Many macroeconomic and financial variables show highly persistent and correlated patterns but not necessarily cointegrated. Recently, Sun, Hsiao and Li (2010) propose using a semiparametric varying coefficient approach to capture correlations between integrated but non cointegrated variables. Due to the complication arising from the integrated disturbance term and the semiparametric functional form, consistent estimation of such a semiparametric model requires stronger conditions than usually needed for consistent estimation for a linear (spurious) regression model, or a semiparametric varying coefficient model with a stationary disturbance. Therefore, it is important to develop a testing procedure to examine for a given data set, whether linear relationship holds or not, while allowing for the disturbance being an integrated process. In this paper we propose two test statistics for detecting linearity against semiparametric varying coefficient alternative specification. Monte Carlo simulations are used to examine the finite sample performances of the proposed tests.

Key Words: Specification test; Spurious regression; varying coefficient; kernel estimation.

JEL Code: C13, C14, C22

*The corresponding author: Li Gan, email: gan@econmail.tamu.edu


1 Introduction

In the study of finance, economists are interested in the co-movements of financial variables among different markets and countries, see Hamao, Masulis and Ng (1990), Koch and Koch (1991), Koutmos and Booth (1995), and etc. Many of models, estimation methods and the related asymptotic theories used in this area of research are based on stationarity assumptions, such as vector ARMA with multivariate GARCH-type models, see, e.g. Bollerslev, Engle, and Wooldridge (1988); Bollerslev (1990); Engle and Kroner (1995); Engle (2002). However, some financial variables like market volatilities show highly-persistent patterns, which leads researchers to consider nonstationarity such as I(1) processes, see Harvey, Ruiz, and Shephard (1994), Hong (2001) and references therein. Also, researchers want to capture the time changing features of the relationship among different variables, as in King, Sentana, and Wadhwani (1994), who use multivariate factor models allowing time-varying conditional volatilities. Recently, Sun et al. (2010) proposed using a flexible semiparametric varying coefficient model specification to capture the volatility spillover effects while allowing for non-stationary and non-cointegrated time series data. Their approach allows for the impact of foreign stock volatility on domestic stock volatility being time-varying. However, consistent estimation of such a semiparametric model requires some strong assumptions. Therefore, in practice one should first conduct some model specification check. If a simple linear model adequately describes the relation of the related economic/financial variables, there is no need to estimate the model using some nonparametric estimation techniques. On the other hand, if a linear model is misspecified, a semiparametric specification can overcome the model misspecification problem. To the best of our knowledge there does not exist any model specification test in a semiparametric regression model framework that allows for correlated integrated but not cointegrated variables. In this paper we propose two test statistics for testing the null hypothesis of a linear regression model against a semiparametric varying coefficient model with correlated but not cointegrated non-stationary variables.

Nonparametric and semiparametric models start to gain popularity due to its flexibility and easy implementation. In the time series literature, nonlinearity gained much attention recently, which is used to capture complicated correlations. Nonparametric and semiparametric regression models is one way to incorporate nonlinearity. There have been many works on nonparametric estimations of stationary regression models. Also, there is a growing interest in applying nonparametric estimation techniques to analyze non-stationary data. Juhl (2005) considered a nonparametric regression model when some variables are generated by unit root process. Recently, estimation of nonparametric and semiparametric cointegration
models have attracted much attention among econometricians and statisticians. Wang and Phillips (2009a, 2009b) considered nonparametric cointegrations. Cai, Li and Park (2009), and Xiao (2009) considered varying coefficient cointegrations with a stationary varying coefficient covariate. Sun, Cai and Li (2010) studied the varying coefficient cointegration with an I(1) varying coefficient covariate.\footnote{Varying coefficient models were investigated by Robinson (1991), Chen and Tsay (1993). There is a rapidly growing literature on it, including Fan, Cai and Li (2000), Cai, Fan and Yao (2000), Li et al. (2002), and Fan, Yao and Cai (2003), among others.} Karlsen, Myklebust and Tjøstheim (2007) considered a nonparametric cointegration type model but with a more general group of nonstationary processes, that is, the null recurrent Markov chains, which include the unit root process. For model specification tests with non-stationary covariates, Gao et al. (2009) considered the problem of testing the null of a parametric regression functional form with an I(1) regressor. Sun et al. (2010) considered the problem of testing the null of a linear regression model against a semiparametric varying coefficient model. Both Gao et al. and Sun et al. considered the cointegration relationship. That is, the error terms in their regression models are stationary I(0) processes. In this paper we consider the case that the error term is a non-stationary I(1) process so that we allow for correlated but not cointegrated I(1) processes.

Sun et al.’s (2010) approach generalizes nonparametric estimation method to the non-stationary and non-cointegrated data case. Many economic relationships concerned by economists can be modeled with this framework. One interesting example to consider is the stock markets’ volatility spill-over effects. Empirical data suggest that stock market volatilities usually follow I(1) or near I(1) processes. While volatilities from different markets are likely to be correlated with each other, it is unlikely that a domestic stock market volatility is cointegrated with a foreign stock market volatility because there are many other (domestic) factors that also affect the domestic market’s volatility. Therefore, financial markets volatilities of two countries are likely to be correlated with each other but not cointegrated. Moreover, this correlation has a time-varying feature which may depend on the varying risk-premium. One reason for the change of risk-premium is the fluctuation in the exchange rate market. Therefore, adding change of the exchange rate as the covariate in the varying coefficient function may give a more precise characterization of the spill-over effects of stock market volatilities. However, if a linear model can adequately describe the relation among economic variables, then one can estimate a linear model more efficiently than by using some
semiparametric estimation method. This paper aims to provide some testing procedures that can be used to examine whether the relationship between two markets’ volatilities follows a linear relationship or not, while allowing for the two markets’ volatilities being correlated but not cointegrated.

