DEPARTMENT OF ECONOMICS
UNIVERSITY OF CYPRUS

DETECTION OF FUNCTIONAL FORM MISSPECIFICATION IN COINTEGRATING RELATIONS

Ioannis Kasparis

Discussion Paper 2007-05

P.O. Box 20537, 1678 Nicosia, CYPRUS Tel.: ++357-2-892430, Fax: ++357-2-892432
Web site: http://www.econ.ucy.ac.cy

DETECTION OF FUNCTIONAL FORM MISSPECIFICATION IN COINTEGRATING RELATIONS

IOANNIS KASPARIS*
University of Cyprus

July 4, 2007

Abstract

A simple specification test based on fully modified residuals and the CUSUM test for cointegration of Xiao and Phillips (Journal of Econometrics, 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. A simulation experiment is carried out to assess 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 reasonably well. However, the Dickey-Fuller test performs poorly under functional form misspecification.

1 INTRODUCTION

Cointegration has probably been the most popular approach in analysing macroeconomic relations since it was introduced, about twenty years ago. Although the concept of cointegration is very appealing from an economic theory point of view, many 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. Corradi, Swanson and White (2000), Teräsvirta and Ellianson (2001) among others,

*This paper is based on Chapter 4 of my Ph.D. thesis at the University of Southampton. I am deeply indebted to Peter Phillips for invaluable advice and encouragement. I am grateful to Grant Hillier, Tassos Magdalinos and Jean-Yves Pitarakis for their support and for useful comments. In addition, I would like to thank the coeditor, Pentti Saikkonen and two referees for comments that have improved the original draft.

consider 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 which provides a theoretical framework 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.

The 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 we address here. Two tests are considered as means of testing for Functional Form (FF hereafter) and lack of cointegration in long-run cointegrating relations. The first is a simple specification test based on fully modified residuals. The second test is the CUSUM test for cointegration proposed by Xiao and Phillips (2002). We show 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, we argue that the Dickey-Fuller test (DF) that is widely used as a cointegration test, performs poorly under FF misspecification in many cases.

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 Arai (2004) and Hong and Phillips (2005) who extend Ramsey's (1969) RESET test to cointegrating relations. Hong and Phillips (2005) consider scalar covariate models linear in parameter and variable, while in Arai (2004) the empirical specification comprises multiple covariate models linear in parameters and variables. The present theoretical framework is more general. We consider additively separable multiple regression models, linear in parameters and nonlinear in variables. 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. Our approach is similar to that of Xiao and Phillips (2002). The fitted model is estimated by a Fully Modified Least Squares (FM-LS) type of estimator and the sample moments of the test statistics are 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 specific to the fitted model, similar to the one reported by Xiao and Phillips (2002). Under the alternative, the residuals of the fitted model are dominated by some transformation of a unit root process, which typically 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.

We expect that other FF tests can be used in this framework. The simulation results of Kim, Lee and Newbold (2005) suggest that several linearity test statistics diverge under lack of cointegration, indicating FF misspecification. Kim, Lee and Newbold (2004) interpret this as "spurious nonlinearity". Nonetheless, this phenomenon is not a nuisance as it implies that various tests for functional form can be used as cointegration tests. After all, lack of cointegration can be seen as FF misspecification.

The DF test performs poorly under FF misspecification. The DF test has been widely used as a cointegration test. When the fitted model is of incorrect FF, it would be desirable that the DF test favours the unit root hypothesis as this would indicate that the fitted model is inadequate. If the DF test is applied to the residuals of model that is misspecified in terms of FF, in many cases the unit root hypothesis is rejected, although the residuals are not stationary. The DF test is designed to detect unit root processes. Under FF misspecification the residuals are dominated by nonlinear transformations 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 the sample size increases. For logarithmic and concave polynomial functions, the limit distribution involves negative components making the test biased towards the alternative of stationarity. Only for convex polynomial transformations the DF tends to 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. 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. If the matrix A is positive definite, that is denoted by $A > \mathbf{0}$. Further, $diag(A_1, ..., A_m)$ is a block diagonal matrix with blocks A_i. For a matrix $A = (a_{ij})$, $\|A\| = (\max_{i,j} |a_{ij}|)$. For a function f, which can be matrix-valued, $\|f\|_C = \sup_{x \in C} \|f(x)\|$. For a possibly matrix-valued random variable x, $\|x\|_p$ $(p \in \mathbb{N})$ is its L_p-norm. As usual for a function f, $\dot{f}$ denotes its first derivative with respect to its argument. Finally $I\{A\}$ is the indicator function of a set A.

2 THEORETICAL FRAMEWORK

We assume two alternative data generating mechanisms for the series $\{y_t\}_{t=1}^n$:

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

or:

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

where f and s belong to the H-regular family, the variables x_t and z_t are unrelated unit root processes, and u_t is some stationary error term which is specified in detail later. The first model postulates that y_t is cointegrated, possibly in a nonlinear way, with some variables of interest (x_{it}'s). On the other hand if (2) holds, y_t is not cointegrated with x_t. 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 y_t is 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 is given by:

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

where g_i's are H-regular functions, possibly different than f_i's. For notational convenience, $f(x_t)$, $s(z_t)$ and $g(x_t)$ in (1), (2) and (3) may be written as f_t, s_t and g_t respectively.

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

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

The following assumption about u_t, v_t and w_t holds:

Assumption 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 $\Pi(L) = diag\left(\Phi(L)_{(1\times1)}, \Psi(L)_{(p\times p)}, \Xi(L)_{(1\times1)}\right)$ satisfies the summability condition $\sum_{j=1}^{\infty} j \left\|\Pi_j\right\| < \infty$.*

(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 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* $l > \min\left(8, 4/(1-2b)\right)$, *with* $0 \leq b < 1/3$. *Further,* ξ_t *has distribution absolutely continuous with respect to Lebesgue measure and has characteristic function* $\varphi(\lambda) = o(\|\lambda\|^{-\delta})$ *as* $\lambda \to \infty$.

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 $0 \leq r \leq 1$. In addition, $(U(r), V'(r), W(r))$ is a $(p+2)$-dimensional Brownian motion with covariance matrix $\Omega = \sum_{k=-\infty}^{\infty} E\left(e_t e_{t+k}'\right)$ and one-sided covariance matrix $\Lambda = \sum_{k=0}^{\infty} E\left(e_t e_{t+k}'\right)$. Under Assumption 1, the strong approximation

$$\sup_{0 \leq r \leq 1} \left\|(U_n(r), V_n'(r), W_n(r)) - (U(r), V'(r), W(r))\right\| = o_{a.s.}(1)$$

holds on some Skorokhod space (e.g. Park and Phillips, 2001). The above result is utilised in the rest of the paper without any further reference. For the purpose of the subsequent analysis, Ω and Λ are conformably partitioned as follows:

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

Next, we specify in detail the functions that appear in (1), (2) and (3). As mentioned earlier, the functions under consideration are confined to the H-regular family of Park and Phillips (1999). The H-regular family comprises of 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 the functions 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 are 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, the asymptotic theory for sample covariance terms is different, than that originally developed by Park and Phillips (1999, 2001). A relevant asymptotic result is provided by de Jong (2002), Saikkonen and Choi (2004) and Ibragimov and Phillips (2004). In order to obtain asymptotic power rates for the tests, we need a certain generalisation of this asymptotic result along the lines of Phillips (1991). This generalisation is provided in Appendix A. The three aforementioned papers impose certain smoothness restrictions on the functions under consideration. A similar approach is followed here. We restrict our functions to a subset of the H-regular class of Park and Phillips (1999). Our smoothness assumptions are shown below:

Assumption 2. *The transformation* $f:\mathbb{R}^p \to \mathbb{R}^p$*, with* $f'(x) = (f_1(x_1), ..., f_p(x_p))$ *satisfies:*

(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)

(a) $\dot{h}_f(x)$ *is continuous,*

or

(b) *For any* $0 < C < \infty$ *and some* $0 < b < 1/3$*, there is a sequence* $j_n \downarrow 0$ *as* $n \to \infty$*, such that*

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

As usual, h_f and k_f are the limit homogenous function and asymptotic order of f respectively. Moreover, note that when f is a p-dimensional vector, k_f and $\dot{f}$ are $(p \times p)$ diagonal matrices. Condition (iiib) in Assumption 2 is a smoothness condition similar to the one in de Jong (2002)[2]. To obtain the limit distribution of the test statistics under correct specification, we employ Assumption 2(i)-(iiia). The particular limit results can be also established under Assumption 2(i)-(ii) and (iiib) with $b = 0$. To obtain power rates under misspecification, we make a stronger smoothness assumption. For this case, we employ Assumption 2(i)-(ii) and (iiib). The convergence rate of the sequence j_n is determined by l, i.e. the order of finite moments of the process ξ_t. In general, for a large b, a large l is required.

Next, FF misspecification and lack of cointegration are defined precisely.

DEFINITION 1. *Suppose f, s and g satisfy Assumption 2.*
(i) *The fitted model (3) is of correct FF, when $g_i(.) = f_i(.)$, for all $i = \{1,..,p\}$ and (1) holds.*
(ii) *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:*
C1: $g_i(.)-f_i(.) = q_i(.)$, *q_i satisfies Assumption 2 and* $k_{q_i}(\lambda)/k_{g_i}(\lambda)$, $k_{q_i}(\lambda)/k_{f_i}(\lambda) \to 0$ as $\lambda \to \infty$, *or*
C2: $k_{g_i}(\lambda)/k_{f_i}(\lambda) \to 0$ *or* ∞ *as* $\lambda \to \infty$.
(ii) *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 $f_1(x_{1t})$,..., $f_p(x_{pt})$, is ruled out. It is obvious from Definition 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 is not attempted here, as it would result in more complexity in our presentation.

Saikkonen and Choi (2004) consider cointegrating models with smooth transition functions, which typically are distribution type of functions. As shown by Park and Phillips (2001), the parameters of these functions lack identification, when the covariates are unit root processes. Saikkonen and Choi (2004) avoid the identification problem by considering models where the covariates are normalised by the square root of the sample size. The present framework does not cover fitted models of this kind, because they are nonlinear in parameters. Limit results for the Nonlinear Least Squares estimator under FF misspecification in models with unit roots have been derived by the author and some extensions of the current results along these lines are possible. Nonetheless, models with normalised variables create extra complications. To obtain asymptotic power rates for this kind of models, development of second order asymptotic theory for H-regular transformations is required.

