PARTIALLY LINEAR MODELS WITH UNIT ROOTS

By

Ted Juhl and Zhijie Xiao

April 2002

COWLES FOUNDATION DISCUSSION PAPER NO. 1359

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

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

PARTIALLY LINEAR MODELS WITH UNIT ROOTS

TED JUHL AND ZHIJIE XIAO

Abstract. 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). It is proven 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 which 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:

$$y_i = \beta' z_i + g(x_i) + \epsilon_i, \tag{1.1}$$

where $g(\cdot)$ is an unknown real function, and β is the vector of unknown parameters that we want to estimate. This type of specification arises when the primary interest is on the parameter β, while the relation of the mean response to additional variables x_i 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. Wabba (1984), Engle et al. (1986), and many others studied the penalized

We thank Bruce Hansen, Roger Koenker, Helmut Lütkepohl, Peter Phillips, 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. Correspondence: Ted Juhl, Department of Economics, University of Kansas (juhl@ukans.edu) and Zhijie Xiao, Department of Economics, University of Illinois at Urbana Champaign (zxiao@uiuc.edu).

least squares method in partially linear regression estimation. Heckman (1986), Chen (1988), Rice (1986), and Speckman (1988) studied $\sqrt{N}$-consistency of β 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 β. Under regularity and smoothness conditions, $\sqrt{N}$-consistency and asymptotic normality are obtained. Robinson also showed that when the errors are *iid* normal, this estimator achieves the semiparametric information bound. A higher order asymptotic analysis of the partially linear regression estimators is given by Linton (1995). In time series, Fan and Li (1999) extended the $\sqrt{N}$-consistency and asymptotic normality results for independent observations to weakly dependent processes. 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 iid 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 fifteen 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 use 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 which focuses on estimating potential trends under the local alternative hypothesis in order 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 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.

While 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, for example, Tong (1990), Granger and Terasvirta (1993), and Granger (1995) among others]. 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 (1999) studied nonlinear autoregressive models and show 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:

$$y_i = \alpha y_{i-1} + g(x_i) + \epsilon_i, \tag{1.2}$$

where α is close to unity and x_i is a stationary covariate. When α is exactly unity, y_i 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 autoregressive root α, but we allow for an unknown nonlinear function of covariates, $g(x)$, to influence the time series. In allowing such a general structure, we hope to improve the efficiency in estimating the 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. Since the form of the nonlinearity is unknown, we estimate this part of the model nonparametrically while modeling the autoregressive component linearly.

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 Robinson's (1988) result to the nonstationary case. In addition, asymptotic distributions of partially linear estimates of the largest autoregressive root and its t-statistic are derived. The limiting distribution of the unit root test is nearly 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 is substantial. Finally, in the course of proving our theorem, we show that nonparametrically regressing an I(1) series on an I(0) series is asymptotically 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 generalization to higher order autoregressions is developed in Section 4. A Monte Carlo experiment is given in Section 5 and Section 6 concludes.

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

2. The Model and The Estimator

We begin with the following time series model with a deterministic component τ_i and a stochastic component s_i:

$$y_i = \tau_i + s_i. \tag{2.1}$$

We write $\tau_i = \varphi' \kappa_i$, where φ is a vector of coefficients and κ_i is a deterministic component of known form. The stochastic component s_i is modeled as

$$\Delta s_i = \delta s_{i-1} + v_i. \tag{2.2}$$

However, there are additional stationary covariates which help explain v_i, so that

$$v_i = [g(x_i) - \mu_g] + \epsilon_i,$$

where $g : \mathbb{R}^q \to \mathbb{R}$ and $E(g(x_i)) = \mu_g$. (A generalization to higher order autoregressions will be discussed in Section 4). Combining (2.1) and (2.2) gives

$$\Delta y_i = \gamma' \kappa_i + \delta y_{i-1} + g(x_i) + \epsilon_i. \tag{2.3}$$

Our primary interest is the estimation and tests on δ. For identification reasons, we absorb the "intercept" term into the nonlinear function $g(x_i)$ and signify the corresponding new term as $g^*(x_i)$. Thus, the "intercept" term is excluded from κ_i and we denote the rest of the deterministic component as $\underline{\kappa}_i$also see discussions at the end of this Section on related issues). Let $z_i = (y_{i-1}, \underline{\kappa}_i)$, (2.3) can be re-written as

$$\Delta y_i = \beta' z_i + g^*(x_i) + \epsilon_i.$$

We estimate β, and thus δ, by the following steps: First, regress Δy_i and z_i nonparametrically on x_i and denote the nonparametric regressions residuals as $\hat{e}_{di}$ and $\hat{e}_{zi}$ respectively. Next, regress the residuals $\hat{e}_{di}$ on $\hat{e}_{zi}$ by least squares method to get an estimate of β.

The leading cases of the deterministic component are (i) $\tau_i = \mu$ or (ii) $\tau_i = \mu + \theta i$ and we focus our attention on these cases. First, consider the case where $\tau_i = \mu$ so that model (2.3) becomes

$$\Delta y_i = \mu^* + \delta y_{i-1} + g(x_i) + \epsilon_i, \tag{2.4}$$

where $\mu^* = -\delta\mu - \mu_g$. In this case, $z_i$is simply y_{i-1} and $\beta = \delta$. Thus in the first step 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

$$\hat{f}_i = \frac{1}{Na^q}\sum_{j=1,j\neq i}^{N} K_{ij}; \qquad \hat{y}_{i-1} = \frac{1}{Na^q}\frac{\sum_{j=1,j\neq i}^{N} K_{ij}y_{j-1}}{\hat{f}_i}; \qquad \hat{\Delta y}_i = \frac{1}{Na^q}\frac{\sum_{j=1,j\neq i}^{N} K_{ij}\Delta y_j}{\hat{f}_i};$$

and the nonparametric regression residuals are

$$\hat{e}_{di} = \Delta y_i - \hat{\Delta y}_i; \;\; \hat{e}_{yi} = y_{i-1} - \hat{y}_{i-1}.$$

The vectors of residuals from regressing Δy_i and y_{i-1} on x_i are denoted $\hat{e}_d$ and $\hat{e}_y$ respectively.

Next, we estimate δ by regressing $\hat{e}_d$ on $\hat{e}_y$. The above 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 $\hat{e}_y$ using OLS and incorporating the density-weighting, so that

$$\hat{\delta} = \left(\left(\hat{e}_y \odot \hat{f}\right)^{\top}\left(\hat{e}_y \odot \hat{f}\right)\right)^{-1}\left(\left(\hat{e}_y \odot \hat{f}\right)^{\top}\left(\hat{e}_d \odot \hat{f}\right)\right), \tag{2.5}$$

where $\odot$ denotes the Hadamard product.

Asymptotic analysis of $\hat{\delta}$ uses the decomposition

$$\hat{\delta} - \delta = \left(\left(\hat{e}_y \odot \hat{f}\right)^{\top}\left(\hat{e}_y \odot \hat{f}\right)\right)^{-1}\left(\left(\hat{e}_y \odot \hat{f}\right)^{\top}\left(\hat{e}_v \odot \hat{f}\right)\right),$$

where $\hat{e}_v$ is the vector of $(\epsilon_i - \widehat{\epsilon}_i + g(x_i) - \widehat{g}(x_i))$ where

$$\widehat{\epsilon}_i = \frac{1}{Na^q}\frac{\sum_{j=1,j\neq i}^{N} K_{ij}\epsilon_j}{\hat{f}_i}; \qquad \widehat{g}(x_i) = \frac{1}{Na^q}\frac{\sum_{j=1,j\neq i}^{N} K_{ij}g(x_j)}{\hat{f}_i}. \tag{2.6}$$

As in Robinson (1988), our result depends on showing that the bias in $g(x_i) - \hat{g}(x_i)$ is negligible. In our case, we need to show that $N^{-1}\sum_{i=1}^{N}(y_{i-1} - \hat{y}_{i-1})(g(x_i) - \hat{g}(x_i))\hat{f}_i^2 = o_p(1)$. Since y_i is nonstationary, a new method of proof is necessary to account for smoothing integrated time series.

Now consider the case where $\tau_i = \mu + \theta i$ so that the model (2.3) becomes

$$\Delta y_i = \mu^* + \theta^* i + \delta y_{i-1} + g(x_i) + \epsilon_i, \tag{2.7}$$

where $\mu^* = \theta(1+\delta) - \delta\mu - \mu_g$ and $\theta^* = -\delta\theta$. Thus, $z_i = (y_{i-1}, i)$. We introduce another term which accounts for the estimated trend,

$$\hat{i} = \frac{1}{Na^q} \frac{\sum_{j=1, j\neq i}^{N} K_{ij} j}{\hat{f}_i}.$$

Let the vector of residuals from nonparametrically regressing the trend on x_i be denoted $\hat{e}_t$. Then we regress $\hat{e}_d \odot \hat{f}$ on $\hat{e}_y \odot \hat{f}$ and $\hat{e}_t \odot \hat{f}$ using OLS so that $\hat{\delta}$ is the coefficient on $\hat{e}_y \odot \hat{f}$.

Notice that by nature of the semiparametric partial linear regression, an intercept term is not identified unless the model is further restricted. Consequently, the estimation of δ on the model with $\tau_i = 0$ (i.e. with no constant term) is the same as that on the model with $\tau_i = \mu$ and thus is given by (2.5). 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 we do not separate the cases of $\tau_i = 0$ and $\tau_i = \mu$ 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.[1]

3. Main Results

We derive the asymptotic properties of the proposed partial 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 hypothesis of a unit root. For purposes of determining asymptotic distributions, we use local to unity asymptotics [Phillips (1987) and Chan and Wei (1987)] 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 assume that the system is initialized by setting $y_0 = 0$ (or, more generally, any random variable

[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.

with finite variance). 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. When $W(r)$ is a Brownian motion, $W^c(r)$ is the conventional Ornstein-Uhlenbeck process.

Following Hansen (1995), we define $\sigma_{v\epsilon f} = Cov(v_i, \epsilon_i f_i^2)$, $\sigma_v^2 = Var(v_i^2)$, $\sigma_{\epsilon f}^2 = E(\epsilon_i^2 f_i^4)$, and

$$\rho^2 = \frac{\sigma_{v\epsilon f}^2}{\sigma_v^2 \sigma_{\epsilon f}^2},$$

with $f_i = f(x_i)$.

The following definitions are given in Robinson (1988), and 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.

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

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

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.

Lemma 3.1. *Let $f(x)$ be the density of x_i. If $\sup_x f(x) < \infty$, $E|g(X)| < \infty$, $\sup_u |k(u)| + \int |k(u)| du < \infty$, and $N^{-1}a^{-q} \to 0$ then*

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

The above result indicates that if one nonparametrically regresses y_{i-1} 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 asymptotic distribution of the semiparametric partially linear regression estimator $\hat{\delta}$ is summarized in the following Theorem.

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 $N \to \infty$, $N^{-1}a^{-2q} \to 0$, $a^{2\min(\lambda+1,\nu)-q} \to 0$; (vi) $k \in \mathcal{K}_{l+n-1}$ for integers l and n such that $l-1 < \lambda \leq l$,$n-1 < \nu \leq n$; (vii) $\sigma_v^2 > 0$ and $\rho^2 > 0$; (viii) μ_g=0 for model 2.4. Then*

$$N(\hat{\delta}-\delta) \Rightarrow b^2 \left(\int (\underline{W}_1^c)^2 \right)^{-1} \left(\rho^2 \int \underline{W}_1^c dW_1 + \sqrt{\rho^2 - \rho^4} \int \underline{W}_1^c dW_2 \right), \tag{3.2}$$

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 (2.4), $\underline{W}_1^c = W_1^c(r) + (6r-4)\int W_1^c(s)ds - (12r-6)\int W_1^c(s)sds$ for model (2.7), and $b^2 = E(f^4)/(E(f^2))^2$.

Assumptions (i)-(v) are similar to Robinson (1988). Just like the parametric ADF (Augmented Dickey-Fuller) regression where the limiting distribution of the AR coefficient estimator (and the unit root tests based on it) is invariant to the parameters of the trend functions and the lag-differenced dependent variables, the proposed semiparametric partial linear regression also has the desirable property that the limiting distribution of $\hat{\delta}$ is invariant to the unknown nonlinear functional form $g(\cdot)$. The limiting distribution given in Theorem 3.2 is similar to Theorem 2 in Hansen (1995). However, our distribution contains an additional factor of b^2 due to the use of density weighting. Moreover, we do not separate the case of $\tau_i = 0$ from the case $\tau_i = \mu$ and thus there is not a third model without a constant term. As we have mentioned in the previous section, the reason is the well known fact that in this form of semiparametric estimation, the intercept term is not identified.

The analog of the Dickey-Fuller test in our model is the t-ratio statistic of $\hat{\delta}$.

Theorem 3.3. *Under the Assumptions of Theorem 3.2, the t-statistic based on $\hat{\delta}$ has limiting distribution*

$$t(\hat{\delta}) \Rightarrow -\frac{c}{b\rho}\left(\int (\underline{W}_1^c)^2\right)^{\frac{1}{2}} + b\left(\int (\underline{W}_1^c)^2\right)^{-\frac{1}{2}}\left(\rho \int \underline{W}_1^c dW_1\right) + b\sqrt{1-\rho^2}N(0,1). \tag{3.3}$$

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,

$$t(\hat{\delta}) \Rightarrow b\left(\int (\underline{W}_1)^2\right)^{-\frac{1}{2}}\left(\rho \int \underline{W}_1 dW_1\right) + b\sqrt{1-\rho^2}N(0,1). \tag{3.4}$$

The limiting distribution in Theorem 3.3 is 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 is employed.