The rest of the paper is organized as follows. Section 2 presents the model and propose two test statistics and examine their asymptotic behaviors. In Section 3 we report Monte Carlo simulation results to examine the finite sample performance of the proposed test statistics. Finally, Section 4 concludes the paper.

2 The Model and the Test Statistic

2.1 The Model and the Testing Problem

We consider the following semiparametric varying coefficient model:

$$Y_t = X_t^T \theta(Z_t) + u_t, \quad (t = 1, ..., n)$$ (2.1)

where $X_t$ and $u_t$ are integrated processes of order one (i.e., $I(1)$ processes) and $Z_t$ is a stationary process. Hence, we have $X_t = X_{t-1} + \eta_t$ and $u_t = u_{t-1} + \epsilon_t$, where $\eta_t$ and $\epsilon_t$, along with $Z_t$, are some weakly dependent stationary processes.

Model (2.1) is a semiparametric model with correlated integrated variables $Y_t$ and $X_t$, but they are not cointegrated because the error term $u_t$ is an $I(1)$ process. As we discussed in the introduction, many macroeconomic and finance variables are correlated integrated processes but they are not necessarily cointegrated with each other. Hence, model (2.1) allows for applied researchers to study the relationship of correlated but not cointegrated economic/finance variables without imposing linearity functional form assumption.

Sun et al. (2010) suggest a two step estimation procedure to consistently estimate the unknown function $\theta(\cdot)$. Sun et al. first decompose $\theta(z)$ into $\theta(z) = \alpha(z) + c_0$, where $\alpha(z) = \theta(z) - E[\theta(Z_t)]$ and $c_0 = E[\theta(Z_t)]$. Let $\hat{\theta}(z)$ denote a standard local linear estimator of $\theta(z)$ by ignoring that the error $u_t$ is an $I(1)$ process. Because of the $I(1)$ error term, $\hat{\theta}(z)$ is not a consistent estimator of $\theta(z)$. Sun et al. show that $\hat{\theta}(z)$ converges to $\theta(z) + \tilde{\theta}$, where $\tilde{\theta}$ is a $O_p(1)$ random variable related to Brownian motions generated with the innovations that generate $X_t$ and $u_t$. Sun et al. further show that $n^{-1} \sum_{t=1}^{n} \tilde{\theta}(Z_t)$ converges to $c_0 + \tilde{\theta}$. 4
Hence, one can consistently estimate \( \alpha(z) = \theta(z) - c_0 \) by \( \hat{\alpha}(z) = \hat{\theta}(z) - n^{-1} \sum_{t=1}^{n} \hat{\theta}(Z_t) \).

Under somewhat strong regularity conditions, Sun et al. show that a two step estimation method can be used to consistently estimate \( c_0 \) (see section 2.3 for the detailed estimation procedure).

Let \( \hat{\alpha}(z) \) and \( \hat{c}_0 \) denote the estimators of \( \alpha(z) \) and \( c_0 \) proposed by Sun et al., then one can estimate \( \theta(z) \) by \( \hat{\theta}(z) = \hat{\alpha}(z) + \hat{c}_0 \). Sun et al. derive the rate of convergence and asymptotic distribution of \( \hat{\alpha}(z) - \alpha(z) \) under quite (weak) standard regularity conditions. Consistent estimation of \( c_0 \) and its related asymptotic theory is much more challenging. Under some stronger regularity conditions, Sun et al. (2010) derive the rate of convergence of \( \hat{c}_0 - c_0 \), but did not provide asymptotic distribution of \( \hat{c}_0 - c_0 \). Therefore, it is difficulty to draw inference on \( \hat{\theta}(z) = \hat{\alpha}(z) + \hat{c}_0 \). However, when \( \theta(z) = \theta_0 \) (for all \( z \)), where \( \theta_0 \) is a \( d \times 1 \) vector of constant parameters, it is well known that one can consistently estimate \( \theta_0 \) based on first difference of the data, and the resulting estimator of \( \theta_0 \) is \( \sqrt{n} \)-consistent and asymptotically normally distributed because first differenced data are stationary. Therefore, if the relationship between \( Y_t \) and \( X_t \) is linear (plus an \( I(1) \) disturbance term), one does not have to use any nonparametric estimation method. On the other hand, if the relationship between \( Y_t \) and \( X_t \) is nonlinear and time varying, then estimation result based on a misspecified linear model may lead to erroneous conclusions. Therefore, testing whether \( \theta(z) \) is a vector of constant parameter is of special importance to applied researchers given the complication of the semiparametric estimation procedure and the fact that the (asymptotic) distribution of the semiparametric estimator \( \hat{\theta}(z) = \hat{\alpha}(z) + \hat{c}_0 \) is unknown.

Therefore, the aim of this paper is to develop some testing procedures to test the null hypothesis that

\[
P[\theta(Z_t) = \theta_0] = 1 \text{ for some } \theta_0 \in \mathbb{R}^d. \quad (2.2)
\]

Ideally, one would hope to have a consistent test, that is, when the null hypothesis is false, the test can reject null with probability approaching one as the sample diverges to infinity. In this paper we consider two test statistics. One test statistic has an asymptotic standard normal distribution under the null hypothesis. But it lacks power in certain directions when the null hypothesis is false. Hence, it is not a consistent test. The second test statistic we propose comes from the principle of constructing a consistent test, however, its asymptotic null distribution is difficult to establish without imposing some high level assumptions. We
will use some bootstrap procedures to approximate the null distribution of the second test statistic. Even the first test statistic has a asymptotic standard normal null distribution, it is well known that in finite sample applications, nonparametric estimation based test statistics usually suffer severe size distortions. Therefore, we will also use a bootstrap procedure to approximate the null distribution of the first test statistic.

