Journal of Econometrics 178 (2014) 80–85

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

# Model specification test with correlated but not cointegrated variables

Li Gan[a,*], Cheng Hsiao[b], Shu Xu[c]

[a] Department of Economics, Texas A&M University, College Station, TX 77843-4228, United States

[b] Department of Economics, University of Southern California, Los Angeles, CA 90089-0253, United States

[c] School of Economics, Southern University of Finance & Economics, Chengdu, China

ARTICLE INFO

*Article history:*
Available online 12 August 2013

*JEL classification:*
C13
C14
C22

*Keywords:*
Specification test
Spurious regression
Varying coefficient
Kernel estimation

ABSTRACT

Many macroeconomic and financial variables show highly persistent and correlated patterns but are not necessarily cointegrated. Recently, Sun et al. (2011) 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.



## 1. Introduction

In the study of finance, economists are interested in the co-movements of financial variables among different markets and countries, see Hamao et al. (1990), Koch and Koch (1991), Koutmos and Booth (1995), etc. Many 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 et al. (1988); Bollerslev (1990); Engle and Kroner (1995); Engle (2002). However, some financial variables like market volatilities show highly-persistent patterns, which lead researchers to consider nonstationarity such as $I(1)$ processes, see Harvey et al. (1994), Hong (2001) and references therein. Also, researchers want to capture the time changing features of the relationship among different variables, as in King et al. (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 checks. 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 are 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 the unit root process. Recently, the estimation of nonparametric and semiparametric cointegration models has attracted much attention among econometricians and statisticians. Wang and

* Corresponding author.
*E-mail address:* gan@econmail.tamu.edu (L. Gan).

0304-4076/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.jeconom.2013.08.008

Phillips (2009a,b) considered nonparametric cointegrations. Cai et al. (2009), and Xiao (2009) considered varying coefficient cointegrations with a stationary varying coefficient covariate. Sun et al. (2010) studied the varying coefficient cointegration with an $I(1)$ varying coefficient covariate.[1] Karlsen et al. (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 market 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 proposes 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 and 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:

$$Y_t = X_t^T\theta(Z_t) + u_t, \quad (t = 1, \dots, n) \tag{2.1}$$

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 $\check{\theta}(z)$ 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, $\check{\theta}(z)$ is not a consistent estimator of $\theta(z)$. Sun et al. show that $\check{\theta}(z)$ converges to $\theta(z) + \bar{\theta}$, where $\bar{\theta}$ 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 $n^{-1}\sum_{t=1}^n \check{\theta}(Z_t)$ converges to $c_0 + \bar{\theta}$. Hence, one can consistently estimate $\alpha(z) = \theta(z) - c_0$ by $\hat{\alpha}(z) = \check{\theta}(z) - n^{-1}\sum_{t=1}^n \check{\theta}(Z_t)$. 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 difficult 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, the 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

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

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 an

[1] Varying coefficient models were investigated by Robinson (1991), Chen and Tsay (1993). There is a rapidly growing literature on it, including Fan et al. (2000), Cai et al. (2000), Li et al. (2002), and Fan et al. (2003), among others.