The distribution has the disadvantage that ρ is a remaining nuisance parameter, and in our case, there is an additional nuisance parameter in b. 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. Since there are already critical values for given values of ρ provided in Hansen (1995), we propose dividing the t-ratio by a consistent estimate of b and comparing the resulting statistic with the tabulated critical values. Consider the following consistent estimate of b

$$\hat{b} = \frac{\sqrt{N^{-1}\sum_{i=1}^{N}\hat{f}_i^4}}{N^{-1}\sum_{i=1}^{N}\hat{f}_i^2}$$

and transformation of $t(\hat{\delta})$ based on $\hat{b}$:

$$t^*(\hat{\delta}) = t(\hat{\delta})/\hat{b},$$

then the limiting null distribution of $t^*(\hat{\delta})$ is

$$\rho \left(\int (\underline{W}_1)^2 \right)^{-\frac{1}{2}} \int \underline{W}_1 dW_1 + \sqrt{1-\rho^2} N(0,1). \tag{3.5}$$

If we estimate v_i by the residual from on OLS regression of Δy_i on y_{i-1} and a constant, and estimate $\epsilon_i f_i$ by

$$\tilde{\epsilon}_i \hat{f}_i = (\hat{e}_{di} - \hat{\delta}\hat{e}_{yi})\hat{f}_i,$$

the estimates $\hat{v}_i$ and $\tilde{\epsilon}_i \hat{f}_i$ can then used to obtain a consistent estimate of ρ in the same way as Hansen (1995). Finally, we will employ the statistic $t^*(\hat{\delta})$ to allow comparison to the critical values in Table 1 of Hansen (1995).

4. Generalizations

The results in the previous sections generalize to higher order autoregressions which are often used in practice to remove serial correlation in the residuals. In this Section, we extend the nonstationary partial linear model (2.1) to the case with k-th order lagged dependent variables. Such a model can also be viewed as a partial linear ADF (augmented Dickey-Fuller) regression.

Consider the following AR(k) generalization of (2.2)

$$A(L)\Delta y_i = \delta y_{i-1} + v_i, \tag{4.1}$$

where $A(L) = 1 - a_1 L - \cdots - a_k L^k$ is a k-th order polynomial in the lag operator. We also assume that all roots of $A(L)$ lie outside the unit circle so that $A(L)v_i$ is stationary. Again, we focus our discussion on the case where the largest autoregressive root is close to unity and assume for simplicity of exposition that $y_0 = 0$. The model contains a unit root under the null hypothesis $H_0 : \delta = 0$ and allows for local departures from the hypothesis by setting

$$\delta = -cA(1)/N.$$

Using a BN (Beveridge and Nelson 1981) decomposition for $A(L)$ as $A(L) = A(1) + A^*(L)(1 - L)$ and denote $\zeta_i = A^*(L)\Delta y_i$, we have

$$A(L)\Delta y_i = -\frac{c}{N}A(L)y_{i-1} + \frac{c}{N}\zeta_{i-1} + v_i.$$

If we denote $y_{ai} = A(L)y_i$ and define y^*_{ai} as

$$\Delta y^*_{ai} = -\frac{c}{N}y^*_{a,i-1} + v_i,$$

then, for all c in a compact set,

$$\frac{1}{\sqrt{N}}\sup_{i\leq N}|y_{ai} - y^*_{ai}| \xrightarrow{p} 0.$$

Of course, when $c = 0$, $y_{ai} = y^*_{ai}$. Let $B(L) = A(L)^{-1}$, we have $B(L) = B(1) + B^*(L)(1 - L)$, where $B^*(L)$ has all roots outside the unit root circle because $A^*(L)$ does, then

$$y_i = \sum_{l=0}^{i}\left(1 - \frac{c}{N}\right)^{i-l} B(1)v_i + O_p(\frac{1}{\sqrt{N}}).$$

Following the notation in the previous Section, we denote

$$\Delta\hat{y}_{i-s} = \frac{1}{Na^q}\frac{\sum_{j=1,j\neq i}^{N} K_{ij}\Delta y_{j-s}}{\hat{f}_i}, \ \ s = 1, ..., k.$$

and the corresponding nonparametric regression residuals from regressing Δy_{i-s} on x_i as

$$\hat{e}_{d,i-s} = \Delta y_{i-s} - \Delta\hat{y}_{i-s}.$$

Furthermore, the vectors of residuals $\hat{e}_{d,i-s}$ are denoted $\hat{e}_{d,s}$ and the matrix of $\{\hat{e}_{d,s}, s = 1, ..., k\}$ is denoted as $\hat{e}_s$.

We estimate δ by regressing $\hat{e}_d$ on $\hat{e}_y$ and $\hat{e}_s$ using a density-weighted regression and denote the regression coefficient on $\hat{e}_y$ by $\hat{\delta}$. By partitioned regression we obtain

$$\hat{\delta} = \left(\left(\hat{e}_y \odot \hat{f} \right)^\top (I - P) \left(\hat{e}_y \odot \hat{f} \right) \right)^{-1} \left(\left(\hat{e}_y \odot \hat{f} \right)^\top (I - P) \left(\hat{e}_d \odot \hat{f} \right) \right), \tag{4.2}$$

where

$$P = \left(\hat{e}_s \odot \hat{f} \right) \left[\left(\hat{e}_s \odot \hat{f} \right)^\top \left(\hat{e}_s \odot \hat{f} \right) \right]^{-1} \left(\hat{e}_s \odot \hat{f} \right)^\top$$

is a projection matrix.

To derive the limiting distribution of $\widehat{\delta}$ given by (4.2), notice that

$$\hat{\delta} = \delta + \left(\left(\hat{e}_y \odot \hat{f} \right)^\top (I - P) \left(\hat{e}_y \odot \hat{f} \right) \right)^{-1} \left(\left(\hat{e}_y \odot \hat{f} \right)^\top (I - P) \left(\hat{e}_v \odot \hat{f} \right) \right),$$

where $\hat{e}_v$ is the vector of $(\epsilon_i - \widehat{\epsilon}_i + g(x_i) - \widehat{g}(x_i))$. We show in the Appendix that

$$\left(\hat{e}_y \odot \hat{f} \right)^\top (I - P) \left(\hat{e}_y \odot \hat{f} \right) = \left(\hat{e}_y \odot \hat{f} \right)^\top \left(\hat{e}_y \odot \hat{f} \right) + o_p(N^2)$$

and

$$\left(\hat{e}_y \odot \hat{f} \right)^\top (I - P) \left(\hat{e}_\epsilon \odot \hat{f} \right) = \left(\hat{e}_y \odot \hat{f} \right)^\top \left(\hat{e}_v \odot \hat{f} \right) + o_p(N).$$

Then, we derive the limiting distributions of $\left(\hat{e}_y \odot \hat{f} \right)^\top \left(\hat{e}_y \odot \hat{f} \right)$ and $\left(\hat{e}_y \odot \hat{f} \right)^\top \left(\hat{e}_\epsilon \odot \hat{f} \right)$. It is shown in the Appendix that similar asymptotic results as the previous sections can be obtained. These results are summarized in the following Theorems.

Theorem 4.1. *Under the assumptions of Theorem 3.2,*

$$N(\hat{\delta} - \delta) \Rightarrow A(1) b^2 \left(\int (\underline{W}_1^c)^2 \right)^{-1} \left(\rho^2 \int \underline{W}_1^c dW_1 + \sqrt{\rho^2 - \rho^4} \int \underline{W}_1^c dW_2 \right), \tag{4.3}$$

where W_2 and W_1 are independent standard Brownian motions, and $\underline{W}_1^c$ is defined in the same way as in Theorem 3.2.

Theorem 4.2. *Under the Assumptions of Theorem 3.2, the t-statistic of $\hat{\delta}$ has the same limiting distribution as that of (3.3) in Corollary 3.3 .*

Again, we can test the unit root hypothesis using the modified t-ratio statistic$t^*(\hat{\delta}) = t(\hat{\delta})/\hat{b}$,where $t(\hat{\delta}) = \widehat{\delta}/se(\hat{\delta})$, and $se(\hat{\delta})$ is the OLS standard error for $\hat{\delta}$. The limiting null distribution of $t^*(\hat{\delta})$ is given by (3.5) and the test can be conducted in the same way as that described in Section 3.

5. Monte Carlo

In this section, a small simulation study is conducted to examine the finite sample performance of the nonstationary partial linear estimation and the associated unit root test. 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 test $t^*(\widehat{\delta})$ using the Partial Linear Model which we denote PLMUR test. The data generating process is

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

The different functions are listed below:

$$g_1(x) = 0;\ g_2(x) = 2x;\ g_3(x) = 2x_1x_2;\ g_4(x) = x^2 - 1;\ g_5(x) = x^3 - x.$$

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 since the covariates enter linearly. We include $g_3(x)$ for the purpose of checking the ability of the PLMUR test to use multiple covariates. The nonparametric estimates $\hat{\Delta y}$ 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 that 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 $v_i = g(x_i) + \epsilon_i$. Therefore, we expect

Table 1: Size

		Model (2.4)			Model (2.7)		
		DF	CADF	PLMUR	DF	CADF	PLMUR
	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
$c = 0$	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

more powerful tests if ρ is small. Straightforward calculations show that

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

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 a Gaussian kernel. The bandwidth is set to $N^{-\frac{1}{5}}$ 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 $\delta = -\frac{c}{N}$. For each specification, we generate samples of size 100 and compute 10,000 replications for both Models (2.4) and (2.7).[2] The numerical results for size appear in Table 1 and we provide graphs of the Power Functions in Figures 1 through 5.

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 $g(x)$.

[2]The programs were written in Ox 2.0, see Doornik (1998).

For $c \neq 0$, the departure from the unit root becomes apparent in the increased rejection frequencies. For $g_1(x)$, the power of the CADF test is very 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 of CADF tests when there is no covariate effect. 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. The advantage of the PLMUR test becomes apparent when g_3 is considered. As expected, the covariate is successfully used to reduce the variance of the estimator of δ. As expected, the CADF test has nearly identical power to the Dickey-Fuller test since the coefficients on the covariates tend to zero. For the quadratic function, the difference is more pronounced with the PLMUR test. Finally, using the cubic function, power is again much higher than the competing tests. The CADF test has more power than the Dickey-Fuller since using x linearly does help explain some of the variance of v_i 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 7/11. As shown in Figure 5, the CADF picks up this effect and power is higher than the DF test. 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.

6. Conclusion

We have studied a class of nonstationary partially linear models using local to unity asymptotics. In particular, we have proposed a unit root test based on estimating a partially linear model where a covariate enters the model nonparametrically. Our Monte Carlo experiment suggests that estimating a partially linear model is relatively benign even when the covariates do not have a nonlinear effect. In addition, the results indicate that the proposed test effectively exploits the nonlinear effect to increase power.

There are several extensions of our partial linear unit root model and semiparametric estimator that may be of econometric interest. First, an extension to allow non iid settings for x_i is possible. For example, Fan and Li (1999) and Li (1999) show that one can obtain the usual $\sqrt{N}$ convergence rate in stationary partially linear models when the data is absolutely regular (beta mixing). Incorporating dependence in the unit root model is possible by using the results developed in the two aforementioned papers in combination with the proofs used for our results. As a practical matter, the issue of bandwidth selection needs to be treated carefully, perhaps with the development of a 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.

Appendix A

We define $E_1(\cdot) = E(\cdot|x_1)$ and $E_{X_N}(\cdot) = E(\cdot|X_N)$ where $X_N = (x_1, x_2, \ldots, x_N)$. 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:

$$A = E\left(\frac{1}{N}(\bar{y} - \hat{y}_{i-1})^2 \hat{f}_i^2\right) = E\left(\frac{1}{N}(\bar{d} + \bar{h} - \hat{d}_{i-1} - \hat{h}_{i-1})^2 \hat{f}_i^2\right)$$

where $d_i = \sum_{l=0}^{i} \epsilon_l$ and $h_i = \sum_{l=0}^{i} g(x_l)$. First,

$$
\begin{aligned}
E\left(\frac{1}{N}(\bar{d} - \hat{d}_{i-1})^2 \hat{f}_i^2\right) &= N^{-3}a^{-2q}E\left(\sum_{p\neq i}^{N} K_{ip}(\bar{d} - d_{p-1})\right)^2 \\
&= N^{-3}a^{-2q}E\sum_{p\neq i}^{N} K_{ip}^2(\bar{d} - d_{p-1})^2 \\
&\quad + N^{-3}a^{-2q}E\sum\sum_{p\neq p'\neq i} K_{ip}K_{ip'}(\bar{d} - d_{p-1})(\bar{d} - d_{p'-1}) \\
&= A_1 + A_2.
\end{aligned}
$$

$A_1 = O(N^{-1}a^{-q})$ since $E(\bar{d} - d_{j-1})^2 = O(N)$ and $E(K_{ip})^2 = O(a^q)$ from Lemma 3 of Robinson (1988). For A_2, condition on $X_N = (x_0, \ldots x_N)$ so that

$$
A_2 = N^{-3}a^{-2q}E\left(K_{ip}K_{ip'}\sum\sum_{p\neq p'\neq i} E_{X_N}(\bar{d} - d_{p-1})(\bar{d} - d_{p'-1})\right)
$$

From the identity $E\left(\sum_i^N (d_i - \bar{d})\right)^2 \equiv 0$ we find that $\sum\sum_{j\neq m\neq i} E(\bar{d} - d_{j-1})(\bar{d} - d_{m-1}) = O(N^2)$, so $A_2 = O(N^{-1})$ since $E(K_{ip}K_{ip'}) = O(a^{2q})$.