Our first test statistic is based on nonparametric estimation of the following quantity:

\[(a) \quad I_a = E[\epsilon_t E(\epsilon_t | Z_t) f(Z_t)] \text{, where } \epsilon_t = \Delta Y_t - \Delta X_t^T \theta_0. \text{ Under } H_0 \text{ we have } I_a = 0 \text{ because } E(\epsilon_t | Z_t) = 0. \text{ When null hypothesis is false, the power of this test depends on the correlation between } Z_t \text{ and } \Delta X_t = X_t - X_{t-1}. \]

The second test statistic we propose is based on nonparametric estimation of \[I_b = E[||\alpha(Z_t)||^2] = E[\alpha(Z_t)^T \alpha(Z_t)]. \] Obviously, \[I_b = 0 \text{ under } H_0 \text{ because } \alpha(z) = \theta(z) - c_0 = \theta_0 - \theta_0 = 0 \text{ under } H_0. \] When null hypothesis is false, \[\theta(z) \neq \theta_0 \text{ for any constant vector } \theta_0 \text{ on a set with positive } (z) \text{ measure, which is equivalent to } \alpha(z) \neq 0 \text{ on a set with positive measure}. \] Hence, \[I_b > 0 \text{ is a positive constant when } H_0 \text{ is false}. \] Our second test statistic will be based on sample analogue of \[I_b: \hat{I}_b = n^{-1} \sum_{t=1}^{n} \hat{\alpha}(Z_t)^T \hat{\alpha}(Z_t), \text{ where } \hat{\alpha}(Z_t) \text{ is the nonparametric estimator of } \alpha(Z_t) \text{ proposed by Sun et al. (2010)}. \] Sun et al. show that \[\hat{\alpha}(z) - \alpha(z) = O_p(h^2 + (nh)^{-1/2}). \] Now under \[H_0, \alpha(z) = 0 \text{ and it is easy to show that the estimation bias term } h^2 \text{ which related to derivative function of } \theta(z) \text{ also disappear under } H_0. \] Hence, we have \[\hat{\alpha}(z) = O_p((nh)^{-1/2}). \] Naturally, we expect that \[\hat{I}_b \text{ converges to } 0 \text{ at the rate of } O_p((nh)^{-1}). \] We will rescale of test statistic by \[nh, \text{ i.e., define } \hat{J}_b = (nh)\hat{I}_b \text{ so that } \hat{J}_b = O_p(1) \text{ under } H_0. \] When the null hypothesis is false, \[\hat{J}_b \text{ diverges to } +\infty \text{ at the rate of } nh \text{ because } \hat{I}_b \text{ converges to a positive constant}. \] Hence, \[\hat{J}_b \text{ test is a consistent test provided that one can find some (bootstrap) procedures to approximate the upper percentile of the null distribution of } \hat{J}_b. \]

2.2 The First Test Statistic

We consider the following semiparametric varying coefficient model

\[Y_t = X_t^T \theta(Z_t) + u_t, \tag{2.3}\]

where \(X_t \text{ (of dimension } d \times 1) \text{ and } u_t \text{ (a scalar) are all non-stationary I(1) processes, } Z_t \text{ is a scalar stationary process, and } \theta(\cdot) \text{ is a smooth but otherwise unspecified function.}\]
Note that $X_t$ and $u_t$ are both I(1) processes. Specifically, we assume that

$$X_t = X_{t-1} + \eta_t,$$

with $X_0 = 0$ and $\eta_t$ is a stationary process. Also,

$$u_t = u_{t-1} + \epsilon_t,$$

with $u_0 = 0$ and $\epsilon_t$ is a stationary process. We assume that both $\eta_t$ and $\epsilon_t$ are $\beta$-mixing processes, satisfying the same regularity conditions as listed in Fan and Li (1999).

Under the null hypothesis of a linear regression model, we have

$$Y_t = X^T_t \theta_0 + u_t,$$

(2.4)

where $\theta_0$ is a $d \times 1$ vector of constant coefficients, and as above $X_t$ and $u_t$ are I(1) variables. In most spurious regression analyses, $\theta_0$ is usually assumed to be a zero vector. However, in practice $\theta_0$ may not be necessarily zero. If $\theta_0 = 0$, then $Y_t = u_t$ and $Y_t$ and $X_t$ are uncorrelated. If $\theta_0 \neq 0$, then $Y_t$ is correlated with $X_t$ even though the ordinary least squares estimator (OLS) based on level data will not lead to a consistent estimator of $\theta_0$.

Taking a first difference of the data we obtain a linear model with all the variables being I(0):

$$\Delta Y_t = \Delta X^T_t \theta_0 + \epsilon_t,$$

where $\Delta Y_t = Y_t - Y_{t-1}$ and $\Delta X_t = X_t - X_{t-1}$. The OLS estimator of $\theta_0$ based on the first difference data is $\sqrt{n}$-consistent under quite general conditions,

$$\hat{\theta}_0 - \theta = O_p(n^{-1/2}).$$

We estimate $\epsilon_t$ by the least squares residuals: $\hat{\epsilon}_t = \Delta Y_t - \Delta X^T_t \hat{\theta}_0 = \epsilon_t - \eta^T_t (\hat{\theta}_0 - \theta_0)$, where $\eta_t = \Delta X_t$. Our first test statistic is based on the following sample analogue of $I_a$

$$\hat{I}_a = \frac{1}{n(n-1)} \sum_{t=1}^{n} \sum_{s \neq t} \hat{\epsilon}_t \hat{\epsilon}_s K_{h,ts},$$

where $K_{h,ts} = h^{-1} K((z_t - z_s)/h)$.

