Detection of Functional Form Misspecification in Cointegrating Relations

Ioannis Kasparis
University of Nottingham

18 December 2004

Abstract

A simple specification test based on fully modified residuals and the CUSUM test for cointegration of Xiao and Phillips (2002) are considered as means of testing for functional form in long-run cointegrating relations. It is shown that both tests are consistent under functional form misspecification and lack of cointegration. An extensive simulation study is carried out to asess the properties of the tests in finite samples. The Dickey-Fuller test is also considered. The simulation results reveal that the first two tests perform well. On the other hand the Dickey-Fuller test performs poorly in many cases, not only when there is functional form misspecification, but also when there is lack of cointegration and the spurious regression is nonlinear.

1 Introduction

The theory of cointegration was developed nearly fifteen years ago. Cointegration has probably been, the most popular approach in modeling macroeconomic relations since it was introduced. Although the concept of cointegration is very appealing from an economic theory point of view, a lot of data sets have failed to show evidence supporting the existence of long-run macroeconomic equilibria. Lack of evidence for cointegration in certain data sets has created doubts about the validity of the classical linear cointegration models and led some researchers to consider the possibility of nonlinearities in macroeconomic relations. Recent work e.g. Corradi, Swanson and White (2000), Teräsvirta and Ellianson (2001) among others, consider the case of nonlinear short-run dynamics in Vector Error Correction Models (VECM). Nonetheless the possibility of nonlinear long-run dynamics has been largely ignored. Park and Phillips (1999, 2001) develop limit distribution theory for nonlinear transformations of unit root processes and also provide an approach for modeling nonlinear long-run relations (see also Chang, Park and Phillips (2001)). In their recent work Saikkonen and Choi (2004) follow the Park and Phillips (1999, 2001) exposition to model smooth transitions in long-run cointegrating relations.

Cointegration is generally understood as a property between a set of unit root processes. The nonlinear cointegration models of Park and Phillips (1999, 2001) depart from this approach. Although unit root processes play a central role in these models, nonlinear cointegration is not necessarily a property between unit root processes. Park and Phillips (1999, 2001) consider models the form:

$$y_t = f(x_t) + u_t,$$

where x_t is a unit root process and u_t is some stationary error term. Clearly if $f(.)$ is linear, the variable y_t will be a unit root process as well. Nonetheless for nonlinear $f(.)$, y_t will not be of the ARIMA type. After all unit root processes may not be the only type of relevant nonstationary variables. Having departed from the classical concept of cointegration, lack of cointegration should be reconsidered as well. Absence of cointegration in this nonlinear framework may arise because a) the function $f(.)$ has not been correctly specified i.e. because functional form misspecification has been committed, b) because y_t and x_t do not cointegrate for any choice of $f(.)$. The first case will be referred to as Functional Form (FF hereafter) misspecification and to the second case as lack of cointegration. This distinction is made for technical reasons. FF misspecification itself can be seen as lack of cointegration.

The recent development of Park and Phillips (1999, 2001) enables the applied worker to use wide range of nonlinear specifications. Nonetheless, when it comes to applied work, the ultimate problem is to choose the appropriate model. This is exactly the problem that will be addressed here. Two tests will be considered as means of testing for FF and lack of cointegration in long-run cointegrating relations. The first is a simple specification test based on fully modified residuals. The test statistic resembles the one of the Bierens (1990) Conditional Moment (CM) test for FF. Nonetheless we are not employing any weighting functions as Bierens (1990) does in order to create a fully consistent test. The second test is the CUSUM test for cointegration proposed by Xiao and Phillips (2002). It will be shown that both tests diverge under FF misspecification or lack of cointegration. Using some theoretical results due to Park and Phillips (1998) and with the aid of the simulation evidence provided in this paper, it will be argued that the Dickey-Fuller test (DF) that is widely used as a cointegration test, performs poorly under FF misspecification in many cases. Finally it is pointed out, that in a nonstationary framework, a simple inspection of the behaviour of slope and variance estimates can sometimes provide evidence for FF misspecification. It is shown that under certain types of FF misspecification, the slope estimators converge to zero or diverge as the sample size increases. Moreover if FF is committed, the estimator for the variance of the errors of the model diverges as the sample size increases. This behaviour is not evident in the stationary framework and can provide useful information about the adequacy of the fitted model.

The present theoretical framework is similar to that of Park and Phillips (1999) and Chang et al. (2001). The only work that is closely related to the present, that the author is aware of, is that of Hong and Phillips (2004) who extend Ramsey's (1969) RESET test to a framework similar to this. Hong and Phillips (2004) consider scalar covariate fitted models linear in parameter and variable. The present theoretical framework is more general. Multiple regression models, additively separable, linear in parameters and nonlinear in variables are considered. Therefore we treat linearity vs. nonlinearity as a special case. The nonlinear functions under consideration belong to the H-regular class of Park and Phillips (1999). Park and Phillips (1999, 2001) assume that the error of the model is a martingale difference sequence. We relax this assumption. Correlated errors and endogeneity are introduced by assuming the errors of the model and the errors that drive the unit root variables is a vector *linear process*. A semiparametric approach is followed for both tests to induce a limit distribution, under the null hypothesis, free of nuisance parameters. The approach is similar to the one of Xiao and Phillips (2002). The fitted model is estimated by a Fully Modified Least Squares (FM-LS) type of estimator and the sample moment of the test statistics is corrected for endogeneity bias.

We derive the limit distribution of the tests under the null hypothesis (correct FF) and we obtain divergence rates under the alternative hypothesis (incorrect FF or lack of cointegration). Under the null hypothesis, the first test (CM) has a chi-square limit distribution while the CUSUM test (CS) has a limit distribution similar to the one reported by Xiao and Phillips (2002). Under the alternative, the residuals of the fitted model will be dominated by some H-regular transformation $u(.)$ say of a unit root process, which is of a higher order of magnitude than the residuals of a correctly specified model. The underlying feature of the tests under consideration is that they can detect abnormal fluctuation in the residuals. The divergence rates under the alternative hypothesis depend on the bandwidth used for the estimation of long-run covariance matrices. The test statistics (CM_n and CS_n) are of the form:

$$\frac{SM_n}{VN_n},$$

where SM_n is some sample moment, VN_n is a variance normalisation term and n the sample size. Under the alternative hypothesis we have:

$$\begin{aligned} CM_n &= \frac{O_p(nk_u(\sqrt{n})^2)}{O_p(Mk_u(\sqrt{n})^2)} = O_p(n/M) \text{ and} \\ CS_n &= \frac{O_p(\sqrt{n}k_u(\sqrt{n}))}{O_p(\sqrt{M}k_u(\sqrt{n}))} = O_p(\sqrt{n/M}), \end{aligned}$$

as $n \to \infty$, where M is the bandwidth parameter used in the estimation of the long-run covariances and k_u is asymptotic order of the H-regular component $u(.)$, that dominates the regression residuals. We expect that other FF and cointegration tests can be used in this framework e.g. White's (1981) Hauseman and Information

Equality (IE) tests, the KPSS test. For instance the idea behind the RESET test and other FF tests is to use some kind of series approximation, e.g. Taylor's theorem, to approximate (possibly nonlinear) regressor remainings in the residuals of the model, when that is misspecified. When the model is misspecified in terms of FF, the information matrix will involve extra terms and this is what the IE test aims to detect.

The DF test performs poorly, when there is FF misspecification. The DF test has been widely used as a cointegration test. Ideally, when the fitted model is of incorrect FF, it would be desirable if the DF test rejected the hypothesis of cointegration(/correct specification). If the DF test is applied to the residuals of model that is misspecified in terms of FF, in many cases the alternative of cointegration will be favoured, although the residuals are not stationary. The DF test is designed to detect unit root processes. In our case however, the residuals will be a nonlinear transformation of unit root processes.

An explanation for the poor performance of the DF can be found in the work of Park and Phillips (1998). Park and Phillips (1998) analyse the limit behaviour of the DF test statistic, when it is applied to a series, which is a nonlinear transformation of a unit root process. In particular they consider integrable and three H-regular transformations namely, indicator, logarithmic and polynomial functions. For integrable and indicator functions they find that the DF test statistic diverges to minus infinity, therefore favouring the alterative of stationarity with probability approaching one as $n \to \infty$. For logarithmic and concave polynomial functions, the limit distribution will involve negative components making the test biased towards the alternative of stationarity. Only for convex polynomial transformations the DF will favour nonstationarity. These theoretical results are confirmed by our simulation experiment.

The rest of this paper is organised as follows. In Section 2 our theoretical framework is specified and some preliminary results are provided. In Section 3 our testing procedures are presented and their properties derived. Section 4 provides some simulation results and Section 5 concludes. Before proceeding to the next section some notation is introduced. For a vector $x = (x_i)$ or a matrix $A = (a_{ij})$, $|x|$ and $|A|$ denotes the vector and matrix respectively of the moduli of their elements. The maximum of the moduli is denoted as $\|.\|$. The transpose of a matrix A is will be written as A'. Also a matrix A of dimensions $n \times m$ may be written as $A_{(n \times m)}$. As usual for a function $f : \mathbb{R} \to \mathbb{R}$, $\dot{f}$ will denote its first derivative with respect to its argument. Finally $I\{A\}$ will denote the indicator function of a set A.

2 Theoretical Framework and Preliminary Results

Assume that the series $\{y_t\}_{t=1}^n$ is generated by:

$$\begin{aligned} y_t &= \theta_{o1} f_1(x_{1t})+,...,+\theta_{op} f_p(x_{pt}) + u_t \\ &= f'(x_t)\theta_o + u_t \end{aligned} \tag{1}$$

or by:

$$y_t = s(z_t), \tag{2}$$

where $f(.)$ and $s(.)$ belong to the H-regular family, the variables x_t and z_t are unit root processes, and u_t is an error term that will be specified in detail later. Our purpose is to examine the case of incorrect FF and the case of no cointegration. For this reason two possible data generating mechanisms will be considered for the dependent variable. The model in (1) will be the true specification when there is cointegration (possibly nonlinear) between y_t and the variables of interest x_{it}. The specification (2) will be the data generating mechanism when there is no cointegrating relationship between y_t and the variables of interest. When the latter is the case, it is usually assumed in the literature (e.g. Xiao and Phillips (2002)) that y_t is a unit root process, z_t say, that is unrelated to the regressors (x_{it}'s). Here it will be assumed that y_t is possibly a nonlinear transformation of such a process. In this way y_t is allowed to be of different order of magnitude than z_t. Clearly when $s(.)$ is linear, y_t is a unit root process. The fitted model will be given by:

$$\begin{aligned} \hat{y}_t &= \hat{a}_1 g_1(x_{1t})+,...,+\hat{a}_p g_p(x_{pt}) + \hat{u}_t \\ &= g'(x_t)\hat{a} + \hat{u}_t \end{aligned} \tag{3}$$

For notational convenience, the vectors $f(x_t)$ and $g(x_t)$ in (1) and (3) may be written as f_t and g_t respectively.

Next the variables and the functions that appear in (1), (2) and (3) will be specified in detail. The variables $x_t' = (x_{1t}, ..., x_{pt})$ and z_t are a unit root processes given by:

$$x_t = x_{t-1} + v_t \textit{ and } z_t = z_{t-1} + w_t.$$

The following assumption about u_t, v_t and w_t holds:

ASSUMPTION 2.1: *The sequence $e_t' = (u_t, v_t', w_t)$ is a linear process given by:*

$$e_t = \sum_{j=1}^{\infty} \Pi_j \xi_{t-j} = \Pi(L)\xi_t,$$

and the following hold:
(i) *The matrix lag polynomial $\Pi(L) = diag\left(\Phi(L)_{(1\times 1)}, \Psi(L)_{(p\times p)}, \Xi(L)_{(1\times 1)}\right)$ satisfies the summability condition $\sum_{j=1}^{\infty} j^{\alpha} \|\Pi_j\| < \infty$ with $\alpha > 1$.*

(ii) *The random sequence* ξ_t *satisfies the following conditions:*

(a) $\left\{\xi_t' = \left(\varepsilon_t, \eta_{t+1}', \omega_{t+1}\right), \mathcal{F}_t = \sigma\left(\xi_s, -\infty \leq s \leq t\right)\right\}$ *is a stationary and ergodic martingale difference sequence with* $\mathbf{E}\left[\xi_t \xi_t' \mid \mathcal{F}_{t-1}\right] = \Sigma$.

(b) *The sequence* ξ_t *is i.i.d. with* $\mathbf{E}\left\|\xi_t\right\|^l < \infty$ *for some* $l > 4$ *and its distribution is absolutely continuous with respect to Lebesgue measure and has characteristic function* $\varphi(\lambda) = o(\|\lambda\|^{-\delta})$ *as* $\lambda \to \infty$.

For the purpose of the subsequent analysis, the covariance matrix Σ will be conformably partition as follows:

$$\Sigma = \begin{pmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\eta} & \Sigma_{\varepsilon\omega} \\ \Sigma_{\eta\varepsilon} & \Sigma_{\eta\eta} & \Sigma_{\eta\omega} \\ \Sigma_{\omega\varepsilon} & \Sigma_{\omega\eta} & \Sigma_{\omega\omega} \end{pmatrix}.$$

For v_t, u_t and w_t define the usual partial sum processes: $(U_n(r), V_n'(r), W_n(r)) = n^{-1/2}\sum_{t=1}^{[nr]}(u_t, v_t', w_t)$ with $0 \leq r \leq 1$. Under Assumption 2.1, it follows from a multivariate extension of the results of Phillips and Solo (1992) that

$$(U_n(r), V_n'(r), W_n(r)) \xrightarrow{d} (U(r), V'(r), W(r)),$$

with $(U(r), V'(r), W(r))$ being an $(p+2)$-dimensional Brownian motion with covariance matrix Ω conformably partitioned as

$$\Omega = \begin{pmatrix} \Omega_{uu} & \Omega_{uv} & \Omega_{uw} \\ \Omega_{vu} & \Omega_{vv} & \Omega_{vw} \\ \Omega_{wu} & \Omega_{wv} & \Omega_{ww} \end{pmatrix}.$$

Under Assumption 2.1, strong approximations hold for the vector $(U_n(r), V_n'(r), W_n(r))$ that allow the use of embedding arguments (see Phillips (1999)). Such embedding arguments are extensively utilised by Park and Phillips (1999, 2001) and will be used here as well. So when convergence in probability or all almost sure convergence arguments are used, those should be interpreted as convergence in distribution unless the limit is nonstochastic. Moreover the usual long-run covariance matrices will be employed. Note that Ω can be expressed as:

$$\Omega = \sum_{k=-\infty}^{\infty} \mathbf{E}\left(e_t e_{t+k}'\right)$$

and the one sided long-run covariance matrix, say Λ is:

$$\Lambda(h) = \sum_{k=0}^{\infty} \mathbf{E}\left(e_t e_{t+k-h}'\right) = \begin{pmatrix} \Lambda_{uu} & \Lambda_{uv} & \Lambda_{uw} \\ \Lambda_{vu} & \Lambda_{vv} & \Lambda_{vw} \\ \Lambda_{wu} & \Lambda_{wv} & \Lambda_{ww} \end{pmatrix}(h),$$

where $h \in \mathbb{Z}$.