3 TESTS

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. Both tests are residual based. When the errors of the model are martingale differences (e.g. Park and Phillips, 2001), the first test under consideration belongs to the Conditional Moment (CM) class of Newey (1985). Under Assumption 1 though, such moment conditions do not hold, because the covariates are endogenous. Nonetheless, for purposes of brevity, we call the first test CM. The second test is the CUSUM test (CS) for cointegration of Xiao and Phillips (2002), generalised to cope with fitted models that are nonlinear in variables. The underling feature of the CM and CS tests is that they can detect abnormal fluctuation in the residuals that typically arises under misspecification. The limit properties of the tests are derived under correct FF, incorrect FF and lack of cointegration.

Because under our assumptions the limit distribution theory is not mixed normal, t-tests and the usual likelihood based tests do not have pivotal distributions under the null hypothesis[3]. In addition the limit distribution of non standard tests like the CUSUM test 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:

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

where $\kappa(.)$ is some kernel on $[-1,1]$ such that $\kappa(0) = 1$ and M_n is a bandwidth such that $M_n \to \infty$, $n/M_n \to 0$ 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 $1 \leq t$, $t + h \leq n$ (e.g. $C_{uu}(h) = n^{-1} \sum_{t=h+1}^{n} \hat{u}_{t-h} \hat{u}_t$ for $h \geq 0$ and $C_{uu}(h) = n^{-1} \sum_{t=-h+1}^{n} \hat{u}_t \hat{u}_{t+h}$, for $h < 0$). Consistency results for this kind of kernel estimators can be found in Jansson (2002).

The estimator under consideration is due to de Jong (2002) and resembles the original FM-LS estimator introduced by Phillips and Hansen (1990). Before the estimator is presented, the following quantities need to be defined:

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

Our estimator 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, the following result holds:

LEMMA 1. (de Jong, 2002) *Let $U(r)^+ = U(r) - V'(r)\Omega_{vv}^{-1}\Omega_{vu}$ and suppose that Assumption 1 and Assumption 2 hold. Then under correct FF, as $n \to \infty$*

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

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:

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

Before we present the test statistics, we introduce some notation. Let $\mathbf{w}$ be a weight function of the form $\mathbf{w}(x_1, ..., x_p) = \sum_{i=1}^{p} \mathbf{w}_i(x_i)$ (additively separable), with each $\mathbf{w}_i$ being H-regular of asymptotic order $k_{\mathbf{w}}$, satisfying Assumption 2. Define the matrices $\hat{A}_n$, $\hat{B}_n$, A, B, $\hat{\Omega}^+$, Ω^+ and their inverses, when they exist, as follows:

$$\tfrac{1}{n} k_g^{-1}(\sqrt{n}) k_{\mathbf{w}}^{-1}(\sqrt{n}) \hat{A}_n = \tfrac{1}{n} k_g^{-1}(\sqrt{n}) k_{\mathbf{w}}^{-1}(\sqrt{n}) \sum_{t=1}^{n} \mathbf{w}(x_t) g(x_t) \xrightarrow{p} A,$$
$$\tfrac{1}{n} k_g^{-1}(\sqrt{n}) \hat{B}_n k_g^{-1}(\sqrt{n}) = \tfrac{1}{n} k_g^{-1}(\sqrt{n}) \sum_{t=1}^{n} g(x_t) g'(x_t) k_g^{-1}(\sqrt{n}) \xrightarrow{p} B,$$
$$\hat{\Omega}^+ = \hat{\Omega}_{uu} - \hat{\Omega}_{uv} \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu} \text{ and } \Omega^+ = \Omega_{uu} - \Omega_{uv} \Omega_{vv}^{-1} \Omega_{vu}.$$

The CM and the CS test statistics are:

$$CM_n = \frac{\left[\sum_{t=1}^{n} \hat{u}_t^+ \mathbf{w}(x_t) - \dot{\mathbf{w}}_n\right]^2}{\hat{\Omega}^+ \sum_{t=1}^{n} \left[\hat{A}_n' \hat{B}_n^{-1} g(x_t) - \mathbf{w}(x_t)\right]^2} \quad \text{and} \quad CS_n = \max_{m=1,\dots,n} \frac{\left|\sum_{t=1}^{m} \hat{u}_t^+\right|}{\sqrt{n\hat{\Omega}^+}},$$

with $\dot{\mathbf{w}}_n = \sum_{i=1}^{p} \sum_{t=1}^{n} \dot{\mathbf{w}}_i(x_{it}) \left(\hat{\Lambda}_{v_i u} - \hat{\Lambda}_{v_i v} \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu}\right)$.

Remark:

(a) The use of the weight functions ($\mathbf{w}$) in the CM test statistic creates additional nuisance terms in the limit, that need to be corrected to obtain a pivotal test. The term $\dot{\mathbf{w}}_n$ is employed to correct the particular nuisance terms.

(b) One may wish to employ weights in order to enhance power against certain alternatives (see for example Bierens, 1990, p.1446). Nonetheless, weight functions are particularly useful, when an intercept is included in the model. For this case, the CM test would be invalid if no weights were employed, as the sum of the Least Squares (LS) residuals is zero by the first order conditions of the LS problem[4].

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

THEOREM 1. *Let* B, $\Omega^+ > \mathbf{0}$ *and suppose that Assumption 1 and Assumption 2 hold. Then under correct FF we have, as* $n \to \infty$

$$CM_n \xrightarrow{d} \chi_1^2 \quad \textit{and} \quad CS_n \xrightarrow{d} \sup_{0 \le \nu \le 1} \frac{\left|\bar{U}(\nu)\right|}{\sqrt{\Omega^+}},$$

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

The limit distribution of the CUSUM test is free of any nuisance parameters and 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 linearity vs. nonlinearity test. On the other hand the CM test has standard limit distribution irrespective of the empirical model employed.

Next we examine the asymptotic power of the tests. The behaviour of the test statistics under the alternative hypothesis is shown by the following result:

THEOREM 2. *Let B, $\Omega^+ > \mathbf{0}$, $M_n = \lfloor n^b \rfloor$ and suppose that Assumption 1 and Assumption 2 hold. Then under incorrect FF or no cointegration we have, as $n \to \infty$:*

$$\mathbf{P}(CM_n > \beta_{1n}),\ \mathbf{P}(CS_n > \beta_{2n}) \to 1,$$

for any nonstochastic sequences β_{1n} and β_{2n} such that

$$\beta_{1n} = o(n/M_n),\ \beta_{2n} = o\left((n/M_n)^{1/2}\right).$$

The divergence rates in both cases are bandwidth dependent. For the first test, the divergence rate is the same as the rate of the RESET test of Hong and Phillips (2005). The divergence rate of the CUSUM test is the same as that reported by Xiao and Phillips (2002), when there is lack of cointegration in the linear framework.

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). In addition, it can be shown that the CM and CS test statistics are bounded in probability under the alternative hypothesis. The regression residuals, under this kind of misspecification, are driven by integrable components, which are known to exhibit weaker signal than that of a stationary process (see Park and Phillips, 1999). Consequently, the fluctuation in the residuals under the alternative hypothesis is the same as that under the null hypothesis and as result none of the two tests under consideration can detect this kind of misspecification.

The simulation study of Xiao and Phillips (2002) reveals that, when it comes to the choice of the bandwidth parameter, there is a trade-off between size and power. Andrews (1991) proposes automatic bandwidth methods. A similar bandwidth method is considered by Xiao and Phillips (2002), where $M_n = 1.447(\hat{\delta}n)^{1/3}$ with $\hat{\delta} = 4\hat{\rho}^2/\left(1-\hat{\rho}^2\right)^2$ and $\hat{\rho}$ is the LS estimator from the residuals autoregression. This method is inappropriate in our case. As Xiao and Phillips (2002) point out this kind of procedures are appropriate for stationary processes. In our case the regression residuals are stationary only under the null hypothesis. Xiao and Phillips (2002) suggest that, when this bandwidth method is used under lack of cointegration, $M_n \sim n$

and as result the CUSUM test has no power. The following lemma demonstrates, that this is true here as well.

LEMMA 2. *Suppose Assumption 1 holds and $\dot{f}$, $\dot{g}$ and $\dot{s}$ satisfy Assumption 2. Then under incorrect FF or no cointegration we have, as $n \to \infty$:*

$$(\hat{\delta}n)^{1/3} = O_p(n).$$

4 SIMULATION EVIDENCE

In this section a Monte Carlo experiment is performed to assess the finite sample properties of the CM, CS and DF tests. First, we examine the size properties of the CM and CS tests and secondly, the ability of the CM, CS and DF tests to detect lack of cointegration and FF misspecification. Clearly, for CM and CS this corresponds to the power of the tests. The DF test is commonly used as a linear cointegration test. FF misspecification and lack of cointegration cannot be rigorously embedded in the hypothesis structure of the test. Nonetheless, it would be desirable in the presence of FF misspecification that the DF test favours 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 with 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 conventionally refer to it as the "power" of the DF test. All the experiments use 10,000 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. Under lack of cointegration and incorrect FF the data is generated by the specifications listed below:

$y_t = z_t$	$(R1)$	$y_t = sign(z_t)\,\lvert z_t\rvert^{0.5}$	$(R2)$
$y_t = sign(x_t)\,\lvert x_t\rvert^{0.75} + u_t$	$(R3)$	$y_t = sign(x_t)\,\lvert x_t\rvert^{1.25} + u_t$	$(R4)$
$y_t = \ln(1 + \lvert x_t\rvert) + u_t$	$(R5)$	$y_t = x_t + \lvert x_t\rvert^{0.5} + u_t$	$(R6)$
$y_t = 0.4x_t 1\{x_t < 0\} + 1.8x_t 1\{x_t \geq 0\} + u_t$	$(R7)$	$y_t = x_t + 1.8\frac{x_t}{1+\exp\left(-x_t/\sqrt{n}-2\right)} + u_t$	$(R8)$
$y_t = x_t + z_t + u_t$	$(R9)$	$y_t = sign(x_t)\left(\lvert x_t\rvert\,\lvert z_t\rvert\right)^{0.5} + u_t$	$(R10)$

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

$u_t = \rho u_{t-1} + \epsilon_t,$
$\Delta x_t = v_t$, with $v_t = \rho v_{t-1} + \eta_t,$
$\Delta z_t = w_t$, with $w_t = 0.3 w_{t-1} + \omega_t,$

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

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

As Xiao and Phillips (2002) point out, when the autoregressive parameter (ρ) is close to unity, the innovation errors become nearly integrated and this adversely affects the size of the test. 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 parameter. In particular, $\rho = 0, 0.2, 0.4, 0.6, 0.8$ and 0.9 has been chosen.