One advantage of the $\hat{I}_a$ test is that it is computationally simple. It only involves a kernel weighted double sum of the least squares residuals. In particular it does not require nonparametric estimation of $\alpha(\cdot)$. 7
Under $H_0$, we have \( \hat{\epsilon}_t = \epsilon_t - \eta_t^T (\bar{\theta}_0 - \theta_0) \), where \( \eta_t = \Delta X_t \). Hence,

\[
\hat{I}_a = \frac{1}{n(n-1)} \sum_{t=1}^{n} n \sum_{s \neq t} \epsilon_t \epsilon_s K_{h,ts} - (\bar{\theta}_0 - \theta_0)^T \frac{2}{n(n-1)} \sum_{t=1}^{n} \sum_{s \neq t} \eta_t \epsilon_s K_{h,ts} + (\bar{\theta}_0 - \theta_0)^T \frac{1}{n(n-1)} \sum_{t=1}^{n} \sum_{s \neq t} \eta_s K_{h,ts} \eta_t^T (\bar{\theta}_0 - \theta_0) = I_{1n} - 2(\bar{\theta}_0 - \theta_0)^T I_{2n} + (\bar{\theta}_0 - \theta_0)^T I_{3n}(\bar{\theta}_0 - \theta_0),
\]

(2.5)

where the definitions of \( I_{1n}, I_{2n} \) and \( I_{3n} \) should be obvious.

Under $H_0$, \( n\sqrt{h}I_n \overset{d}{\rightarrow} N(0, \sigma_b^2) \) follows from Theorem 3.1 of Fan and Li (1999), where \( \sigma_b^2 = 2\sigma_b^2 \nu_0 E[f(Z_t)], \nu_0 = \int K(v)^2 dv \) and \( f(\cdot) \) is the probability density function of \( Z_t \). A consistent estimator of \( \sigma_b^2 \) is given by \( \hat{\sigma}_b^2 = \frac{1}{n(n-1)h} \sum_{t=1}^{n} \sum_{s \neq t} \epsilon_t^2 \epsilon_s^2 K_{ts}^2, K_{ts} = K(\frac{X_t - X_s}{h})^2 \). It is easy to show that \( I_{2n} = O_p(n^{-1/2}) \) and \( I_{3n} = O_p(1) \). These together with \( \bar{\theta}_0 - \theta_0 = O_p(n^{-1/2}) \) imply that the last two terms at the right hand side of (2.5) are \( O_p(n^{-1}) \). Hence, the leading term of \( \hat{I}_a \) is given by \( I_{1n} \) (under $H_0$). Summarizing the above we have the following result.

**Proposition 2.1** Under $H_0$ and regularity conditions on the stationary variables \( \eta_t, Z_t \) and \( \epsilon_t \) similar to those given in Fan and Li (1999), we have

\[
n\sqrt{h}\hat{I}_a / \sqrt{\hat{\sigma}_b^2} \overset{d}{\rightarrow} N(0, 1).
\]

It is now well known that nonparametric kernel based tests often suffer substantial finite sample size distortion. Therefore, we will use the following bootstrap procedure to approximate the finite null distribution of \( \hat{I}_a \).

**Bootstrap steps for the \( \hat{I}_a \) test**

1. First obtain the least squares estimator \( \hat{\theta}_0 = (\eta'\eta)^{-1} \eta' \Delta Y \) based on the first difference data \( \eta = \Delta X_t \), and compute \( \hat{\epsilon}_t = \Delta Y_t - \Delta X_t^T \hat{\theta}_0, t = 1, \ldots, n \). Then generate the two point wild bootstrap error \( \epsilon^*_t \) by \( \epsilon^*_t = a \hat{\epsilon}_t \) with probability \( r \), and \( \epsilon^*_t = b \hat{\epsilon}_t \) with probability \( 1 - r \) with \( a = (1 - \sqrt{5})/2, b = (1 + \sqrt{5})/2 \) and \( r = (1 + \sqrt{5})/(2\sqrt{5}) \).

2. Generate the bootstrap sample \( \Delta Y^*_t = \Delta X_t^T \hat{\theta}_0 + \epsilon^*_t \). Using the bootstrap sample \( \{\Delta Y^*_t, \Delta X_t\}_{t=1}^{n} \) to compute \( \hat{\theta}_0^* \) (by OLS), then obtain bootstrap residual \( \hat{\epsilon}^*_t = \Delta Y_t^* - \Delta X_t^T \hat{\theta}_0^* \), and compute the bootstrap statistic \( \hat{J}_a^* = (n\sqrt{h})(n^2h)^{-1} \sum_{t=1}^{n} \sum_{s \neq t} \epsilon^*_t \epsilon^*_s K_{ts} / \sqrt{\hat{\sigma}^2_a} \), where \( \hat{\sigma}^2_a = (n^2h)^{-1} \sum_{t=1}^{n} \sum_{s \neq t} \epsilon^*_t \epsilon^*_s K_{st}^2 \) with \( K_{st}^2 = K(\frac{X_s - X_t}{h})^2 \).
3. Repeat steps 1 and 2 a large number of times, say $B$ times, obtain the $\alpha$ upper percentile (say, $\hat{J}^*_a,\alpha$) of the empirical distribution of the $B$ bootstrap statistics $\hat{J}^*_a$. Reject $H_0$ if $\hat{J}_a > \hat{J}^*_a,\alpha$, and do not reject $H_0$ otherwise.

Under $H_1$, we have $\hat{\epsilon}_t = \Delta Y_t - \eta_t \hat{\theta}_0 = \epsilon_t + X_t^T \theta(Z_t) - X_{t-1}^T \theta(Z_{t-1}) - \eta_t \hat{\theta}_0$. Sun et al. (2010) show that $\hat{\theta}_0 \overset{d}{\rightarrow} c_0 + \bar{\theta}$, where $c_0 = E[\theta(Z_t)]$ and $\bar{\theta}$ is a $O_p(1)$ random variable related to Brownian motions generated from the innovations $\eta_t = \Delta X_t$ and $\epsilon_t = u_t - u_{t-1}$. The power of the $\hat{I}_a$ test depends on $E[X_t \theta(Z_t) - X_{t-1} \theta(Z_{t-1}) - \eta_t^T (c_0 + \bar{\theta}) | Z_t] \neq 0$. While we cannot construct a simple case that $E[X_t \theta(Z_t) - X_{t-1} \theta(Z_{t-1}) - \eta_t^T (c_0 + \bar{\theta}) | Z_t] = 0$, we cannot prove that this conditional expectation is always non-zero either. Therefore, there might be cases that this conditional expectation is zero and therefore the $\hat{I}_a$ test may have only trivial power under such cases.

