_t^2(\hat{f}_t-f_t)^2f_t^2\\&+\frac{1}{N^2}B(1)^2E\sum\sum_{t\neq j}(d_{t-1}-\bar{d})(d_{j-1}-\bar{d})\epsilon_t\epsilon_j(\hat{f}_t-f_t)(\hat{f}_j-f_j)f_tf_j\\&+\frac{1}{N^2}B(1)^2E\sum\sum_{t\neq j}(h_{t-1}-\bar{h})(h_{j-1}-\bar{h})\epsilon_t\epsilon_j(\hat{f}_t-f_t)(\hat{f}_j-f_j)f_tf_j\\&+\frac{1}{N^2}B(1)^2E\sum\sum_{t\neq j}(h_{t-1}-\bar{h})(d_{j-1}-\bar{d})\epsilon_t\epsilon_j(\hat{f}_t-f_t)(\hat{f}_j-f_j)f_tf_j\\&+\frac{1}{N^2}B(1)^2E\sum\sum_{t\neq j}(d_{t-1}-\bar{d})(h_{j-1}-\bar{h})\epsilon_t\epsilon_j(\hat{f}_t-f_t)(\hat{f}_j-f_j)f_tf_j\\=&C_{21}+C_{22}+C_{23}+C_{24}+C_{25}\end{aligned}$$

and $C_{21}=O(a^{2\lambda}+N^{-1}a^{-q})$ because $E(\hat{f}_t-f_t)^2=O(a^{2\lambda}+N^{-1}a^{-q})$ from a direct calculation of expectation. The term $C_{23}=0$ because $E(\epsilon_t\epsilon_j)=0$. Conditioning on X_N gives

$$C_{22}=B(1)^2E\left((\hat{f}_t-f_t)(\hat{f}_j-f_j)f_tf_j\frac{1}{N^2}E_{X_N}\sum\sum_{t\neq j}(d_{t-1}-\bar{d})(d_{j-1}-\bar{d})\epsilon_t\epsilon_j\right).$$

The inner expectation is $O(N^2)$ and $E((\hat{f}_t-f_t)(\hat{f}_j-f_j)f_tf_j)=O(a^{2\lambda}+N^{-1}a^{-q})$ by Cauchy–Schwarz so that $C_{23}=O(a^{2\lambda}+N^{-1}a^{-q})$. Next,

$$C_{24}=B(1)^2E\left(\frac{1}{N^2}\sum\sum_{t\neq j}(h_{t-1}-\bar{h})(\hat{f}_t-f_t)(\hat{f}_j-f_j)f_tf_jE_{X_N}(d_{j-1}-\bar{d})\epsilon_t\epsilon_j\right).$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

The inner expectation is zero so that $C_{24} = 0$ and $C_{25} = 0$. The proof of EC_3^2 is similar except that we use $E(\hat{f}_1 - f_1)^4 = O(a^{4\lambda} + N^{-2}a^{-2q})$. ■

Proof of Proposition 9.

$$E\left(\frac{1}{N}\sum_{t=1}^{N}(\bar{y}^* - \hat{y}_{t-1}^*)\epsilon_t \hat{f}_t^2\right)^2 = \frac{1}{N^2}E\sum_{t=1}^{N}(\bar{y}^* - \hat{y}_{t-1}^*)^2\epsilon_t^2\hat{f}_t^4$$

$$+ \frac{1}{N^2}B(1)^2E\sum\sum_{t\neq j}(\bar{h} - \hat{h}_{t-1})(\bar{h} - \hat{h}_{j-1})\hat{f}_t^2\hat{f}_j^2\epsilon_t\epsilon_j$$

$$+ \frac{1}{N^2}B(1)^2E\sum\sum_{t\neq j}(\bar{d} - \hat{d}_{t-1})(\bar{d} - \hat{d}_{j-1})\hat{f}_t^2\hat{f}_j^2\epsilon_t\epsilon_j$$

$$+ \frac{1}{N^2}B(1)^2E\sum\sum_{t\neq j}(\bar{h} - \hat{h}_{t-1})(\bar{d} - \hat{d}_{j-1})\hat{f}_t^2\hat{f}_j^2\epsilon_t\epsilon_j$$