The performance of the tests depends on the sample size and the bandwidth parameter. To achieve good size properties a large bandwidth parameter will be required, if the innovation errors exhibit strong intensity. Moreover it is apparent from our theoretical results, that a large bandwidth adversely affects the power of the tests. Further, the employment of the automatic bandwidth methods proposed by Andrews (1991) results in inconsistent tests. This is because under FF or lack of cointegration $\hat{\rho}$ converges very quickly to one (n-consistent). Alternatively, Sul, Phillips, and Choi (2004) propose the following rule for choosing the autoregressive parameter:

$$\tilde{\rho} = \min\left(\hat{\rho},\ 1 - n^{-\varphi}\right),\ \varphi = 0.5.$$

If $\tilde{\rho}$ is employed instead of $\hat{\rho}$, we get $CM_n \sim n^{\frac{2}{3}(1-2\varphi)}$ and $CS_n \sim n^{\frac{1}{3}(1-2\varphi)}$. In general, a small φ improves power but makes the size properties of the tests worse. In order to assess the extent of the trade off between size and power, three values for the bandwidth are considered: $M1 = n^{1/5}$, $M2 = n^{1/3}$ and $M3 = 1.447(\hat{\delta} n)^{1/3}$, where $\tilde{\rho}$ with $\varphi = 0.1$ is employed[5] instead of $\hat{\rho}$. In addition, we 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 $\rho \leq 0.2$, while for $M = M2$ the performance is good as long as $\rho \leq 0.4$. The best performance is attained, when $M3$ is employed. In particular, the size properties for both tests are quite good for $\rho \leq 0.6$. If the autoregressive parameters are restricted within this range, the performance of both tests is comparable. For larger autoregressive coefficients, severe overrejection of the null hypothesis occurs, when $M1$ is used, with the CM test performing better.

Table 2 shows the empirical power performance of the CM, CS and DF i.e. the ability of the tests to detect lack of cointegration and FF. All three tests perform well under lack of cointegration. The DF test has the best performance in small samples while the CS test outperforms the other two tests in large samples, when the spurious regression is nonlinear and $M1$ and $M2$ are employed. Under functional form misspecification the DF test performs very poorly. Although the residuals are nonstationary in this case, the DF test favours stationarity. The CM and CS tests perform reasonably well under FF misspecification. Obviously the power of the tests varies with choice of the bandwidth parameter. Best performance is attained when $M1$ is used, while in most cases the employment of $M3$ does not result in a severe reduction in power.

The relative performance of the tests varies with the type of misspecification. The simulation results seem to suggest that the CM test performs better for logarithmic and threshold alternatives. On the other hand the CS test performs better for polynomial alternatives and under lack of cointegration. Interestingly, for half of the cases considered in Table 2, the CS test outperforms the CM test, despite the fact the latter attains faster divergence rates. Note that under misspecification the regression residuals are dominated by some H-regular transformation, $u(.)$ say, of a unit root process. We see that CM performs very well, when the regression residuals are dominated by some component, $u(x_t)$, that does not change sign. If the residual process is allowed to change sign, then the CS test tends to perform better. Clearly, the larger the sample moment of the test statistic is, the better the test performs. If the function $u(.)$ is allowed to change sign, $u(x_t)$ exhibits the typical random walk type of behaviour. Lengthy periods in which the term is positive, alternate with lengthy periods in which the term in negative, undermining the magnitude of the sample moment of the test statistic. 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 achieved. This becomes more apparent if one considers the limit expressions for the test statistics. As shown in the Appendix, when misspecification is committed the test statistics behave asymptotically as follows:

$$CM_n \sim (n/M_n) \left[\int_0^1 h_u(V(r))dr \right]^2 \text{ and } CS_n \sim (n/M_n)^{1/2} \sup_{0 \le \nu \le 1} \left| \int_0^\nu h_u(V(r))dr \right|.$$

Clearly, the magnitude of the integral terms in the expressions above affect the power of the test. The CS test maximises the integral term with respect to the integration horizon. We expect that other specification tests may also perform well. For instance, the underling principle behind a MOSUM test that uses fully modified residuals is similar to that of the CS test. The KPSS test (e.g. Shin, 1994) can detect abnormal fluctuation in residuals as well. Just like the CS test, Shin's KPSS is based on partial sums of regression residuals. The CS test opts the maximal partial sum, while the KPSS test averages over all partial sums. Using similar arguments as those for the

CS test, it can be shown that the relevant KPSS statistic[6] behaves as follows:

$KPSS_n \xrightarrow{d} \int_0^1 \bar{U}(r)^2/\Omega^+ dr$, under correct specification,
$KPSS_n \sim (n/M_n) \int_0^1 \left(\int_0^\nu h_u(V(r))dr\right)^2 d\nu$, under incorrect specification.

5 Conclusion

We have considered two residual based tests as means of testing for functional form in long-run cointegrating relations. 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_n for the first test and $\sqrt{n/M_n}$ for the second.

The Monte Carlo experiment suggests that both tests perform reasonably well. The choice of the bandwidth parameter plays important role. If a small bandwidth parameter is selected, the tests have relatively good power properties but can be severely oversized, when the intensity of the errors is strong. Although Andrews' (1991) automatic methods are inappropriate in our framework, when combined with the Sul, Phillips and Choi (2005) rule for the selection of the autoregressive parameter, they provide a good compromise between size and power. In particular we find that it results in quite good size without causing a big reduction in the power of the tests, when the parameter φ is set equal to 0.1. The simulation results suggest that the CM test performs better for logarithmic and threshold alternatives. The CS test performs better for polynomial alternatives and under lack of cointegration. The simulation study of Hong and Phillips (2005) 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, 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, it favours stationarity, when in fact the residual process is nonstationary. The work of Park and Phillips (1998) provides some useful theoretical results that justify this.

The present theoretical framework is by no means exhaustive. Many specifications that are appealing for applied econometric work are not included. In order to handle the more complicated asymptotic 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. In addition, many specifications of interest are nonlinear in parameters. Extensions to these directions may prove quite challenging. Some extensions to models nonlinear in parameters, are possible and are under development by the author. Moreover we expect that several other FF tests, apart from the CM, CS and RESET tests, are adequate means of testing for FF in a nonstationary framework. Therefore further future work is required, to assess the adequacy and relative performance of all these tests.

NOTES

1. Pötscher (2004) generalises these results to all locally integrable transformations under the assumption, that the process $t^{-1/2}x_t$ possess uniformly bounded density functions.
2. See de Jong (2002) pages 5 and 21 and Remark A1 in Appendix A.
3. Under similar assumptions, Hong and Phillips (2005) show that quadratic form statistics have noncentral mixed chi-square limit distributions rather than ordinary chi-square distributions.
4. I would like to thank an anonymous referee for pointing out this problem.
5. In a preliminary simulation experiment we tried $\tilde{\rho}$ with $\varphi = 0.5$. We found that the size performance of the tests was very good, but power was poor.
6. $KPSS_n = n^{-2}\sum_{m=1}^{n}\left(\sum_{t=1}^{m}\hat{u}_t^{+}\right)^2/\hat{\Omega}^{+}$.

REFERENCES

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

Arai, Y. (2004) Testing linearity in regressions with I(1) processes. Mimeo.

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, 1-36.

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. Mimeo.

——————— (2004) Addendum to "Asymptotics for nonlinear transformations of integrated time series". *Econometric Theory* 20, 627-635.

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 (2005) Testing linearity in cointegrating relations with an application to PPP. Cowles Foundation Discussion Paper 1541.

Ibragimov, R. and P.C.B. Phillips (2004) Regression asymptotics using martingale convergence methods. Cowles Foundation Discussion Paper 1473.

Jansson, M (2002) Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449-1459.

Lee, Y-S, Kim, T-H. and P. Newbold (2004) Spurious nonlinear regressions in econometrics. *Economics Letters* 87(3), 301-306.

Newey, W.K. (1985) Maximum likelihood specification testing and conditional moment tests, *Econometrica,* 53*,* 1047-1070.

Park, J.Y. and P.C.B. Phillips (1998) Unit roots in nonlinear transformations of integrated time series. Mimeo.

———— (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.

Phillips, P.C.B. (1991) Spectral regression for cointegrated time series. *Non-parametric and semiparametric methods in econometrics and statistics*, edited by W. Barnett, J. Powell and G. Tauchen. Cambridge: Cambridge University Press.

Phillips, P.C.B. and B. E. Hansen (1990) Statistical inference in instrumental variables regressions with I(1) Processes. *Review of Economic Studies*, 53, 473-496.

Phillips, P.C.B. 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, 1-22.

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.

Shin, Y. (1994) A residual based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91-115.

Sul, D., Phillips P.C.B. and C-Y. Choi (2005) Prewhitening bias in HAC estimation. *Oxford Bulletin of Economics and Statistics*, 67(4), 517-546.

Xiao, Z. and P.C.B. Phillips (2002) A CUSUM test for cointegration using regression residuals. *Journal of Econometrics* 108, 43-61.