In the next subsection we construct our second test statistic which has power in all direction of deviations from the null model.

### 2.3 The Second Test Statistic

Different from the way we construct the first test statistic, the second test statistic is based on estimation of the varying coefficient function $\hat{\alpha}(\cdot)$. We now describe the estimation procedure for the smooth coefficient function $\theta(\cdot)$ defined via the following varying coefficient model:

$$Y_t = X_t^T \theta(Z_t) + u_t,$$

where as before $Y_t$, $X_t$ and $u_t$ are all I(1) variables, $Z_t$ is an I(0) variable. For the same reason as in the linear model case, a standard semiparametric estimator of $\theta(z)$ will not be a consistent estimator due to the I(1) error $u_t$.

In the linear regression model case, one can estimate the coefficient parameter by estimating a first difference equation. Now taking a first difference of (2.6) leads to

$$\Delta Y_t = X_t^T \theta(Z_t) - X_{t-1}^T \theta(Z_{t-1}) + \epsilon_t.$$  

The error $\epsilon_t$ in (2.7) is an I(0) process. Moreover, equation (2.7) is an additive model with the two coefficient functions $\theta(Z_t)$ and $-\theta(Z_{t-1})$, respectively. One may be tempted to think that $\theta(\cdot)$ can be consistently estimated based on (2.7), say, using some nonparametric
(backfitting) estimation methods. Unfortunately, the two I(1) regressors $X_t$ and $X_{t-1}$ are asymptotically perfectly collinear (because the correlation coefficient between $X_t$ and $X_{t-1}$ approaches one as $t \to \infty$). Therefore, (2.7) cannot lead to consistent estimate of $\theta(z)$ due to the collinearity problem. Sun et al. (2010) show that one can obtain a consistent estimate for $\alpha(z) = \theta(z) - E[\theta(Z_t)]$ by a two step estimation method. Let $c_0 = E[\theta(Z_t)]$ and $\alpha(Z_t) = \theta(Z_t) - c_0$. We re-write (2.6) as

$$Y_t = X_T^T c_0 + X_T^T \alpha(Z_t) + u_t.$$  \hspace{1cm} (2.8)

By construction $\alpha(Z_t)$ has a zero mean. Sun et al. show that one can consistently estimate $\alpha(z)$ by a de-meaned semiparametric estimator. A standard local constant estimator of $\theta(z)$ is given by

$$\hat{\theta}(z) = \left[ \sum_{t=1}^{n} X_t X_t^T K_{tz} \right]^{-1} \sum_{t=1}^{n} X_t Y_t K_{tz},$$  \hspace{1cm} (2.9)

where $K_{tz} = K((Z_t - z)/h)$, $K(.)$ is the kernel function and $h$ is the smoothing parameter.

$\hat{\theta}(z)$ is not a consistent estimator for $\theta(z)$ due to the I(1) error term $u_t$. This is similar to the OLS estimator in the linear regression model case. Under quite general regularity conditions, it is well established that

$$\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{[n]} \eta_i \right) \overset{d}{\to} \left( W_\eta(.) \right),$$

$$\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{[n]} \epsilon_i \right) \overset{d}{\to} \left( W_\epsilon(.) \right),$$

where $[a]$ denotes the largest integer that is less than or equal to $a$, $W_\eta(.)$ is a $d \times 1$ vector of Brownian motion with variance given by $\text{Var}(\eta_t)$ and $W_\epsilon$ is a scalar Brownian motion with variance given by $\sigma^2_\epsilon$.

Sun et al. establish the following result:

$$\hat{\theta}(z) - \theta(z) \overset{d}{\to} \left[ \int_0^1 W_\eta(r) W_\eta(r)^T dr \right]^{-1} \int_0^1 W_\eta(r) W_\epsilon(r) dr \equiv \bar{\theta}.$$  \hspace{1cm} (2.10)

The source of inconsistency comes from the I(1) error term. Sun et al. have shown that, by subtracting the sample mean from $\hat{\theta}(z)$, one obtains a consistent estimator for $\alpha(z)$, i.e., one can estimate $\alpha(.)$ by de-meaning the semiparametric estimator defined by

$$\hat{\alpha}(z) = \hat{\theta}(z) - \frac{1}{n} \sum_{t=1}^{n} \hat{\theta}(Z_t).$$  \hspace{1cm} (2.11)
Hence, our second test statistic is given by

\[ \hat{J}_b = nh \hat{I}_b = nh \left[ \frac{1}{n} \sum_{t=1}^{n} \hat{\alpha}(Z_t)^T \hat{\alpha}(Z_t) \right]. \]

The asymptotic distribution of \( \hat{\alpha}(\cdot) \) is established in Sun et al. (2010), we summarize it in a proposition below.

**Proposition 2.2** Under Assumptions A1-A6 in Sun et al. (2011), we have

\[
\sqrt{nh} [\hat{\alpha}(z) - \alpha(z) - h^2(B(z) - E(B(Z_t)))] \xrightarrow{d} \frac{\sqrt{\nu_2}}{f(z)} \left[ \int_0^1 W_n(r)W_n(r)^T dr \right]^{-1} \int_0^1 W_n(r)W_r(r)dW_w(r),
\]

where \( \nu_2 = \int K(v)2v^2dv \), \( W_w(\cdot) \) is a standard Brownian motion Brownian motion. \( h^2\{B(z) - E[B(Z_t)]\} \) is the leading bias term which related to derivative functions of \( \theta(\cdot) \).