Next,

$$
\begin{aligned}
E\left(\frac{1}{N}(\bar{h} - \hat{h}_{i-1})^2 \hat{f}_i^2\right) &= N^{-3}a^{-2q}E\sum_{p\neq i}^{N} K_{ip}^2(\bar{h} - h_{p-1})^2 \\
&\quad + N^{-3}a^{-2q}E\sum\sum_{p\neq p'\neq i} K_{ip}K_{ip'}(\bar{h} - h_{p-1})(\bar{h} - h_{p'-1}) \\
&= A_3 + A_4.
\end{aligned}
$$

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

$$
\begin{aligned}
E(K_{ip}K_{ip'}\bar{h}^2) = &\left(\frac{1}{N^2}\sum_{r=0}^{N} r^2\right) E(K_{ip}K_{ip'}g(x_s)^2) + \frac{(N-i)^2}{N^2}E(K_{ip}K_{ip'}g(x_i)^2) \\
&+ \frac{(N-p)^2}{N^2}E(K_{ip}K_{ip'}g(x_p)^2) + \frac{(N-p')^2}{N^2}E(K_{ip}K_{ip'}g(x_{p'})^2).
\end{aligned}
$$

If $p' < p$ and $i < p$,

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

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

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

Therefore,

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

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 $E\sum\sum_{p\neq p'\neq i}(\bar{h}-h_{p-1})(\bar{h}-h_{p'-1})$ which is $O(N^2)$. $E(|K_{ip}K_{ip'}|g(x_s)^2) = O(a^{2q})$ and $E(|K_{ip}K_{ip'}|g(x_p)^2) = O(a^q)$. Thus, we have $|A_4| = O(N^{-1}) + O(N^{-1}a^{-q})$. Using Cauchy-Schwartz, we can show that

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

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

APPENDIX B

We prove Theorem 3.2 for model (2.4) in this appendix. Note that $N(\hat{\delta}-\delta) = B_1^{-1}(B_2 + B_3))$ where

$$B_1 = \frac{1}{N^2}\sum_{i=1}^{N}(y_{i-1} - \hat{y}_{i-1})^2 \hat{f}_i^2$$

$$B_2 = \frac{1}{N}\sum_{i-1}^{N}(y_{i-1} - \hat{y}_{i-1})(\hat{\epsilon}_i + g(x_i) - \hat{g}(x_i))\hat{f}_i^2$$

$$B_3 = \frac{1}{N}\sum_{i-1}^{N}(y_{i-1} - \hat{y}_{i-1})\epsilon_i\hat{f}_i^2,$$

with

$$\hat{\epsilon}_i = \frac{1}{Na^q}\frac{\sum_{p=1,p\neq i}K_{ip}\epsilon_p}{\hat{f}_i} \quad \text{and} \quad \hat{g}(x_i) = \frac{1}{Na^q}\frac{\sum_{p=1,p\neq i}K_{ip}g(x_p)}{\hat{f}_i}.$$

Notice that under the assumptions of the theorem, $\sigma_{v\epsilon f} = \sigma_\epsilon^2 E(f^2)$ and $\sigma_{\epsilon f} = \sigma_\epsilon\sqrt{E(f^4)}$ where $\sigma_\epsilon^2 = E(\epsilon^2)$. This means that

$$\rho = \frac{\sigma_\epsilon}{\sigma_v}\frac{E(f^2)}{\sqrt{E(f^4)}}.$$

Then the theorem holds if we can show that

$$B_1 \Rightarrow E(f^2)\sigma_v^2\int(\underline{W}_1^c))^2,$$

$$B_2 \xrightarrow{p} 0,$$

$$B_3 \Rightarrow \sigma_v\sigma_{\epsilon f}\left(\rho\int\underline{W}_1^c dW_1 + \sqrt{1-\rho^2}\int\underline{W}_1^c dW_2\right).$$

We prove these results in a series of seven propositions.

Proposition 1.

$$\frac{1}{N}\sum_{i=1}^{N}(y_{i-1} - \bar{y})\epsilon_i\hat{f}_i^2 \Rightarrow \sigma_v\sigma_{\epsilon f}\left(\rho\int\underline{W}_1^c dW_1 + \sqrt{1-\rho^2}\int\underline{W}_1^c dW_2\right) \tag{B.1}$$

Proof of Proposition 1:

$$\frac{1}{N}\sum_{i=1}^{N}(y_{i-1}-\bar{y})\epsilon_i\hat{f}_i^2 = \frac{1}{N}\sum_{i=1}^{N}(y_{i-1}-\bar{y})\epsilon_i(\hat{f}_i - f_i + f_i)^2$$
$$= \frac{1}{N}\sum_{i=1}^{N}(y_{i-1}-\bar{y})\epsilon_i(f_i^2 + (\hat{f}_i - f_i)f_i + (\hat{f}_i - f_i)^2)$$
$$= C_1 + C_2 + C_3.$$

Notice that C_1 converges to the expression given in (B.1) by Lemma 3 of Hansen (1995).

We show that the remaining parts converge to zero in mean square. First,

$$EC_2^2 = \frac{1}{N^2}E\sum_{i=1}^{N}(y_{i-1}-\bar{y})^2\epsilon_i^2(\hat{f}_i - f_i)^2 f_i^2$$
$$+ \frac{1}{N^2}E\sum\sum_{i\neq j}(d_{i-1}-\bar{d})(d_{j-1}-\bar{d})\epsilon_i\epsilon_j(\hat{f}_i - f_i)(\hat{f}_j - f_j)f_i f_j$$
$$+ \frac{1}{N^2}E\sum\sum_{i\neq j}(h_{i-1}-\bar{h})(h_{j-1}-\bar{h})\epsilon_i\epsilon_j(\hat{f}_i - f_i)(\hat{f}_j - f_j)f_i f_j$$
$$+ \frac{1}{N^2}E\sum\sum_{i\neq j}(h_{i-1}-\bar{h})(d_{j-1}-\bar{d})\epsilon_i\epsilon_j(\hat{f}_i - f_i)(\hat{f}_j - f_j)f_i f_j$$
$$+ \frac{1}{N^2}E\sum\sum_{i\neq j}(d_{i-1}-\bar{d})(h_{j-1}-\bar{h})\epsilon_i\epsilon_j(\hat{f}_i - f_i)(\hat{f}_j - f_j)f_i f_j$$
$$= C_{21} + C_{22} + C_{23} + C_{24} + C_{25}.$$

$C_{21} = O(a^{2\lambda} + N^{-1}a^{-q})$ since $E(\hat{f}_i - f_i)^2 = O(a^{2\lambda} + N^{-1}a^{-q})$ from the proof of Proposition 4 in Robinson (1988). $C_{23} = 0$ since $E(\epsilon_i\epsilon_j) = 0$. Conditioning on X_N gives

$$C_{22} = E\left((\hat{f}_i - f_i)(\hat{f}_j - f_j)f_i f_j \frac{1}{N^2}E_{X_N}\sum\sum_{i\neq j}(d_{i-1}-\bar{d})(d_{j-1}-\bar{d})\epsilon_i\epsilon_j\right).$$

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

$$C_{24} = E\left(\frac{1}{N^2}\sum\sum_{i\neq j}(h_{i-1}-\bar{h})(\hat{f}_i - f_i)(\hat{f}_j - f_j)f_i f_j E_{X_N}(d_{j-1}-\bar{d})\epsilon_i\epsilon_j\right)..$$

The inner expectation is zero so that $C_{24} = 0$ and $C_{25} = 0$. The proof of EC_3^2 is similar except we need to find the order of $E(\hat{f}_1 - f_1)^4$. Define $\tilde{f}_1 = E_1\hat{f}_1$. Then

$$E(\hat{f}_1 - f_1) \leq 8E(\hat{f}_1 - \tilde{f}_1)^4 + 8E(\tilde{f}_1 - f_1)^4.$$

$$\begin{aligned} E(\tilde{f}_1 - f_1)^4 &= E(a^{-q}E_1(K_{12}) - f_1)^4 \\ &= O(a^{4\lambda}) \end{aligned}$$

from Lemma 4 in Robinson (1988). Now

$$\begin{aligned} E(\hat{f}_1 - \tilde{f}_1)^4 &= E\left(N^{-1}a^{-q}\sum_{j\neq 1}^{N}(K_{1j} - E_1(K_{1j}))\right)^4 \\ &\leq N^{-4}a^{-4q}\sum_{j\neq 1}E(K_{1j})^4 + N^{-4}a^{-4q}\sum\sum_{j\neq m\neq 1}EK_{ij}^2K_{1m}^2 \\ &= O(N^{-2}a^{-2q}) \end{aligned}$$

so that $E(\hat{f}_1 - f_1)^4 = O(a^{4\lambda} + N^{-2}a^{-2q})$ and $EC_3^2 = O(a^{4\lambda} + N^{-2}a^{-2q})$. □

Proposition 2.

(B.2) $$\frac{1}{N}\sum_{i=1}^{N}(\bar{y} - \hat{y}_{i-1})\epsilon_i\hat{f}_i^2 \xrightarrow{p} 0$$

Proof of Proposition 2:

$$
\begin{aligned}
E\left(\frac{1}{N}\sum_{i=1}^{N}(\bar{y}-\hat{y}_{i-1})\epsilon_i\hat{f}_i^2\right)^2 &= \frac{1}{N^2}E\sum_{i=1}^{N}(\bar{y}-\hat{y}_{i-1})^2\epsilon_i^2\hat{f}_i^4 \\
&+ \frac{1}{N^2}E\sum\sum_{i\neq j}(\bar{h}-\hat{h}_{i-1})(\bar{h}-\hat{h}_{j-1})\hat{f}_i^2\hat{f}_j^2\epsilon_i\epsilon_j \\
&+ \frac{1}{N^2}E\sum\sum_{i\neq j}(\bar{d}-\hat{d}_{i-1})(\bar{d}-\hat{d}_{j-1})\hat{f}_i^2\hat{f}_j^2\epsilon_i\epsilon_j \\
&+ \frac{1}{N^2}E\sum\sum_{i\neq j}(\bar{h}-\hat{h}_{i-1})(\bar{d}-\hat{d}_{j-1})\hat{f}_i^2\hat{f}_j^2\epsilon_i\epsilon_j \\
&+ \frac{1}{N^2}E\sum\sum_{i\neq j}(\bar{d}-\hat{d}_{i-1})(\bar{h}-\hat{h}_{j-1})\hat{f}_i^2\hat{f}_j^2\epsilon_i\epsilon_j \\
&= D_1+D_2+D_3+D_4+D_5.
\end{aligned}
$$

$D_1 = O(N^{-1}a^{-q})$ using the proof of Lemma 3.1. Conditioning on X_N, $D_2 = 0$ since $E(\epsilon_i\epsilon_j) = 0$. Conditioning on X_N, $D_4 = 0$ and $D_5 = 0$ since $E_{X_N}(\bar{d}-\hat{d}_{i-1})\epsilon_i\epsilon_j = 0$. For D_3, we treat the individual summands to find

$$
\begin{aligned}
&E(\bar{d}-\hat{d}_{i-1})(\bar{d}-\hat{d}_{j-1})\hat{f}_i\hat{f}_j\epsilon_i\epsilon_j \\
&\qquad = (Na^q)^{-2}E\sum_{p\neq i\neq j}^{N}K_{ip}K_{jp}(\bar{d}-d_{p-1})^2\epsilon_i\epsilon_j \\
&\qquad\quad + (Ta^q)^{-2}E\left(K_{ip}K_{jp'}E\left(\sum\sum_{p\neq p'\neq i\neq j}(\bar{d}-d_{p-1})(\bar{d}-d_{p'-1})\epsilon_i\epsilon_j|X_N\right)\right) \\
&\qquad = a^{-2q}E(K_{ip}K_{jp})D_{31} + a^{-2q}E(K_{ip}K_{jp'})D_{32}.
\end{aligned}
$$

Matching indices, we find that

$$
\begin{aligned}
D_{32} = {}& 2\frac{(N-i)(N-j)}{N^2} + 2\frac{\max(i,j)^2}{N^2} - 2\frac{\max(i,j)}{N^2} + 2\frac{\min(i,j)}{N^2} \\
&- 2\frac{(N-\max(i,j))(N-\min(i,j))}{N^2} - 2\frac{(N-\max(i,j))(N-\max(i,j))}{N^2} \\
&+ 1 - \frac{ij}{N^2}.
\end{aligned}
$$

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

Proposition 3.

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

Proof of Proposition 3:

$$\text{(B.3)} \qquad \frac{1}{N^4}E\left(\sum_{i=1}^{N}(y_{i-1}-\bar{y})(\bar{y}-\hat{y}_{i-1})\hat{f}_i^2\right)^2 \leq \frac{1}{N^2}\sum_{i=1}^{N}\sqrt{E(y_{i-1}-\bar{y})^4 E\frac{1}{N}(\bar{y}-\hat{y}_i)^4\hat{f}_i^4}$$

by Loève's c_r inequality and Cauchy-Schwartz. Using a similar proof to Lemma 3.1, we have $E\frac{1}{N}(\bar{y}-\hat{y}_{i-1})^4\hat{f}_i^4 = O(N^{-1}a^{-2q})$ and $E(y_{i-1}-\bar{y})^4 = O(N^2)$, then (B.3) is $O(N^{-1/2}a^{-q})$.
□

Proposition 4.

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

Proof of Proposition 4: The proof follows from $E\frac{1}{N}(\bar{y}-\hat{y}_{i-1})^4\hat{f}_i^4 = O(N^{-1}a^{-2q})$.

Proposition 5.

$$\text{(B.4)} \qquad \frac{1}{N^2}\sum_{i=1}^{N}(y_{i-1}-\hat{y}_{i-1})^2\hat{f}_i^2 \Rightarrow E(f^2)\sigma_v^2\int(W_1-\bar{W}_1)^2$$

Proof of Proposition 5:

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

The second and third terms on the right hand side converge to zero by Propositions 3 and 4. Using the method in the proof of Proposition 1, the first term converges to the right hand side of (B.4). □

Proposition 6.

$$\frac{1}{N}\sum_{i=1}^{N}(y_{i-1}-\hat{y}_{i-1})\hat{\epsilon}_i\hat{f}_i^2 \xrightarrow{p} 0.$$

Proof of Proposition 6:

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

Note that

$$E\left(\hat{\epsilon}_i^2\hat{f}_i^2\right) = E\left((Na^q)^{-2}\sum_{p\neq i}^{N}K_{ip}^2\epsilon_p^2 + (Na^q)^{-2}\sum\sum_{p\neq p'\neq i}K_{ip}K_{ip'}\epsilon_p\epsilon_{p'}\right)$$
$$= O(N^{-1}a^{-q}), \tag{B.5}$$

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

$$E(F_1^2) = E\left(\frac{1}{N^2}\sum_{i=1}^{N}(y_{i-1}-\bar{y})^2\hat{\epsilon}_i^2\hat{f}_i^4\right)$$
$$+ E\left(\frac{1}{N^2}\sum\sum_{i\neq j}(y_{i-1}-\bar{y})(y_{j-1}-\bar{y})\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2\right)$$
$$= F_{11} + F_{12}.$$

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

$$E(y_{i-1}-\bar{y})(y_{j-1}-\bar{y})\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2 = E(\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2v_s^2)F_{13} + E(\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2v_{s'}v_j)F_{14}$$
$$+ E(\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2v_{t'}v_t)F_{15} + o(1),$$

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

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

$$F_{14} = (N-i)\left(1 - \frac{(i-1)}{N} - \frac{(j-1)}{N}\right) + (N-j)\left(1 - \frac{(j-1)}{N} - \frac{(i-1)}{N}\right) - \frac{N}{2} + \min(i-1, j-1)$$