Next the functions that appear in (1), (2) and (3) are specified. As mentioned earlier the functions under consideration will be confined to the H-regular family of Park and Phillips (1999). The H-regular family comprises transformations that are asymptotically homogeneous. An H-regular transformation $f(.)$ say, behaves as

$$f(\lambda x) \sim k_f(\lambda)h_f(x) \text{ for large } \lambda,$$

where h_f and k_f are the so called limit homogenous function and asymptotic order of f respectively. The limit homogenous function satisfies certain regularity conditions. Functions that do so are called by Park and Phillips (1999) "*regular*". The asymptotic results provided by Park and Phillips (1999) for *regular* transformations have been extended by de Jong (2004) to a more general class of transformations that comprise of locally integrable functions with finite many poles and which are monotone between poles[1]. Due to the introduction of weak dependence in the error structure of the model, some smoothness will need to be imposed on the first derivatives of the functions. We will restict our functions to a subset of the H-regular class of Park and Phillips (1999). The present class of functions will be called H_1-regular with H_1-regularity defined as follows:

DEFINITION 2.1:
The transformation $f : \mathbb{R}^p \to \mathbb{R}^p$, *such that* $f'(x) = (f_1(x_1), .., f_p(x_p))$ *will be called* H_1*-regular if:*
(i) $f(\lambda x) = k_f(\lambda)h_f(x) + R_f(x, \lambda)$ *with* $h_f(.)$ *regular and*
(a) $|R_f(x, \lambda)| \leq a_f(\lambda)P_f(x)$, *with* $\limsup_{\lambda\to\infty} \left\| a_f(\lambda)k_f^{-1}(\lambda)\right\| = 0$ *and* $P_f(.)$ *locally integrable, or*
(b) $|R_f(x, \lambda)| \leq b_f(\lambda)Q_f(\lambda x)$, *with* $\limsup_{\lambda\to\infty} \left\| b_f(\lambda)k_f^{-1}(\lambda)\right\| < \infty$ *and* $Q_f(.)$ *locally integrable and vanishing at infinity.*
(ii) $\lambda\dot{f}(\lambda x) = k_f(\lambda)\dot{h}_f(x) + \dot{R}_f(x, \lambda)$ *with* $\dot{h}_f(.)$ *regular and*
(a) $\left|\dot{R}_f(x, \lambda)\right| \leq \dot{a}_f(\lambda)\dot{P}_f(x)$, *with* $\limsup_{\lambda\to\infty} \left\|\lambda\dot{a}_f(\lambda)k_f^{-1}(\lambda)\right\| = 0$ *and* $\dot{P}_f(.)$ *locally integrable, or*
(b) $\left|\dot{R}_f(x, \lambda)\right| \leq \dot{b}_f(\lambda)\dot{Q}_f(\lambda x)$, *with* $\limsup_{\lambda\to\infty} \left\|\lambda\dot{b}_f(\lambda)k_f^{-1}(\lambda)\right\| < \infty$ *and* $\dot{Q}_f(.)$ *locally integrable and vanishing at infinity.*
(iii) *For any* $0 < K < \infty$ *and some* $0 \leq b < 1/2$ *there is a sequence* $s_n \downarrow 0$ *as* $n \to \infty$, *such that*

$$\limsup_{n\to\infty} \left\|n^{1/2+b}k_f(\sqrt{n})^{-1}\right\| \sup_{\|x_1\|\leq K} \sup_{\|x_1-x_2\|\leq s_n} \left\|\dot{f}(\sqrt{n}x_1) - \dot{f}(\sqrt{n}x_2)\right\| = 0.$$

As usual h_f and k_f will be called the limit homogenous functions and asymptotic order of f respectively. Moreover note that when f is a p-dimensional vector, k_f

[1] Pötscher (2004) provides the same asymptotic results under less restrictive assumptions about the functions but more restrictive conditions about the innovations of the unit root process.

and $\dot{f}$ will be $(p \times p)$ diagonal matrices. Condition (iii) in the definition above is the same smoothness condition employed by de Jong (2002)[2]. The convergence rate of the sequence s_n will be determined by l, i.e. the order of finite moments of the process ξ_t. In general the larger the parameter b is, a larger l will be required.

For linear models it is well known (e.g. Phillips (1986, 1988)) that when the errors of the model are weakly dependent, the covariance asymptotics involve extra terms. In the limit, apart from the stochastic integral, a long-run covariance matrix appears. For nonlinear models the long-run covariance matrix is weighted by functionals of Brownian motion. This result was originally shown by de Jong (2002), when the errors are near epoch depended. Below a similar result, for errors that are linear processes, is provided.

THEOREM 2.1:
Let $f'(x) = (f_1(x_1), ..., f(x_p))$ be H_1-regular. Under Assumption 2.1, for $1 \leq t$, $t + h \leq n$, $|h| \leq n^b$, we have

$$\text{(i)}\ \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} f(x_{t+h}) u_t \xrightarrow{d} \int_0^1 h_f(V(r))dU(r) + \int_0^1 \dot{h}_f(V(r))dr\Lambda_{vu}(h)$$

and

$$\text{(ii)}\ \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} f(x_{t+h}) v_t' \xrightarrow{d} \int_0^1 h_f(V(r))dV'(r) + \int_0^1 \dot{h}_f(V(r))dr\Lambda_{vv}(h),$$

as $n \to \infty$.

For the derivation of the limit distribution theory under correct specification, the above covariance asymptotic results will be employed with the parameter h set equal to zero. In this case the condition (iii) of Definition 2.1 can be relaxed by setting b equal to zero as well. Under FF misspecification however, the limit behaviour of the estimators is determined by sample sums like those above with arbitrary value for the parameter h. Moreover the results of Theorem 2.1 are useful, for the study of spectral regressions (see for example Phillips (1991)), under the current theoretical framework.

Next FF misspecification and lack of cointegration will be defined precisely.

DEFINITION 2.2:
For g and f H_1-regular:
(i) *We will say that the fitted model (3) is of correct FF, when $g_i(.) = f_i(.)$ for all $i = \{1, .., p\}$ and (1) holds.*
(ii) *We will say that the fitted model (3) is of incorrect FF, when the true model is given by (1) and $g_i(.) \neq f_i(.)$ for some $i = \{1, ..., p\}$ and one of the following conditions hold:*

[2] See de Jong (2002) page 21.

C1: $g_i - f_i = q_i$ *with* q_i H_1*-regular such that* $k_{q_i}(\lambda)/k_{g_i}(\lambda)$, $k_{q_i}(\lambda)/k_{f_i}(\lambda) \to 0$ as $\lambda \to \infty$, *or*

C2: $k_{g_i}(\lambda)/k_{f_i}(\lambda) \to 0$ *or* ∞ *as* $\lambda \to \infty$.

(ii) *We will say that there is no cointegration, when the fitted model is given by (3) and the true model by (2).*

Condition **C1** postulates that some term is correctly specified up to some lower order H-regular component, while **C2** postulates that a fitted component does not agree in asymptotic order with its counterpart at all. The possibility of having a second cointegrating relationship between $f_1(x_{1t})$,..., $f_p(x_{pt})$, will be ruled out. It is obvious from Definition 2.2 that the present theoretical framework does not allow for omitted or redundant variables. An extension of the subsequent results in that direction is possible but will not be attempted here, as it would result in more complexity in our presentation.

Notice that the linear model that is commonly used in cointegrating relationships is H_1-regular. In practice functional form misspecification could arise from neglected lower order components (lower order than the linear specification). Consider for instance the case where $f(x) = x + |x|^{1/2}$ and $g(x) = x$. If the errors of the model are martingale differences as in Park and Phillips (1999), it is possible to obtain power rates under this kind of FF misspecification. Under the current framework however we are unable to obtain explicit power rate results as the component $|x|^{1/2}$ is not H_1*-regular.* Below some examples of H_1-regular transformations are provided:

EXAMPLE:

(i) $f(x) = |x|^c$, $c > 2$.

(ii) $f(x) = x \exp(x) / (1 + \exp(x))$.

(iii) $f_n(x) = \left(\frac{n^{1/4}c^{1/2}}{2} + \frac{n^{-1/4}c^{-1/2}}{2}x\right) I\{x \leq n^{1/2}c\} + x^{1/2} I\{x > n^{1/2}c\}$, $c > 0$.

(iv) $f_n(x) = \left(\ln\left(cn^{1/2}\right) - 1 + \frac{1}{n^{1/2}c}x\right) I\{x \leq n^{1/2}c\} + \ln(x) I\{x > n^{1/2}c\}$, $c > 0$.

Saikkonen and Choi (2004) have recently analysed cointegrating Smooth Transition Regression (STR) models. Their specification is comprised by a linear component multiplied by a transition function. They explicitly consider a logistic function. Because distribution type of functions lack identification, when the covariates are unit root processes (see Park and Phillips (2001)), Saikkonen and Choi (2004) actually consider models where the covariates are normalised by the square root of the sample size. In practice one might want to test whether the transition function used is correctly specified. The present framework does not cover fitted models of this kind because they are nonlinear in parameters. Limit results for the behaviour of the Nonlinear Least Squares estimator under FF misspecification in models with unit roots, have been derived by the author (Kasparis (2004)) and some extensions of the current results along these lines are possible. Nonetheless models with normalised variables create an extra complication. It has been assumed (Definition 2.2) that if the fitted model g, is of incorrect FF and of the same order as the true model f, then g and f agree up to some lower order component q, say. Actually to the best of the author's knowledge, this has to be the case (when both f and g are *H-regular*). Now if the model involves normalised variables one can find examples of f, and g that are of the same asymptotic order but do not agree at all, for example let $f_n(x) = 1\{x/\sqrt{n} > c\}$ and $g_n(x,a) = \exp(ax/\sqrt{n})\,(1+\exp(ax/\sqrt{n}))^{-1}$. Development of second order asymptotic theory for H-regular transformations is required, to obtain asymptotic power rates for this type of models.

3 Detection of Functional Form Misspecification

The main focus in this section is to develop two specification tests as means of testing for FF in the theoretical framework of Section 2. The first test statistic resembles the Bierens (1990) CM test. The Bierens (1990) test is based on a conditional moment condition that holds under the null hypothesis (correct FF). A similar condition can be shown to hold for the first test, when the errors of the model are martingale differences as in Park and Phillips (1999, 2001). Under the current framework though, such condition does not hold, because the covariates are endogenous. Nonetheless for purposes of brevity we will call the first test CM. The second test is the CUSUM test for cointegration of Xiao and Phillips (2001) generalised to cope with fitted models that are nonlinear in variables. The limit properties of the tests are derived under correct FF, incorrect FF and lack of cointegration.

The asymptotic behaviour of the Least Squares (LS) estimator and the statistical tests is determined by sample covariances like those in Theorem 2.1. Because the limit distribution theory is not mixed normal, the usual likelihood based and t-tests tests do not have standard distributions under the null hypothesis. Moreover the limit distribution of non standard tests like the CUSUM test will involve nuisance parameters. To resolve this problem the model is fitted by a FM-LS type of estimator and an endogeneity correction term is introduced in the statistic. To obtain the estimator and the correction term, kernel estimators for Ω_{uu}, Ω_{vv}, Ω_{vu}, Λ_{vu} and Λ_{vv} are used:

$$\begin{array}{ll}\hat{\Omega}_{uu} = \sum_{h=-M}^{M} \kappa\left(\frac{h}{M}\right) C_{uu}(h), & \hat{\Omega}_{vv} = \sum_{h=-M}^{M} \kappa\left(\frac{h}{M}\right) C_{vv}(h), \\ \hat{\Omega}_{vu} = \sum_{h=-M}^{M} \kappa\left(\frac{h}{M}\right) C_{vu}(h), & \hat{\Lambda}_{vv} = \sum_{h=0}^{M} \kappa\left(\frac{h}{M}\right) C_{vv}(h), \\ \hat{\Lambda}_{vu} = \sum_{h=0}^{M} \kappa\left(\frac{h}{M}\right) C_{vu}(h), & \end{array}$$

where $\kappa(.)$ is the lag window defined on $[-1,1]$ such that $\kappa(0) = 1$ and M is a bandwidth such that $M \to \infty$, $n/M \to 0$ as $n \to \infty$. Moreover $C_{uu}(h)$, $C_{vv}(h)$, and $C_{vu}(h)$ are sample covariances defined by $C_{uu}(h) = n^{-1}\sum_t' \hat{u}_t\hat{u}_{t+h}$, $C_{vv}(h) = n^{-1}\sum_t' v_t v_{t+h}'$ and $C_{vu}(h) = n^{-1}\sum_t' v_t\hat{u}_{t+h}$, where $\hat{u}$ are the residuals from LS estimation and $\sum_t'$ is summation over $1 \le t,\ t+h \le n$. Consistency results for this kind of kernel estimators can be found in Andrews (1991), when the processes satisfy mixing conditions. Under the current framework consistency results are provided by Phillips (1995).

The estimator under consideration closely resembles the original FM-LS estimator introduced by Phillips and Hansen (1990). Before the estimator is presented, the following quantities need to be defined:

$$y_t^+ = y_t - v_t' \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu} \text{ and } \hat{\Lambda}_{vu}^+ = \hat{\Lambda}_{vu} - \hat{\Lambda}_{vv} \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu}.$$

The FM-LS estimator under consideration is:

$$\hat{a} = \left[\sum_{t=1}^{n} g(x_t) g'(x_t) \right]^{-1} \left[\sum_{t=1}^{n} g(x_t) y_t^+ - \dot{g}_n \hat{\Lambda}_{vu}^+ \right],$$

with $\dot{g}_n = \sum_{t=1}^{n} \dot{g}(x_t)$. Under correct specification, he following result holds:

LEMMA 3.1:
Under correct FF as $n \to \infty$

$$\sqrt{n} k_g (\hat{a} - \theta_o) \xrightarrow{d} \left[\int_0^1 h_g (V(r)) \, h_g' (V(r)) \, dr \right]^{-1} \int_0^1 h_g (V(r)) \, dU(r)^+,$$

where $U(r)^+ = U(r) - V'(r) \Omega_{vv}^{-1} \Omega_{vu}$.