APPENDIX A (TECHNICAL RESULTS)

In this Appendix we provide some auxiliary results required to prove the main results of the paper. The proofs of our main results are provided in Appendix B below. For notational convenience, the sequence $k_f(\sqrt{n})$ relating to the asymptotic order of any H-regular function f will be written $k_{n,f}$. Further, K, $\overline{K}$ and $\underline{K}$ are defined as follows:

$$K = \int_{-1}^{1} \kappa(s)ds,\ \overline{K} = \int_{0}^{1} \kappa(s)ds,\ \underline{K} = \int_{-1}^{0} \kappa(s)ds.$$

LEMMA A1: *Set $M_n = \lfloor n^b \rfloor$ with b as in Assumption 1. Then we have, as $n \to \infty$:*

(a) *Under Assumption 1:*

$$\sup_{r\in[0,1]} \frac{1}{\sqrt{n}} \sum_{|h|\le\lfloor n^b\rfloor} \left\| v_{\lfloor nr\rfloor+h} \right\| = o_{a.s.}(1).$$

(b) *Suppose that* h_f is *continuous. Under Assumption 1 and Assumption 2(i):*

$$\frac{1}{M_n} \sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n}\right) \frac{1}{n} k_{n,f}^{-1} \sum_{t=1}^{n} f(x_t) f'(x_{t+h}) k_{n,f}^{-1} \xrightarrow{p} K \int_0^1 h_f(V(r)) h_f'(V(r)) dr.$$

(c) *Under Assumption 1 with* $b=0$ *and Assumption 2(i)-(iiia):*

$$\frac{1}{\sqrt{n}} k_{n,f}^{-1} \sum_{t=1}^{n} f(x_t)\left(u_t, v_t'\right) \xrightarrow{p} \int_0^1 h_f(V(r)) d\left(U(r), V'(r)\right) + \int_0^1 \dot{h}_f(V(r)) dr \left(\Lambda_{vu}, \Lambda_{vv}\right).$$

(d) *Suppose that* $\left\|\dot{h}_f(x)\right\|^4 = O\left(e^{c\|x\|}\right)$*, as* $\|x\| \to \infty$*, for some* $c>0$*. Under Assumption 1 and Assumption 2(i)-(ii), (iiib):*

$$\frac{1}{M_n} \sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n}\right) \frac{1}{\sqrt{n}} k_{n,f}^{-1} \sum_{t=1}^{n} f(x_{t+h})\left(u_t, v_t'\right)$$
$$\xrightarrow{p} K \int_0^1 h_f(V(r)) d\left(U(r), V'(r)\right) + \overline{K} \int_0^1 \dot{h}_f(V(r)) dr \left(\Omega_{vu}, \Omega_{vv}\right).$$

Remark A1:

(a) The limit result of Lemma A1(c) is also shown by de Jong (2002), Saikkonen and Choi (2004) and Ibragimov and Phillips (2004), under different smoothness conditions. Here we assume continuously differentiable functions, which is a weaker condition than those used by Saikkonen and Choi (2004) and Ibragimov and Phillips (2005). Further, the result can be established, when continuity is replaced by Assumption (iiib) with $b=0$. The latter is a weaker smoothness condition than that of de Jong (2002), although the assumptions about the processes are more general in the aforementioned paper.

(b) The following one-sided analogs of Lemma A1(d) also hold:

(i) $$\frac{1}{M_n} \sum_{h=0}^{M_n} \kappa\left(\frac{h}{M_n}\right) \frac{1}{\sqrt{n}} k_{n,f}^{-1} \sum_{t=1}^{n} f(x_{t+h}) u_t \xrightarrow{p} \overline{K} \int_0^1 h_f(V(r)) dU(r) + \overline{K} \int_0^1 \dot{h}_f(V(r)) dr \Omega_{vu},$$

(ii) $$\frac{1}{M_n} \sum_{h=-M_n}^{0} \kappa\left(\frac{h}{M_n}\right) \frac{1}{\sqrt{n}} k_{n,f}^{-1} \sum_{t=1}^{n} f(x_{t+h}) u_t \xrightarrow{p} \underline{K} \int_0^1 h_f(V(r)) dU(r).$$

Proof of Lemma A1. For part (a) notice that

$$\sup_{r\in[0,1]} \frac{1}{\sqrt{n}} \sum_{|h|\leq \lfloor n^b \rfloor} \left\| v_{\lfloor nr \rfloor + h} \right\| \leq \max_{|j|\leq n+\lfloor n^b \rfloor} \frac{2\lfloor n^b \rfloor}{\sqrt{n}} \left\| v_j \right\|.$$

Now, for any $\delta > 0$,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbf{P}\left(\max_{|j|\leq n+\lfloor n^b \rfloor} \|v_j\| > \delta n^{\frac{1}{2}-b} \right) &\leq \sum_{n=1}^{\infty} \sum_{|j|\leq n+\lfloor n^b \rfloor} \mathbf{P}\left(\|v_j\| > \delta n^{\frac{1}{2}-b} \right) \leq \sum_{n=1}^{\infty} \sum_{|j|\leq n+\lfloor n^b \rfloor} \frac{\mathbf{E}\|v_j\|^l}{\delta^l n^{l\left(\frac{1}{2}-b\right)}} \\
&\leq \sum_{n=1}^{\infty} \left(n + \lfloor n^b \rfloor + 1 \right) \frac{2\mathbf{E}\|v_j\|^l}{\delta^l n^{l\left(\frac{1}{2}-b\right)}} \leq \sum_{n=1}^{\infty} \frac{6\mathbf{E}\|v_j\|^l}{\delta^l n^{l\left(\frac{1}{2}-b\right)-1}} < \infty,
\end{aligned}$$

and the result follows.

Next, we prove part (b). First, we show that

$$\left\| \frac{1}{M_n} \sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n} \right) \frac{1}{n} k_{n,f}^{-1} \sum_{t=1}^{n} f(x_t) \left\{ f'(x_{t+h}) - f'(x_t) \right\} k_{n,f}^{-1} \right\| = o_p(1).$$

Given this the result follows easily, since by Theorem 3.3 in Park and Phillips (2001),

$$\frac{1}{M_n} \sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n} \right) \frac{1}{n} k_{n,f}^{-1} \sum_{t=1}^{n} f(x_t) f'(x_t) k_{n,f}^{-1} \xrightarrow{p} K \int_0^1 h_f(V(r)) h_f'(V(r)) dr.$$

By part (a) we get

$$\sup_{r\in[0,1]} \sup_{|h|\leq M_n} \left\| \frac{x_{\lfloor nr \rfloor + h}}{\sqrt{n}} - \frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \right\| = o_{a.s.}(1). \tag{A1}$$

Set $C = \sup_{r\in[0,1]} \|V(r)\| + 1$. Then by (A1) and the strong approximation (e.g. Park and Phillips, 1999, Lemma 2.3) we have, for n large enough,

$$\sup_{r\in[0,1]} \sup_{|h|\leq M_n} \left\| x_{\lfloor nr \rfloor + h} / \sqrt{n} \right\|, \ \sup_{r\in[0,1]} \left\| x_{\lfloor nr \rfloor} / \sqrt{n} \right\| \leq C \ a.s. \tag{A2}$$

Further notice, that

$$\begin{aligned}
&\left\| \frac{1}{M_n} \sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n} \right) \frac{1}{n} k_{n,f}^{-1} \sum_{t=1}^{n} f(x_t) \left\{ f'(x_{t+h}) - f'(x_t) \right\} k_{n,f}^{-1} \right\| \\
\leq & \left| \frac{1}{M_n} \sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n} \right) \right| \left\| k_{n,f}^{-1} \right\| \left\| k_{n,\dot{f}}^{-1} \right\| \sup_{r\in[0,1]} \sup_{|h|\leq M_n} \left\| f\left(x_{\lfloor nr \rfloor + h} \right) f'(x_{\lfloor nr \rfloor + h}) - f'(x_{\lfloor nr \rfloor}) \right\| \\
\leq & \ 2 \left\| k_{n,f}^{-1} \right\| \left\| k_{n,\dot{f}}^{-1} \right\| \sup_{r\in[0,1]} \sup_{|h|\leq M_n} \left\| f(x_{\lfloor nr \rfloor + h}) f'(x_{\lfloor nr \rfloor + h}) - f'(x_{\lfloor nr \rfloor}) \right\| \\
\leq & \ 2 \left\| h_f \right\|_C \sup_{r\in[0,1]} \sup_{|h|\leq M_n} \left\| h_f\left(\frac{x_{\lfloor nr \rfloor + h}}{\sqrt{n}} \right) - h_f\left(\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \right) \right\| + o_{a.s.}(1),
\end{aligned}$$

for n large enough. The last inequality above follows from (A2). Further, $\|h_f\|_C < \infty$ $a.s.$, as h_f is locally bounded (see for example Park and Phillips, 2001 p.159). Therefore, by the uniform continuity of $h_f(x)$ on $\|x\| \leq C$ and (A1) we get

$$\sup_{r\in[0,1]} \sup_{|h|\leq M_n} \left\| h_f\left(\frac{x_{\lfloor nr\rfloor+h}}{\sqrt{n}}\right) - h_f\left(\frac{x_{\lfloor nr\rfloor}}{\sqrt{n}}\right)\right\| = o_{a.s.}(1),$$

and this completes the proof.

Part (c) can be proved using similar arguments to those used in the proof of part (b) above and part (d) below. A proof can be provided by the author upon request.

Finally, we show part (d). For purposes of brevity, we show the result for the left block of the sample sum. The proof for right block is identical. Write

$$\begin{aligned} &\frac{1}{M_n}\sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n}\right)\frac{1}{\sqrt{n}}k_{n,f}^{-1}\sum_{t=1}^{n} f(x_{t+h})u_t \\ = &\frac{1}{M_n}\sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n}\right)\frac{1}{\sqrt{n}}k_{n,f}^{-1}\left\{\sum_{t=1}^{n} f(x_t)u_t + \sum_{t=1}^{n}\left(f(x_{t+h}) - f(x_t)\right)u_t\right\} \equiv S_n^1 + S_n^2. \end{aligned}$$

By part (c), the first summand

$$S_n^1 \xrightarrow{p} K\int_0^1 h_f(V(r))dU(r) + K\int_0^1 \dot{h}_f(V(r))dr\Lambda_{vu}. \tag{A3}$$

Set $\sum_{\nu}^{*,h} = \sum_{\nu=1}^{h} 1\{h>0\} - \sum_{\nu=h-1}^{0} 1\{h<0\}$ and $\bar{x}_t^h = x_t + \gamma_t \sum_{\nu}^{*,h} v_{t+\nu}$, $\gamma_t = diag\left(\gamma_{1t}, ..., \gamma_{pt}\right)$ with γ_{it}'s $\in [-1,1]$. Then, by the mean value theorem, the second summand is

$$\begin{aligned} S_n^2 &= \frac{1}{M_n}\sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n}\right)\frac{1}{\sqrt{n}}k_{n,f}^{-1}\sum_{t=1}^{n} \dot{f}\left(\bar{x}_t^h\right)\sum\nolimits_{\nu}^{*,h} v_{t+\nu}u_t \\ &= \frac{1}{M_n}\sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n}\right)\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \dot{f}\left(x_t\right)\sum\nolimits_{\nu}^{*,h} v_{t+\nu}u_t + o_p(1), \end{aligned} \tag{A4}$$

where the last equality can be established as follows. By part (a),

$$j_n = \sup_{r\in[0,1]} \sup_{|h|\leq\lfloor n^b\rfloor} \frac{1}{\sqrt{n}}\left\|\bar{x}_{\lfloor nr\rfloor}^h - x_{\lfloor nr\rfloor})\right\| = o_{a.s.}(1).$$