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

Due to cancellations, we have

$$\frac{1}{N^2}\sum\sum_{i \neq j} F_{13} = O(1)$$

$$\frac{1}{N^2}\sum\sum_{i \neq j} F_{14} = O(1)$$

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

$E(\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2 v_s^2) = O(N^{-1}a^{-q})$ and $E(\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2 v_{s'}v_j) = O(N^{-1}a^{-q})$ by Cauchy-Schwartz so it remains to show that $E(\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2 v_{t'}v_t) = o(N^{-1})$.

$$E(\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2 v_{t'}v_t) = E\left((Na^q)^{-4}\sum_{p\neq i}\sum_{p'\neq i}\sum_{u\neq j}\sum_{u'\neq j} K_{ip}K_{ju}K_{ip'}K_{ju'}\epsilon_{p'}\epsilon_{u'}v_{t'}v_t\right)$$

which is nonzero when $t' = p'$ and $t = u'$, or $t' = p$ and $t = u$, or $t' = p'$ and $t = u$ so that

$$E(\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2 v_{t'}v_t) = 2E\left((Na^q)^{-4}\sum_{p\neq i}\sum_{u\neq j} K_{ip}K_{ju}K_{it'}K_{jt}\epsilon_{t'}v_{t'}\epsilon_t v_t\right)$$
$$+ 2E\left((Na^q)^{-4}\sum_{p'\neq i}\sum_{u'\neq j} K_{it'}K_{jt}K_{ip'}K_{ju'}\epsilon_{p'}\epsilon_{u'}v_{t'}v_t\right)$$
$$+ 2E\left((Na^q)^{-4}\sum_{p\neq i}\sum_{u'\neq j} K_{ip}K_{jt}K_{it'}K_{ju'}\epsilon_{t'}\epsilon_{u'}v_{t'}v_t\right)$$

which is $O(N^{-2}a^{-2q})$ by Cauchy-Schwartz so that $F_{12} = O(N^{-1}a^{-2q})$ and

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

□

Proposition 7.

$$\frac{1}{N}\sum_{i=1}^{N}(y_{i-1} - \hat{y}_{i-1})(g(x_i) - \hat{g}(x_i))\hat{f}_i^2 \xrightarrow{p} 0.$$

Proof of Proposition 7: Let $\xi = \min(\lambda + 1, \nu)$. The proof follows Proposition 6. However, we need to find the order of $E(g(x_i) - \hat{g}(x_i))^2 \hat{f}_i^2$ and $E((g(x_i) - \hat{g}(x_i))(g(x_j) - \hat{g}(x_j))\hat{f}_i^2 \hat{f}_j^2 v_{t'} v_t)$. First,

$$\begin{aligned} E(g(x_i) - \hat{g}(x_i))^2 \hat{f}_i^2 &= (Na^q)^{-2} E\left(\sum_{p\neq i}^{N} K_{ip}(g(x_i) - g(x_p))\right)^2 \\ &= O(N^{-1}a^{-q} + a^{2\xi}) \end{aligned} \tag{B.6}$$

from the proof of Proposition 1 in Robinson (1988). Next,

$$\begin{aligned} E((g(x_i) - \hat{g}(x_i))(g(x_j) - \hat{g}(x_j))\hat{f}_i^2 \hat{f}_j^2 v_{t'} v_t) &= 2E\left((Na^q)^{-4}\sum_{p\neq i}\sum_{u\neq j} K_{ip}K_{ju}K_{it'}K_{jt}\eta_{it'}v_{t'}\eta_{jt}v_t\right) \\ &+ 2E\left((Na^q)^{-4}\sum_{p'\neq i}\sum_{u'\neq j} K_{it'}K_{jt}K_{ip'}K_{ju'}\eta_{ip'}\eta_{ju'}v_{t'}v_t\right) \\ &+ 2E\left((Na^q)^{-4}\sum_{p\neq i}\sum_{u'\neq j} K_{ip}K_{jt}K_{it'}K_{ju'}\eta_{it'}\eta_{ju'}v_{t'}v_t\right) \end{aligned}$$

with $\eta_{jt} = g(x_j) - g(x_t)$. The above term is $O(N^{-2}a^{-2q})$. Combining these results and using the proof of Proposition 6, we find

$$\frac{1}{N}\sum_{i=1}^{N}(y_{i-1} - \hat{y}_{i-1})(g(x_i) - \hat{g}(x_i))\hat{f}_i^2 = O_p(N^{-\frac{1}{2}}a^{-q} + a^{\xi - \frac{q}{2}}).$$

APPENDIX C

We prove Theorem 3.2 for model (2.7). 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. Let $\hat{e}_w$ and $\hat{e}_{\Delta w}$ denote the vectors of nonparametric residuals from regressing w_{i-1} and Δw_i on x_i. The joint estimators for (δ, θ^*) are given by

$$\begin{pmatrix} \hat{\delta} \\ \hat{\theta}^* \end{pmatrix} = \begin{pmatrix} (\hat{e}_w \odot \hat{f})^\top (\hat{e}_w \odot \hat{f}) & (\hat{e}_w \odot \hat{f})^\top (\hat{e}_t \odot \hat{f}) \\ (\hat{e}_t \odot \hat{f})^\top (\hat{e}_w \odot \hat{f}) & (\hat{e}_t \odot \hat{f})^\top (\hat{e}_t \odot \hat{f}) \end{pmatrix}^{-1} \begin{pmatrix} (\hat{e}_y \odot \hat{f})^\top (\hat{e}_{\Delta w} \odot \hat{f}) \\ (\hat{e}_t \odot \hat{f})^\top (\hat{e}_{\Delta w} \odot \hat{f}) \end{pmatrix}$$

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.

$$\begin{pmatrix} N & 0 \\ -\theta N^{\frac{3}{2}} & N^{\frac{3}{2}} \end{pmatrix} \begin{pmatrix} \hat{\delta} \\ \hat{\theta}^* \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}^{-1} \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}$$

with

$$G_{11} = \frac{1}{N^2} \sum_{i=1}^{N} (\hat{e}_{wi} - \theta \hat{e}_{ti})^2 \hat{f}_i^2$$

$$G_{12} = \frac{1}{N^{\frac{5}{2}}} \sum_{i=1}^{N} (\hat{e}_{wi} - \theta \hat{e}_{ti}) \hat{e}_{ti} \hat{f}_i^2$$

$$G_{22} = \frac{1}{N^3} \sum_{i=1}^{N} \hat{e}_{ti}^2 \hat{f}_i^2$$

$$H_1 = \frac{1}{N} \sum_{i=1}^{N} (\hat{e}_{wi} - \theta \hat{e}_{ti}) \hat{e}_{\Delta wi} \hat{f}_i^2$$

$$H_2 = \frac{1}{N^{\frac{3}{2}}} \sum_{i=1}^{N} \hat{e}_{ti} \hat{e}_{\Delta wi} \hat{f}_i^2$$

The proof of Theorem 3.2 for model (2.7) follows using partitioned inverses and showing that

$$\begin{aligned}
G_{11} &\Rightarrow E(f^2)\sigma_v^2 \int \left(W_1(s) - \bar{W}_1\right)^2 \\
G_{12} &\Rightarrow E(f^2)\sigma_v \int \left(W_1(s) - \bar{W}_1\right) sds \\
G_{22} &\Rightarrow E(f^2)\frac{1}{12} \\
H_1 &\Rightarrow \sigma_v\sigma_{\epsilon f}\left(\rho\int (W_1(s) - \bar{W}_1)dW_1 + \sqrt{1-\rho^2}\int (W_1(s) - \bar{W}_1)dW_2\right) \\
H_2 &\Rightarrow \sigma_v\sigma_{\epsilon f}\left(\rho\int (s - \frac{1}{2})dW_1 + \sqrt{1-\rho^2}\int (s - \frac{1}{2})dW_2\right).
\end{aligned}$$

Notice that using the rotation above, we generate

$$\begin{aligned}
\hat{e}_{wi} - \theta\hat{e}_{ti} &= w_{i-1} - \hat{w}_{i-1} - \theta((i-1) - (\widehat{i-1})) \\
&= \theta(i-1) + y_{i-1} - \theta(\widehat{i-1}) - \hat{y}_{i-1} - \theta(i-1) + \theta(\widehat{i-1}) \qquad \text{(C.1)} \\
&= y_{i-1} - \hat{y}_{i-1}.
\end{aligned}$$

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

Proposition 8.

$$\frac{1}{N^2}\sum_{i=1}^{N}(\hat{e}_{wi} - \theta\hat{e}_{ti})^2\hat{f}_i^2 \Rightarrow E(f^2)\sigma_v^2\int \left(W_1 - \bar{W}_1\right)^2 \qquad \text{(C.2)}$$

Proof of Proposition 8: The result follows by Proposition 5.

Proposition 9.

$$E\left(\frac{\hat{i}}{N} - \frac{1}{2}\right)^2\hat{f}_i^2 = O_p(N^{-1}a^{-q}) \qquad \text{(C.3)}$$

Proof of Proposition 9

$$E\left(\frac{\hat{i}}{N}-\frac{1}{2}\right)^2\hat{f}_i^2=(Na^q)^{-2}E\sum_{p=1,p\neq i}^{N}K_{ip}^2\left(\frac{p}{N}-\frac{1}{2}\right)^2$$
$$+(Na^q)^{-2}E\sum\sum_{p\neq q\neq i}K_{ip}K_{iq}\left(\frac{p}{N}-\frac{1}{2}\right)\left(\frac{q}{N}-\frac{1}{2}\right)$$
$$=J_1+J_2.$$

$J_1=O(N^{-1}a^{-q})$ and we know that $E|K_{ip}K_{iq}|=O(a^{2q})$ and $\sum\sum_{p\neq q\neq i}\left(\frac{p}{N}-\frac{1}{2}\right)\left(\frac{q}{N}-\frac{1}{2}\right)=O(N)$ so that $J_2=O(N^{-1})$. □

Proposition 10.

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

Proof of Proposition 10

$$N^{-\frac{5}{2}}\sum_{i=1}^{N}(\hat{e}_{wi}-\theta\hat{e}_{ti})\hat{e}_{ti}\hat{f}_i^2=N^{-\frac{5}{2}}\sum_{i=1}^{N}(y_{i-1}-\hat{y}_{i-1})i\hat{f}_i^2-N^{-\frac{5}{2}}\sum_{i=1}^{N}(y_{i-1}-\hat{y}_{i-1})\hat{i}\hat{f}_i^2$$
$$=K_1+K_2$$

and

$$K_1=N^{-\frac{5}{2}}\sum_{i=1}^{N}(y_{i-1}-\bar{y})i\hat{f}_i^2+N^{-\frac{5}{2}}\sum_{i=1}^{N}(\bar{y}-\hat{y}_{i-1})i\hat{f}_i^2$$
$$=K_{11}+K_{12}.$$

Using the well known result that

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

and the proof of Proposition 1, we have

$$K_{11}\Rightarrow E(f^2)\sigma_v\int(W_1-\bar{W}_1)sds.$$

Next,

$$E(K_{12}^2) = N^{-5}\sum_{i=1}^{N}(\bar{y}-\hat{y}_{i-1})^2 i^2 \hat{f}_i^4 + N^{-5}\sum\sum_{i\neq j}(\bar{y}-\hat{y}_{i-1})(\bar{y}-\hat{y}_{j-1}) ij \hat{f}_i^2 \hat{f}_j^2$$
$$= O(N^{-1}a^{-q})$$

Now consider K_2. Using the proof of Proposition 1, it is easy to see that

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

so it is enough to show that

$$K_3 = N^{-\frac{3}{2}}\sum_{i=1}^{N}(y_{i-1}-\hat{y}_{i-1})(\frac{\hat{i}}{N}-\frac{1}{2})\hat{f}_i^2 \xrightarrow{p} 0.$$

We write

$$K_3 = N^{-\frac{3}{2}}\sum_{i=1}^{N}(y_{i-1}-\bar{y})(\frac{\hat{i}}{N}-\frac{1}{2})\hat{f}_i^2 + N^{-\frac{3}{2}}\sum_{i=1}^{N}(\bar{y}-\hat{y}_{i-1})(\frac{\hat{i}}{N}-\frac{1}{2})I_i$$
$$= K_{31} + K_{32}.$$

Using Lemma 3.1 and Proposition 9, we have

$$E(K_{31}^2) = N^{-3}\sum_{i=1}^{N}E(y_{i-1}-\bar{y})^2(\frac{\hat{i}}{N}-\frac{1}{2})^2\hat{f}_i^4$$
$$+ N^{-3}\sum\sum_{i\neq j}E(y_{i-1}-\bar{y})(y_{j-1}-\bar{y})(\frac{\hat{i}}{N}-\frac{1}{2})(\frac{\hat{j}}{N}-\frac{1}{2})\hat{f}_i^2\hat{f}_j^2$$
$$= O(N^{-1}a^{-q}),$$

$$E(L_{32}^2) = N^{-3}\sum_{i=1}^{N}E(\bar{y}-\hat{y}_{i-1})^2(\frac{\hat{i}}{N}-\frac{1}{2})^2\hat{f}_i^4$$
$$+ N^{-3}\sum\sum_{i\neq j}E(\bar{y}-\hat{y}_{i-1})(\bar{y}-\hat{y}_{j-1})(\frac{\hat{i}}{N}-\frac{1}{2})(\frac{\hat{j}}{N}-\frac{1}{2})\hat{f}_i^2\hat{f}_j^2$$
$$= O(N^{-2}a^{-2q}).$$

□

Proposition 11.

$$N^{-3}\sum_{i=1}^{N}\hat{e}_{ti}^{2}\hat{f}_i^2 \xrightarrow{p} E(f^2)\frac{1}{12}$$

Proof of Proposition 11

$$N^{-3}\sum_{i=1}^{N}\hat{e}_{ti}^{2}\hat{f}_i^2 = N^{-1}\sum_{i=1}^{N}\left(\left(\frac{i}{N}-\frac{1}{2}\right)^2+\left(\left(\frac{i}{N}-\frac{1}{2}\right)\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)\right)+\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)^2\right)\hat{f}_i^2$$
$$= L_1+L_2+L_3$$

Following the proof of Proposition 1, we have

$$L_1 \xrightarrow{p} E(f^2)\frac{1}{12}.$$

Next,

$$E(L_2^2) = N^{-4}E\sum_{i=1}^{N}\left(\frac{i}{N}-\frac{1}{2}\right)^2\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)^2\hat{f}_i^2$$
$$+N^{-4}E\sum\sum_{i\neq j}\left(\frac{i}{N}-\frac{1}{2}\right)\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)\left(\frac{j}{N}-\frac{1}{2}\right)\left(\frac{1}{2}-\frac{\hat{j}}{N}\right)\hat{f}_i^2\hat{f}_j^2$$

which is $O(N^{-1}a^{-q})$ by Proposition 9 and Cauchy-Schwartz. Then

$$E(L_3^2) = N^{-4}E\sum_{i=1}^{N}\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)^4\hat{f}_i^4 + N^{-4}E\sum\sum_{i\neq l}\left(\frac{1}{2}-\frac{\hat{i}}{N}\right)^2\left(\frac{1}{2}-\frac{\hat{l}}{N}\right)^2\hat{f}_i^2\hat{f}_j^2$$
$$= L_{31}+L_{32}.$$

Following the proof of Proposition 9, we can show that $L_{31} = O(N^{-2}a^{-2q})$. $L_{32} = O(N^{-4}a^{-2q})$ by Cauchy-Schwartz. □

Proposition 12.

$$N^{-\frac{3}{2}}\sum_{i=1}^{N}(i-\hat{i})\epsilon_i\hat{f}_i^2 \xrightarrow{d} \int(s-\frac{1}{2})\left(\sigma_{\epsilon f}(\rho dW_1(s)+\sqrt{1-\rho^2}dW_2(s))\right)$$

Proof of Proposition 12: The proof is completed by showing that

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

Conditioning on X_N and using the independence of ϵ_i gives

$$E\left(N^{-\frac{1}{2}}\sum_{i=1}^{N}\left(\frac{\hat{i}}{N}-\frac{1}{2}\right)\epsilon_i\hat{f}_i^2\right)^2 = N^{-1}E\sum_{i=1}^{N}\left(\frac{\hat{i}}{N}-\frac{1}{2}\right)^2\epsilon_i^2\hat{f}_i^4$$

which is $O(N^{-1}a^{-q})$. □

Proposition 13.