Notice that the limit distribution of the estimator is mixed normal as V and U^+ are independent. Both of the tests under consideration are residual based. An endogeneity bias correction term is introduced in the residuals of the fitted model giving the so called fully modified residuals $\hat{u}_t^+$ defined as

$$\hat{u}_t^+ = y_t - \hat{a}' g(x_t) - v_t' \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu}.$$

Next the test statistics will be presented. First define the matrices $\hat{A}_n$, $\hat{B}_n$, A and B and their inverses, when they exist, as follows:

$$\frac{1}{n} k_g^{-1} \hat{A}_n = \frac{1}{n} k_g^{-1} \sum_{t=1}^{n} g(x_t) \xrightarrow{p} A,$$

and

$$\frac{1}{n} k_g^{-1} \hat{B}_n k_g^{-1} = \frac{1}{n} k_g^{-1} \sum_{t=1}^{n} g(x_t) g'(x_t) k_g^{-1} \xrightarrow{p} B.$$

The CM test statistic is:

$$CM_n = \frac{\left[\sum_{t=1}^{n} \hat{u}_t^+ \right]^2}{\left(\hat{\Omega}_{uu} - \hat{\Omega}_{uv} \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu} \right) \sum_{t=1}^{n} \left[\hat{A}_n' \hat{B}_n^{-1} g(x_t) - 1 \right]^2},$$

and the CUSUM test statistic:

$$CS_n = \max_{k=1,..,n} \frac{\left|\sum_{t=1}^{k} \hat{u}_t^+\right|}{\sqrt{n\left(\hat{\Omega}_{uu} - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right)}}.$$

The behaviour of the tests under the null hypothesis is shown in the theorem below.

THEOREM 3.1:
Under correct FF as $n \to \infty$,

$$CM_n \xrightarrow{d} \chi_1^2$$

and

$$CS_n \xrightarrow{d} \sup_{0 \le s \le 1} \left|\bar{U}(s)\right| / \sqrt{\Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}},$$

where

$$\bar{U}(s) = U(s)^+ - \left[\int_0^1 dU(r)^+ h_g(V(r))'\right]\left[\int_0^1 h_g(V(r))h_g(V(r))'dr\right]^{-1}\left[\int_0^s h_g(V(r))dr\right].$$

The limit distribution of the CUSUM test is resembles the one derived by Xiao and Phillips (2002). If $h_g(.)$ is allowed to be linear in the expression above, $\bar{U}(s)$ will be as in Xiao and Phillips (2002). Note that the distribution of the CUSUM test is not standard and simulations are required to obtain critical values. Moreover, the limit distribution is specific to the fitted model and therefore different critical values are required for different models. This makes the test somewhat impractical when the fitted model is nonlinear. Nonetheless it can be easily implemented as a linear vs. nonlinear test. On the other hand the CM test has standard limit distribution irrespective of the empirical model employed.

Next we will examine the asymptotic power of the tests. Note that under the alternative some of the kernel estimators mentioned earlier will be inconsistent. Before their limit behaviour is considered, some notation needs to be introduced. Define $d = f - g$ with f, g as in (1) and (3). Moreover denote by "$*$" the index of the leading element(s) of d, which can be expressed as $d_* = f_* - g_*$, and k_{d^*}, k_{f^*}, k_{g^*} are the relevant asymptotic orders. We will consider two scenarios:

S1: $k_{d^*} < k_{f^*}$ and k_{g^*},
S2: $k_{d^*} = k_{f^*}$ or k_{g^*}.

Under **S1** the leading misspecified component behaves as in **C1**, while under **S2** the behaviour of the leading misspecified component is given by **C2**. Denote by $\hat{a}_{LS}$ the least squares estimator corresponding to the fitted model. Under FF misspecification f we will be partitioned as $f'_{(1\times p)} = \left(f^{1\prime}_{(1\times p_1)},\ f^{2\prime}_{(1\times p_2)}\right)$ with f^2 being the components of f that have not been correctly specified. The leading element(s) of f^2 will be denoted as f^{2*} and its asymptotic order is $k_{f^{2*}}$. The vector θ_o is also partitioned as

$\theta_o = \left(\theta^{1\prime}_{o(1\times p_1)},\ \theta^{2\prime}_{o(1\times p_2)}\right)$, where θ^1_o and θ^2_o are the coefficients of f^1 and f^2 respectively. Aslo $\bar{\theta}_o$ is defined by $\bar{\theta}'_{o(1\times p)} = \left(\theta^{1\prime}_{o(1\times p_1)},\ \mathbf{0}'_{(1\times p_2)}\right)$. Finally some further notation is introduced by Definition 3.1:

DEFINITION 3.1:
Define:
(i) *The vectors ζ_1, ζ_2 and ζ_3 are the following limits:*

$$\begin{aligned}
&\tfrac{k_g}{k_{d*}}\left(\hat{a}_{LS} - \theta_o\right) \xrightarrow{p} \zeta_1, \textit{ under incorrect FF when } \mathbf{S1} \textit{ holds},\\
&\tfrac{k_g}{k_{f^2*}}\left(\hat{a}_{LS} - \bar{\theta}_o\right) \xrightarrow{p} \zeta_2, \textit{ under incorrect FF when } \mathbf{S2} \textit{ holds},\\
&\tfrac{k_g}{k_s}\hat{a}_{LS} \xrightarrow{p} \zeta_3, \textit{ under no cointegration.}
\end{aligned}$$

(ii) *The vectors $h_{\bar{d}}(.)'_{(1\times p)}$, $h(.)'_{\bar{f}^2(1\times p_2)}$ and the matrices $\dot{h}_{\bar{d}}(.)_{(p\times p)}$, $\dot{h}_{\bar{f}^2}(.)_{(p_2\times p_2)}$ by:*

$$\begin{aligned}
(nk_{d^*})^{-1}\textstyle\sum_{t=1}^n d_t \xrightarrow{p} \int_0^1 h_{\bar{d}}(V(r))dr, &\quad \left(\sqrt{n}k_{d^*}\right)^{-1}\textstyle\sum_{t=1}^n \dot{d}_t \xrightarrow{p} \int_0^1 \dot{h}_{\bar{d}}(V(r))dr,\\
(nk_{f^{2*}})^{-1}\textstyle\sum_{t=1}^n f_t^2 \xrightarrow{p} \int_0^1 h_{\bar{f}^2}(V(r))dr, &\quad \left(\sqrt{n}k_{f^{2*}}\right)^{-1}\textstyle\sum_{t=1}^n \dot{f}_t \xrightarrow{p} \int_0^1 \dot{h}_{\bar{f}^2}(V(r))dr.
\end{aligned}$$

(iii) *The vectors $\bar{\zeta}_1$, $\bar{\zeta}_2$, $\bar{\zeta}_3$, $\bar{h}_1$, $\bar{h}_2$, $\bar{h}_3$ and the matrices $\dot{H}_1$, $\dot{H}_2$, $\dot{H}_3$, $\bar{\Omega}$ by:*

$$\begin{aligned}
&\bar{\zeta}'_1 = \left(\theta'_o, -\zeta'_1\right),\ \bar{\zeta}'_2 = \left(\theta^{2\prime}_o, -\zeta'_2\right),\ \bar{\zeta}'_3 = \left(1, -\zeta'_3\right),\\
&\bar{h}'_1 = \left(h'_{\bar{d}}, h'_g\right),\ \bar{h}'_2 = \left(h'_{\bar{f}^2}, h'_g\right),\ \bar{h}'_3 = \left(h_s, h'_g\right),\\
&\dot{H}'_1 = \left(\dot{h}'_{\bar{d}}, \dot{h}'_g\right),\ \dot{H}'_2 = \left(\dot{h}'_{\bar{f}^2}, \dot{h}'_g\right),\ \dot{H}'_3 = \left(\dot{h}_s, \dot{h}'_g\right),\\
&\bar{\Omega} = \left(\Omega_{vw}, \Omega_{vv}\right).
\end{aligned}$$

The expressions in Definition 3.1(i) characterise the limit behaviour of the LS estimator under FF misspecification and lack of cointegration. Explicite expressions for the limits ζ_1, ζ_2 and ζ_3 are given in the Appendix. It is apparent from Definition 3.1(i) that under incorrect FF, the slope estimators will not always converge to the parameter of interest. For instance, when **S1** holds, an individual slope estimator, $\hat{\theta}_{LS_i}$, will converge to θ_{oi} only if g_i dominates d_* in terms of asymptotic order. Generally under FF misspecification one of the following holds: a) The estimator may converge to the parameter of interest. b) It may converge to functionals of Brownian motion. c) It may vanish i.e. converge to zero. d) It may be unbounded in probability.

Next, the limits of the covariance estimators under FF misspecification and lack of cointegration will be presented. Let $K(s) = \lim_{n\to\infty}(2\pi M)^{-1}\sum_{h=-M}^{M}\kappa\left(h/M\right)e^{ihs}$ and $K_1(s)$ is its one-sided version. The limit behaviour of the kernel estimators under the alternative hypothesis is given in the following result.

LEMMA 3.2:
Let Assumption 2.1 hold. As $n \to \infty$ we have:
(i) *Under incorrect FF when* **S1** *holds,*

$$\begin{aligned}
\frac{n^{1/2}}{Mk_{d^*}}\hat{\Omega}_{vu} &\xrightarrow{p} 2\pi K(0)\int_0^1 dV(r)\bar{h}_1'(V(r))\bar{\zeta}_1 + \Omega_{vv}\int_0^1 \dot{H}_1'(V(r))\bar{\zeta}_1 dr,\\
\frac{n^{1/2}}{Mk_{d^*}}\hat{\Lambda}_{vu} &\xrightarrow{p} 2\pi K_1(0)\int_0^1 dV(r)\bar{h}_1'(V(r))\bar{\zeta}_1 + \Omega_{vv}\int_0^1 \dot{H}_1'(V(r))\bar{\zeta}_2 dr,\\
\frac{1}{Mk_{d^*}^2}\hat{\Omega}_{uu} &\xrightarrow{p} 2\pi K(0)\int_0^1 \bar{\zeta}_1'\bar{h}_1(V(r))\bar{h}_1'(V(r))\bar{\zeta}_1 dr.
\end{aligned}$$

(ii) *Under incorrect FF when* **S2** *holds,*

$$\begin{aligned}
\frac{n^{1/2}}{Mk_{f2^*}}\hat{\Omega}_{vu} &\xrightarrow{p} 2\pi K(0)\int_0^1 dV(r)\bar{h}_2'(V(r))\bar{\zeta}_2 + \Omega_{vv}\int_0^1 \dot{H}_2'(V(r))\bar{\zeta}_2 dr,\\
\frac{n^{1/2}}{Mk_{f2^*}}\hat{\Lambda}_{vu} &\xrightarrow{p} 2\pi K_1(0)\int_0^1 dV(r)\bar{h}_2'(V(r))\bar{\zeta}_2 + \Omega_{vv}\int_0^1 \dot{H}_2'(V(r))\bar{\zeta}_2 dr,\\
\frac{1}{Mk_{f2^*}^2}\hat{\Omega}_{uu} &\xrightarrow{p} 2\pi K(0)\int_0^1 \bar{\zeta}_2'\bar{h}_2(V(r))\bar{h}_2'(V(r))\bar{\zeta}_2 dr.
\end{aligned}$$

(iii) *Under no cointegration*

$$\begin{aligned}
\frac{n^{1/2}}{Mk_s}\hat{\Omega}_{vu} &\xrightarrow{p} 2\pi K(0)\int_0^1 dV(r)\bar{h}_3'(W(r),V(r))\bar{\zeta}_3 + \bar{\Omega}\int_0^1 \dot{H}_3'\left(W(r),V(r)\right)\bar{\zeta}_3 dr\\
\frac{n^{1/2}}{Mk_s}\hat{\Lambda}_{vu} &\xrightarrow{p} 2\pi K_1(0)\int_0^1 dV(r)\bar{h}_3'(W(r),V(r))\bar{\zeta}_3 + \bar{\Omega}\int_0^1 \dot{H}_3'\left(W(r),V(r)\right)\bar{\zeta}_3 dr\\
\frac{1}{Mk_s^2}\hat{\Omega}_{uu} &\xrightarrow{p} 2\pi K(0)\int_0^1 \bar{\zeta}_3'\dot{H}_3\left(W(r),V(r)\right)\dot{H}_3'\left(W(r),V(r)\right)\bar{\zeta}_3 dr.
\end{aligned}$$

The behaviour of the test statistics under the alternative is given by the following result:

THEOREM 3.2:
Under incorrect FF or no cointegration as $n \to \infty$ we have

$$\mathbf{P}(CM_n > A_n),\ \mathbf{P}(CS_n > B_n) \to 1,$$

for any nonstochastic sequences A_n and B_n such that

$$A_n = o(n/M),\ B_n = o\left((n/M)^{1/2}\right).$$

The divergence rates in both cases are bandwidth depended. For the first test, the divergence rate is the same with the rate of the KPSS test and the RESET test of Hong and Phillips (2004). The divergence rate of the CUSUM test is the same with that reported by Xiao and Phillips (2002), when there is lack of cointegration in the linear framework.

From the simulation study of Xiao and Phillips (2002) it is obvious that when it comes to the choice of the bandwidth parameter, there is a trade-off between size and power. Andrews (1991) proposes an automatic bandwidth method where

$M = 1.447(\hat{\delta}n)^{1/3}$, with $\hat{\delta} = 4\hat{\rho}/\left(1-\hat{\rho}^2\right)^2$ and $\hat{\rho}$ is the LS estimator from the residuals autoregression. Methods like this one are inappropriate in our case. As Xiao and Phillips (2002) point out these kind of procedures were developed for stationary processes. The regression residuals are stationary only under the null hypothesis. Under the alternative hypothesis they are not stationary. It is apparent from the simulation results of Xiao and Phillips (2002) that when this bandwidth method is used, the CUSUM test has no power. Xiao and Phillips (2002) suggest that under the alternative of their test $M \sim n$. As the following result shows, this is true in here as well.