Therefore we have,

$$\begin{aligned} & \left\| \frac{1}{M_n} \sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n}\right) \frac{1}{\sqrt{n}} k_{n,\dot f}^{-1} \sum_{t=1}^{n} \left[\dot f\left(\bar x_t^h\right) - \dot f\left(x_t\right)\right] \sum\nolimits_{\nu}^{*,h} v_{t+\nu} u_t \right\| \\ \leq & \left\| n^{1/2+b} k_{n,\dot f}^{-1} \right\| \left(\sup_{\|x_1\| \leq C} \sup_{\|x_1 - x_2\| \leq j_n} \left\| \dot f(\sqrt{n} x_1) - \dot f(\sqrt{n} x_2) \right\| \right) \\ & \times \frac{1}{M_n n^{1+b}} \sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n}\right) \sum_{t=1}^{n} \sum\nolimits_{\nu}^{*,h} \left\| v_{t+\nu} u_t \right\| \xrightarrow{p} 0, \end{aligned} \tag{A5}$$

as $n \to \infty$, given Assumption 2(iiib) and the fact that the expectation of the last term above is bounded by $4\left\|v_t\right\|_2 \left\|u_t\right\|_2 < \infty$. In view of (A4) we can write,

$$\begin{aligned} S_n^2 &= \frac{1}{M_n} \sum_{h=-M_n}^{M_n} \kappa\left(\frac{h}{M_n}\right) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \dot f\left(x_t\right) \sum\nolimits_{\nu}^{*,h} \left\{\mathbf{E}\left(v_{t+\nu} u_t\right) + \left(v_{t+\nu} u_t - \mathbf{E}\left(v_{t+\nu} u_t\right)\right)\right\} + o_p(1) \\ &\equiv S_n^3 + S_n^4 + o_p(1). \end{aligned}$$

By Toeplitz's lemma and Theorem 3.3 of Park and Phillips (2001), the first summand above

$$S_n^3 \xrightarrow{p} \int_0^1 \dot h_f(V(r)) dr \left\{ \overline{K} \sum_{\nu=1}^{\infty} \mathbf{E}(v_t u_{t-\nu}) - \underline{K} \sum_{\nu=0}^{\infty} \mathbf{E}(v_t u_{t+\nu}) \right\}. \tag{A6}$$

Moreover, it can be shown that

$$\sup_{|h| \leq n^b} \left\| \frac{1}{n} k_{n,\dot f} \sum_{t=1}^{n} f(x_t) \sum\nolimits_{\nu}^{*,h} \left(v_{t+\nu} u_t - \mathbf{E}\left(v_{t+\nu} u_t\right)\right) \right\| = o_p(1). \tag{A7}$$

Notice that (A7) implies that $S_n^4 = o_p(1)$. In view of this the requisite result follows from (A3) and (A6). Finally, we show (A7). For notational brevity we consider the case $h > 0$. Using the convention that Ψ_j, Φ_j are zero for negative index, we can write (see for example Phillips and Solo, 1992, Remark 3.8)

$$\frac{1}{n} k_{n,\dot f}^{-1} \sum_{t=1}^{n} \dot f(x_t) \sum_{\nu=1}^{h} \left(v_{t+\nu} u_t - \mathbf{E}\left(v_{t+\nu} u_t\right)\right) =$$

$$\frac{1}{n} k_{n,\dot f}^{-1} \sum_{t=1}^{n} \dot f(x_t) \sum_{\nu=1}^{h} \sum_{j=0}^{\infty} \Psi_{j+\nu-1} \Phi_j \left(\eta_{t-j+1} \varepsilon_{t-j} - \Sigma_{\eta\epsilon}\right)$$

$$+ \frac{1}{n} k_{n,\dot f}^{-1} \sum_{t=1}^{n} \dot f(x_t) \sum_{\nu=1}^{h} \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \Psi_{j+\nu-1+r} \Phi_j \eta_{t-j-r+1} \varepsilon_{t-j}$$

$$+ \frac{1}{n} k_{n,\dot f}^{-1} \sum_{t=1}^{n} \dot f(x_t) \sum_{\nu=1}^{h} \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \Psi_{j+\nu-1-r} \Phi_j \eta_{t-j+1+r} \varepsilon_{t-j} \equiv I_{1n,h} + I_{2n,h} + I_{3n,h}.$$

It can be shown that $I_{1n,h}$, $I_{2n,h}$, and $I_{3n,h}$ are asymptotically negligible uniformly over $|h| \leq n^b$. The proof for $I_{1n,h}$ and $I_{2n,h}$ is relatively easy. We show the result for $I_{3n,h}$. Define the lag polynomials $B_{\nu r}(L)$, $\tilde{B}_{\nu r}(L)$ by:

$$
\begin{aligned}
B_{\nu r}(L) &= \sum_{j=0}^{\infty} B_{\nu r j} L^j, \text{ with } B_{\nu r j} = \Psi_{j+\nu-1-r}\Phi_j, \\
\tilde{B}_{\nu r}(L) &= \sum_{j=0}^{\infty} \tilde{B}_{\nu r j} L^j, \text{ with } \tilde{B}_{\nu r j} = \sum_{s=j+1}^{\infty} B_{\nu r s}.
\end{aligned}
$$

Let $\zeta_{ht} = \sum_{r=1}^{\infty}\sum_{\nu=1}^{h} B_{\nu r}(1)\eta_{t+1+r}\varepsilon_t$ and $\tilde{\zeta}_{ht} = \sum_{\nu=1}^{h}\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\tilde{B}_{\nu r j}\eta_{t-j+1+r}\varepsilon_{t-j}$. Then, from second order Beveridge-Nelson decomposition (e.g. Phillips and Solo, 1992 equation (23)) on $I_{3n,h}$ we get

$$
\begin{aligned}
I_{3n,h} &= \frac{1}{n}k_{n,\dot{f}}^{-1}\sum_{t=1}^{n}\dot{f}(x_t)\sum_{\nu=1}^{h}\sum_{j=0}^{\infty}\sum_{r=1}^{\infty} B_{\nu r j}\eta_{t-j+1+r}\varepsilon_{t-j} \\
&= \frac{1}{n}k_{n,\dot{f}}^{-1}\sum_{t=1}^{n}\dot{f}(x_t)\zeta_{ht} - \frac{1}{n}k_{n,\dot{f}}^{-1}\sum_{t=1}^{n}\dot{f}(x_t)\Delta\tilde{\zeta}_{ht}. \qquad \text{(A8)}
\end{aligned}
$$

First, we show that the first summand in (A8) is $o_p(1)$ uniformly in $|h| \leq n^b$. Without loss of generality assume that $\dot{f}(x_t)$ and ζ_{ht} are scalars. Then,

$$
\begin{aligned}
&\mathbf{P}\left(\max_{|h|\leq n^b}\left|\frac{1}{n}k_{n,\dot{f}}^{-1}\sum_{t=1}^{n}\dot{f}(x_t)\zeta_{ht}\right| > \epsilon\right) \leq \epsilon^{-2}\sum_{|h|\leq n^b}\mathbf{E}\left(\frac{1}{n}k_{n,\dot{f}}^{-1}\sum_{t=1}^{n}\dot{f}(x_t)\zeta_{ht}\right)^2 \\
&= \epsilon^{-2}\sum_{|h|\leq n^b}\mathbf{E}\left(\frac{1}{n^2}k_{n,\dot{f}}^{-2}\sum_{t=1}^{n}\dot{f}(x_t)^2\zeta_{ht}^2\right) + \epsilon^{-2}\sum_{|h|\leq n^b}\mathbf{E}\left(\frac{1}{n^2}k_{n,\dot{f}}^{-2}\sum_{s\neq l}\dot{f}(x_s)\zeta_{hs}\dot{f}(x_l)\zeta_{hl}\right) \\
&= T_{1n} + T_{2n}.
\end{aligned}
$$

The first term,

$$
T_{1n} \leq \frac{\epsilon^{-2}}{n^2}\sum_{|h|\leq n^b}\sum_{t=1}^{n}\left\|k_{n,\dot{f}}^{-1}\dot{f}(x_t)\right\|_4^2 \|\zeta_{ht}\|_4^2 \leq \frac{\epsilon^{-2}}{n^2}\left(\mathbf{E}\left\|h_{\dot{f}}\right\|_C^4\right)^{1/2}\sum_{|h|\leq n^b}\sum_{t=1}^{n}\|\zeta_{ht}\|_4^2,
$$

where the last inequality holds for n large enough. In addition, under our assumptions we have $\mathbf{E}\left\|h_{\dot{f}}\right\|_C^4 < \infty$ (e.g. Park and Phillips, 2001, p.147). Next,

$$
\|\zeta_{ht}\|_4 \leq \lim\inf_{s_1,s_2\to\infty}\left\|\sum_{\nu=1}^{h}\sum_{r=1}^{s_1}\sum_{j=0}^{s_2}\Psi_{j+\nu-1-r}\Phi_j\eta_{t+1+r}\varepsilon_t\right\|_4 \text{ (by Fatou's lemma)}
$$

$$
\leq \sum_{\nu=1}^{h}\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\left\|\Psi_{j+\nu-1-r}\Phi_j\eta_{t+1+r}\varepsilon_t\right\|_4 \text{ (by Minkowski's inequality)}
$$

$$\leq \sum_{\nu=1}^{h}\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\|\Psi_{j+\nu-1-r}\Phi_j\|\left\|\eta_{t+1}\varepsilon_t\right\|_4 \leq \|\eta_t\|_8\|\varepsilon_t\|_8\sum_{\nu=1}^{h}\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\|\Psi_{j+\nu-1-r}\Phi_j\|$$

$$\leq h\|\eta_t\|_8\|\varepsilon_t\|_8\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\|\Psi_i\|\|\Phi_j\| \leq h\|\eta_t\|_8\|\varepsilon_t\|_8\sum_{i,j=0}^{\infty}\|\Psi_i\|\|\Phi_j\|.$$

Therefore, we get

$$n^{-2}\sum_{|h|\leq n^b}\sum_{t=1}^{n}\|\zeta_{ht}\|_4^2 \leq \left(\|\eta_t\|_8\|\varepsilon_t\|_8\sum_{i,j=0}^{\infty}\|\Psi_i\|\|\Phi_j\|\right)^2\sup_{|h|\leq n^b}\frac{h^2\left(2n^b+1\right)}{n} \to 0,$$

as $n\to\infty$. Hence, $T_{1n}=o_p(1)$.