Note that under \( H_0 \), \( \theta(z) = \theta_0 \), \( \alpha(z) = 0 \) and the bias term is identically zero because \( \frac{d}{dz} \theta(z) = \frac{d}{dz} \theta_0 = 0 \) under \( H_0 \). With \( B(z) = 0 \) we obtain \( \hat{\alpha}(z) = O_p((nh)^{-1/2}) \) and \( \hat{\alpha}(z)^T \hat{\alpha}(z) = O_p((nh)^{-1}) \). We conjecture that \( n^{-1} \sum_{t=1}^{n} \hat{\alpha}(Z_t)^T \hat{\alpha}(Z_t) = O_p(1) \) under \( H_0 \). Hence, we normalize it by \( nh \) and this leads to \( \hat{J}_b = (nh) \left[n^{-1} \sum_{t=1}^{n} \hat{\alpha}(Z_t)^T \hat{\alpha}(Z_t)\right] = O_p(1) \) under \( H_0 \). Indeed our simulations show that under \( H_0 \), \( \hat{J}_b \) seems to have a stable distribution. In particular, its mean and variance do not vary much for different values of \( n \), supporting our conjecture that \( \hat{J}_b = O_p(1) \) under \( H_0 \).

When the null hypothesis is false, \( \hat{\alpha}(z)^T \hat{\alpha}(z) = \alpha(z)^T \alpha(z) + o_p(1) = O_p(1) \). Here the \( O_p(1) \) is an exact order of \( O_p(1) \), i.e., it is not \( o_p(1) \). Hence, the test statistic \( \hat{J}_b \) diverges to \( +\infty \) at the rate of \( nh \). Thus, we expect \( \hat{J}_b \) to be a consistent test.

Since we do not know the asymptotic distribution of the \( \hat{J}_b \) test, we will use some bootstrap procedures to approximate the unknown null distribution of \( \hat{J}_b \). In order to generate bootstrap errors, we need, among other things, to estimate \( \sigma^2_\epsilon \), which in turn requires that we obtain residual estimates \( \hat{\epsilon}_t \). From \( u_t = Y_t - X_t^T \theta(Z_t) = Y_t - X_t^T(\alpha(Z_t) + c_0) \), naturally, we will estimate \( u_t \) by \( \hat{\epsilon}_t = \hat{u}_t - \hat{u}_{t-1} \) and \( \hat{\sigma}^2_\epsilon = n^{-1} \sum_{t=1}^{n} \epsilon_t^2 \). However, simulation results show that this method often overestimate \( \sigma^2_\epsilon \) and this leads to the test under sized under \( H_0 \). The reason for the overestimate of \( \sigma^2_\epsilon \) is that \( c_0 \) is difficult to estimate, especially when \( \eta_h \) and \( \epsilon_t \) are correlated. We list the specific bootstrap procedure below.
Bootstrap steps for the $\hat{J}_b$ test

Step (i) First obtain the nonparametric estimator $\hat{\alpha}(Z_t)$ as discussed earlier for $t = 1, ..., n$. Define $\tilde{Y}_t = Y_t - X_t^T \hat{\alpha}(Z_t)$, then we estimate $c_0$ by regressing $\Delta \tilde{Y}_t$ on $\Delta X_t = \eta_t$, i.e.,

$$\hat{c}_0 = \left[ \sum_{t=1}^{n} \eta_t \eta_t^T \right]^{-1} \sum_{t=1}^{n} \eta_t \Delta \tilde{Y}_t,$$

where $\Delta \tilde{Y}_t = Y_t - X_t^T \hat{\alpha}(Z_t) - [Y_{t-1} + X_{t-1}^T \hat{\alpha}(Z_{t-1})]$. This gives us an estimator for $\theta(Z_t)$ given by $\hat{\theta}(Z_t) = \hat{\alpha}(Z_t) + \hat{c}_0$. We then obtain $\hat{u}_t = Y_t - X_t^T \hat{\theta}(Z_t)$ for $t = 1, ..., n$, $\hat{\epsilon}_t = \hat{u}_t - \hat{u}_{t-1}$ for $t = 2, ..., n$, and $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{t=2}^{n} \hat{\epsilon}_t^2$.

Step (ii) Generate iid $\epsilon_i^*$ from $N(0, \hat{\sigma}_\epsilon)$, and $u_i^* = \sum_{s=1}^{t} \epsilon_s^*$ for $t = 1, ..., n$. Compute $Y_i^* = X_i^T \hat{\theta}(Z_i) + u_i^*$. Using the bootstrap sample $\{X_t, Z_t, Y_t^*\}_{t=1}^{n}$ to compute the bootstrap statistic $\hat{J}_b^* = (nh) \left[ n^{-1} \sum_{t=1}^{n} \hat{\alpha}^*(Z_t)^T \hat{\alpha}^*(Z_t) \right]$, where $\hat{\alpha}^*(Z_t)$ is the same as $\hat{\alpha}(Z_t)$ except that $Y$ is replaced by $Y^*$.

Step (iii) Repeat steps (i) and (ii) a large number of times, say $B$ times, and use upper $a$-percentile $\hat{J}_{b,(a)}^*$ (of $\{\hat{J}_b^*\}_{t=1}^{B}$) to approximate the upper $a$-percentile critical value of the null distribution of $\hat{J}_b$. We reject $H_0$ if $\hat{J}_b \geq \hat{J}_{b,(a)}^*$, and we do not reject $H_0$ otherwise.

Our simulation results reported in the next section show that the $\hat{J}_b$ test is undersized. Moreover, as sample size $n$ increases, it gets more undersized. Although an undersized test has small typo I error, it may hurt the power of the test.