LEMMA 3.3:
Let Assumption 2.1 hold with f, g and s having H_1-regular derivatives and $M = \left(\hat{\delta}n\right)^{1/3}$. Then under incorrect FF or no cointegration we have

$$M = O_p(n).$$

It is apparent from the limit expressions in Definition 3.1 that if FF misspecification is committed, the LS estimator will diverge or vanish as the sample size increases, is some cases. Moreover under incorrect FF the long-run covariance estimator $\hat{\Omega}_{uu}$ diverges. Such kind of behaviour would indicate that the fitted model is not correctly specified. Therefore an informal test that can be easily implemented would be to estimate the model for several sample sizes and check whether the slope and the variance estimates behave in the way described above. Note that this kind of behaviour is not evident to stationary models. For stationary models, the LS estimator will typically converge to a finite quantity. Moreover the variance estimator will be bounded in probability (see White (1981)).

4 Simulation Evidence

In this section a Monte Carlo experiment is performed to asses the finite sample properties of the CM, CS and DF tests. In particular, first the size properties of the CM and CS tests are examined and secondly the ability of CM, CS and DF to detect lack of cointegration and FF misspecification. Clearly for CM and CS this corresponds to the power of the tests. The DF test can be used as a linear cointegration test. FF misspecification and lack of cointegration in the wider, nonlinear sense, cannot be rigorously embedded in the hypothesis structure of the test. Nonetheless it would be desirable if in the presence of FF misspecification, the DF test favoured the unit root hypothesis, as this would be an indication that the regression residuals are nonstationary and therefore the fitted model inadequate. For this reason, the frequency which the DF test favours the unit root hypothesis, will be used as a measure of its ability to detect incorrect FF or lack of cointegration in the nonlinear sense. We will conventionally refer to it as the "power" of the DF test. All the experiments use 1000 simulations and significance level is set at 5%. The Barlett spectral window is employed for the kernel estimators.

The fitted model used in the experiment is linear with a scalar covariate given by:

$$\hat{y}_t = \hat{a}x_t + \hat{u}_t.$$

For the data generating mechanism, a wide range of H-regular specifications including threshold, polynomial, logarithmic and smooth transition models are considered. An integrable specification is also included in the experiment. Under lack of cointegration and incorrect FF the data is generated by the specifications shown in Table A and Table B respectively:

Table A: $(y_t = s(z_t))$

$y_t = z_t$	(1)	$y_t = \frac{z_t}{\log(1+\lvert z_t\rvert)}$	(4)
$y_t = z_t/\left(1+\lvert z_t\rvert^{0.5}\right)$	(2)	$y_t = z_t 1\{z_t \geq 0\} + 1.3 z_t 1\{z_t < 0\}$	(5)
$y_t = sign(z_t)\lvert z_t\rvert^{1.5}$	(3)	$y_t = sign(z_t)\lvert z_t\rvert^{0.75}$	(6)

Table B: $(y_t = f(x_t) + u_t)$

$y_t = \ln(1+\lvert x_t\rvert) + u_t$	$(1')$	$y_t = x_t + \lvert x_t\rvert^{0.5} + u_t$	$(7')$
$y_t = 1.8x_t I\{x_t \geq 0\} + 0.4x_t I\{x_t < 0\} + u_t$	$(2')$	$y_t = sign(x_t)\lvert x_t\rvert^{1.5} + u_t$	$(8')$
$y_t = x_t + 1.8\frac{x_t}{1+\lvert x_t\rvert^{0.5}} + u_t$	$(3')$	$y_t = x_t^2 + u_t$	$(9')$
$y_t = x_t + \log(1+\lvert x_t\rvert) + u_t$	$(4')$	$y_t = sign(x_t)\lvert x_t\rvert^{1.25} + u_t$	$(10')$
$y_t = x_t\log(1+\lvert x_t\rvert) + u_t$	$(5')$	$y_t = 0.5\left(n^{1/4} + n^{-1/4}x_t\right) I\{\frac{x_t}{\sqrt{n}} \leq 1\}$ $+x_t^{0.5} I\{\frac{x_t}{\sqrt{n}} > 1\} + u_t$	$(11')$
$y_t = x_t + 1.8\frac{x_t}{1+\exp\left(-x_t/\sqrt{n}-2\right)} + u_t$	$(6')$	$y_t = \exp(-x_t^2) + u_t$	$(12')$

The variables x_t, z_t and u_t are constructed as follows:

$$\begin{aligned} &u_t = \psi u_{t-1} + \epsilon_t, \\ &\Delta x_t = v_t, \text{ with } v_t = \psi v_{t-1} + \eta_t, \\ &\Delta z_t = w_t, \text{ with } w_t = 0.3 w_{t-1} + \omega_t, \end{aligned}$$

and $\left(\epsilon_t,\ \eta_{t+1},\ \omega_{t+1}\right)' = r'_{t(1\times 3)} A'_{(3\times 3)}$, where

$$A = \begin{pmatrix} 1 & 0.2 & 0.1 \\ 0.3 & 2 & 0 \\ 0 & 0.1 & 1.2 \end{pmatrix} \text{ and } r_t \sim i.i.d.\ N(\mathbf{0}, \mathbf{I}).$$

As Xiao and Phillips (2002) point out, when the autoregressive parameters (ψ) are close to unity, the innovation errors become nearly integrated and this adversely affects the size of the test. Actually, the tests will overreject the null hypothesis.

In order to investigate how sensitive the size of the tests is to the intensity of the innovation errors, a wide range of values is used for the autoregressive parameters. In particular, $\psi = 0, 0.2, 0.4, 0.6, 0.8$ and 0.9 has been chosen.

It is apparent from the theoretical results, that the performance of the tests depends on the sample size and the bandwidth parameter. In particular to achieve good size properties a large bandwidth parameter will be required, if the innovation errors exhibit strong intensity. On the other hand a large bandwidth will adversely affect the power of the tests. In order to asses the extent of the trade off between size and power, two values for the bandwidth are considered: $M1 = n^{1/5}$ and $M2 = n^{1/3}$. Moreover we will consider several sample sizes: $n = 50, 100, 200, 300$ and 500.

Table 1 shows the empirical size of the CM and CS tests for several sample sizes. The findings are similar to those reported by Xiao and Phillips (2002). As seen in Table 1, the size performance of the tests is good for $M = M1$ as long as $\psi \leq 0.2$, while for $M = M2$ good performance is attained as long as $\psi \leq 0.4$. If the autoregressive parameters are restricted to this range, the performance of both tests is comparable. For larger autoregressive coefficients, severe overrejection of the null hypothesis occurs with the CM test performing better.

Table 1: *Correct FF*, empirical size for CM & CS *(5% level)*

	$n = 100$				$n = 200$			
	$M1$	$M1$	$M2$	$M2$	$M1$	$M1$	$M2$	$M2$
ψ	CM	CS	CM	CS	CM	CS	CM	CS
0	0.0290	0.0260	0.0270	0.0220	0.0290	0.0240	0.0380	0.0370
0.2	0.0440	0.0450	0.0430	0.0340	0.0560	0.0600	0.0510	0.0510
0.4	0.0980	0.0950	0.0620	0.0500	0.1010	0.1170	0.0710	0.0630
0.6	0.1850	0.1930	0.1120	0.0950	0.1800	0.2190	0.1060	0.0990
0.8	0.3440	0.4160	0.2180	0.2040	0.3620	0.4740	0.2070	0.2100
0.9	0.4700	0.5830	0.3440	0.3400	0.4880	0.6780	0.3330	0.3750
	$n = 300$				$n = 500$			
	$M1$	$M1$	$M2$	$M2$	$M1$	$M1$	$M2$	$M2$
ψ	CM	CS	CM	CS	CM	CS	CM	CS
0	0.0260	0.0260	0.0300	0.0320	0.0420	0.0400	0.0480	0.0540
0.2	0.0500	0.0490	0.0480	0.0460	0.0630	0.0700	0.0570	0.0610
0.4	0.0920	0.0780	0.0650	0.0570	0.0880	0.1080	0.0680	0.0770
0.6	0.1460	0.1510	0.0910	0.0750	0.1560	0.1990	0.0930	0.1030
0.8	0.2870	0.3600	0.1790	0.1740	0.2920	0.4150	0.1740	0.1980
0.9	0.4340	0.6010	0.2990	0.3500	0.4490	0.6670	0.2820	0.3720

Tables 2 and 3 show the empirical power performance of the CM, CS and DF i.e. the ability of the tests to detect lack of cointegration. Interestingly the DF test performs very poorly in some cases. For certain nonlinearities the ability of the DF test to detect lack of cointegration deteriorates as the sample size increases. The performance of the CS test is quite good for both choices of the bandwidth parameter, while the CM test performs moderately well for large M.

Table 2: *No Cointegration, empirical power for* CM, CS & DF *(*$M = M1$*, 5% level)*

n	(1) CM	(1) CS	(1) DF	(2) CM	(2) CS	(2) DF	(3) CM	(3) CS	(3) DF
50	0.6180	0.6520	0.9660	0.6770	0.6880	0.3280	0.5830	0.6320	0.9930
100	0.7590	0.8660	0.9660	0.7840	0.8810	0.3010	0.7440	0.8670	1.0000
200	0.8390	0.9770	0.9900	0.8580	0.9810	0.2710	0.8260	0.9740	1.0000
300	0.8210	0.9630	0.9960	0.8470	0.9720	0.2640	0.8020	0.9630	1.0000
500	0.8620	0.9960	0.9940	0.8840	0.9930	0.2120	0.8490	0.9960	1.0000
n	(4) CM	(4) CS	(4) DF	(5) CM	(5) CS	(5) DF	(6) CM	(6) CS	(6) DF
50	0.6600	0.6840	0.4460	0.7830	0.6980	0.9470	0.6560	0.6860	0.8010
100	0.7760	0.8720	0.4020	0.8760	0.8830	0.9700	0.7740	0.8740	0.8300
200	0.8550	0.9780	0.4040	0.9300	0.9750	0.9810	0.8540	0.9800	0.8020
300	0.8420	0.9660	0.3950	0.9260	0.9710	0.9850	0.8400	0.9680	0.7730
500	0.8700	0.9930	0.3460	0.9480	0.9930	0.9880	0.8720	0.9940	0.7400

Table 3: *No Cointegration, empirical power for* CM, CS & DF *(*$M = M2$*, 5% level)*

n	(1) CM	(1) CS	(1) DF	(2) CM	(2) CS	(2) DF	(3) CM	(3) CS	(3) DF
50	0.5330	0.4900	0.9660	0.5880	0.5340	0.3280	0.5100	0.4660	0.9930
100	0.6440	0.6420	0.9660	0.6920	0.6810	0.3010	0.6190	0.6280	1.0000
200	0.7390	0.8360	0.9900	0.7700	0.8390	0.2710	0.7280	0.8210	1.0000
300	0.7450	0.8740	0.9960	0.7840	0.8760	0.2640	0.7290	0.8710	1.0000
500	0.7990	0.9390	0.9940	0.8250	0.9510	0.2120	0.7850	0.9350	1.0000
n	(4) CM	(4) CS	(4) DF	(5) CM	(5) CS	(5) DF	(6) CM	(6) CS	(6) DF
50	0.5720	0.5210	0.4460	0.5620	0.5080	0.9470	0.5700	0.5160	0.8010
100	0.6820	0.6660	0.4020	0.6700	0.6520	0.9700	0.6810	0.6660	0.8300
200	0.7630	0.8450	0.4040	0.7490	0.8390	0.9810	0.7650	0.8460	0.8020
300	0.7720	0.8770	0.3950	0.7530	0.8670	0.9850	0.7740	0.8750	0.7730
500	0.8170	0.9540	0.3460	0.8000	0.9390	0.9880	0.8180	0.9500	0.7400

The Monte Carlo results under FF misspecification are presented in Tables 4 and 5. It is obvious that the performance of the DF test is very poor in most of the cases under consideration. The CM test outperforms the CS test, when the regression residuals are dominated by some component, $u(x_t)$, that is either positive or negative. If $u(.)$ is allowed to change sign, the CS test performs better despite the fact that the CM test attains a better divergence rate. Clearly the larger the sample moment is, the better the test performs. Now if the component $u(.)$ is allowed to change sign, $u(x_t)$ will have the typical random walk type of behaviour. Lengthy periods in which the term is positive, will alternate with lengthy periods in which the term in negative, undermining the magnitude of the sample moment of the test. The CS test is more adequate in this case. The CS test adjusts the summation horizon in a way that maximises the sample moment and as result, better power performance is attained. This is confirmed by the simulation results. The underling principle behind a MOSUM test that uses fully modified residuals is similar to that of the CS test and it is therefore expected that this kind of test will also perform well. The test statistic of a MOSUM test involves a sum of regression residuals that belong to subsample windows of prespecified length and the window that maximises the sample moment of the test will be eventually opted (see Xiao and Phillips (2002) for some further discussion). Note that none of the tests are consistent when the true model is an integrable transformation of a unit root process. The inconsistency of the DF test can be explained by the results of Park and Phillips (1998). Moreover it is shown in Kasparis (2004) that in this case the CM statistic is bounded in probability under the alternative hypothesis. The regression residuals under this type of misspecification are driven by integrable components which are known to be of the same order of magnitude as a stationary process (see Park and Phillips (1999)). The fluctuation in the residuals under the alternative hypothesis is the same with that under the null hypothesis and as result none of the two tests under consideration can pick up this kind of misspecification.