Next, we show that $T_{2n}=0$. Assume $s>l$, without loss of generality. Then we have

$$\mathbf{E}\left(\frac{1}{n^2}k_{n,\dot{f}}^{-2}\sum_{s\neq l}\dot{f}(x_s)\zeta_{hs}\dot{f}(x_l)\zeta_{hl}\right) = \left(\frac{1}{n^2}k_{n,\dot{f}}^{-2}\sum_{s\neq l}\mathbf{E}\left[\dot{f}(x_s)\dot{f}(x_l)\mathbf{E}\left[\zeta_{hs}\zeta_{hl}\mid\mathcal{F}_s\right]\right]\right)$$

Next,

$$\mathbf{E}\left[\zeta_{hs}\zeta_{hl}\mid\mathcal{F}_s\right] = \sum_{r_1,r_2=1}^{\infty}\sum_{\nu_1,\nu_2=1}^{h}B_{\nu_1 r_1}(1)B_{\nu_2 r_2}(1)\varepsilon_s\varepsilon_l\mathbf{E}\left[\eta_{s+1+r_1}\eta_{l+1+r_2}\mid\mathcal{F}_s\right]$$

$$= \sum_{r_1=1}^{\infty}\sum_{r_2=1}^{r_1+(s-l)-1}\sum_{\nu_1,\nu_2=1}^{h}B_{\nu_1 r_1}(1)B_{\nu_2 r_2}(1)\varepsilon_s\varepsilon_l\eta_{l+1+r_2}\overbrace{\mathbf{E}\left[\eta_{s+1+r_1}\mid\mathcal{F}_s\right]}^{0}$$

$$+\sum_{r_1=1}^{\infty}\sum_{r_2=r_1+(s-l)+1}^{\infty}\sum_{\nu_1,\nu_2=1}^{h}B_{\nu_1 r_1}(1)B_{\nu_2 r_2}(1)\varepsilon_s\varepsilon_l\mathbf{E}\left[\eta_{s+1+r_1}\eta_{l+1+r_2}\mid\mathcal{F}_s\right]$$

$$+\sum_{r_1=1}^{\infty}\sum_{\nu_1,\nu_2=1}^{h}B_{\nu_1,r_1}(1)B_{\nu_2,s-l+1+r_1}(1)\varepsilon_s\varepsilon_l\mathbf{E}\left[\eta_{s+1+r_1}^2\mid\mathcal{F}_s\right]$$

$$= \sum_{r_1=1}^{\infty}\sum_{r_2=r_1+(s-l)+1}^{\infty}\sum_{\nu_1,\nu_2=1}^{h}B_{\nu_1 r_1}(1)B_{\nu_2 r_2}(1)\varepsilon_s\varepsilon_l\mathbf{E}\left[\eta_{s+1+r_1}\overbrace{\mathbf{E}\left[\eta_{l+1+r_2}\mid\mathcal{F}_{s+r_1}\right]}^{0}\mid\mathcal{F}_s\right]$$

$$+\varepsilon_s\varepsilon_l\sum_{r_1=1}^{\infty}\sum_{\nu_1,\nu_2=1}^{h}B_{\nu_1,r_1}(1)B_{\nu_2,s-l+1+r_1}(1)\mathbf{\Sigma}_{\eta\eta} = \varepsilon_s\varepsilon_l\tilde{\mathbf{\Sigma}}_{\eta\eta},$$

where the first equality above follows from dominated convergence. Therefore

$$\frac{1}{n^2}k_{n,\dot{f}}^{-2}\sum_{s\neq l}\mathbf{E}\left[\dot{f}(x_s)\dot{f}(x_l)\mathbf{E}\left[\zeta_{hs}\zeta_{hl}\mid\mathcal{F}_s\right]\right]=\tilde{\boldsymbol{\Sigma}}_{\eta\eta}\frac{1}{n^2}k_{n,\dot{f}}^{-2}\sum_{s\neq l}\mathbf{E}\left[\dot{f}(x_s)\dot{f}(x_l)\varepsilon_l\overbrace{\mathbf{E}\left[\varepsilon_s\mid\mathcal{F}_{s-1}\right]}^{0}\right]=0$$

In view of this it follows that the first term in (A8) is $o_p(1)$ uniformly over $|h|\leq n^b$.

Next, we show that the second term in (A8) is negligible uniformly over $|h|\leq n^b$. Write

$$\frac{1}{n}k_{n,\dot{f}}^{-1}\sum_{t=1}^{n}\dot{f}(x_t)\Delta\tilde{\zeta}_{ht}=\frac{1}{n}k_{n,\dot{f}}^{-1}\dot{f}(x_n)\tilde{\zeta}_{hn}-\frac{1}{n}k_{n,\dot{f}}^{-1}\sum_{t=1}^{n}\Delta\dot{f}(x_t)\tilde{\zeta}_{ht}. \tag{A9}$$

Further, $n^{-b}\sup_{|h|\leq n^b}\mathbf{E}\left\|\tilde{\zeta}_{ht}\right\|<\infty$, because

$$\begin{aligned}
n^{-b}\sup_{|h|\leq n^b}\mathbf{E}\left\|\tilde{\zeta}_{ht}\right\| &\leq n^{-b}\sup_{|h|\leq n^b}\lim\inf_{s_1,s_2\to\infty}\mathbf{E}\left\|\sum_{\nu=1}^{h}\sum_{r=1}^{s_1}\sum_{j=0}^{s_2}\tilde{B}_{\nu rj}\eta_{t-j+1+r}\varepsilon_{t-j}\right\| \\
&\leq \left\|\eta_t\right\|_2\left\|\varepsilon_t\right\|_2 n^{-b}\sup_{|h|\leq n^b}h\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty}\left\|\Psi_{s-1-r}\Phi_s\right\| \\
&\leq \left\|\eta_t\right\|_2\left\|\varepsilon_t\right\|_2\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\left\|s\Psi_{s-1-r}\Phi_s\right\|\leq\left\|\eta_t\right\|_2\left\|\varepsilon_t\right\|_2\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\left\|\Psi_r\right\|\left\|s\Phi_s\right\|<\infty.
\end{aligned}$$

In view of this and using the same arguments as those in (A5), it follows that the second terms in (A8) converges in probability to zero, uniformly in $|h|\leq n^b$. ■

Note that under the alternative hypothesis, some of the kernel estimators mentioned earlier are 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 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, we partition f as follows: $f'_{(1\times p)}=\left(f^{1\prime}_{(1\times p_1)},\ f^{2\prime}_{(1\times p_2)}\right)$ with f^2 being the components of f that have not been correctly specified. The leading element(s) of f^2 is 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. Also $\bar{\theta}_o$ 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:

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

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

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

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

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

$$\begin{aligned}
&\bar{\zeta}_1' = \left(\theta_o', -\zeta_1'\right),\ \bar{\zeta}_2' = \left(\theta_o^{2\prime}, -\zeta_2'\right),\ \bar{\zeta}_3' = \left(1, -\zeta_3'\right),\\
&\bar{h}_1' = \left(h_{\bar{d}}', h_g'\right),\ \bar{h}_2' = \left(h'_{\bar{f}^2}, h_g'\right),\ \bar{h}_3' = \left(h_s, h_g'\right),\\
&\dot{H}_1' = \left(\dot{h}_{\bar{d}}', \dot{h}_g'\right),\ \dot{H}_2' = \left(\dot{h}'_{\bar{f}^2}, \dot{h}_g'\right),\ \dot{H}_3' = \left(\dot{h}_s, \dot{h}_g'\right),\\
&\bar{\Omega} = \left(\Omega_{vw}, \Omega_{vv}\right).
\end{aligned}$$

Remark A2:

The expressions in Definition A(i) characterise the limit behaviour of the LS estimator under FF misspecification and lack of cointegration. It is apparent from Definition A(i) that under incorrect FF, the slope estimators do not always converge to the parameter of interest. For instance, when **S1** holds, an individual slope estimator, $\hat{a}_{LS_i}$, converges 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.

LEMMA A2. *Let Assumption 1 hold. Then we have as,* $n\to\infty$ *:*
(i) *Under incorrect FF, when* **S1** *holds,*

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

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

$$\begin{array}{l}\frac{n^{1/2}}{Mk_{n,f^{2*}}}\hat{\Omega}_{vu}\xrightarrow{p}K\int_0^1 dV(r)\bar{h}_2'(V(r))\bar{\zeta}_2+\overline{K}\Omega_{vv}\int_0^1\dot{H}_2'(V(r))\bar{\zeta}_2dr,\\ \frac{n^{1/2}}{Mk_{n,f^{2*}}}\hat{\Lambda}_{vu}\xrightarrow{p}\overline{K}\int_0^1 dV(r)\bar{h}_2'(V(r))\bar{\zeta}_2+\overline{K}\Omega_{vv}\int_0^1\dot{H}_2'(V(r))\bar{\zeta}_2dr,\\ \frac{1}{Mk_{n,f^{2*}}^2}\hat{\Omega}_{uu}\xrightarrow{p}K\int_0^1\bar{\zeta}_2'\bar{h}_2(V(r))\bar{h}_2'(V(r))\bar{\zeta}_2dr.\end{array}$$

(iii) *Under no cointegration*

$$\begin{array}{l}\frac{n^{1/2}}{Mk_{n,s}}\hat{\Omega}_{vu}\xrightarrow{p}K\int_0^1 dV(r)\bar{h}_3'(W(r),V(r))\bar{\zeta}_3+\overline{K}\bar{\Omega}\int_0^1\dot{H}_3'\left(W(r),V(r)\right)\bar{\zeta}_3dr\\ \frac{n^{1/2}}{Mk_{n,s}}\hat{\Lambda}_{vu}\xrightarrow{p}\overline{K}\int_0^1 dV(r)\bar{h}_3'(W(r),V(r))\bar{\zeta}_3+\overline{K}\bar{\Omega}\int_0^1\dot{H}_3'\left(W(r),V(r)\right)\bar{\zeta}_3dr\\ \frac{1}{Mk_{n,s}^2}\hat{\Omega}_{uu}\xrightarrow{p}K\int_0^1\bar{\zeta}_3'\dot{H}_3\left(W(r),V(r)\right)\dot{H}_3'\left(W(r),V(r)\right)\bar{\zeta}_3dr.\end{array}$$