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

Proof of Proposition 13: Again, we break this term into two parts:

$$\begin{aligned}N^{-\frac{3}{2}}\sum_{i=1}^{N}(i-\hat{i})\hat{\epsilon}_i\hat{f}_i^2 &= N^{-\frac{1}{2}}\sum_{i=1}^{N}(\frac{i}{N}-\frac{1}{2})\hat{\epsilon}_i\hat{f}_i^2 + N^{-\frac{1}{2}}\sum_{i=1}^{N}(\frac{1}{2}-\frac{\hat{i}}{N})\hat{\epsilon}_i\hat{f}_i^2 \\ &= M_1 + M_2.\end{aligned}$$

For M_1,

$$\begin{aligned}E(M_1^2) &= N^{-1}E\sum_{i=1}^{N}(\frac{i}{N}-\frac{1}{2})^2\hat{\epsilon}_i^2\hat{f}_i^2 \\ &\quad + N^{-1}E\sum\sum_{i\neq j}(\frac{i}{N}-\frac{1}{2})(\frac{i}{N}-\frac{1}{2})\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2 \\ &= M_{11} + M_{12}.\end{aligned}$$

Using (B.5), M_{11} is $O(N^{-1}a^{-q})$. Since $E\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2$ is identical for $i \neq j$ and $\sum\sum_{i\neq j}(\frac{i}{N}-\frac{1}{2})(\frac{j}{N}-\frac{1}{2}) = O(N)$, M_{12} is also $O(N^{-1}a^{-q})$. For M_2, we have

$$\begin{aligned}E(M_2^2) &= N^{-1}E\sum_{i=1}^{N}(\frac{1}{2}-\frac{\hat{i}}{N})^2\hat{\epsilon}_i^2\hat{f}_i^2 \\ &\quad + N^{-1}E\sum\sum_{i\neq j}(\frac{1}{2}-\frac{\hat{i}}{N})(\frac{1}{2}-\frac{\hat{j}}{N})\hat{\epsilon}_i\hat{\epsilon}_j\hat{f}_i^2\hat{f}_j^2\end{aligned}$$

Using Cauchy-Schwartz, (B.5), and Proposition 9, $E(M_2^2) = O(N^{-1}a^{-2q})$. □

Proposition 14.

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

Proof of Proposition 14: The proof follows the proof of Proposition 13 with the exception that (B.6) is applied to $(g(x_i)-\hat{g}(x_i))\hat{f}_i$ as (B.5) was applied to $\hat{\epsilon}_i\hat{f}_i$. The order is $O_p(N^{-1/2}a^{-q/2}+a^{\zeta})$ where $\zeta=\min(\lambda+1,\nu)$ as in Proposition 7.

APPENDIX D

We provide the proof for the generalization to higher order autoregressive models in this appendix. Again, for simplicity of exposition and without loss of generality, we demonstrate the proofs for the exact unit root (i.e. $c=0$) case, local to unit root case being similar. In this case:

$$\begin{aligned} y_i &\approx \sum_{s=1}^{i} B(L)v_s = \sum_{s=1}^{i}\left[B(L)g(x_s)+B(L)\epsilon_s\right], \\ \Delta y_{i-s} &= B(L)v_{i-s} = B(L)g(x_{i-s})+B(L)\epsilon_{i-s}, B(L)=\sum_{l=0}^{\infty} b_l L^l. \end{aligned}$$

Proposition 15. *For* $s=1,....,k$,

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

Proof of Proposition 15:

Notice that

$$\begin{aligned}
&\sum_{i=1}^{N}(y_{i-1}-\hat{y}_{i-1})(\Delta y_{i-s}-\widehat{\Delta y}_{i-s})\hat{f}_i^2 \\
=\;& \sum_{i=1}^{N}(Na^q)^{-2}\left(\sum_{p\neq i}^{N}K_{ip}(y_{i-1}-y_{p-1})\right)\left(\sum_{r\neq i}^{N}K_{ir}(\Delta y_{i-s}-\Delta y_{r-s})\right) \\
=\;& \sum_{i=1}^{N}(Na^q)^{-2}\left(\sum_{p\neq i}^{N}K_{ip}(\sum_{t=1}^{i-1}B(L)v_t-\sum_{t=1}^{p-1}B(L)v_t)\right)\left(\sum_{r\neq i}^{N}K_{ir}(B(L)v_{i-s}-B(L)v_{r-s})\right) \\
=\;& \sum_{i=1}^{N}(Na^q)^{-2}\left(\sum_{p\neq i}^{N}K_{ip}(\sum_{t=1}^{i-1}B(L)\left[g(x_t)+\epsilon_t\right]-\sum_{t=1}^{p-1}B(L)\left[g(x_t)+\epsilon_t\right])\right)\times \\
&\left(\sum_{r\neq i}^{N}K_{ir}(B(L)\left[g(x_{i-s})+\epsilon_{i-s}\right]-B(L)\left[g(x_{r-s})+\epsilon_{r-s}\right])\right)
\end{aligned}$$

$$\begin{aligned}
=\;& \sum_{i=1}^{N}(Na^q)^{-2}\left(\sum_{p\neq i}^{N}K_{ip}(\sum_{t=1}^{i-1}\sum_{l=0}^{\infty}b_l\left[g(x_{t-l})+\epsilon_{t-l}\right]-\sum_{t=1}^{p-1}\sum_{l=0}^{\infty}b_l\left[g(x_{t-l})+\epsilon_{t-l}\right])\right)\times \\
&\left(\sum_{r\neq i}^{N}K_{ir}(\sum_{\nu=0}^{\infty}b_\nu\left[g(x_{i-s-\nu})+\epsilon_{i-s-\nu}\right]-\sum_{\nu=0}^{\infty}b_\nu\left[g(x_{r-s-\nu})+\epsilon_{r-s-\nu}\right])\right)
\end{aligned}$$

For the term

$$\begin{aligned}
\text{(D.1)}\qquad &\sum_{i=1}^{N}(Na^q)^{-2}\left(\sum_{p\neq i}^{N}K_{ip}(\sum_{t=1}^{i-1}\sum_{l=0}^{\infty}b_l g(x_{t-l}))\right)\left(\sum_{r\neq i}^{N}K_{ir}(\sum_{\nu=0}^{\infty}b_\nu g(x_{i-s-\nu}))\right) \\
=\;& \sum_{i=1}^{N}(Na^q)^{-2}\sum_{p\neq i}^{N}\sum_{r\neq i}^{N}\sum_{l=0}^{\infty}\sum_{\nu=0}^{\infty}\sum_{t=1}^{i-1}K_{ip}K_{ir}b_l b_\nu g(x_{t-l})g(x_{i-s-\nu}),
\end{aligned}$$

the second moment of (D.1) is

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

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

$$\left(\frac{1}{Na^q}\right)^4 \sum_{i=1}^{N} \sum_{p\neq i}^{N} \sum_{r\neq i}^{N} \sum_{l=0}^{\infty} \sum_{\nu=0}^{\infty} \sum_{\alpha=1}^{N} \sum_{j\neq\alpha}^{N} \sum_{k\neq\alpha}^{N} \sum_{\beta=0}^{\infty} \sum_{\mu=0}^{\infty} b_l b_\nu b_\beta b_\mu K_{ip} K_{ir} K_{\alpha j} K_{\alpha k} g(x_{i-s-\nu})^2 g(x_{\alpha-s-\mu})^2$$

is of order $O(N^2)$ since

$$\sum_{l=0}^{\infty} \sum_{\nu=0}^{\infty} \sum_{\beta=0}^{\infty} \sum_{\mu=0}^{\infty} b_l b_\nu b_\beta b_\mu = O(1)$$

and $E[K_{ip}K_{ir}K_{\alpha j}K_{\alpha k} g(x_{i-s-\nu})^2 g(x_{\alpha-s-\mu})^2] = O(a^{4q})$. Thus, (D.1) is of order $O_p(N)$. For

$$\text{(D.2)} \qquad \sum_{i=1}^{N} (Na^q)^{-2} \left(\sum_{p\neq i}^{N} K_{ip} (\sum_{t=1}^{i-1} \sum_{l=0}^{\infty} b_l g(x_{t-l})) \right) \left(\sum_{r\neq i}^{N} K_{ir} (\sum_{\nu=0}^{\infty} b_\nu \epsilon_{i-s-\nu}) \right),$$

the second moment

$$\left(\frac{1}{Na^q}\right)^4 \sum_{i=1}^{N} \sum_{p\neq i}^{N} \sum_{r\neq i}^{N} \sum_{l=0}^{\infty} \sum_{\nu=0}^{\infty} \sum_{t=1}^{i-1} \sum_{\alpha=1}^{N} \sum_{j\neq\alpha}^{N} \sum_{k\neq\alpha}^{N} \sum_{\beta=0}^{\infty} \sum_{\mu=0}^{\infty} \sum_{\tau=1}^{\alpha-1}$$
$$K_{ip} K_{ir} K_{\alpha j} K_{\alpha k} b_l b_\nu b_\beta b_\mu g(x_{t-l}) g(x_{\tau-\beta}) \epsilon_{i-s-\nu} \epsilon_{\alpha-s-\mu}.$$

Again, it can be verified that the order of magnitude of the leading terms are $O(N^2)$: when $\alpha = i + \mu - \nu$ and $\tau = t + \beta - l$ we have

$$\left(\frac{1}{Na^q}\right)^4 \sum_{i=1}^{N} \sum_{p\neq i}^{N} \sum_{r\neq i}^{N} \sum_{l=0}^{\infty} \sum_{\nu=0}^{\infty} \sum_{\beta=0}^{\infty} \sum_{\mu=0}^{\infty} \sum_{j\neq i+\mu-\nu}^{N} \sum_{k\neq i+\mu-\nu}^{N} \sum_{t=1}^{i-1}$$
$$K_{ip} K_{ir} K_{(i+\mu-\nu)j} K_{(i+\mu-\nu)k} b_l b_\nu b_\beta b_\mu g(x_{t-l})^2 \epsilon_{i-s-\nu}^2,$$

which is of order $O(N^2)$ since $E[K_{ip}K_{ir}K_{(i+\mu-\nu)j}K_{(i+\mu-\nu)k} g(x_{t-l})^2 \epsilon_{i-s-\nu}^2] = O(a^{4q})$ and $\sum_l \sum_\nu \sum_\beta \sum_\mu b_l b_\nu b_\beta b_\mu = O(1)$. Consequently, (D.2) is of order $O_p(N)$. By similar methods, we can verify that the remaining terms are $O_p(N)$. □

Proposition 16.

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

Proof of Proposition 16: This is a direct result from Proposition 15.