Table 4: *Incorrect FF, empirical power for CM, CS & DF ($M = M1$, 5% level)*

	$(1')$	$(1')$	$(1')$	$(2')$	$(2')$	$(2')$	$(3')$	$(3')$	$(3')$
n	CM	CS	DF	CM	CS	DF	CM	CS	DF
50	0.4230	0.3570	0.1180	0.4010	0.3420	0.2820	0.1130	0.1000	0.0400
100	0.7620	0.6640	0.0090	0.5610	0.4990	0.2640	0.2510	0.2310	0
200	0.9660	0.8860	0	0.6760	0.6110	0.2600	0.4560	0.4770	0
300	0.9940	0.9490	0	0.7050	0.6560	0.2630	0.5340	0.5890	0
500	1.0000	0.9910	0	0.7820	0.7240	0.2450	0.6520	0.7730	0
	$(4')$	$(4')$	$(4')$	$(5')$	$(5')$	$(5')$	$(6')$	$(6')$	$(6')$
n	CM	CS	DF	CM	CS	DF	CM	CS	DF
50	0.4230	0.3570	0.1180	0.5000	0.3870	0.6520	0.2460	0.1920	0.1810
100	0.7620	0.6640	0.0090	0.6210	0.6550	0.6240	0.4050	0.3640	0.1470
200	0.9660	0.8860	0	0.7050	0.8460	0.6430	0.6160	0.6120	0.1380
300	0.9940	0.9490	0	0.7150	0.8740	0.6540	0.6880	0.7090	0.1550
500	1.0000	0.9910	0	0.7780	0.9440	0.6680	0.8030	0.8350	0.1410

n	(7′) CM	(7′) CS	(7′) DF	(8′) CM	(8′) CS	(8′) DF	(9′) CM	(9′) CS	(9′) DF
50	0.5480	0.4530	0.2040	0.5280	0.4460	0.8010	0.7160	0.6170	0.9970
100	0.8530	0.7510	0.0650	0.6430	0.6930	0.8520	0.8160	0.8440	1.0000
200	0.9800	0.9230	0.0080	0.7070	0.8690	0.9020	0.8890	0.9470	1.0000
300	0.9960	0.9650	0.0010	0.7230	0.8840	0.9340	0.8590	0.9460	1.0000
500	1.0000	0.9970	0	0.7840	0.9520	0.9560	0.9000	0.9890	1.0000

n	(10′) CM	(10′) CS	(10′) DF	(11′) CM	(11′) CS	(11′) DF	(12′) CM	(12′) CS	(12′) DF
50	0.3470	0.2440	0.2110	0.0870	0.6300	0.5750	0.0880	0.0660	0.0140
100	0.5590	0.5140	0.1760	0.2230	0.7810	0.5040	0.1050	0.0890	0
200	0.6690	0.7810	0.1870	0.4070	0.9300	0.4550	0.0940	0.0920	0
300	0.6800	0.8400	0.2180	0.4640	0.9750	0.4210	0.0810	0.0710	0
500	0.7600	0.9200	0.2610	0.4790	0.9950	0.3850	0.1030	0.0910	0

Table 5: *Incorrect FF, empirical power for CM, CS & DF ($M = M2$, 5% level)*

n	(1′) CM	(1′) CS	(1′) DF	(2′) CM	(2′) CS	(2′) DF	(3′) CM	(3′) CS	(3′) DF
50	0.3770	0.3120	0.1180	0.3730	0.3070	0.2820	0.0890	0.6500	0.0400
100	0.7020	0.5920	0.0090	0.5250	0.4550	0.2640	0.2170	0.1660	0
200	0.9520	0.8370	0	0.6510	0.5680	0.2600	0.4160	0.3880	0
300	0.9900	0.9250	0	0.6910	0.6280	0.2630	0.4960	0.3880	0
500	1.0000	0.9820	0	0.7620	0.6920	0.2450	0.5990	0.6680	0

n	(4′) CM	(4′) CS	(4′) DF	(5′) CM	(5′) CS	(5′) DF	(6′) CM	(6′) CS	(6′) DF
50	0.3770	0.3120	0.1180	0.4300	0.2850	0.6520	0.1950	0.1480	0.1810
100	0.7020	0.5920	0.0090	0.5370	0.4640	0.6240	0.3260	0.2880	0.1470
200	0.9520	0.8370	0	0.6150	0.6530	0.6430	0.5480	0.4990	0.1380
300	0.9900	0.9250	0	0.6240	0.7280	0.6540	0.6390	0.6220	0.1550
500	1.0000	0.9820	0	0.6890	0.8240	0.6680	0.7570	0.7630	0.1410

n	(7′) CM	(7′) CS	(7′) DF	(8′) CM	(8′) CS	(8′) DF	(9′) CM	(9′) CS	(9′) DF
50	0.5010	0.4090	0.2040	0.4540	0.3030	0.8010	0.6320	0.4740	0.9970
100	0.8080	0.6740	0.0650	0.5420	0.4830	0.8520	0.7290	0.6330	1.0000
200	0.9710	0.8790	0.0080	0.6140	0.6460	0.9020	0.8070	0.7910	1.0000
300	0.9940	0.9510	0.0010	0.6290	0.7300	0.9340	0.8080	0.8390	1.0000
500	1.0000	0.9870	0	0.6970	0.8230	0.9560	0.8500	0.9120	1.0000

	(10′)	(10′)	(10′)	(11′)	(11′)	(11′)	(12′)	(12′)	(12′)
n	CM	CS	DF	CM	CS	DF	CM	CS	DF
50	0.2880	0.1710	0.2110	0.0680	0.5950	0.5750	0.0500	0.0410	0.0140
100	0.4700	0.3820	0.1760	0.1700	0.7270	0.5040	0.0480	0.0410	0
200	0.5800	0.5970	0.1870	0.3620	0.8920	0.4550	0.0600	0.0600	0
300	0.6030	0.6940	0.2180	0.4400	0.9580	0.4210	0.0540	0.0510	0
500	0.6840	0.8070	0.2610	0.4620	0.9900	0.3850	0.0600	0.0680	0

5 Conclusion

We have considered two tests as means for testing for functional form in long-run cointegrating relations. The first test resembles the Bierens (1990) test for functional form. The second test is a cointegration test. A semiparametric approach was followed to induce limit distributions free of nuisance parameters. The limit distribution of the CM statistic is chi-square, while the limit distribution of the CS test statistic involves functionals of Brownian motion and is specific to the fitted model. We have shown that both test statistics diverge under FF misspecification or lack of cointegration and explicit asymptotic power rates have been obtained. Divergence rates are bandwidth dependent and are n/M for the first test and $\sqrt{n/M}$ for the second.

The Monte Carlo experiment carried out here suggests that both tests perform well in terms of size and power. For most type of nonlinearities considered in our simulation study, the CUSUM test performs better in terms of power than the CM test, particularly when large bandwidth parameters are used in the estimation of long-run covariance matrices. The simulation study in Hong and Phillips (2004) seems to suggest that the performance of the RESET test is comparable to performance of the tests considered here.

A finding of this paper that is of importance to empirical work, is that the DF test which is widely used as a cointegration test, performs very poorly under FF misspecification. If the DF test is applied to the residuals of a model misspecified in terms of FF, in many cases it will suggest that the residual process is stationary, when in fact it is driven by nonstationary components. Surprisengly, the DF test may also perform quite poorly under lack of cointegration, if the spurious regression is nonlinear. The work of Park and Phillips (1998) provides some useful theoretical results that can justify this.

The present theoretical framework is by no means exhaustive. A lot of specifications that are appealing for applied econometric work are not included. In order to handle the more complicated asymtotic theory resulting from the introduction of weak dependence in the error structure of the model, the theoretical framework has been confined to continuously differentiable transformations. Moreover a lot of specifications of interest are nonlinear in parameters. Extentions to these directions may prove quite challenging. Some extensions to models nonlinear in parameters, are possible and are under development by the author.

Appart from the CM, CS and RESET tests, several other FF and cointegration tests that have been developed over time that may prove to be adequate means of testing for FF in a nonstationary framework. For instance the simulation results of Kim, Lee and Newbold (2004) suggest that various FF tests can detect lack of cointegration. It will not be surprising if these tests can detect FF misspecification as well. There are many forms in which FF misspecification can manifest itself and therefore a lot of future work will be required to asses the adequacy and relative performance of all these tests.

6 Appendix

LEMMA A:

Let Assumption 2.1 hold and f H_1-regular. Then for $|h| \leq n^b$, $0 \leq b < 1$, we have

$$\frac{1}{n} k_f^{-1} \sum_{t=1}^{n} f(x_t) f'(x_{t+h}) k_f^{-1} \xrightarrow{p} \int_0^1 h_f(V(r)) h'_f(V(r)) dr, \; 1 \leq t, \; t+h \leq n,$$

as $n \to \infty$.

PROOF OF LEMMA A:

Let $\bar{n} = n - h$ and $V_{\bar{n}}(r) = \frac{1}{\sqrt{\bar{n}}} \sum_{t=1}^{[r\bar{n}]} v_t$. Now

$$\begin{aligned} \frac{1}{\bar{n}} k_f^{-1} \sum_{t=1}^{\bar{n}} f(x_t) f'(x_{t+h}) k_f^{-1} &\sim \frac{1}{\bar{n}} \sum_{t=1}^{\bar{n}} h_f\left(\frac{x_t}{\sqrt{n}}\right) h'_f\left(\frac{x_{t+h}}{\sqrt{n}}\right) \\ &\text{(from the definition of } H_1\text{-}\textit{regularity}) \\ &= \int_0^1 h_f(V_{\bar{n}}(r)) h'_f\left(V_{\bar{n}}\left(r + \frac{h}{\bar{n}}\right)\right) dr \xrightarrow{p} \int_0^1 h_f(V(r)) h'_f(V(r)) dr, \end{aligned}$$

as $n \to \infty$, from an application of the continuous mapping theorem since,

$$V_{\bar{n}}\left(r + \frac{h}{\bar{n}}\right) - V_{\bar{n}}(r) = O_p\left(\frac{h^{1/2}}{\bar{n}^{1/2}}\right) = o_p(1).$$

■

PROOF OF THEOREM 2.1:

First we start with the proof of (i) for $h = 0$. From the Beveridge-Nelson (BN) decomposition we have

$$\begin{aligned} &\frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} f(x_t) u_t \\ &= \frac{1}{\sqrt{n}} k_f(\sqrt{n}) \sum_{t=1}^{n} f(x_t) \Phi(1) \varepsilon_t - \frac{1}{\sqrt{n}} k_f(\sqrt{n}) \sum_{t=1}^{n} f(x_t) \Delta \tilde{\varepsilon}_t, \end{aligned}$$

where $\tilde{\varepsilon}_t = \sum_{j=1}^{\infty} \left(\sum_{i=j+1}^{\infty} \Phi_i \right) \varepsilon_{t-j}$. Now $\frac{1}{\sqrt{n}} k_f(\sqrt{n}) \sum_{t=1}^{n} f(x_t)\Phi(1)\varepsilon_t \xrightarrow{p} \int_0^1 h_f(V(r))dU(r)$ by Theorem 3.3 in Park and Phillips (2001). The term

$$\begin{aligned}
\frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} f(x_t)\Delta\tilde{\varepsilon}_t &= \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) f(x_n)\tilde{\varepsilon}_n - \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \Delta f(x_t)\tilde{\varepsilon}_{t-1} \\
&= O_p(n^{-1/2}) - \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}(x_{t-1} + \gamma_t v_t) v_t \tilde{\varepsilon}_{t-1},
\end{aligned}$$

where $\gamma_t = diag\left(\gamma_{1t}, .., \gamma_{pt}\right)$ with $\gamma_{it} \in [-1, 1]$.

Set $\bar{x}_{t-1} = x_{t-1} + \gamma_t v_t$, and $V_{n-1}(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[rn]} v_{t-1}$. First we will show that

$$\frac{1}{\sqrt{n}} \left\| k_f^{-1}(\sqrt{n}) \left(\sum_{t=1}^{n} \dot{f}(\bar{x}_{t-1}) v_t \tilde{\varepsilon}_{t-1} - \sum_{t=1}^{n} \dot{f}(x_{t-1}) v_t \tilde{\varepsilon}_{t-1} \right) \right\| \xrightarrow{a.s.} 0.$$

Note that $\sup_{r\in[0,1]} \|V_{n-1}(r) - V(r)\| = o_{a.s.}(s_{1n})$, for some sequence $s_{1n} \downarrow 0$, by Lemma D in Phillips (1999). Also $\sup_{r\in[0,1]} \left\|\bar{x}_{[nr]} - V(r)\right\| \le \sup_{r\in[0,1]} \|V_{n-1}(r) - V(r)\| +$ $\sup_{r\in[0,1]} \left\| \frac{v_{[nr]}}{\sqrt{n}} \right\| = o_{a.s.}(s_{2n})$ as well for some $s_{2n} \downarrow 0$, because for any $\delta > 0$ and $0 \le \mu < 1/2$,

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathbf{P} \left(\max_{1\le t\le n} \|v_t\| \ge \delta n^{1/2-\mu} \right) \\
\le\ &\sum_{n=1}^{\infty} \sum_{t=1}^{n} \mathbf{P} \left(\|v_t\| \ge \delta n^{1/2-\mu} \right) \le \sum_{n=1}^{\infty} \mathbf{E} \left(\frac{\|v_t\|^l}{\delta n^{l(1/2-\mu)-1}} \right) < \infty, \text{ for } l > 4/(1-2\mu).
\end{aligned}$$

Also set $\sup_{r\in[0,1]} \|V(r)\| = K$ and note that $K < \infty$ $a.s.$ Let s_{1n} and s_{2n} be as in condition (iii) of Definition 2.2. We therefore get

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \left\| k_f^{-1}(\sqrt{n}) \left(\sum_{t=1}^{n} \dot{f}(\bar{x}_{t-1}) v_t \tilde{\varepsilon}_{t-1} - \sum_{t=1}^{n} \dot{f}(x_{t-1}) v_t \tilde{\varepsilon}_{t-1} \right) \right\| \\
\le\ &\left\| \sqrt{n} k_f^{-1}(\sqrt{n}) \right\| \left(\sup_{r\in[0,1]} \left\| \dot{f}(\sqrt{n}\bar{x}_{[nr]}) - \dot{f}(\sqrt{n}V(r)) \right\| + \sup_{r\in[0,1]} \left\| \dot{f}(\sqrt{n}\bar{V}_n(r)) - \dot{f}(\sqrt{n}V(r)) \right\| \right) \\
&\times \frac{1}{n} \sum_{t=1}^{n} \left\| v_t \tilde{\varepsilon}_{t-1} \right\| \ a.s. \\
\le\ &\sqrt{n} \left\| k_f^{-1}(\sqrt{n}) \right\| \sup_{\|x_1\|\le K} \sup_{\|x_1 - x_2\| \le s_{1n}} \left\| \dot{f}(\sqrt{n}x_1) - \dot{f}(\sqrt{n}x_2) \right\| \times \frac{1}{n} \sum_{t=1}^{n} \|v_t \tilde{\varepsilon}_{t-1}\| \\
&+\sqrt{n} \left\| k_f^{-1}(\sqrt{n}) \right\| \sup_{\|x_1\|\le K} \sup_{\|x_1 - x_2\| \le s_{2n}} \left\| \dot{f}(\sqrt{n}x_1) - \dot{f}(\sqrt{n}x_2) \right\| \times \frac{1}{n} \sum_{t=1}^{n} \|v_t \tilde{\varepsilon}_{t-1}\|
\end{aligned}$$

$$\stackrel{a.s.}{\to} 0 \text{ as } n \to \infty,$$

given condition (iii) of Definition 2.2 and the fact that $\mathbf{E}\left\|v_t\tilde{\varepsilon}_{t-1}\right\|$ is bounded by a finite constant that can be easily checked. Therefore we have shown that

$$\frac{1}{\sqrt{n}}k_f^{-1}(\sqrt{n})\sum_{t=1}^{n} f(x_t)\Delta\tilde{\varepsilon}_t = o_p(1) - \frac{1}{\sqrt{n}}k_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})v_t\tilde{\varepsilon}_{t-1}.$$

But,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}k_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})v_t\tilde{\varepsilon}_{t-1} \\
=\;& \frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\mathbf{E}\left(v_t\tilde{\varepsilon}_{t-1}\right) - \frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\left(v_t\tilde{\varepsilon}_{t-1} - \mathbf{E}\left(v_t\tilde{\varepsilon}_{t-1}\right)\right) \\
=\;& \int_0^1 \dot{h}_f(V(r))dr\mathbf{E}\left(v_t\tilde{\varepsilon}_{t-1}\right) - \frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\left(v_t\tilde{\varepsilon}_{t-1} - \mathbf{E}\left(v_t\tilde{\varepsilon}_{t-1}\right)\right) + o_p(1).
\end{aligned}$$

In view of the fact $\mathbf{E}\left(v_t\tilde{\varepsilon}_{t-1}\right) = \Lambda_{vu}$ it would suffice to show that the last term above is asymptotically negligible, to complete the proof of part (i). This is what we set out to do next.