Proof of Lemma A2. We start with the proof of part (i). Under incorrect FF the LS estimator can be written as

$$\frac{k_{n,g}}{k_{n,d^*}}\left(\hat{a}_{LS}-\theta_o\right)=\left[\frac{1}{n}k_{n,g}^{-1}\sum_{t=1}^n g_tg_t'k_{n,g}^{-1}\right]^{-1}\frac{1}{nk_{n,d^*}}\left[\sum_{t=1}^n g_td_t'\theta_o+k_{n,g}^{-1}\sum_{t=1}^n g_tu_t\right]$$

Hence, by Theorem 3.3 of Park and Phillips (2001) and Lemma A1(c) we get

$$\begin{aligned}\frac{k_{n,g}}{k_{n,d^*}}\left(\hat{a}_{LS}-\theta_o\right)&=\left[\int_0^1 h_g\left(V(r)\right)h_g'\left(V(r)\right)dr\right]^{-1}\int_0^1 h_g\left(V(r)\right)h_{d^*}'\left(V(r)\right)\theta_odr+O_p(1/\sqrt{n}k_{n,d^*})\\&=\zeta_1+o_p(1).\end{aligned}$$

Define the normalising matrix $N_{n,d^*}=diag\left(I_{(p\times p)},\ k_{n,g}/k_{n,d^*}\right)$. In what follows the regression residuals (from OLS estimation) will be written in the following form:

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

Hence

$$\begin{aligned}\frac{n^{1/2}}{Mk_{n,d^*}}\hat{\Omega}_{vu}&=\frac{1}{M_nk_{n,d^*}}\sum_{h=-M_n}^{M_n}\kappa\left(\frac{h}{M_n}\right)\left(\frac{1}{n}\sum_{t=1}^n v_t\left(\begin{array}{cc}d_{t+h}'&g_{t+h}'\end{array}\right)N_{n,d^*}^{-1}N_{n,d^*}\left(\begin{array}{c}\theta_o\\-\left(\hat{a}-\theta_o\right)\end{array}\right)\right)\\&\quad+\frac{n^{1/2}}{Mk_{n,d^*}}\Omega_{vu}+o_p(1)\\&=K\int_0^1 dV(r)\left(h_{d^*}'(V(r))\theta_o-h_g'(V(r))\zeta_1\right)\\&\quad+\overline{K}\Omega_{vv}\int_0^1\left[\dot{h}_{d^*}'(V(r))\theta_o-\dot{h}_g'(V(r))\zeta_1\right]dr+\frac{n^{1/2}}{M_nk_{n,d^*}}\Omega_{vu}+o_p(1)\\&=K\int_0^1 dV(r)\bar{h}_1'(V(r))\bar{\zeta}_1+\overline{K}\Omega_{vu}\int_0^1\dot{H}_1'(V(r))\bar{\zeta}_1dr+\frac{n^{1/2}}{M_nk_{n,d^*}}\Omega_{vu}+o_p(1)\end{aligned}\quad\text{(A10)}$$

where the second equality above is due to Lemma A1(d). By Remark A1(b) and using similar arguments as above, we get

$$\frac{n^{1/2}}{M_n k_{n,d^*}}\hat{\Lambda}_{vu} = \overline{K}\int_0^1 dV(r)\bar{h}_1'(V(r))\bar{\zeta}_1 + \overline{K}\Omega_{vv}\int_0^1 \dot{H}_1'(V(r))\bar{\zeta}dr + \frac{n^{1/2}}{M_n k_{n,d^*}}\Lambda_{vu} + o_p(1) \tag{A11}$$

Next, we find the order of $\hat{\Omega}_{uu}$. By Lemma A1(b,d), it easy to show that

$$\begin{aligned}\frac{1}{M_n k_{n,d^*}^2}\hat{\Omega}_{uu} &= K\int_0^1 \bar{\zeta}_1'\bar{h}_1(V(r))\bar{h}_1'(V(r))\bar{\zeta}_1 dr + \frac{2}{M_n\sqrt{n}k_{n,d^*}}K\int_0^1 \bar{\zeta}_1'\bar{h}_1(V(r))dU(r) \\ &\quad + \frac{2}{M_n\sqrt{n}k_{n,d^*}}\overline{K}\int_0^1 \bar{\zeta}_1'\dot{H}_1(V(r))dr\Omega_{vu} + o_p(1). \end{aligned}\tag{A12}$$

By (A9) and (A11) we have

$$\begin{aligned}\hat{\Omega}_{uu} - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu} &= O_p\left(M_n k_{n,d^*}^2\right) + O_p\left(\frac{M_n^2 k_{n,d^*}^2}{n}\right) \\ &= O_p\left(M_n k_{n,d^*}^2\right). \end{aligned}\tag{A13}$$

Next, we prove part (ii). For any two H-regular transformations T_1 and T_2 let $I_{T_1T_2} = \int_0^1 T_1(V(r))T_2(V(r))'dr$. 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

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

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

$P^1 = \left(I_{h_{f^{2*}}h_{f^1}} - I_{h_{f^{2*}}h_{g^2}}I_{h_{g^2}h_{g^2}}^{-1}I_{h_{g^2}h_{f^1}}\right)(P^3)^{-1}$,
$P^2 = I_{h_{f^{2*}}h_{g^2}}I_{h_{g^2}h_{g^2}}^{-1} - \left(I_{h_{f^{2*}}h_{f^1}} - I_{h_{f^{2*}}h_{g^2}}I_{h_{g^2}h_{g^2}}^{-1}I_{h_{g^2}h_{f^1}}\right)(P^3)^{-1}I_{h_{f^1}h_{g^2}}I_{h_{g^2}h_{g^2}}^{-1}$
and $P^3 = I_{h_{f^1}h_{f^1}} - I_{h_{f^1}h_{g^2}}I_{h_{g^2}h_{g^2}}^{-1}I_{h_{g^2}h_{f^1}}$.

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

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

Now similar arguments as those above give

$$\frac{n^{1/2}}{Mk_{n,f^{2*}}}\hat{\Omega}_{vu} = K\int_0^1 dV(r)\bar{h}_2'(V(r))\bar{\zeta}_2 + \overline{K}\Omega_{vv}\int_0^1 \dot{H}_2'(V(r))\bar{\zeta}_2 dr + o_p(1),$$

$$\frac{n^{1/2}}{Mk_{n,f^{2*}}}\hat{\Lambda}_{vu} = \overline{K}\int_0^1 dV(r)\bar{h}_2'(V(r))\bar{\zeta}_2 + \overline{K}\Omega_{vv}\int_0^1 \dot{H}_2'(V(r))\bar{\zeta}_2 dr + o_p(1),$$

and

$$\frac{1}{Mk_{n,f^{2*}}^2}\hat{\Omega}_{uu} = \overline{K}\int_0^1 \bar{\zeta}_2'\bar{h}_2(V(r))\bar{h}_2'(V(r))\bar{\zeta}_2 dr + o_p(1).$$

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

APPENDIX B (PROOFS OF MAIN RESULTS)

Proof of Theorem 1. First, we show the result for the CM test. Set $\tilde{A}_n = \frac{1}{n}k_{n,g}^{-1}k_{n,\mathbf{w}}^{-1}\hat{A}_n$, $\tilde{B}_n = \frac{1}{n}k_{n,g}^{-1}\hat{B}_n k_{n,g}^{-1}$. By Lemma A1(c)

$$n^{-1/2}k_{n,\mathbf{w}}^{-1}\left\{\sum_{t=1}^n \left(u_t - v_t'\Omega_{vv}^{-1}\Omega_{vu}\right)\mathbf{w}_t - \dot{\mathbf{w}}_n\right\} = \int_0^1 h_{\mathbf{w}}(V(r))dU^+(r) + o_p(1). \quad \text{(A14)}$$

Next notice that

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

$$= \frac{\left[\tilde{A}_n'\tilde{B}_n^{-1}k_{n,g}^{-1}\frac{1}{\sqrt{n}}\left[\sum_{t=1}^n g_t u_t^+ - \dot{g}_n\hat{\Lambda}_{vu}^+\right] - \frac{k_{n,\mathbf{w}}^{-1}}{\sqrt{n}}\sum_{t=1}^n \left(u_t - v_t'\Omega_{vv}^{-1}\Omega_{vu}\right)\mathbf{w}_t - \frac{k_{n,\mathbf{w}}^{-1}}{\sqrt{n}}\dot{\mathbf{w}}_n\right]^2}{\left(\Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}\right)\frac{1}{n}\sum_{t=1}^n \left[\tilde{A}_n'\tilde{B}_n^{-1}k_{n,g}^{-1}g_t - k_{n,\mathbf{w}}^{-1}\mathbf{w}_t\right]^2}$$

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

where the last line is due to Theorem 3.3 of P&P, Lemma A1(c) and (A14). The result follows from the fact that V and U^+ are independent.