Proposition 17. *For $s, t = 1, ..., k$,*

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

Proof of Proposition 17: Notice that

$$\Delta y_{i-s} = B(L)v_{i-s} = B(L)g(x_{i-s}) + B(L)\epsilon_{i-s}, B(L) = \sum_{l=0}^{\infty} b_l L^l$$

$$\begin{aligned}
& \sum_{i=1}^{N} (\Delta y_{i-s} - \hat{\Delta y}_{i-s})(\Delta y_{i-t} - \hat{\Delta y}_{i-t})\hat{f}_i^2 \\
= \; & \sum_{i=1}^{N} (Na^q)^{-2} \left(\sum_{p \neq i}^{N} K_{ip}(\Delta y_{i-s} - \Delta y_{p-s}) \right) \left(\sum_{r \neq i}^{N} K_{ir}(\Delta y_{i-t} - \Delta y_{r-t}) \right) \\
= \; & \sum_{i=1}^{N} (Na^q)^{-2} \left(\sum_{p \neq i}^{N} K_{ip} B(L)([g(x_{i-s}) - g(x_{p-s})] + [\epsilon_{i-s} - \epsilon_{p-s}]) \right) \\
& \times \left(\sum_{r \neq i}^{N} K_{ir} B(L)([g(x_{i-t}) - g(x_{r-t})] + [\epsilon_{i-t} - \epsilon_{r-t}]) \right) \\
= \; & \sum_{i=1}^{N} (Na^q)^{-2} \left(\sum_{p \neq i}^{N} K_{ip} \sum_{l=0}^{\infty} b_l([g(x_{i-s-l}) - g(x_{p-s-l})] + [\epsilon_{i-s-l} - \epsilon_{p-s-l}]) \right) \\
& \times \left(\sum_{r \neq i} K_{ir} \sum_{\nu=0}^{\infty} b_\nu([g(x_{i-t-\nu}) - g(x_{r-t-\nu})] + [\epsilon_{i-t-\nu} - \epsilon_{r-t-\nu}]) \right) \\
= \; & \sum_{i=1}^{N} (Na^q)^{-2} \sum_{l=0}^{\infty} b_l \sum_{\nu=0}^{\infty} b_\nu \\
& \sum_{p \neq i} \sum_{r \neq i} K_{ip} K_{ir} ([g(x_{i-s-l}) - g(x_{p-s-l})] + [\epsilon_{i-s-l} - \epsilon_{p-s-l}]) \\
& \times (([g(x_{i-t-\nu}) - g(x_{r-t-\nu})] + [\epsilon_{i-t-\nu} - \epsilon_{r-t-\nu}]))
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{N}(Na^q)^{-2}\sum_{l=0}^{\infty}b_l\sum_{\nu=0}^{\infty}b_\nu\sum_{p\neq i}^{N}\sum_{r\neq i}^{N}K_{ip}K_{ir}\left[g(x_{i-s-l})-g(x_{p-s-l})\right]\left[g(x_{i-t-\nu})-g(x_{r-t-\nu})\right]\\
&\quad+\sum_{i=1}^{N}(Na^q)^{-2}\sum_{l=0}^{\infty}b_l\sum_{\nu=0}^{\infty}b_\nu\sum_{p\neq i}^{N}\sum_{r\neq i}^{N}K_{ip}K_{ir}\left[g(x_{i-s-l})-g(x_{p-s-l})\right]\left[\epsilon_{i-t-\nu}-\epsilon_{r-t-\nu}\right]\\
&\quad+\sum_{i=1}^{N}(Na^q)^{-2}\sum_{l=0}^{\infty}b_l\sum_{\nu=0}^{\infty}b_\nu\sum_{p\neq i}^{N}\sum_{r\neq i}^{N}K_{ip}K_{ir}\left[\epsilon_{i-s-l}-\epsilon_{p-s-l}\right]\left[g(x_{i-t-\nu})-g(x_{r-t-\nu})\right]\\
&\quad+\sum_{i=1}^{N}(Na^q)^{-2}\sum_{l=0}^{\infty}b_l\sum_{\nu=0}^{\infty}b_\nu\sum_{p\neq i}^{N}\sum_{r\neq i}^{N}K_{ip}K_{ir}\left[\epsilon_{i-s-l}-\epsilon_{p-s-l}\right]\left[\epsilon_{i-t-\nu}-\epsilon_{r-t-\nu}\right]\\
&= P_1+P_2+P_3+P_4
\end{aligned}$$

Similar to the proof of Proposition 15, we can verify that all these terms are $O_p(N)$. □

Proposition 18.

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

Proof of Proposition 18: This is a direct result from Proposition 17.

Proposition 19.

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

Proof of Proposition 19: This can be obtained by using the results of Propositions 16 and 18.

Proposition 20.

$$\frac{1}{N^2}\left(\hat{e}_y\odot\hat{f}\right)^\top\left(\hat{e}_y\odot\hat{f}\right)\Rightarrow E(f^2)\sigma_v^2\int(W_1-\overline{W}_1)^2/A(1)^2.$$

Proof of Proposition 20: Notice that $B(L)=B(1)+B^*(L)(1-L)$

$$y_i=\sum_{l=0}^{i}B(L)v_l=B(1)\sum_{l=0}^{i}v_l+v_i^*-v_0^*=\sum_{l=0}^{i}B(1)v_l+O_p(1),$$

where $v_i^*=B^*(L)v_i$.

$$\widehat{y}_{i-1}\widehat{f}_i=(Na^q)^{-1}\left(\sum_{p\neq i}^{N}K_{ip}\sum_{t=1}^{p-1}B(L)v_t\right)=(Na^q)^{-1}\left(\sum_{p\neq i}^{N}K_{ip}\left[B(1)\sum_{l=0}^{p-1}v_l+v_{p-1}^*-v_0^*\right]\right)$$

$$
\begin{aligned}
& \left(\hat{e}_y \odot \hat{f}\right)^\top \left(\hat{e}_y \odot \hat{f}\right) \\
= & \sum_{i=1}^N \left(y_{i-1} - \widehat{y}_{i-1}\right)^2 \widehat{f}_i^2 \\
= & \sum_{i=1}^N \left(B(1) \sum_{l=0}^{i-1} v_l + v_{i-1}^* - v_0^* - \frac{1}{Na^q} \left(\sum_{p \neq i}^N K_{ip} \left[B(1) \sum_{l=0}^{p-1} v_l + v_{p-1}^* - v_0^* \right] \right) / \widehat{f}_i \right)^2 \widehat{f}_i^2 \\
= & \sum_{i=1}^N \left(y_{i-1}^* - \widehat{y}_{i-1}^* + v_{i-1}^* - v_0^* - \frac{1}{Na^q} \left(\sum_{p \neq i}^N K_{ip} \left[v_{p-1}^* - v_0^* \right] \right) / \widehat{f}_i \right)^2 \widehat{f}_i^2 \\
= & \sum_{i=1}^N \left(y_{i-1}^* - \widehat{y}_{i-1}^* \right)^2 \widehat{f}_i^2 + \sum_{i=1}^N \left(\left(v_{i-1}^* - v_0^* \right) - \frac{1}{Na^q} \left(\sum_{p \neq i}^N K_{ip} \left[v_{p-1}^* - v_0^* \right] \right) / \widehat{f}_i \right)^2 \widehat{f}_i^2 \\
& + 2 \sum_{i=1}^N \left(y_{i-1}^* - \widehat{y}_{i-1}^* \right) \left(\left(v_{i-1}^* - v_0^* \right) - \frac{1}{Na^q} \left(\sum_{p \neq i}^N K_{ip} \left[v_{p-1}^* - v_0^* \right] \right) / \widehat{f}_i \right) \widehat{f}_i^2
\end{aligned}
$$

where

$$
y_i^* = B(1) \sum_{l=0}^{i} v_l .
$$

It can be shown that

$$
\begin{aligned}
\sum_{i=1}^N \left(\left(v_{i-1}^* - v_0^* \right) - \frac{1}{Na^q} \left(\sum_{p \neq i}^N K_{ip} \left[v_{p-1}^* - v_0^* \right] \right) / \widehat{f}_i \right)^2 \widehat{f}_i^2 &= O_p(N) \\
\sum_{i=1}^N \left(y_{i-1}^* - \widehat{y}_{i-1}^* \right) \left(\left(v_{i-1}^* - v_0^* \right) - \frac{1}{Na^q} \left(\sum_{p \neq i}^N K_{ip} \left[v_{p-1}^* - v_0^* \right] \right) / \widehat{f}_i \right) \widehat{f}_i^2 &= O_p(N)
\end{aligned}
$$

thus

$$
\left(\hat{e}_y \odot \hat{f}\right)^\top \left(\hat{e}_y \odot \hat{f}\right) = \sum_{i=1}^N \left(y_{i-1}^* - \widehat{y}_{i-1}^* \right)^2 \widehat{f}_i^2 + o_p(N^2)
$$

By a similar analysis as the proof of Theorem 1, we can show that

$$
\frac{1}{N^2} \sum_{i=1}^N (y_{i-1}^* - \widehat{y}_{i-1}^*)^2 \widehat{f}_i^2 \Rightarrow B(1)^2 E(f^2) \sigma_v^2 \int (W_1 - \overline{W}_1)^2 .
$$

□

Proposition 21.

$$\frac{1}{N^2}\left(\hat{e}_y\odot\hat{f}\right)^\top(I-P)\left(\hat{e}_y\odot\hat{f}\right)\Rightarrow E(f^2)\sigma_v^2\int(W_1-\overline{W}_1)^2/A(1)^2.$$

Proof of Proposition 21:

$$\begin{aligned}
&\left(\hat{e}_y\odot\hat{f}\right)^\top(I-P)\left(\hat{e}_y\odot\hat{f}\right)\\
=&\left(\hat{e}_y\odot\hat{f}\right)^\top\left(\hat{e}_y\odot\hat{f}\right)\\
&-\left(\hat{e}_y\odot\hat{f}\right)^\top\left(\hat{e}_s\odot\hat{f}\right)\left[\left(\hat{e}_s\odot\hat{f}\right)^\top\left(\hat{e}_s\odot\hat{f}\right)\right]^{-1}\left(\hat{e}_s\odot\hat{f}\right)^\top\left(\hat{e}_y\odot\hat{f}\right)\\
=&\left(\hat{e}_y\odot\hat{f}\right)^\top\left(\hat{e}_y\odot\hat{f}\right)+o_p(N^2)
\end{aligned}$$

Then, by the result of Proposition 20, we have the proof. □

Proposition 22. *For* $s=1,....,k$,

$$\sum_{i=1}^N(\Delta y_{i-s}-\widehat{\Delta y}_{i-s})(\epsilon_i-\widehat{\epsilon}_i+g(x_i)-\widehat{g}(x_i))\hat{f}_i^2=o_p(N).$$

Proof of Proposition 22: By definition,

$$\begin{aligned}
&\sum_{i=1}^N(\Delta y_{i-s}-\widehat{\Delta y}_{i-s})(\epsilon_i-\widehat{\epsilon}_i+g(x_i)-\widehat{g}(x_i))\hat{f}_i^2\\
=&\sum_{i=1}^N\hat{f}_i^2B(L)v_{i-s}\epsilon_i-\sum_{i=1}^N\hat{f}_i\left(\frac{1}{Na^q}\sum_{p\neq i}^N K_{ip}B(L)v_{p-s}\right)\epsilon_i\\
&-\sum_{i=1}^N\hat{f}_i^2B(L)v_{i-s}\widehat{\epsilon}_i+\sum_{i=1}^N\hat{f}_i\left(\frac{1}{Na^q}\sum_{p\neq i}^N K_{ip}B(L)v_{p-s}\right)\widehat{\epsilon}_i\\
&+\sum_{i=1}^N\hat{f}_i^2B(L)v_{i-s}(g(x_i)-\widehat{g}(x_i))-\sum_{i=1}^N\hat{f}_i\left(\frac{1}{Na^q}\sum_{p\neq i}^N K_{ip}B(L)v_{p-s}\right)(g(x_i)-\widehat{g}(x_i))\\
=&Q_1+Q_2+Q_3+Q_4+Q_5+Q_6
\end{aligned}$$

$$Q_1=\sum_{i=1}^N\hat{f}_i^2B(L)v_{i-s}\epsilon_i=\sum_{i=1}^N\hat{f}_i^2B(L)g(x_{i-s})\epsilon_i+\sum_{i=1}^N\hat{f}_i^2B(L)\epsilon_{i-s}\epsilon_i$$

$$\begin{aligned}
Q_2 &= \sum_{i=1}^{N} \hat{f}_i \left(\frac{1}{Na^q} \sum_{p \neq i}^{N} K_{ip} B(L) v_{p-s} \right) \epsilon_i \\
&= \sum_{i=1}^{N} \hat{f}_i \left(\frac{1}{Na^q} \sum_{p \neq i}^{N} K_{ip} B(L) \left[g(x_{p-s}) + \epsilon_{p-s} \right] \right) \epsilon_i \\
&= \sum_{i=1}^{N} \hat{f}_i \left(\frac{1}{Na^q} \sum_{p \neq i}^{N} K_{ip} B(L) g(x_{p-s}) \right) \epsilon_i + \sum_{i=1}^{N} \hat{f}_i \left(\frac{1}{Na^q} \sum_{p \neq i}^{N} K_{ip} B(L) \epsilon_{p-s} \right) \epsilon_i
\end{aligned}$$

$$\begin{aligned}
Q_3 &= \sum_{i=1}^{N} \hat{f}_i^2 B(L) v_{i-s} \widehat{\epsilon}_i \\
&= \sum_{i=1}^{N} \hat{f}_i^2 B(L) [g(x_{i-s}) + \epsilon_{i-s}] \left(\frac{1}{Na^q} \sum_{p \neq i}^{N} K_{ip} \epsilon_p \right) \\
&= \sum_{i=1}^{N} \hat{f}_i^2 B(L) g(x_{i-s}) \left(\frac{1}{Na^q} \sum_{p \neq i}^{N} K_{ip} \epsilon_p \right) + \sum_{i=1}^{N} \hat{f}_i^2 B(L) \epsilon_{i-s} \left(\frac{1}{Na^q} \sum_{p \neq i}^{N} K_{ip} \epsilon_p \right) \\
&= \sum_{p=1}^{N} \epsilon_p \left(\frac{1}{Na^q} \sum_{i \neq p}^{N} K_{ip} \hat{f}_i^2 B(L) g(x_{i-s}) \right) + \sum_{i=1}^{N} \hat{f}_i^2 B(L) \epsilon_{i-s} \left(\frac{1}{Na^q} \sum_{p \neq i}^{N} K_{ip} \epsilon_p \right)
\end{aligned}$$

$$\begin{aligned}
Q_4 &= \sum_{i=1}^{N} \hat{f}_i \left(\frac{1}{Na^q} \sum_{p \neq i}^{N} K_{ip} B(L) v_{p-s} \right) \widehat{\epsilon}_i \\
&= \sum_{i=1}^{N} \hat{f}_i \left(\frac{1}{Na^q} \sum_{p \neq i}^{N} K_{ip} B(L) v_{p-s} \right) \left(\frac{1}{Na^q} \sum_{j \neq i}^{N} K_{ij} \epsilon_j \right) \\
&= \left(\frac{1}{Na^q} \right)^2 \sum_{i=1}^{N} \hat{f}_i \left(\sum_{p \neq i} \sum_{j \neq i} K_{ip} K_{ij} \epsilon_j \left[B(L) \left[g(x_{p-s}) + \epsilon_{p-s} \right] \right] \right) \\
&= \left(\frac{1}{Na^q} \right)^2 \sum_{i=1}^{N} \hat{f}_i \left(\sum_{p \neq i} \sum_{j \neq i} K_{ip} K_{ij} \epsilon_j \left[B(L) g(x_{p-s}) \right] \right) \\
&\quad + \left(\frac{1}{Na^q} \right)^2 \sum_{i=1}^{N} \hat{f}_i \left(\sum_{p \neq i} \sum_{j \neq i} K_{ip} K_{ij} \epsilon_j \left[B(L) \epsilon_{p-s} \right] \right)
\end{aligned}$$