Consider the term

$$\begin{aligned}
&\frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\left(v_t\tilde{\varepsilon}_{t-1} - \mathbf{E}\left(v_t\tilde{\varepsilon}_{t-1}\right)\right) \\
=\;& \frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\sum_{j=0}^{\infty}\Psi_j\tilde{\Phi}_{j+1}\left(\eta_{t-j}\varepsilon_{t-j-1} - \mathbf{\Sigma}_{\eta\varepsilon}\right) \\
&+\frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\Psi_j\tilde{\Phi}_{j+r+1}\eta_{t-j}\varepsilon_{t-j-r-1} \\
&+\frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\Psi_{j+r-1}\tilde{\Phi}_j\eta_{t-j-r}\varepsilon_{t-j-1} \\
:\;& = I_{1n} + I_{2n} + I_{3n}.
\end{aligned}$$

We will show that I_{1n}, I_{2n}, and I_{3n} are asymptotically negligible. First define the lag polynomials $A(L)$, $\tilde{A}(L)$, $B_r(L)$, $\tilde{B}_r(L)$, $C_r(L)$, $\tilde{C}_r(L)$ with $r \in \mathbb{N}$. by:

$$\begin{aligned}
A(L) &= \sum_{j=0}^{\infty}A_jL^j,\ \tilde{A}(L) = \sum_{j=0}^{\infty}\tilde{A}_jL^j \\
\text{with } A_j &= \Psi_j\tilde{\Phi}_{j+1} \text{ and } \tilde{A}_j = \sum_{s=j+1}^{\infty}A_s,
\end{aligned}$$

$$
\begin{aligned}
B_r(L) &= \sum_{j=0}^{\infty} B_{rj} L^j, \ \tilde{B}_r(L) = \sum_{j=0}^{\infty} \tilde{B}_{rj} L^j \\
\text{with } B_{rj} &= \Psi_j \tilde{\Phi}_{j+1+r} \text{ and } \tilde{B}_{rj} = \sum_{s=j+1}^{\infty} B_{rs},
\end{aligned}
$$

and

$$
\begin{aligned}
C_r(L) &= \sum_{j=0}^{\infty} C_{rj} L^j, \ \tilde{C}_r(L) = \sum_{j=0}^{\infty} \tilde{C}_{rj} L^j \\
\text{with } C_{rj} &= \Psi_{j+r-1} \tilde{\Phi}_j \text{ and } \tilde{C}_{rj} = \sum_{s=j+1}^{\infty} C_{rs}.
\end{aligned}
$$

Also define ζ_{tj}, ζ_{tjr} and $\bar{\zeta}_{tjr}$ by

$$
\begin{aligned}
\zeta_{tj} &= \eta_{t-j}\varepsilon_{t-j-1} - \mathbf{\Sigma}_{\eta\varepsilon}, \\
\zeta_{tjr} &= \eta_{t-j}\varepsilon_{t-j-r-1}, \\
\bar{\zeta}_{tjr} &= \eta_{t-j-r+1}\varepsilon_{t-j-1}.
\end{aligned}
$$

We will apply BN decomposition to the lag polynomials $A(L)$, $B_r(L)$, $C_r(L)$:

$$
A(L) = A(1) - (1-L)\,\tilde{A}(L),
$$

$$
B_r(L) = B_r(1) - (1-L)\,\tilde{B}_r(L)
$$

and

$$
C_r(L) = C_r(1) - (1-L)\,\tilde{C}_r(L).
$$

We start with I_{1n}. Define $\tilde{\zeta}_t = \sum_{j=0}^{\infty} \tilde{A}_j \zeta_{tj}$. Then using BN on I_{1n}

$$
\begin{aligned}
I_{1n} &= \frac{1}{n} \dot{k}_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}(x_{t-1}) \sum_{j=0}^{\infty} A_j \zeta_{tj} \\
&= \frac{1}{n} \dot{k}_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}(x_{t-1}) A(1) \zeta_{t0} - \frac{1}{n} \dot{k}_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}(x_{t-1}) \Delta \tilde{\zeta}_t.
\end{aligned}
$$

Note that $\{\zeta_{t0}, \mathcal{F}_{t-1}\}_{t=1}^{n}$ is a martingale difference sequence. Hence $\frac{1}{n}\dot{k}_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}(x_{t-1}) A(1) \zeta_{t0}$ $= O_p(1/\sqrt{n})$ by Theorem 3.3 in Park and Phillips (2001). The second term above

$$
\begin{aligned}
&\frac{1}{n} \dot{k}_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}(x_{t-1}) \Delta \tilde{\zeta}_t \\
&= \frac{1}{n} \dot{k}_f^{-1}(\sqrt{n}) \dot{f}(x_{n-1}) \tilde{\zeta}_n - \frac{1}{n} \dot{k}_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \Delta \dot{f}(x_{t-1}) \tilde{\zeta}_{t-1} \\
&= O_p(1/n) - \frac{1}{n} \dot{k}_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \Delta \dot{f}(x_{t-1}) \tilde{\zeta}_{t-1}.
\end{aligned}
$$

Now given the H_1-*regularity* of f and the fact that $\mathbf{E}\left\|\tilde{\zeta}_{t-1}\right\| < \infty$, using the same arguments as above we can show that $\frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\Delta\dot{f}(x_{t-1})\tilde{\zeta}_{t-1} = o_p(1)$. Hence $I_{1n} = o_p(1)$.

Next will show the result for I_{2n}. First define $\zeta_t^B = \eta_t\sum_{r=1}^{\infty}B_r(1)\varepsilon_{t-r-1}$ and $\tilde{\zeta}_t^B = \sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\tilde{B}_{jr}\zeta_{tjr}$. Hence an application of the BN decomposition on I_{2n} gives

$$\begin{aligned} I_{2n} &= \frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}B_{jr}\zeta_{tjr} \\ &= \frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\zeta_t^B - \frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\Delta\tilde{\zeta}_t^B. \end{aligned}$$

Note that $\left\{\zeta_t^B, \mathcal{F}_{t-1}\right\}_{t=1}^{n}$ is a martingale difference sequence, therefore $\frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\zeta_{t-1}^B$ $= O_p(1/\sqrt{n})$ by Theorem 3.3 in Park and Phillps (2001). The second term above

$$\begin{aligned} &\frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\dot{f}(x_{t-1})\Delta\tilde{\zeta}_t^B \\ =\;& \frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\dot{f}(x_{n-1})\tilde{\zeta}_n^B - \frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\Delta\dot{f}(x_{t-1})\tilde{\zeta}_{t-1}^B \\ =\;& O_p(1/n) - \frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\Delta\dot{f}(x_{t-1})\tilde{\zeta}_{t-1}^B. \end{aligned}$$

Now note that $\mathbf{E}\left\|\tilde{\zeta}_{t-1}^B\right\| < \infty$ because,

$$\begin{aligned} \mathbf{E}\left\|\tilde{\zeta}_{t-1}^B\right\| &= \mathbf{E}\left\|\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\tilde{B}_{jr}\zeta_{tjr}\right\| \le \lim\inf_{s_1\to\infty}\lim\inf_{s_2\to\infty}\mathbf{E}\left\|\sum_{r=1}^{s_1}\sum_{j=0}^{s_2}\tilde{B}_{jr}\zeta_{tjr}\right\| \\ &\le \sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\mathbf{E}\left\|\tilde{B}_{jr}\zeta_{tjr}\right\| \le \mathbf{E}\left\|\zeta_{tjr}\right\|\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty}\left\|\Psi_s\tilde{\Phi}_{s+r}\right\| \\ &= \mathbf{E}\left\|\zeta_{tjr}\right\|\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\sum_{j=0}^{s-1}\left\|\Psi_s\tilde{\Phi}_{s+r}\right\| \le \mathbf{E}\left\|\zeta_{tjr}\right\|\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\left\|s\Psi_s\right\|\left\|\tilde{\Phi}_{s+r}\right\| \\ &\le \mathbf{E}\left\|\zeta_{tjr}\right\|\left(\sum_{s=1}^{\infty}\left\|s\Psi_s\right\|\right)\left(\sum_{r=1}^{\infty}\left\|\tilde{\Phi}_r\right\|\right) < \infty. \end{aligned}$$

Therefore using the same arguments as before it follows that $\frac{1}{n}\dot{k}_f^{-1}(\sqrt{n})\sum_{t=1}^{n}\Delta\dot{f}(x_{t-1})\tilde{\zeta}_{t-1}^B = o_p(1)$ as well, which establishes that $I_{2n} = o_p(1)$. The proof that I_{3n} is negligible is the same with the proof for I_{2n} and therefore omitted.

Next we will consider the case $0 < h < n^b$. The proof for $-n^b < h < 0$ is the same and therefore will be omitted. For $1 \leq t,\ t+h \leq n$, note that

$$\begin{aligned}
& \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} f(x_{t+h}) u_t \\
= \ & \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} f(x_t) u_t + \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \left(f(x_{t+h}) - f(x_t)\right) u_t \\
: \ & = I_{4n} + I_{5n}.
\end{aligned}$$

Next we find that

$$I_{4n} \xrightarrow{p} \int_0^1 h_f(V(r)) dU(r) + \int_0^1 \dot{h}_f(V(r)) dr \Lambda_{vu}.$$

On the other hand an application of the mean value theorem to I_{5n} gives:

$$\begin{aligned}
I_{5n} &= \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}\left(x_t + \gamma_t \sum_{i=1}^{h} v_{t+i}\right) \sum_{i=1}^{h} v_{t+i} u_t \\
&= \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}\left(x_t\right) \sum_{i=1}^{h} v_{t+i} u_t + o_p(1),
\end{aligned}$$

where the last equality can be established as follows. Setting $\bar{x}_t^h = x_t + \gamma_t \sum_{i=1}^{h} v_{t+i}$ and $\bar{n} = n - h$ we can show that $\sup_{r\in[0,1]} \left\| \bar{x}_{[\bar{n}r]}^h - V(r) \right\| = o_{a.s.}(s_{1n})$ and $\sup_{r\in[0,1]} \| V_{\bar{n}}(r) - V(r) \| = o_{a.s.}(s_{2n})$ for some sequences s_{1n}, s_{2n} that decrease to zero. Therefore

$$\left\| \frac{1}{\sqrt{n}} k_f^{-1}\left(\sqrt{n}\right) \sum_{t=1}^{n} \left[\dot{f}\left(x_t + \gamma_t \sum_{i=1}^{h} v_{t+i} \right) - \dot{f}\left(x_t\right) \right] \sum_{i=1}^{h} v_{t+i} u_t \right\|$$

$$\leq \left\| n^{1/2+b} k_f^{-1}\left(\sqrt{n}\right) \right\| \left(\sup_{r\in[0,1]} \left\| \dot{f}(\sqrt{n}\bar{x}_{[\bar{n}r]}^h) - \dot{f}(\sqrt{n}V(r)) \right\| + \sup_{r\in[0,1]} \left\| \dot{f}(\sqrt{n}V_{\bar{n}}(r)) - \dot{f}(\sqrt{n}V(r)) \right\| \right)$$

$$\times \frac{1}{n^{1+b}} \sum_{t=1}^{n} \sum_{i=1}^{h} \| v_{t+i} u_t \| \xrightarrow{a.s.} 0 \text{ as } n \to \infty,$$

given condition (iii) of Definition 2.2 and the fact that $\mathbf{E} \| v_{t+i} u_{t-1} \|$ is bounded by a finite constant. Therefore we get

$$I_{5n} = \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}\left(x_t\right) \sum_{i=1}^{h} \left(v_{t+i} u_t\right) + o_p(1).$$

Moreover,

$$
\begin{aligned}
I_{5n} &= \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}\left(x_t\right) \sum_{i=1}^{h} \mathbf{E}\left(v_{t+i} u_t\right) \\
& \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}\left(x_t\right) \sum_{i=1}^{h}\left(v_{t+i} u_t-\mathbf{E}\left(v_{t+i} u_t\right)\right)+o_p(1) \\
&= \int_0^1 \dot{h}_f(V(r)) d r \sum_{i=1}^{h} \mathbf{E}\left(v_{t+i} u_t\right)+o_p(1)
\end{aligned}
$$

where the last equality can be established by using higher order BN decomposition. Combining the limit results for I_{4n} and I_{5n} and given the fact that $\Lambda_{vu}(h) = \sum_{i=1}^{h} \mathbf{E}\left(v_{t+i} u_t\right) + \Lambda_{vu}$ the result follows.