One can also use some sub-sampling method to approximate the null distribution of the test statistic $\hat{J}_b$. Specifically, we select a sub-sample of size $[n^\alpha]$ for some $0 < \alpha < 1$ to compute the test statistic; call it $\hat{J}_{b, sub}$. There are $n - [n^\alpha]$ such samples, we use the resulting empirical distribution of $\hat{J}_{b, sub}$ to approximate the null distribution of the statistic $\hat{J}_a$. As will be shown in the next section, the sub-sampling method can solve the undersize problem of the original $\hat{J}_b$ based on the full-sample re-sampling bootstrap method. Although the sub-sampling method can resolve the undersize problem (as it does not require the estimation of $c_0$), the power of the $\hat{J}_{b, sub}$ is hurt by the fact that the upper percentile values of the $\hat{J}_{b, sub}$ also diverges to $\infty$ as $n$ increases, although at a slower rate than the original statistic $\hat{J}_b$ test. So the test is still a consistent test, but the finite sample power of the test suffers from the fact that the null hypothesis is not imposed at the sub-sampling process. We describe the subsampling procedure below.

12
Sub-sampling steps for the $\hat{J}_b$ test

Step (i) First chose the block size $b = [n^a]$. Pick up the blocks \{(Y_t, X_t, Z_t)\}_{t=s-b}^{s+b-1}$, $s = 1, \ldots, n - b + 1$ sequentially. Use these subsamples to estimate the model and get the estimators $\hat{\alpha}_s(z)$ for $s = 1, \ldots, n - b + 1$. Use these estimators to compute the corresponding sub-sampling statistic $\hat{J}_{b,s} = (bh_b) \left[ b^{-1} \sum_{t=s}^{s+b-1} \hat{\alpha}_s(Z_t)^T \hat{\alpha}_s(Z_t) \right]$, $s = 1, \ldots, n - b + 1$.

Step (ii) Sort $\{\hat{J}_{b,s}\}_{s=1}^{n-b+1}$ and use upper a-percentile $\hat{J}_{b,(a)}$ (of $\{\hat{J}_{b,s}\}_{s=1}^{n-b+1}$) to approximate the upper a-percentile critical value of the null distribution of $\hat{J}_b$.

Step (iii) Estimate the model by full sample and calculate the test statistic $\hat{J}_b$. We reject $H_0$ if $\hat{J}_b \geq \hat{J}_{b,(a)}$, and we do not reject $H_0$ otherwise.

We use simulations to examine the finite sample performances of the test statistics $\hat{J}_a$, $\hat{J}_b$ and $\hat{J}_{b,sub}$ in the next section.

3 Monte Carlo Simulations

In this section we examine the finite sample performances of our proposed test statistics $\hat{J}_a$, $\hat{J}_b$ and $\hat{J}_{b,sub}$. We consider the following data generating process (DGP):

$$Y_t = X_t \theta(Z_t) + u_t, \quad (t = 1, \ldots, n),$$

where $X_t$ and $u_t$ are both I(1) variables and $Z_t$ is a stationary covariate. Specifically, $X_t = X_{t-1} + \eta_t$ with $X_0 = 0$ and $\eta_t$ is i.i.d. $N(0, \sigma_\eta^2)$; $u_t = u_{t-1} + \epsilon_t$ with $u_0 = 0$ and $\epsilon_t$ is i.i.d. $N(0, \sigma_\epsilon^2)$; $z_t = v_t + v_{t-1} + \eta_t$ with $v_t$ is i.i.d. uniform $[0, 2]$. We choose $\sigma_\epsilon = 1$, $\sigma_\eta = 2$ or 3.

When $H_0$ is true, we select $\theta(z) = \theta_0 = 1$. When $H_0$ is false, we choose $\theta(z) = \theta_1(z)$ or $\theta(z) = \theta_2(z)$, where $\theta_1(z) = z - 0.5z^2$ and $\theta_2(z) = 1/(1 + e^{-z})$. In both cases, we have $\alpha(z) = \theta(z) - c_0$ with $c_0 = E[\theta(Z_t)]$. We choose two different combinations for $(\sigma_\epsilon, \sigma_\eta) = (1, 2)$, $(1, 3)$.

The case of $(1, 3)$ means that the relative variance of $X_t$ over the variance of $u_t$ gets larger (smaller noise to signal ratio) compared to the case of $(\sigma_\epsilon, \sigma_\eta) = (1, 2)$. The sample sizes are $n = 50, 100$ and $200$. The number of replications is 1,000, and within each replication, 400 bootstrap statistics are generated to yield 1%, 5%, 10% and 20% upper percentile values of the bootstrap statistics. The smoothing parameter is selected via $h = \hat{\sigma}_z n^{-1/5}$, where $\hat{\sigma}_z$ is the sample standard deviation of $\{Z_t\}_{t=1}^n$. 


The estimated sizes of the $\hat{J}_a$ is given in Table 1 and the estimated powers of the $\hat{J}_a$ test are given in Tables 2 and 3.

From Table 1 we observe that the estimated sizes are quite close to their nominal sizes for all cases considered. From Tables 2 and 3 we observe that the power of the test improves as sample size $n$ increases, and the power also increases when signal-to-noise ratio increases ($\sigma_{\eta}/\sigma_\epsilon$ increases) for a fixed value of $n$. Finally, we observe that the $\hat{J}_a$ test is more powerful against DGP2 than against DGP1.

Next, the estimated sizes and powers of the $\hat{J}_b$ test are given in Tables 4 to 6. From Table 4 we observe the $\hat{J}_b$ is undersized, moreover, the size distortion gets more severe as sample size increases. The reason for the size distortion is that $c_0$ is not accurately estimated. Consequently, $\sigma_\epsilon^2$ is over estimated and this makes the bootstrap statistic $\hat{J}_b^*$ larger than the original statistic $\hat{J}_b$, resulting in a undersized result. Even $\hat{J}_b$ test is undersized, Tables 2 and 3 show that the $\hat{J}_b$ test is very powerful against both DGP1 and DGP2. In particular, the $\hat{J}_b$ test is more powerful than the $\hat{J}_a$ test even without size adjustment. Certainly, the size adjusted power advantage of $\hat{J}_b$ is even more substantial over that of the $\hat{J}_a$ test.