$$\begin{aligned}
Q_5 &= \sum_{i=1}^{N} \hat{f}_i^2 B(L) v_{i-s}(g(x_i) - \widehat{g}(x_i)) \\
&= \sum_{i=1}^{N} \hat{f}_i^2 \left[B(L)[g(x_{i-s}) + \epsilon_{i-s}]\right] \left(\frac{1}{Na^q}\sum_{j=1}^{N} K(\frac{x_i - x_j}{h}) \left[g(x_i) - g(x_j)\right] / \widehat{f}(x_i)\right) \\
&= \sum_{i=1}^{N} \hat{f}_i \left[B(L) g(x_{i-s})\right] \left(\frac{1}{Na^q}\sum_{j=1}^{N} K(\frac{x_i - x_j}{h}) \left[g(x_i) - g(x_j)\right]\right) \\
&\quad + \sum_{i=1}^{N} \hat{f}_i \left[B(L) \epsilon_{i-s}\right] \left(\frac{1}{Na^q}\sum_{j=1}^{n} K(\frac{x_i - x_j}{h}) \left[g(x_i) - g(x_j)\right]\right)
\end{aligned}$$

$$\begin{aligned}
Q_6 &= \sum_{i=1}^{N} \hat{f}_i \left(\frac{1}{Na^q}\sum_{p\neq i}^{N} K_{ip} B(L) v_{p-s}\right) (g(x_i) - \widehat{g}(x_i)) \\
&= \sum_{i=1}^{N} \hat{f}_i \left(\frac{1}{Na^q}\sum_{p\neq i}^{N} K_{ip} B(L) \left[g(x_{p-s}) + \epsilon_{p-s}\right]\right) \left(\frac{1}{Na^q}\sum_{j=1}^{N} K(\frac{x_i - x_j}{h}) \left[g(x_i) - g(x_j)\right] / \widehat{f}(x_i)\right) \\
&= \sum_{i=1}^{N} \left(\frac{1}{Na^q}\sum_{p\neq i}^{N} K_{ip} \left[B(L) g(x_{p-s})\right]\right) \left(\frac{1}{Na^q}\sum_{j=1}^{N} K(\frac{x_i - x_j}{h}) \left[g(x_i) - g(x_j)\right]\right) \\
&\quad + \sum_{i=1}^{N} \left(\frac{1}{Na^q}\sum_{p\neq i} K_{ip} \left[B(L) \epsilon_{p-s}\right]\right) \left(\frac{1}{Na^q}\sum_{j=1}^{n} K(\frac{x_i - x_j}{h}) \left[g(x_i) - g(x_j)\right]\right)
\end{aligned}$$

By staightforward but tedious moment verification we can show that all of the above terms are of order $o_p(N)$. □

Proposition 23.

$$\left(\hat{e}_s \odot \hat{f}\right)^\top \left(\hat{e}_\epsilon \odot \hat{f}\right) = o_p(N)$$

Proof of Proposition 23: This is a direct result of Lemma 22.

Proposition 24.

$$\left(\hat{e}_y \odot \hat{f}\right)^\top \left(\hat{e}_s \odot \hat{f}\right) \left[\left(\hat{e}_s \odot \hat{f}\right)^\top \left(\hat{e}_s \odot \hat{f}\right)\right]^{-1} \left(\hat{e}_s \odot \hat{f}\right)^\top \left(\hat{e}_\epsilon \odot \hat{f}\right) = o_p(N)$$

Proof of Proposition 24: The results can be obtained by the results of Propositions 16, 18, and 23.

Proposition 25.

$$\frac{1}{N}\left(\hat{e}_y\odot\hat{f}\right)^\top\left(\hat{e}_\epsilon\odot\hat{f}\right)\Rightarrow B(1)\sigma_v\sigma_{\epsilon f}\left(\rho\int W_1^\tau dW_1+\sqrt{1-\rho^2}\int W_1^\tau dW_2\right).$$

Proof of Proposition 25:

$$\begin{aligned}
&\left(\hat{e}_y\odot\hat{f}\right)^\top\left(\hat{e}_\epsilon\odot\hat{f}\right)\\
=&\sum_{i=1}^N\left(y_{i-1}-\widehat{y}_{i-1}\right)\widehat{f}_i^2\left(\epsilon_i-\widehat{\epsilon}_i+g(x_i)-\widehat{g}(x_i)\right)\\
=&\sum_{i=1}^N\left(B(1)\sum_{l=0}^{i-1}v_l+v_{i-1}^*-v_0^*-\frac{1}{Na^q}\left(\sum_{p\neq i}^N K_{ip}\left[B(1)\sum_{l=0}^{p-1}v_l+v_{p-1}^*-v_0^*\right]\right)/\widehat{f}_i\right)\\
&\times\left(\epsilon_i-\widehat{\epsilon}_i+g(x_i)-\widehat{g}(x_i)\right)\widehat{f}_i^2\\
=&\sum_{i=1}^N\left(y_{i-1}^*-\widehat{y}_{i-1}^*+v_{i-1}^*-v_0^*-\frac{1}{Na^q}\left(\sum_{p\neq i}^N K_{ip}\left[v_{p-1}^*-v_0^*\right]\right)/\widehat{f}_i\right)\\
&\times\left(\epsilon_i-\widehat{\epsilon}_i+g(x_i)-\widehat{g}(x_i)\right)\widehat{f}_i^2\\
=&\sum_{i=1}^N\left(y_{i-1}^*-\widehat{y}_{i-1}^*\right)\left(\epsilon_i-\widehat{\epsilon}_i+g(x_i)-\widehat{g}(x_i)\right)\widehat{f}_i^2\\
&+\sum_{i=1}^N\left(v_{i-1}^*-v_0^*-\frac{1}{Na^q}\left(\sum_{p\neq i}^N K_{ip}\left[v_{p-1}^*-v_0^*\right]\right)/\widehat{f}_i\right)\left(\epsilon_i-\widehat{\epsilon}_i+g(x_i)-\widehat{g}(x_i)\right)\widehat{f}_i^2
\end{aligned}$$

It can be shown that

$$\sum_{i=1}^N\left(v_{i-1}^*-v_0^*-\frac{1}{Na^q}\left(\sum_{p\neq i}^N K_{ip}\left[v_{p-1}^*-v_0^*\right]\right)/\widehat{f}_i\right)\left(\epsilon_i-\widehat{\epsilon}_i+g(x_i)-\widehat{g}(x_i)\right)\widehat{f}_i^2=o_p(N)$$

thus

$$\begin{aligned}
&\frac{1}{N}\left(\hat{e}_y\odot\hat{f}\right)^\top\left(\hat{e}_\epsilon\odot\hat{f}\right)\\
=&\frac{1}{N}\sum_{i=1}^N\left(y_{i-1}^*-\widehat{y}_{i-1}^*\right)\left(\epsilon_i-\widehat{\epsilon}_i+g(x_i)-\widehat{g}(x_i)\right)\widehat{f}_i^2+o_p(1)\\
=&\frac{1}{N}\sum_{i=1}^N\left(y_{i-1}^*-\widehat{y}_{i-1}^*\right)\epsilon_i\widehat{f}_i^2+o_p(1)\\
\Rightarrow&B(1)\sigma_v\sigma_{\epsilon f}\left(\rho\int W_1^\tau dW_1+\sqrt{1-\rho^2}\int W_1^\tau dW_2\right).
\end{aligned}$$

□

Proposition 26.

$$\frac{1}{N}\left(\hat{e}_y \odot \hat{f}\right)^\top (I-P)\left(\hat{e}_\epsilon \odot \hat{f}\right) \Rightarrow B(1)\sigma_v\sigma_{\epsilon f}\left(\rho\int W_1^\tau dW_1 + \sqrt{1-\rho^2}\int W_1^\tau dW_2\right).$$

Proof of Proposition 26:

$$\begin{aligned}
& \left(\hat{e}_y \odot \hat{f}\right)^\top (I-P)\left(\hat{e}_\epsilon \odot \hat{f}\right) \\
= \; & \left(\hat{e}_y \odot \hat{f}\right)^\top \left(\hat{e}_\epsilon \odot \hat{f}\right) \\
& -\left(\hat{e}_y \odot \hat{f}\right)^\top \left(\hat{e}_s \odot \hat{f}\right)\left[\left(\hat{e}_s \odot \hat{f}\right)^\top \left(\hat{e}_s \odot \hat{f}\right)\right]^{-1}\left(\hat{e}_s \odot \hat{f}\right)^\top \left(\hat{e}_\epsilon \odot \hat{f}\right) \\
= \; & \left(\hat{e}_y \odot \hat{f}\right)^\top \left(\hat{e}_\epsilon \odot \hat{f}\right) + o_p(N)
\end{aligned}$$

Thus the result follows immediately. □

References

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

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*. 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.

ENGLE, R. F., C. W. J. GRANGER, J. RICE, AND A. WEISS (1986): "Semiparametric Estimates for the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81, 310–320.

FAN, Y., AND Q. LI (1999): "Root-N-Consistent Estimation of Partially Linear Time Series Models," *Journal of Nonparametric Statistics*, 11, 251–269.

FULLER, W. A. (1976): *Introduction to Statistical Time Series*. New York: Wiley.

GRANGER, C. W. J. (1995): "Modeling Nonlinear Relationships Between Extended-Memory Variables," *Econometrica*, 63, 265–279.

GRANGER, C. W. J., AND T. TERASVIRTA (1993): *Modelling Nonlinear Economic Relationships*. New York: Oxford University Press.

HANSEN, B. E. (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.

——— (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.

HECKMAN, N. E. (1986): "Spline Smoothing in a Partly Linear Model," *Journal of the Royal Statistical Society, Series B*, 48, 244–248.

LI, Q. (1999): "Consistent Model Specification Tests for Time Series Econometric Models," *Journal of Econometrics*, 92, 101–147.

LINTON, O. B. (1995): "Second Order Approximation in the Partially Linear Regression Model," *Econometrica*, 63, 1079–1112.

LINTON, O. B., AND J. NIELSEN (1995): "A kernel method of estimating structured nonparametric regression based on marginal integration," *Biometrica*, 82, 93–100.

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

POWELL, J. L., J. H. STOCK, AND 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*, pp. 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. (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.

SHILLER, R. J. (1984): "Smoothness Priors and Nonlinear Regression," *Journal of the American Statistical Association*, 72, 420–423.

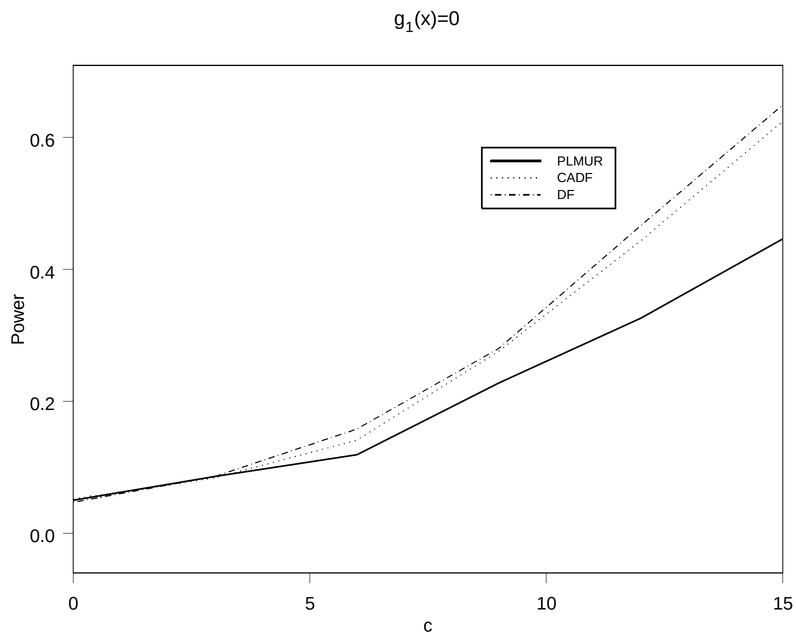
SHIN, D. W., AND B. S. SO (1999): "Unit Root Tests Based on Adaptive Maximum Likelihood Estimation," *Econometric Theory*, 15, 1–23.

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*. New York: Oxford University Press.

WAHBA, G. (1984): "Partial Spline Models for the Semiparametric Estimation of Functions of Sevearl Variables," in *Statistical Analysis of Time Series*, pp. 319–329. Tokyo: Institute of Statistical Mathematics.