The CS test statistic is:

$$CS_n = \frac{\sup_{0\le\nu\le1}\frac{1}{\sqrt{n}}\left|\sum_{t=1}^{[\nu n]}\left(g_t'(\hat{a}-a_o) - u_t + v_t'\Omega_{vv}^{-1}\Omega_{vu}\right)\right|}{\sqrt{\Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}}} + o_p(1)$$

The numerator above is:

$$\sup_{0\le\nu\le1}\left|\frac{1}{n}\sum_{t=1}^{[\nu n]} g_t'k_{n,g}^{-1}\left[\frac{1}{n}k_{n,g}^{-1}\sum_{t=1}^n g_t g_t'k_{n,g}^{-1}\right]^{-1}k_{n,g}^{-1}\frac{1}{\sqrt{n}}\left[\sum_{t=1}^n g_t u_t^+ - \dot{g}_n\hat{\Lambda}_{vu}^+\right] - \frac{1}{\sqrt{n}}\sum_{t=1}^{[\nu n]}\left(u_t - v_t'\Omega_{vv}^{-1}\Omega_{vu}\right)\right|$$

$$= \sup_{0\le\nu\le1} \left|\bar{U}(\nu)\right| + o_p(1)$$

by Theorem 3.3 of P&P and Lemma A2, the result follows. ■

Proof of Theorem 2. We will prove the result under incorrect FF, when **S1** holds. The proof for the other cases is similar and will be omitted. Rearranging the expression for the FM-LS estimator and in view of (A10) and (A11), we get:

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

Recall that the CM test statistic is:

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

Consider first the numerator rescaled by $(nk_{n,d^*}k_{n,\mathbf{w}})^2$:

$$\begin{aligned}
&\left(\frac{1}{nk_{n,d^*}k_{n,\mathbf{w}}}\right)^2\left[\sum_{t=1}^{n}\left\{\left(y_t^{+} - g_t{}'\hat{a}\right)\mathbf{w}_t - v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\mathbf{w}_t\right\} - \dot{\mathbf{w}}_n\right]^2 \\
=\;& \left(\frac{1}{nk_{n,d^*}k_{n,\mathbf{w}}}\right)^2\left[\sum_{t=1}^{n}\left\{\begin{pmatrix} d_t' & g_t'\end{pmatrix} N_{n,d^*}^{-1}N_{n,d^*}\begin{pmatrix}\theta_o \\ -\left(\hat{a}-\theta_o\right)\end{pmatrix} + \left(u_t - v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right)\mathbf{w}_t\right\} - \dot{\mathbf{w}}_n\right]^2 \\
=\;& \left[\int_0^1 \bar{h}_1'(V(r))\bar{\zeta}_1 h_{\mathbf{w}}(V(r))dr + o_p(1) + O_p\left(\frac{1}{\sqrt{n}k_{n,d^*}}\right) + O_p\left(\frac{M_n}{n}\right)\right]^2. \qquad \text{(A15)}
\end{aligned}$$

Next, by Theorem 3.3 in Park and Phillips (2001) and (A13), the denominator rescaled by n:

$$\left(\hat{\Omega}_{uu} - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right)\int_0^1\left[A'B^{-1}h_g(V(r)) - h_{\mathbf{w}}(V(r))\right]^2 dr + o_p(1) = O_p\left(M_n k_{n,d^*}^2 k_{n,\mathbf{w}}^2\right). \qquad \text{(A16)}$$

In view of (A15) and (A16) we have $CM_n \sim (n/M_n)\left[\int_0^1 \bar{h}_1'(V(r))\bar{\zeta}_1 h_{\mathbf{w}}(V(r))dr\right]^2$, which gives the requisite result.

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

$$\begin{aligned}
&\frac{\max_{1\le k\le n}}{nk_{n,d^*}}\left|\sum_{t=1}^{k}\left\{\begin{pmatrix} d_t' & g_t'\end{pmatrix} N_{n,d^*}^{-1}N_{n,d^*}\begin{pmatrix}\theta_o \\ -\left(\hat{a}-\theta_o\right)\end{pmatrix} + \left(u_t - v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right)\right\}\right| \\
=\;& \sup_{0\le\nu\le1}\left|\int_0^{\nu} \bar{h}_1'(V(r))\bar{\zeta}_1 dr + o_p(1) + O_p\left(\frac{1}{\sqrt{n}k_{n,d^*}}\right) + O_p\left(\frac{M_n}{n}\right)\right|.
\end{aligned}$$

By (A12) the denominator rescaled by $\sqrt{n}$ is:

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

Therefore $CS_n \sim (n/M_n)^{1/2}\sup_{0\leq\nu\leq 1}\left|\int_0^{\nu}\bar{h}_1'(V(r))\bar{\zeta}_1 dr\right|$, which completes the proof. ■

Proof of Lemma 2. We will show that the result under FF misspecification, 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 A3 we have that $(nk_{n,d})^{-1}\sum_{t=1}^{n}u(x_t)=\int_0^1\bar{h}_1'(V(r))\bar{\zeta}_1dr+o_p(1)=\int_0^1 h_u(V(r))dr+o_p(1)$ and similarly define $\dot{u}(x_t)$, $\ddot{u}(x_t)$, $\dot{h}_u$ and $\ddot{h}_u$. First consider

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

Hence by Lemma A1(c) and Lemma A2 we have

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

Since $k_{n,d}/k_{n,\dot{d}}=\sqrt{n}$, the result follows easily. ■

TABLE 1. Empirical size for CM, CS (5% level)

$\rho\|n$	100						200					
	$M1$		$M2$		$M3$		$M1$		$M2$		$M3$	
	CM	CS	CM	CS	CM	CS	CM	CS	CM	CS	CM	CS
0	0.030	0.028	0.026	0.028	0.032	0.031	0.030	0.029	0.030	0.031	0.035	0.031
0.2	0.047	0.041	0.037	0.037	0.041	0.037	0.049	0.048	0.040	0.040	0.035	0.031
0.4	0.076	0.071	0.056	0.505	0.046	0.037	0.079	0.079	0.056	0.052	0.045	0.046
0.6	0.137	0.134	0.089	0.081	0.065	0.057	0.141	0.150	0.087	0.082	0.064	0.058
0.8	0.270	0.282	0.181	0.165	0.129	0.109	0.281	0.336	0.165	0.171	0.118	0.108
0.9	0.406	0.441	0.296	0.280	0.225	0.189	0.425	0.538	0.287	0.307	0.213	0.202
$\rho\|n$	300						500					
	$M1$		$M2$		$M3$		$M1$		$M2$		$M3$	
	CM	CS	CM	CS	CM	CS	CM	CS	CM	CS	CM	CS
0	0.029	0.031	0.031	0.034	0.033	0.032	0.034	0.034	0.038	0.038	0.038	0.035
0.2	0.044	0.046	0.040	0.041	0.033	0.032	0.049	0.051	0.047	0.048	0.038	0.035
0.4	0.069	0.068	0.054	0.053	0.047	0.048	0.074	0.075	0.058	0.058	0.053	0.052
0.6	0.115	0.124	0.077	0.080	0.059	0.058	0.118	0.133	0.078	0.081	0.064	0.060
0.8	0.238	0.279	0.152	0.162	0.101	0.101	0.236	0.308	0.144	0.157	0.101	0.090
0.9	0.383	0.487	0.268	0.298	0.186	0.184	0.381	0.541	0.251	0.303	0.173	0.186

TABLE 2. Empirical power for CM, CS & DF (5% level)

n		50	100	200	300	500	50	100	200	300	500
				(R1)					(R2)		
	DF	0.916	0.921	0.936	0.940	0.943	0.854	0.861	0.850	0.855	0.845
$M1$	CM	0.566	0.701	0.788	0.800	0.848	0.632	0.744	0.819	0.827	0.873
	CS	0.527	0.762	0.920	0.936	0.984	0.578	0.790	0.930	0.947	0.984
$M2$	CM	0.486	0.600	0.698	0.733	0.787	0.565	0.665	0.741	0.770	0.817
	CS	0.414	0.589	0.757	0.828	0.912	0.473	0.635	0.788	0.849	0.924
$M3$	CM	0.411	0.519	0.619	0.658	0.723	0.493	0.593	0.679	0.707	0.762
	CS	0.330	0.463	0.627	0.692	0.811	0.386	0.518	0.672	0.728	0.827
				(R3)					(R4)		
	DF	0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.002	0.005	0.007
$M1$	CM	0.079	0.202	0.398	0.502	0.603	0.268	0.491	0.627	0.660	0.741
	CS	0.078	0.180	0.377	0.515	0.698	0.186	0.430	0.706	0.770	0.902
$M2$	CM	0.062	0.170	0.363	0.454	0.554	0.208	0.419	0.552	0.597	0.654
	CS	0.050	0.133	0.302	0.423	0.609	0.133	0.321	0.536	0.643	0.771
$M3$	CM	0.055	0.154	0.339	0.432	0.526	0.116	0.358	0.500	0.541	0.592
	CS	0.044	0.114	0.263	0.377	0.544	0.099	0.248	0.421	0.505	0.640
				(R5)					(R6)		
	DF	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$M1$	CM	0.518	0.830	0.980	0.997	1.000	0.396	0.742	0.958	0.993	1.000
	CS	0.423	0.706	0.901	0.956	0.993	0.334	0.626	0.862	0.937	0.989
$M2$	CM	0.469	0.788	0.970	0.996	1.000	0.352	0.689	0.943	0.990	0.999
	CS	0.378	0.638	0.856	0.934	0.985	0.296	0.563	0.813	0.908	0.979
$M3$	CM	0.440	0.763	0.970	0.995	1.000	0.324	0.659	0.933	0.987	0.997
	CS	0.348	0.599	0.856	0.911	0.975	0.274	0.528	0.781	0.886	0.967
				(R7)					(R8)		
	DF	0.017	0.012	0.011	0.010	0.011	0.013	0.015	0.014	0.013	0.013
$M1$	CM	0.385	0.547	0.655	0.698	0.754	0.179	0.363	0.565	0.663	0.804
	CS	0.335	0.485	0.597	0.644	0.704	0.148	0.327	0.557	0.657	0.825
$M2$	CM	0.356	0.517	0.629	0.680	0.738	0.139	0.301	0.506	0.621	0.751
	CS	0.299	0.444	0.551	0.608	0.671	0.113	0.260	0.455	0.580	0.754
$M3$	CM	0.333	0.494	0.613	0.669	0.728	0.113	0.257	0.466	0.582	0.717
	CS	0.274	0.406	0.515	0.566	0.637	0.088	0.214	0.391	0.503	0.681

TABLE 2. (continued)

		(R9)					(R10)				
	DF	0.336	0.329	0.346	0.343	0.345	0.001	0.001	0.001	0.000	0.000
$M1$	CM	0.545	0.692	0.786	0.799	0.847	0.148	0.227	0.497	0.534	0.647
	CS	0.508	0.753	0.915	0.935	0.983	0.156	0.411	0.702	0.753	0.904
$M2$	CM	0.471	0.597	0.694	0.731	0.786	0.106	0.221	0.349	0.420	0.509
	CS	0.405	0.584	0.753	0.827	0.911	0.107	0.244	0.441	0.541	0.707
$M3$	CM	0.401	0.514	0.618	0.659	0.735	0.091	0.164	0.257	0.317	0.403
	CS	0.325	0.459	0.625	0.693	0.810	0.094	0.166	0.293	0.367	0.517