Finally, we examine the result of the sub-sampling method based test $\hat{J}_{b,\text{sub}}$. We choose $\alpha = 1/2$, i.e., the sub-sample size is $\lceil n^{1/2} \rceil$. From table 7 we observe that it has good estimated sizes. The reason that the $\hat{J}_{b,\text{sub}}$ does not suffer the substantial size distortion of the $\hat{J}_b$ test is that there is no need to estimate $c_0$ when computing the sub-sampling statistic $\hat{J}_{b,\text{sub}}$.

From Tables 8 and 9, we see that for DGP1, the $\hat{J}_{b,\text{sub}}$ test is not as powerful as the $\hat{J}_a$ or the $\hat{J}_b$ tests. For DGP2 the $\hat{J}_{b,\text{sub}}$ test is more powerful than the $\hat{J}_a$ test, but less powerful than the $\hat{J}_b$ test. The reason that the $\hat{J}_{b,\text{sub}}$ test is less powerful than the $\hat{J}_b$ test is that, when the null hypothesis is false, like the $\hat{J}_b$ test, the $\hat{J}_{b,\text{sub}}$ statistic also diverges to $\infty$ as $n$ increases. The consistency testing property of the $\hat{J}_{n,\text{sub}}$ test is that it uses a sub-sample size much smaller than the full sample size. Hence, $\hat{J}_{b,\text{sub}}$ diverges to $\infty$ much slower than that of the $\hat{J}_b$ test.

Summarizing the above we observe that the $\hat{J}_b$ has both smaller type I error (due to its undersize) and smaller type II error (it has better power) than both the $\hat{J}_a$ test and the sub-sampling based $\hat{J}_{b,\text{sub}}$. Therefore, we recommend the use of the $\hat{J}_b$ test in practice.
References


Table 1: Estimated Size of the $\hat{J}_a$ test ($\theta(z) = \theta_0 = 1$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma_\eta = 2$</th>
<th>$\sigma_\eta = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>50</td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>100</td>
<td>.008</td>
<td>.048</td>
</tr>
<tr>
<td>200</td>
<td>.014</td>
<td>.049</td>
</tr>
</tbody>
</table>

Table 2: Estimated Power of the $\hat{J}_a$ test ($\theta(z) = z^2 - 0.5z^2$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma_\eta = 2$</th>
<th>$\sigma_\eta = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>50</td>
<td>.148</td>
<td>.362</td>
</tr>
<tr>
<td>100</td>
<td>.474</td>
<td>.633</td>
</tr>
<tr>
<td>200</td>
<td>.704</td>
<td>.794</td>
</tr>
</tbody>
</table>

Table 3: Estimated Power of the $\hat{J}_a$ test ($\theta(z) = 1/(1 + e^{-z})$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma_\eta = 2$</th>
<th>$\sigma_\eta = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>50</td>
<td>.004</td>
<td>.030</td>
</tr>
<tr>
<td>100</td>
<td>.030</td>
<td>.102</td>
</tr>
<tr>
<td>200</td>
<td>.102</td>
<td>.254</td>
</tr>
</tbody>
</table>

Table 4: Estimated Size of the $\hat{J}_b$ test ($\theta(z) = \theta_0 = 1$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma_\eta = 2$</th>
<th>$\sigma_\eta = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>50</td>
<td>.001</td>
<td>.004</td>
</tr>
<tr>
<td>100</td>
<td>.000</td>
<td>.002</td>
</tr>
<tr>
<td>200</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 5: Estimated Power of the $\hat{J}_b$ test ($\theta(z) = z - 0.5z^2$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma_\eta = 2$</th>
<th>$\sigma_\eta = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>50</td>
<td>.788</td>
<td>.858</td>
</tr>
<tr>
<td>100</td>
<td>.835</td>
<td>.924</td>
</tr>
<tr>
<td>200</td>
<td>.894</td>
<td>.958</td>
</tr>
</tbody>
</table>
Table 6: Estimated Power of the $\hat{J}_b$ test ($\theta(z) = 1/[1 + e^{-z}]$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma_\eta = 2$</th>
<th>$\sigma_\eta = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>50</td>
<td>.466</td>
<td>.582</td>
</tr>
<tr>
<td>100</td>
<td>.592</td>
<td>.682</td>
</tr>
<tr>
<td>200</td>
<td>.662</td>
<td>.738</td>
</tr>
</tbody>
</table>

Table 7: Estimated Size of the $\hat{J}_c$ test ($\theta(z) = \theta_0 = 1$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\gamma = 0$</th>
<th>$\gamma = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>50</td>
<td>.026</td>
<td>.050</td>
</tr>
<tr>
<td>100</td>
<td>.010</td>
<td>.030</td>
</tr>
<tr>
<td>200</td>
<td>.008</td>
<td>.038</td>
</tr>
</tbody>
</table>

Table 8: Estimated Size of the $\hat{J}_c$ test ($\theta(z) = z - .5z^2$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\gamma = 0$</th>
<th>$\gamma = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>50</td>
<td>.140</td>
<td>.222</td>
</tr>
<tr>
<td>100</td>
<td>.190</td>
<td>.304</td>
</tr>
<tr>
<td>200</td>
<td>.216</td>
<td>.354</td>
</tr>
</tbody>
</table>

Table 9: Estimated Power of the $\hat{J}_c$ test ($\theta(z) = 1/[1 + e^{-z}]$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\gamma = 0$</th>
<th>$\gamma = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>50</td>
<td>.160</td>
<td>.224</td>
</tr>
<tr>
<td>100</td>
<td>.196</td>
<td>.304</td>
</tr>
<tr>
<td>200</td>
<td>.248</td>
<td>.398</td>
</tr>
</tbody>
</table>