FIGURE 1


g1(x)=0
PLMUR
CADF
DF
0.6
0.4
0.2
0.0
Power
0
5
10
15
c

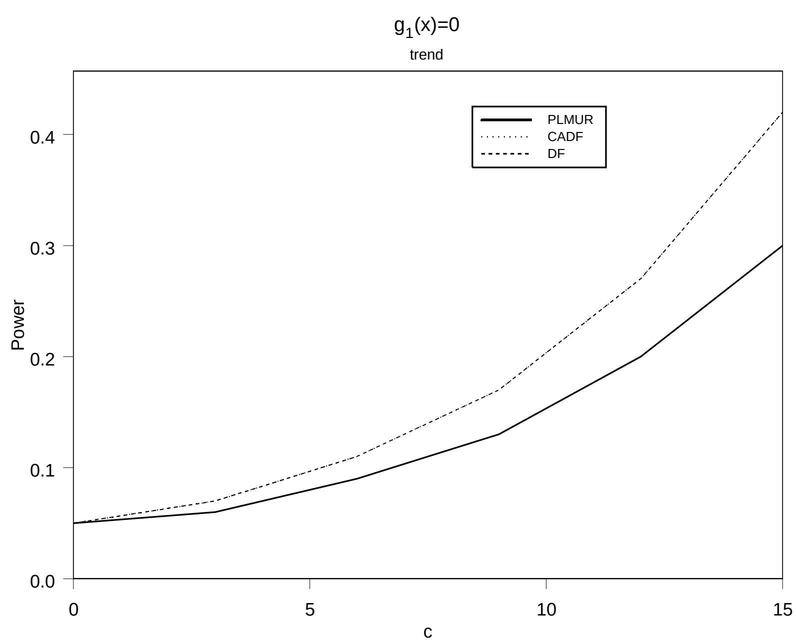
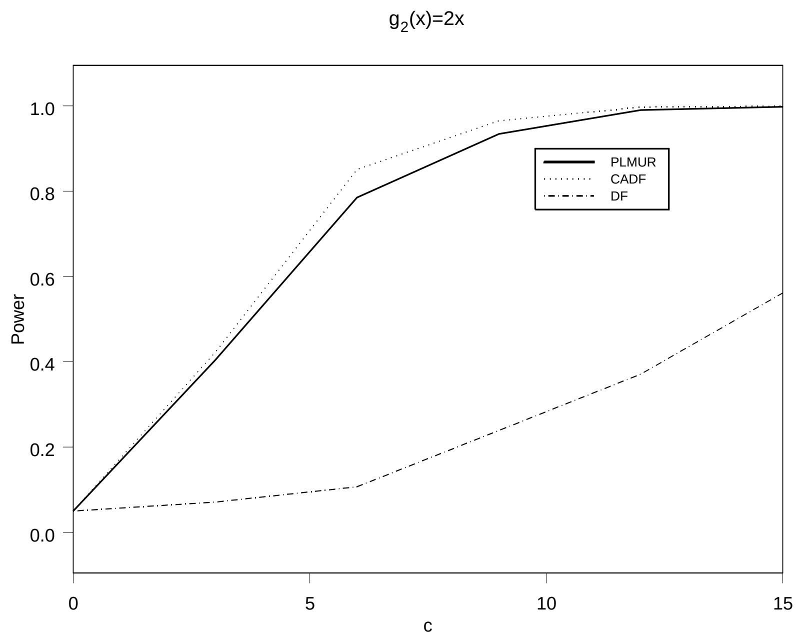

g1(x)=0
trend
PLMUR
CADF
DF
0.4
0.3
0.2
0.1
0.0
Power
0
5
10
15
c

Figure 2


g2(x)=2x
Power
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
c
PLMUR
CADF
DF

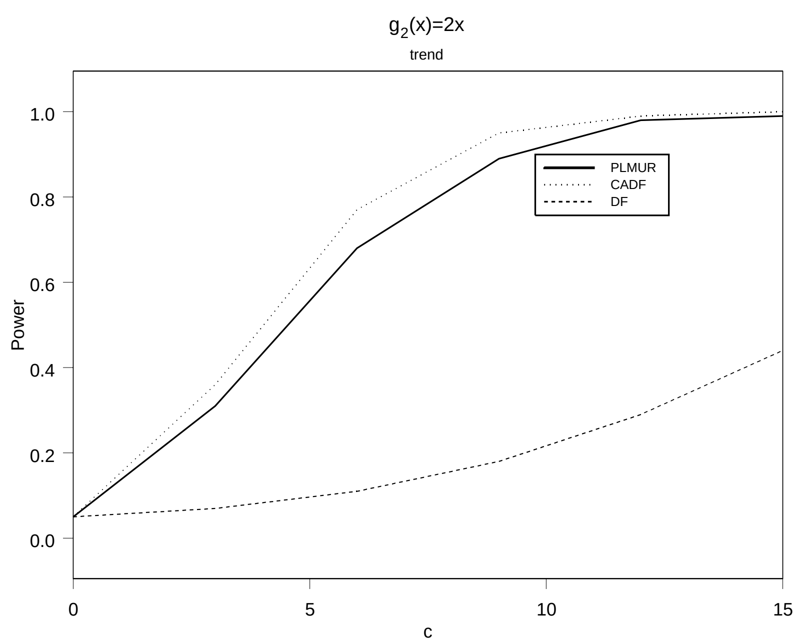
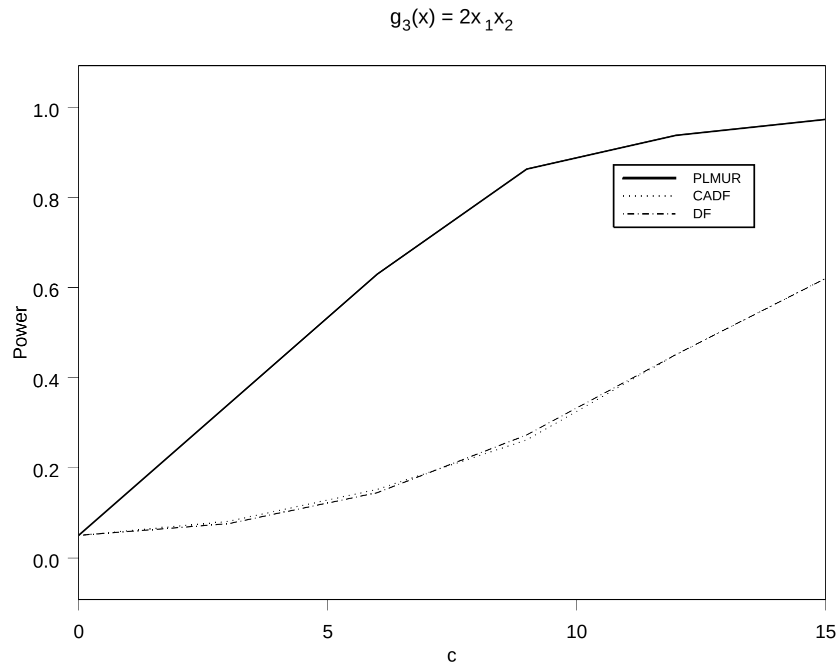

g2(x)=2x
trend
Power
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
c
PLMUR
CADF
DF

Figure 3


$g_3(x) = 2x_1x_2$
Power
c
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
PLMUR
CADF
DF

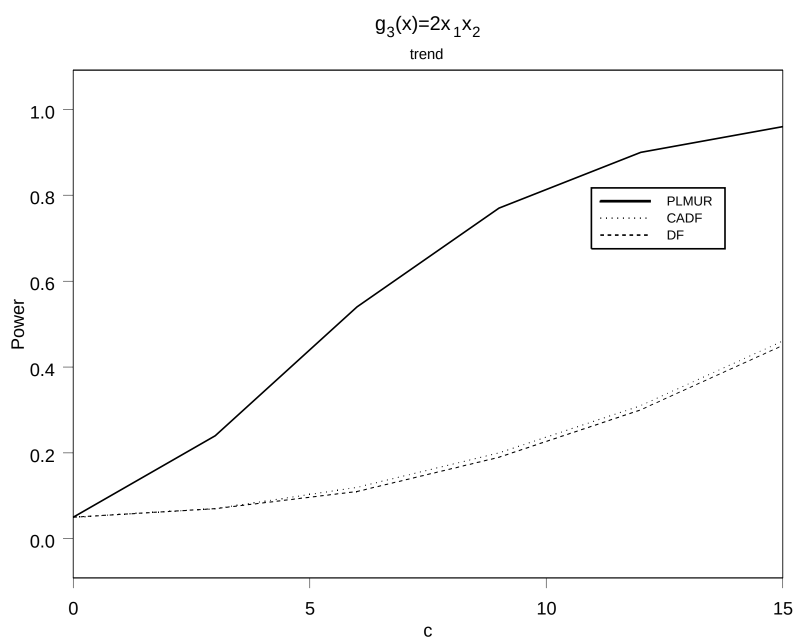
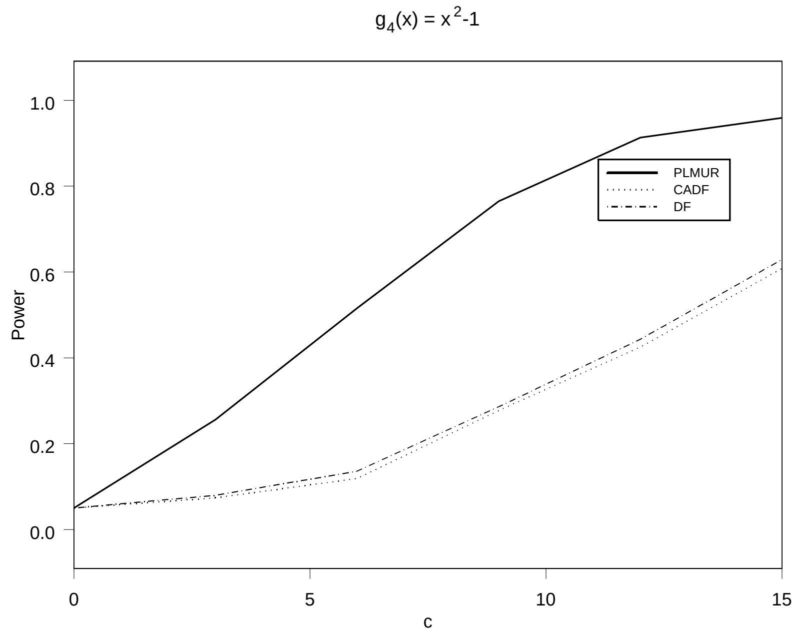

$g_3(x)=2x_1x_2$
trend
Power
c
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
PLMUR
CADF
DF

Figure 4


g4(x) = x2-1
Power
c
PLMUR
CADF
DF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15

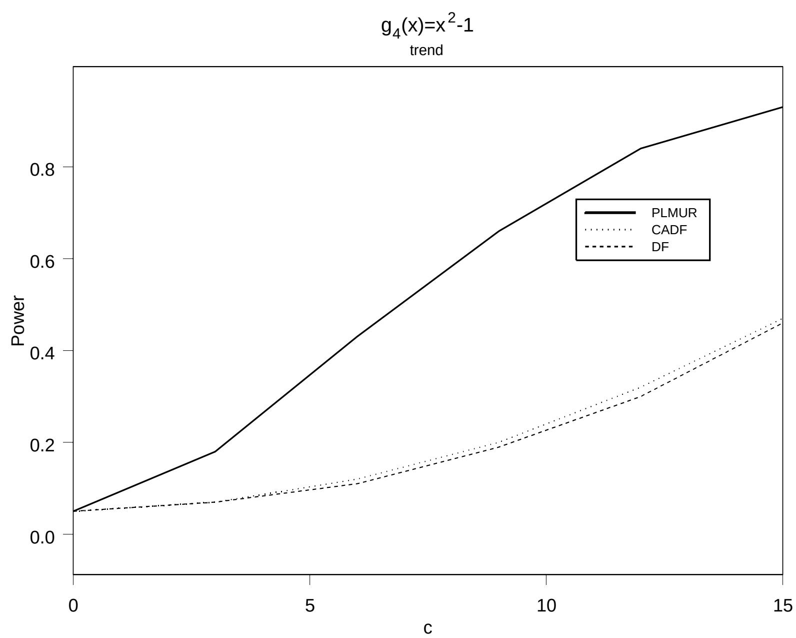
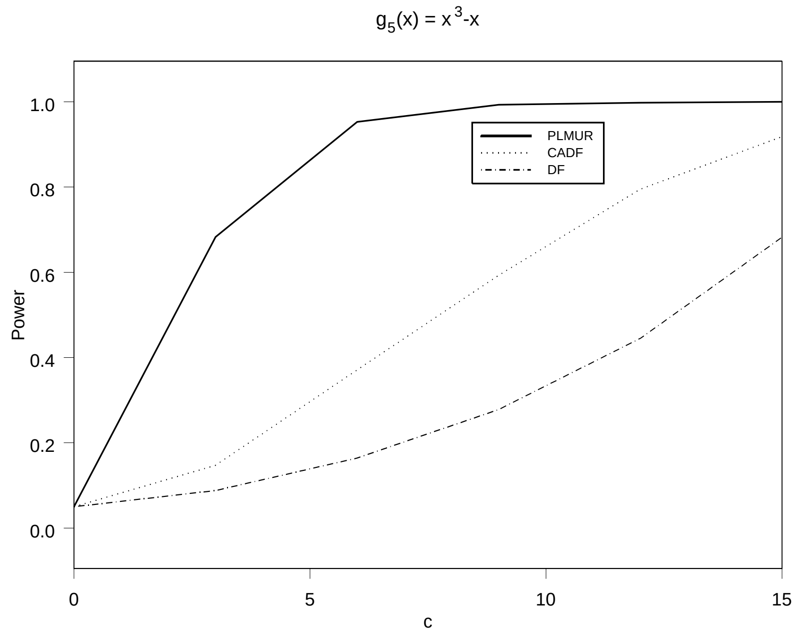
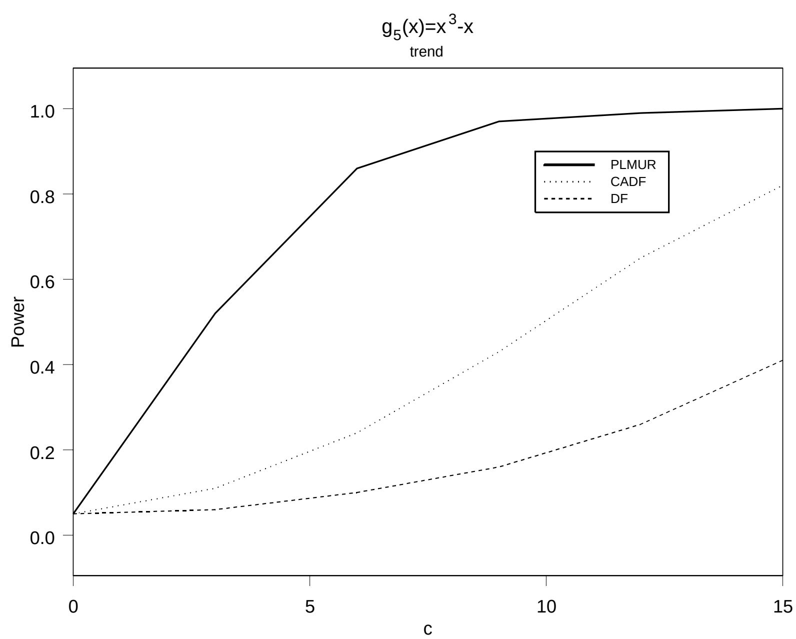

g4(x)=x2-1
trend
Power
c
PLMUR
CADF
DF
0.8
0.6
0.4
0.2
0.0
0
5
10
15

Figure 5



g5(x) = x3-x
1.0
0.8
0.6
0.4
0.2
0.0
Power
PLMUR
CADF
DF
0
5
10
15
c



g5(x)=x3-x
trend
1.0
0.8
0.6
0.4
0.2
0.0
Power
PLMUR
CADF
DF
0
5
10
15
c