$$+ \frac{1}{N^2}B(1)^2E\sum\sum_{t\neq j}(\bar{d} - \hat{d}_{t-1})(\bar{h} - \hat{h}_{j-1})\hat{f}_t^2\hat{f}_j^2\epsilon_t\epsilon_j$$

$$= D_1 + D_2 + D_3 + D_4 + D_5.$$

By using a proof similar to the proof of Lemma 3.1, we show that $D_1 = O(N^{-1}a^{-q})$. Conditioning on X_N, $D_2 = 0$ because $E(\epsilon_t\epsilon_j) = 0$. Conditioning on X_N, $D_4 = 0$ and $D_5 = 0$ because $E_{X_N}(\bar{d} - \hat{d}_{t-1})\epsilon_t\epsilon_j = 0$. For D_3,

$$E(\bar{d} - \hat{d}_{t-1})(\bar{d} - \hat{d}_{j-1})\hat{f}_t\hat{f}_j\epsilon_t\epsilon_j$$

$$= (Na^q)^{-2}E\sum_{p\neq t\neq j}^{N}K_{tp}K_{jp}(\bar{d} - d_{p-1})^2\epsilon_t\epsilon_j$$

$$+ (Ta^q)^{-2}E\left(K_{tp}K_{jp'}E\left(\sum\sum_{p\neq p'\neq t\neq j}(\bar{d} - d_{p-1})(\bar{d} - d_{p'-1})\epsilon_t\epsilon_j|X_N\right)\right)$$

$$= a^{-2q}E(K_{ip}K_{jp})D_{31} + a^{-2q}E(K_{tp}K_{jp'})D_{32}.$$

We find that

$$D_{32} = 2\frac{(N-t)(N-j)}{N^2} + 2\frac{\max(t,j)^2}{N^2} - 2\frac{\max(t,j)}{N^2} + 2\frac{\min(t,j)}{N^2}$$

$$- 2\frac{(N - \max(t,j))(N - \min(t,j))}{N^2} - 2\frac{(N - \max(t,j))(N - \max(t,j))}{N^2}$$

$$+ 1 - \frac{tj}{N^2}.$$

After algebra, $N^{-2}\sum\sum_{t\neq j}^{N}D_{32} = O(N^{-1})$, and similarly $N^{-2}\sum\sum_{t\neq j}^{N}D_{31} = O(N^{-1})$, so that $D_3 = O(N^{-1})$. ■

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

Proof of Proposition 10. Notice that

$$\frac{1}{N^2} E\left| \sum_{t=1}^{N} (y_{t-1}^* - \bar{y}^*)(\bar{y}^* - \hat{y}_{t-1}^*)\hat{f}_t^2 \right|$$

$$\leq \frac{1}{N^{3/2}} \sum_{t=1}^{N} \sqrt{E(y_{t-1}^* - \bar{y}^*)^2 \hat{f}_t^2} \sqrt{E \frac{1}{N} (\bar{y}^* - \hat{y}_t^*)^2 \hat{f}_t^2}$$

by Cauchy–Schwarz. Because $E(y_{t-1}^* - \bar{y}^*)^2 = O(N)$ and $E(1/N)(\bar{y}^* - \hat{y}_{t-1}^*)^2 \hat{f}_t^2 = O(N^{-1}a^{-q})$, thus the right term is $O(N^{-1/2}a^{-q/2})$. ■

Proof of Proposition 11.

$$\frac{1}{N^2} \sum_{t=1}^{N} (y_{t-1}^* - \hat{y}_{t-1}^*)^2 \hat{f}_t^2 = \frac{1}{N^2} \sum_{t=1}^{N} (y_{t-1}^* - \bar{y}^*)^2 \hat{f}_t^2 + 2 \frac{1}{N^2} \sum_{t=1}^{N} (y_{t-1}^* - \bar{y}^*)(\bar{y}^* - \hat{y}_{t-1}^*)\hat{f}_t^2$$

$$+ \frac{1}{N^2} \sum_{t=1}^{N} (\bar{y}^* - \hat{y}_{t-1}^*)^2 \hat{f}_t^2.$$

The second and third terms on the right-hand side converge to zero by Propositions 10 and a proof similar to that of Lemma 3.1. Using the method in the proof of Proposition 8, the first term converges to the right-hand side of (A.3). ■