For part (ii) first we write

$$
\frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} f(x_t) v_t' = \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} f(x_{t-1}) v_t' + \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \left(f(x_t)-f(x_{t-1})\right) v_t'.
$$

Let $\tilde{\eta}_t = \sum_{j=1}^{\infty}\left(\sum_{i=j+1}^{\infty} \Psi_i\right) \eta_{t-j}$. Now applying BN decomposition to the first term above and using the same arguments with those in part (i) we get

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} f(x_{t-1}) v_t' \\
= & \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} f(x_{t-1}) \eta_t' \Psi'(1) + \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}(x_{t-1}) \mathbf{E}\left(v_{t-1} \tilde{\eta}_{t-1}'\right) + o_p(1) \\
= & \int_0^1 h_f(V(r)) d V'(r) + \int_0^1 \dot{h}_f(V(r)) d r \mathbf{E}\left(v_{t-1} \tilde{\eta}_{t-1}'\right) + o_p(1).
\end{aligned}
$$

Now the term

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \left(f(x_t)-f(x_{t-1})\right) v_t' \\
= & \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}(x_{t-1}) \mathbf{E}(v_t v_t') + \frac{1}{\sqrt{n}} k_f^{-1}(\sqrt{n}) \sum_{t=1}^{n} \dot{f}(x_{t-1}) \left(v_t v_t' - \mathbf{E}(v_t v_t')\right) \\
= & \int_0^1 \dot{h}_f(V(r)) d r \mathbf{E}(v_t v_t') + o_p(1)
\end{aligned}
$$

and in view of the fact that $\Lambda_{vv} = \mathbf{E}(v_t v_t') + \mathbf{E}\left(v_t \tilde{\eta}_t'\right)$, this completes the proof for the case $h=0$. For $h \neq 0$ the proof is the same with that for part (i) and therefore will be omitted.■

PROOF OF LEMMA 3.1:
Note that

$$\sqrt{n}k_g\left(\hat{a}-a^*\right)=\left[\frac{1}{n}k_g^{-1}\sum_{t=1}^{n}g(x_t)g'(x_t)k_g^{-1}\right]^{-1}\frac{k_g^{-1}}{\sqrt{n}}\left[\sum_{t=1}^{n}g(x_t)u_t^+-\dot{g}_n\hat{\Lambda}_{vu}^+\right].$$

Now,

$$\frac{1}{n}k_g^{-1}\sum_{t=1}^{n}g(x_t)g'(x_t)\dot{k}_g^{-1}\xrightarrow{p}\int_0^1 h_g\left(V(r)\right)h_g'\left(V(r)\right)dr \tag{A1}$$

and

$$\frac{k_g^{-1}}{\sqrt{n}}\dot{g}_n\hat{\Lambda}_{vu}^+\xrightarrow{p}\int_0^1\dot{h}_g(V(r))dr\Lambda_{vu}^+. \tag{A2}$$

Also,

$$\frac{k_g^{-1}}{\sqrt{n}}\sum_{t=1}^{n}g(x_t)u_t^+=\frac{k_g^{-1}}{\sqrt{n}}\sum_{t=1}^{n}g(x_t)u_t-\frac{k_g^{-1}}{\sqrt{n}}\sum_{t=1}^{n}g(x_t)v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}.$$

The first term

$$\frac{k_g^{-1}}{\sqrt{n}}\sum_{t=1}^{n}g(x_t)u_t=\int_0^1 h_g\left(V(r)\right)dU(r)+\int_0^1\dot{h}_g(V(r))dr\Lambda_{vu}+o_p(1),$$

by Theorem 2.1(i). The second term on the right hand side above is

$$\begin{aligned}&\frac{k_g^{-1}}{\sqrt{n}}\sum_{t=1}^{n}g(x_t)v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\\ =&\int_0^1 h_g\left(V(r)\right)d\left(V'(r)\Omega_{vv}^{-1}\Omega_{vu}\right)+\int_0^1\dot{h}_g(V(r))dr\Lambda_{vv}\Omega_{vv}^{-1}\Omega_{vu}+o_p(1),\end{aligned}$$

by Theorem 2.1 (ii). Hence

$$\frac{k_g^{-1}}{\sqrt{n}}\sum_{t=1}^{n}g(x_t)u_t^+\xrightarrow{p}\int_0^1 h_g\left(V(r)\right)dU^+(r)+\int_0^1\dot{h}_g(V(r))dr\left(\Lambda_{vu}-\Lambda_{vv}\Omega_{vv}^{-1}\Omega_{vu}\right) \tag{A3}$$

In view of (A1), (A2) and (A3) the result follows.■

PROOF OF THEOREM 3.1:

For CM note that

$$CM_n=\frac{\left[\sum_{t=1}^{n}\left(g_t'\left(\hat{a}-a_o\right)-u_t+v_t'\Omega_{vv}^{-1}\Omega_{vu}\right)\right]^2}{\left(\Omega_{uu}-\Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}\right)\sum_{t=1}^{n}\left[\hat{A}_n'\hat{B}_n^{-1}g_t-1\right]^2}+o_p(1)$$

$$= \frac{\left[\frac{1}{n}\sum_{t=1}^{n} g_t' k_g^{-1}\left[\frac{1}{n}k_g^{-1}\sum_{i=1}^{n} g_i g_i' k_g^{-1}\right]^{-1} k_g^{-1}\frac{1}{\sqrt{n}}\left[\sum_{j=1}^{n} g_j u_j^+ - \dot{g}_n \hat{\Lambda}_{vu}^+\right] - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(u_t + v_t'\Omega_{vv}^{-1}\Omega_{vu}\right)\right]^2}{\left(\Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}\right)\frac{1}{n}\sum_{t=1}^{n}\left[\hat{A}_n' k_g^{-1} k_g \hat{B}_n^{-1} k_g k_g^{-1} g_t - 1\right]^2}$$

$$= \frac{\left[\int_0^1 \left[A'B^{-1}h_g(V(r)) - 1\right] dU^+(r)\right]^2}{\left(\Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}\right)\int_0^1 \left[A'B^{-1}h_g(V(r)) - 1\right]^2 dr} + o_p(1),$$

where the last line is due to Theorem 2.1. In view of the fact that V and U^+ are independent the result follows.

For CS with $0 \leq s \leq 1$,

$$CS_n = \frac{\sup_{0\leq s\leq 1}\frac{1}{\sqrt{n}}\left|\sum_{t=1}^{[sn]}\left(g_t'\left(\hat{a} - a_o\right) - u_t + v_t'\Omega_{vv}^{-1}\Omega_{vu}\right)\right|}{\sqrt{\Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}}} + o_p(1)$$

The numerator of the above expression

$$\sup_{0\leq s\leq 1}\left|\frac{1}{n}\sum_{t=1}^{[sn]} g_t' k_g^{-1}\left[\frac{1}{n}k_g^{-1}\sum_{i=1}^{n} g_i g_i' k_g^{-1}\right]^{-1} k_g^{-1}\frac{1}{\sqrt{n}}\left[\sum_{j=1}^{n} g_j u_j^+ - \dot{g}_n\hat{\Lambda}_{vu}^+\right] - \frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left(u_t + v_t'\Omega_{vv}^{-1}\Omega_{vu}\right)\right|$$

$$= \sup_{0\leq s\leq 1}\left|\bar{U}(s)\right| + o_p(1)$$

by Theorem 2.1 and the result follows.■

Proof of Lemma 3.2:

We start with the proof of part (i). The arguments we use are similar to the ones of Phillips (1991). Under incorrect FF the LS estimator can be written as

$$\frac{k_g}{k_{d^*}}\left(\hat{a}_{LS} - \theta_o\right) = \left[\frac{1}{n}k_g^{-1}\sum_{t=1}^{n} g_t g_t' k_g^{-1}\right]^{-1}\frac{1}{nk_{d^*}}\left[\sum_{t=1}^{n} g_t d_t'\theta_o + k_g^{-1}\sum_{t=1}^{n} g_t u_t\right]$$

Hence by Theorem 2.1

$$\frac{k_g}{k_{d^*}}\left(\hat{a}_{LS} - \theta_o\right) = \left[\int_0^1 h_g\left(V(r)\right) h_g'\left(V(r)\right) dr\right]^{-1}\int_0^1 h_g\left(V(r)\right) h_{d^*}'\left(V(r)\right)\theta_o dr + O_p(1/\sqrt{n}k_{d^*})$$

$$= \zeta_1 + o_p(1).$$

Define the normalising matrix $N_{d^*,n} = \begin{pmatrix} I_p & \mathbf{0} \\ \mathbf{0} & k_g/k_{d^*} \end{pmatrix}$. In what follows the regression residuals (from OLS estimation) will be written in the following form:

$$\begin{aligned} \hat{u}_t &= f_t'\theta_o - g_t'\hat{a} + u_t \\ &= d_t'\theta_o - g_t'\left(\hat{a} - \theta_o\right) + u_t \end{aligned}$$

Hence

$$\begin{aligned}
\frac{n^{1/2}}{Mk_{d^*}}\hat{\Omega}_{vu} &= \frac{1}{Mk_{d^*}}\sum_{h=-M}^{M}\kappa\left(\frac{h}{M}\right)\left(\frac{1}{n}\sum_{t=1}^{n}v_t\left(\begin{array}{cc} d'_{t+h} & g'_{t+h}\end{array}\right)N_{d^*,n}^{-1}N_{d^*,n}\left(\begin{array}{c}\theta_o\\ -(\hat{a}-\theta_o)\end{array}\right)\right)\\
&\quad+\frac{n^{1/2}}{Mk_{d^*}}\Omega_{vu}+o_p(1)\\
&= 2\pi K(0)\int_0^1 dV(r)\left(h'_{d^*}(V(r))\theta_o - h'_g(V(r))\zeta_1\right)\\
&\quad+\Omega_{vv}\int_0^1\left[\dot{h}'_{d^*}(V(r))\theta_o-\dot{h}'_g(V(r))\zeta_1\right]dr+\frac{n^{1/2}}{Mk_{d^*}}\Omega_{vu}+o_p(1)\\
&= 2\pi K(0)\int_0^1 dV(r)\bar{h}'_1(V(r))\bar{\zeta}_1+\Omega_{vu}\int_0^1\dot{H}'_1(V(r))\bar{\zeta}_1 dr+\frac{n^{1/2}}{Mk_{d^*}}\Omega_{vu}+o_p(1),
\end{aligned}$$

where the second equality is by Theorem 2.1 with $n^b = M$. Therefore

$$\hat{\Omega}_{vu} = O_p\left(\frac{Mk_{d^*}}{n^{1/2}}\right)+O_p(1). \tag{A4}$$

Now using similar arguments as above it turns out that

$$\frac{n^{1/2}}{Mk_{d^*}}\hat{\Lambda}_{vu} = 2\pi K_1(0)\int_0^1 dV(r)\bar{h}'_1(V(r))\bar{\zeta}_1+\Omega_{vv}\int_0^1\dot{H}'_1(V(r))\bar{\zeta}dr+\frac{n^{1/2}}{Mk_{d^*}}\Lambda_{vu}+o_p(1)$$

So,

$$\hat{\Lambda}_{vu} = O_p\left(\frac{Mk_{d^*}}{n^{1/2}}\right) \tag{A5}$$

Next we will find the order of $\hat{\Omega}_{uu}$. Note that

$$\hat{\Omega}_{uu} = \sum_{h=-M}^{M}\kappa\left(\frac{h}{M}\right)C_{uu}(h)$$

and

$$\begin{aligned}
\frac{1}{k_{d^*}^2}C_{uu}(h) &= \frac{1}{nk_{d^*}^2}\sum_{t=1}^{n}\left(\begin{array}{cc}\theta_o & -(\hat{a}'-\theta'_o)\frac{k_g}{k_{d^*}}\end{array}\right)N_{d^*,n}^{-1}\left(\begin{array}{cc} d_t d'_{t+h} & d_t g'_{t+h}\\ g_t d'_{t+h} & g_t g'_{t+h}\end{array}\right)N_{d^*,n}^{-1}\left(\begin{array}{c}\theta_o\\ -\frac{k_g}{k_{d^*}}(\hat{a}-\theta_o)\end{array}\right)\\
&+\frac{1}{nk_{d^*}^2}\sum_{t=1}^{n}\left(\begin{array}{cc}\theta'_o & -(\hat{a}'-\theta'_o)\frac{k_g}{k_{d^*}}\end{array}\right)\left(\begin{array}{c} d_t\\ g_t\end{array}\right)u_{t+h}+\frac{1}{nk_{d^*}^2}\sum_{t=1}^{n}\left(\begin{array}{cc} d'_{t+h} & g'_{t+h}\end{array}\right)\left(\begin{array}{c}\theta_o\\ -(\hat{a}-\theta_o)\end{array}\right)u_t\\
&= \int_0^1\bar{\zeta}'_1\bar{h}_1(V(r))\bar{h}'_1(V(r))\bar{\zeta}_1 dr\\
&\quad+\frac{1}{\sqrt{n}k_{d^*}}\left\{2\int_0^1\bar{\zeta}'_1\bar{h}_1(V(r))dU(r)+\int_0^1\bar{\zeta}'_1\dot{H}_1(V(r))dr\left(\Lambda_{vu}(-h)+\Lambda_{vu}(h)\right)\right\}+o_p(1),
\end{aligned}$$

where the last line is due to Lemma A and Theorem 2.1. Hence,

$$\frac{1}{Mk_{d^*}^2}\hat{\Omega}_{uu} = 2\pi K(0)\int_0^1 \bar{\zeta}_1'\bar{h}_1(V(r))\bar{h}_1'(V(r))\bar{\zeta}_1 dr + O_p\left(\frac{1}{M\sqrt{n}k_{d^*}}\right)$$

Hence we have

$$\hat{\Omega}_{uu} = O_p\left(Mk_{d^*}^2\right) + O_p\left(\frac{k_{d^*}}{\sqrt{n}}\right) = O_p\left(Mk_{d^*}^2\right) \qquad \text{(A6)}$$

Consequently under incorrect FF (A4) and (A6) give

$$\begin{aligned}\hat{\Omega}_{uu} - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu} &= O_p\left(Mk_{d^*}^2\right) + O_p\left(\frac{M^2k_{d^*}^2}{n}\right)\\ &= O_p\left(Mk_{d^*}^2\right) \text{ when } M/n \to 0 \text{ as } n \to \infty. \qquad \text{(A7)}\end{aligned}$$