Proof of Proposition 12.

$$\frac{1}{N} \sum_{t=1}^{N} (y_{t-1}^* - \hat{y}_{t-1}^*)\hat{\epsilon}_t \hat{f}_t^2 = \frac{1}{N} \sum_{t=1}^{N} (y_{t-1}^* - \bar{y}^*)\hat{\epsilon}_t \hat{f}_t^2 + \frac{1}{N} \sum_{t=1}^{N} (\bar{y}^* - \hat{y}_{t-1}^*)\hat{\epsilon}_t \hat{f}_t^2 = F_1 + F_2.$$

Note that $E(\hat{\epsilon}_t^2 \hat{f}_t^2) = O(N^{-1}a^{-q})$ so that $E(F_2^2) = O(N^{-1}a^{-2q})$ from Loéve's inequality. Now

$$E(F_1^2) = E\left(\frac{1}{N^2} \sum_{t=1}^{N} (y_{t-1}^* - \bar{y}^*)^2 \hat{\epsilon}_t^2 \hat{f}_t^4 \right)$$

$$+ E\left(\frac{1}{N^2} \sum \sum_{t \neq j} (y_{t-1}^* - \bar{y}^*)(y_{j-1}^* - \bar{y}^*)\hat{\epsilon}_t \hat{\epsilon}_j \hat{f}_t^2 \hat{f}_j^2 \right)$$

$$= F_{11} + F_{12};$$

$F_{11} = O(N^{-1}a^{-q})$ because $E(y_{t-1}^* - \bar{y}^*) = O(N)$. Consider the summands in F_{12}. It is easy to verify that

$$E(y_{t-1}^* - \bar{y}^*)(y_{j-1}^* - \bar{y}^*)\hat{\epsilon}_t \hat{\epsilon}_j \hat{f}_t^2 \hat{f}_j^2 = B(1)^2 E(\hat{\epsilon}_t \hat{\epsilon}_j \hat{f}_t^2 \hat{f}_j^2 v_s^2) F_{13}$$

$$+ B(1)^2 E(\hat{\epsilon}_t \hat{\epsilon}_j \hat{f}_t^2 \hat{f}_j^2 v_{s'} v_j) F_{14}$$

$$+ B(1)^2 E(\hat{\epsilon}_t \hat{\epsilon}_j \hat{f}_t^2 \hat{f}_j^2 v_{i'} v_i) F_{15} + o(1),$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

where $s \neq t \neq j$, $s' \neq t \neq j$, and $i \neq i' \neq t \neq j$. In addition,

$$F_{13} = \min(t-1, j-1) - \frac{1}{N}\sum_{r=0}^{t-1}(N-r) - \frac{1}{N}\sum_{r=0}^{j-1}(N-r) + \frac{1}{N^2}\sum_{r=0}^{N} r^2,$$

$$F_{14} = (N-t)\left(1 - \frac{(t-1)}{N} - \frac{(j-1)}{N}\right) + (N-j)\left(1 - \frac{(j-1)}{N} - \frac{(t-1)}{N}\right)$$

$$- \frac{N}{2} + \min(t-1, j-1),$$

$$F_{15} = (t-1)(j-1) - \frac{(t-1)(N+1)}{2} - \frac{(j-1)(N+1)}{2} + \frac{(N+1)^2}{4} - F_{13} - F_{14}.$$

As a result of cancellations, we have

$$\frac{1}{N^2}\sum\sum_{t\neq j} F_{13} = O(1), \frac{1}{N^2}\sum\sum_{t\neq j} F_{14} = O(1), \frac{1}{N^2}\sum\sum_{t\neq j} F_{15} = O(N).$$