For (ii) without loss of generality partition $g' = (f^{1\prime},\ g^{2\prime})$ and $k_g = diag\,(k_{f^1},\ k_{g^2})$. Using results for partitioned matrices it follows after some lengthy but straightforward algebraic manipulations that

$$\frac{k_g}{k_{f^{2*}}}\left(\hat{a}_{LS} - \bar{\theta}_o\right) = \begin{pmatrix} P_n^1 \\ P_n^2 \end{pmatrix}\theta_o^2 + O_p(1/k_{f^{2*}}\sqrt{n})$$

where $P_n^1 \xrightarrow{p} P^1$ and $P_n^2 \xrightarrow{p} P^2$ with

$$P^1 = \left(\int_0^1 h_{f^{2*}}(V)h_{f^1}(V)' - \int_0^1 h_{f^{2*}}(V)h_{g^2}(V)'\left(\int_0^1 h_{g^2}(V)h_{g^2}(V)'\right)^{-1}\int_0^1 h_{g^2}(V)h_{f^1}(V)'\right)(P^3)^{-1},$$

$$\begin{aligned}P^2 = &\int_0^1 h_{f^{2*}}(V)h_{g^2}(V)'\left(\int_0^1 h_{g^2}(V)h_{g^2}(V)'\right)^{-1}\\ &+\left(\int_0^1 h_{f^{2*}}(V)h_{g^2}(V)'\left(\int_0^1 h_{g^2}(V)h_{g^2}(V)'\right)^{-1}\int_0^1 h_{g^2}(V)h_{f^1}(V)' - \int_0^1 h_{f^{2*}}(V)h_{f^1}(V)'\right)\\ &\times (P^3)^{-1}\int_0^1 h_{f^1}(V)h_{g^2}(V)'\left(\int_0^1 h_{g^2}(V)h_{g^2}(V)'\right)^{-1}\end{aligned}$$

and

$$P^3 = \int_0^1 h_{f^1}(V)h_{f^1}(V)' - \int_0^1 h_{f^1}(V)h_{g^2}(V)'\left(\int_0^1 h_{g^2}(V)h_{g^2}(V)'\right)^{-1}\int_0^1 h_{g^2}(V)'h_{f^1}(V).$$

Setting $\zeta_2' = \theta_o^{2\prime}\left(P^{1\prime}, P^{2\prime}\right)$ the LS residuals

$$\begin{aligned}&\frac{1}{nk_{f^{2*}}}\sum_{t=1}^n (y_t - g(x_t)'\hat{a}_{LS}) = \frac{1}{nk_{f^{2*}}}\sum_{t=1}^n f_t'\theta_o - \frac{1}{nk_{f^{2*}}}\sum_{t=1}^n g(x_t)'\hat{a}_{LS} + o_p(1)\\ =\ &\frac{1}{nk_{f^{2*}}}\sum_{t=1}^n f_t^{2\prime}\theta_o^2 - \frac{1}{n}\sum_{t=1}^n g(x_t)'k_g^{-1}\frac{k_g}{k_{f^{2*}}}\left(\hat{a}_{LS} - \bar{\theta}_o\right) + o_p(1)\\ =\ &\int_0^1 \left(h_{f^{2*}}'\left(V(r)\right)\theta_o - h_g'\left(V(r)\right)\zeta_2\right)dr + o_p(1) = \int_0^1 \bar{h}_2'\left(V(r)\right)\bar{\zeta}_2 dr + o_p(1).\end{aligned}$$

Now similar arguments as those above give

$$\frac{n^{1/2}}{Mk_{f^{2*}}}\hat{\Omega}_{vu} = 2\pi K(0)\int_0^1 dV(r)\bar{h}_2'(V(r))\bar{\zeta}_2 + \Omega_{vv}\int_0^1 \dot{H}_2'(V(r))\bar{\zeta}_2 dr + O_p\left(\frac{1}{M\sqrt{n}k_{f^*}}\right),$$

$$\frac{n^{1/2}}{Mk_{f^{2*}}}\hat{\Lambda}_{vu} = 2\pi K_1(0)\int_0^1 dV(r)\bar{h}_2'(V(r))\bar{\zeta}_2 + \Omega_{vv}\int_0^1 \dot{H}_2'(V(r))\bar{\zeta}_2 dr + O_p\left(\frac{1}{M\sqrt{n}k_{f^*}}\right)$$

and

$$\frac{1}{Mk_{f^{2*}}^2}\hat{\Omega}_{uu} = 2\pi K(0)\int_0^1 \bar{\zeta}_2'\bar{h}_2(V(r))\bar{h}_2'(V(r))\bar{\zeta}_2 dr + O_p\left(\frac{1}{M\sqrt{n}k_{f^*}}\right).$$

The proof for (iii) is similar to that of (i) and (ii) and therefore omitted.■

Proof of Theorem 3.2:

We will prove the result under incorrect FF when **S1** holds. The proof for the other cases is similar and will be omitted. Note that the FM-LS estimator

$$\hat{a} = \left[\sum_{t=1}^n g(x_t)g'(x_t)\right]^{-1}\left[\sum_{t=1}^n g(x_t)y_t^+ - \dot{g}_n\hat{\Lambda}_{vu}^+\right],$$

rearranging

$$\begin{aligned}\frac{k_g}{k_{d^*}}(\hat{a}-\theta_o) &= \left[\frac{1}{n}k_g^{-1}\sum_{t=1}^n g_t g_t' k_g^{-1}\right]^{-1} \\ &\quad \times\frac{1}{nk_{d^*}}k_g^{-1}\left[\sum_{t=1}^n g_t d_t'\theta_o + \sum_{t=1}^n g_t u_t - \sum_{t=1}^n g_t v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu} - \dot{g}_n\hat{\Lambda}_{vu}^+\right] \\ &= \zeta_1 + O_p\left(\frac{1}{\sqrt{n}k_{d^*}}\right)\end{aligned}$$

where we have used the fact that $\hat{\Lambda}_{vu}, \hat{\Omega}_{vu} = O_p\left(\frac{Mk_{d^*}}{\sqrt{n}}\right)$ (equations (A4),(A5)).

Recall that the CM test statistic is

$$CM_n = \frac{\left[\sum_{t=1}^n\left(y_t^+ - g(x_t)'\hat{a} - v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right)\right]^2}{\left(\hat{\Omega}_{uu} - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right)\sum_{t=1}^n\left[\hat{A}_n'\hat{B}_n^{-1}g(x_t) - 1\right]^2},$$

Consider first the numerator rescaled by $(nk_{d^*})^2$:

$$\begin{aligned}&\left[\frac{1}{nk_{d^*}}\sum_{t=1}^n\left\{\left(y_t^+ - g(x_t)'\hat{a}\right) - v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right\}\right]^2 \\ =\ &\left[\frac{1}{nk_{d^*}}\sum_{t=1}^n\left\{\begin{pmatrix} d_t' & g_t'\end{pmatrix} N_{d^*,n}^{-1}N_{d^*,n}\begin{pmatrix}\theta_o \\ -(\hat{a}-\theta_o)\end{pmatrix} + \left(u_t - v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right)\right\}\right]^2 \\ =\ &\left[\int_0^1 \bar{h}_1'(V(r))\bar{\zeta}_1 dr + O_p\left(\frac{1}{\sqrt{n}k_{d^*}}\right) + O_p\left(\frac{M}{n}\right)\right]^2. \qquad \text{(A8)}\end{aligned}$$

Now the the denominator rescaled by n

$$\begin{aligned}
& \left(\hat{\Omega}_{uu}-\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right)\frac{1}{n}\sum\nolimits_{t=1}^{n}\left[\hat{A}_n'\hat{B}_n^{-1}g(x_t)-1\right]^2 \\
= & \left(\hat{\Omega}_{uu}-\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right)\int_0^1\left[A'B^{-1}h_g(V(r))-1\right]^2dr+o_p(1) \\
= & O_p\left(Mk_{d^*}^2\right), \qquad \text{(A9)}
\end{aligned}$$

where the last lines due to equation (A7). In view of (A8) and (A9)

$$CM_n\sim(n/M)\left[\int_0^1\bar{h}_1'(V(r))\bar{\zeta}_1dr\right]^2,$$

which gives the requisite result.

For CS test note that the numerator of test statistic rescaled by $\sqrt{n}k_{d^*}$

$$\begin{aligned}
& \frac{\max_{1\le k\le n}}{nk_{d^*}}\left|\sum\nolimits_{t=1}^{k}\left\{\begin{pmatrix} d_t' & g_t'\end{pmatrix}N_{d^*,n}^{-1}N_{d^*,n}\begin{pmatrix}\theta_o \\ -(\hat{a}-\theta_o)\end{pmatrix}+\left(u_t-v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right)\right\}\right| \\
= & \sup_{0\le s\le 1}\left|\int_0^s\bar{h}_1'(V(r))\bar{\zeta}_1dr+O_p\left(\frac{1}{\sqrt{n}k_{d^*}}\right)+O_p\left(\frac{M}{n}\right)\right|.
\end{aligned}$$

The denominator rescaled by $\sqrt{n}$ is

$$\sqrt{\left(\hat{\Omega}_{uu}-\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right)/n}=O_p\left(\sqrt{M}k_{d^*}\right).$$

Therefore

$$CS_n\sim(n/M)^{1/2}\sup_{0\le s\le 1}\left|\int_0^s\bar{h}_1'(V(r))\bar{\zeta}_1dr,\right|$$

which completes the proof.■

PROOF OF LEMMA 3.3:

We will show that the result under FFM when C1 holds. The proof for the other cases is similar and therefore omitted. Denote by $u(x_t)$ the regressions residuals from FM-LS estimation and without loss of generality assume that x_t is scalar. From the proof of Lemma 3.2 (equation (12)) we have that $(nk_d)^{-1}\sum_{t=1}^{n}u(x_t)=\int_0^1\bar{h}_1'(V(r))\bar{\zeta}_1dr+o_p(1)=\int_0^1h_u(V(r))dr+o_p(1)$ and similarly define $\dot{u}(x_t)$, $\ddot{u}(x_t)$, $\dot{h}_u$ and $\ddot{h}_u$. First consider

$$\begin{aligned}
\hat{\rho}^2-1 &= \frac{\left(\sum_{t=2}^{n}u(x_t)u(x_{t-1})\right)^2-\left(\sum_{t=2}^{n}u(x_{t-1})^2\right)^2}{\left\{\sum_{t=2}^{n}u(x_{t-1})^2\right\}^2} \\
&= \frac{\left\{\sum_{t=2}^{n}u(x_t)\left(u(x_t)+u(x_{t-1})\right)\right\}\left\{\sum_{t=2}^{n}u(x_t)\left(u(x_t)-u(x_{t-1})\right)\right\}}{\left\{\sum_{t=2}^{n}u(x_{t-1})^2\right\}^2}
\end{aligned}$$

Hence by Lemma A and Theorem 2.1 we have

$$\begin{aligned}\frac{\sqrt{n}k_d}{k_d}\left(\hat{\rho}^2-1\right) \sim & \left\{\int_0^1 h_u\left(V(r)\right)^2 dr\right\}^{-1} \times \\ & \left\{\int_0^1 h_u(V(r))^2 dV(r)+\int_0^1 \left(\dot{h}_u(V(r))^2 dr + h_u(V(r))\ddot{h}_u\left(V(r)\right)\right) dr\Lambda_{vv}\right\}.\end{aligned}$$

Since $k_d/k_{\dot{d}}=\sqrt{n}$,

$$\left(\hat{\rho}^2-1\right)^2=O_p(1/n^2).$$

Hence

$$\hat{\rho}^2/\left(\hat{\rho}^2-1\right)^2=O_p(n)$$

and therefore

$$\left(\hat{\delta}n\right)^{1/3}=O_p(n),$$

as required.■

7 References

ANDREWS, D.W.K. (1991): “Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817-858.

BIERENS, H.J. (1990): “A Consistent Test of Functional Form”, *Econometrica* 58, 1443-1458.

CHANG, Y., PARK J.Y. AND P.C.B. PHILLIPS (2001): “Nonlinear Econometric Models with Cointegrated and Deterministically Trending Regressors,” *Econometrics Journal*, 4.

CORRADI, V., SWANSON N.R. AND H. WHITE (2000): “Testing for Stationarity-Ergodicity and Comovement between Nonlinear Discrete Time Markov Processes,” *Journal of Econometrics*, 96, 39-73.

DE JONG, R.M. (2002): “Nonlinear Estimators with Integrated Regressors but without Exogeneity,” Working Paper, Department of Economics, Michigan State University.

——— (2004): “Addendum to "Asymptotics for Nonlinear Transformations of Integrated Time Series",” *Econometric Theory,* 20.

TERÄSVIRTA, T. AND A. ELIASSON (2001): “Nonlinear Error Correction Model and the UK Demand for Broad Money,” *Journal of Applied Econometrics*, 16, 277-288.

HONG, S. H. AND P.C.B. PHILLIPS (2004): “Testing Linearity in Cointegrating Relations with an Application to PPP,” *Mimeo*, Yale University.

KASPARIS, I. (2004): “Functional Form Misspecification in Regressions with Integrated Time Series,” Ph.D. Thesis, University of Southampton.

Kim, T-H., Lee, Y-S. and P. Newbold (2004): "Spurious Nonlinear Regressions in Econometrics,", *Economics Letters* (Forthcoming).

Park, J.Y. and P.C.B. Phillips (1998): "Unit Roots in Nonlinear Transformations of Integrated Time Series," *Mimeo* Yale University.

——— (1999): "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory,* 15, 269-298.

——— (2001): "Nonlinear Regressions with Integrated Time Series," *Econometrica,* 69, 117-161.

P.C.B. Phillips. (1986): "Time Series Regression with a Unit Root," *Econometrica,* 55, 277-301.

——— (1988): "Weak Convergence to the Matrix Stochastic Integral $\int_0^1 BdB'$," *Journal of Multivariate Analysis*, 24, 252-264.

——— (1991): "Spectral Regression for Cointegrated Time Series," in *Nonparametric and semiparametric methods in econometrics and statistics*, edited by W. Barnett, J. Powell and G. Tauchen. Cambridge: Cambridge University Press.

——— (1995): "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, 63, 1023-1078.

——— (1999): "Unit Root Log Periodogram Regression," Cowles Foundation Discussion Paper 1244.

P.C.B. Phillips. and B. E. Hansen (1990): "Statistical Inference in Instrumental Variables Regressions with I(1) Processes," *Review of Economic Studies*, 53, 473-496.

P.C.B. Phillips. and V. Solo (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971-1001.

Pötscher B.M. (2004): "Nonlinear Functions and Convergence to Brownian Motion: Beyond the Continuous Mapping Theorem," *Econometric Theory,* 20.

Ramsey, J.B. (1969): "Classical Model Selection Through Specification Error Tests," *Journal of the Royal Statistical Society,* Series B, 32, 350-371.

Saikkonen, P. and I. Choi (2004): "Cointegrating Smooth Transition Regressions," *Econometric Theory*, 20, 301-340.

White, H. (1981): "Consequences and Detection of Misspecified Nonlinear Regression Models," *Journal of American Statisitcal Association*, 76, 419-433.

Xiao, Z. and P.C.B. Phillips (2002): "A CUSUM Test for Cointegration Using Regression Residuals," *Journal of Econometrics*, 108, 43-61.