Using $\hat{f}_t = f_t + (\hat{f}_t - f_t)$, it is easy to see that $E(\hat{\epsilon}_t\hat{\epsilon}_j\hat{f}_t^2\hat{f}_j^2 v_s^2)$ has the same order of magnitude as $E(\hat{\epsilon}_t\hat{\epsilon}_j\hat{f}_t\hat{f}_j f_t f_j v_s^2)$ (which has only two summations), and it can be verified that $E(\hat{\epsilon}_t\hat{\epsilon}_j\hat{f}_t\hat{f}_j f_t f_j v_s^2) = O(N^{-1}a^{-q})$. Similarly, $E(\hat{\epsilon}_t\hat{\epsilon}_j\hat{f}_t^2\hat{f}_j^2 v_{s'} v_j)$ and $E(\hat{\epsilon}_t\hat{\epsilon}_j\hat{f}_t^2\hat{f}_j^2 v_{i'} v_i)$ have the same order of magnitudes as $E(\hat{\epsilon}_t\hat{\epsilon}_j\hat{f}_t\hat{f}_j f_t f_j v_{s'} v_j)$ and $E(\hat{\epsilon}_t\hat{\epsilon}_j\hat{f}_t\hat{f}_j f_t f_j v_{i'} v_i)$, respectively, and it can be verified that $E(\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_t^2\hat{f}_j^2 v_{s'} v_j) = O(N^{-1}a^{-q})$ and $E(\hat{\epsilon}_t\hat{\epsilon}_j\hat{f}_t^2\hat{f}_j^2 v_{i'} v_i) = O(N^{-2}a^{-2q})$ by Cauchy–Schwarz so that $F_{12} = O(N^{-1}a^{-2q})$, and thus

$$\frac{1}{N}\sum_{i=1}^{N}(y_{t-1}^* - \hat{y}_{t-1}^*)\hat{\epsilon}_t\hat{f}_t^2 = O_p(N^{-1/2}a^{-q}).$$

Proof of Proposition 13. Let $\xi = \min(\lambda + 1, \nu)$. The proof follows Proposition 12.

$$\frac{1}{N}\sum_{t=1}^{N}(y_{t-1}^* - \hat{y}_{t-1}^*)(g(x_t) - \hat{g}(x_t))\hat{f}_t^2 = \frac{1}{N}\sum_{t=1}^{N}(y_{t-1}^* - \bar{y}^*)(g(x_t) - \hat{g}(x_t))\hat{f}_t^2$$

$$+ \frac{1}{N}\sum_{t=1}^{N}(\bar{y}^* - \hat{y}_{t-1}^*)(g(x_t) - \hat{g}(x_t))\hat{f}_t^2.$$

However, we need to find the order of $E(g(x_t) - \hat{g}(x_t))^2\hat{f}_t^2$ and

$$E((g(x_t) - \hat{g}(x_t))(g(x_j) - \hat{g}(x_j))\hat{f}_t^2\hat{f}_j^2 v_{i'} v_i).$$

First, $\mathrm{E}(g(x_t) - \hat{g}(x_t))^2\hat{f}_t^2 = O(N^{-1}a^{-q} + a^{2\xi})$, from the proof of Proposition 1 in Robinson. Next, we again use $\hat{f}_t = f_t + (\hat{f}_t - f_t)$ and that $E((g(x_t) - \hat{g}(x_t))(g(x_j) - \hat{g}(x_j))\hat{f}_t^2\hat{f}_j^2 v_{i'} v_i)$ has the same order of magnitude as

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press

$$E((g(x_t) - \hat{g}(x_t))\hat{f}_t(g(x_j) - \hat{g}(x_j))\hat{f}_j f_t f_j v_{i'} v_i),$$

which is $O(N^{-2}a^{-2q})$. Combining these results and using the proof of Proposition 12, we find

$$\frac{1}{N}\sum_{t=1}^{N}(y^*_{t-1} - \hat{y}^*_{t-1})(g(x_t) - \hat{g}(x_t))\hat{f}_t^2 = O_p(N^{-1/2}a^{-q} + a^{\xi-(q/2)}).$$

https://doi.org/10.1017/S0266466605050450 Published online by Cambridge University Press