DEPARTMENT OF ECONOMICS
UNIVERSITY OF CYPRUS

FUNCTIONAL FORM MISSPECIFICATION IN REGRESSIONS WITH A UNIT ROOT

Ioannis Kasparis

Discussion Paper 2008-02

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

FUNCTIONAL FORM MISSPECIFICATION IN REGRESSIONS WITH A UNIT ROOT

IOANNIS KASPARIS
University of Cyprus

January 2008*

Abstract

We examine the limit properties of the Non-linear Least Squares (NLS) estimator under functional form misspecification in regression models with a unit root. Our theoretical framework is the same as that of Park and Phillips, *Econometrica* 2001. We show that the limit behaviour of the NLS estimator is largely determined by the relative order of magnitude of the true and fitted models. If the estimated model is of different order of magnitude than the true model, the estimator converges to boundary points. When the pseudo-true value is on a boundary, standard methods for obtaining rates of convergence and limit distribution results are not applicable. We provide convergence rates and limit distribution results, when the pseudo-true value is an interior point. If functional form misspecification is committed in the presence of stochastic trends, the convergence rates can be slower and the limit distribution different than that obtained under correct specification.

This paper is based on Chapter 2 of my Ph.D. thesis at the University of Southampton. I am deeply indebted to Grant Hillier and Peter Phillips for invaluable advice and encouragement. I am grateful to Tassos Magdalinos and Jean-Yves Pitarakis for their support and for useful comments.

$*$ Previous version, August 2006.

1 INTRODUCTION

This paper is concerned with the Least Squares (LS) approximation to an unknown function in a nonstationary context. Standard estimation and inference analysis relies on the convention that the fitted model is correctly specified. Accepting that any economic model is an abstraction of reality, rather than a "true" data generating mechanism, it is important to know what the estimators' properties are in the presence of misspecification. The asymptotic properties of the NLS estimator, under functional form misspecification (FFM), have been studied by White (1981), for independent and identically distributed data (i.i.d.) and by Domowitz and White (1982) for heterogenous weakly dependent (WD) data (see also Bierens (1984) for similar results). The purpose of this paper is to explore the limit behaviour of the NLS estimator in misspecified models with strongly dependent nonstationary regressors. In particular, we consider non-linear regressions with unit root covariates. The results provided here are not only of theoretical interest, but also useful for the development of specification tests. In order to obtain asymptotic power rates, for certain misspecification tests e.g. Ramsey (1969), Bierens (1990) (tests without specific alternative), knowledge about the asymptotic behaviour of the estimator under misspecification is necessary. In addition, to determine the limit distribution of certain model selection statistics under the null hypothesis, e.g. Cox (1961, 1962), Davidson and MacKinnon (1981) and Voung (1989) (tests with specific alternative), the estimator's limit distribution about the pseudo-true value, is required.

In order to address the issue of FFM, we need to depart from the standard linear framework. The asymptotic properties of estimators for nonlinear models with stationary and weakly dependent data have been explored twenty five years ago (e.g. Hansen (1981), White and Domowitz (1984)). Nevertheless, no well developed limit distribution theory existed, for nonlinear models with strongly dependent nonstationary regressors, prior to the recent development of Park and Phillips (1999, 2001)[1]. Park and Phillips (2001) (P&P hereafter) consider nonlinear models with an exogenous unit root covariate and martingale difference errors. They focus on two classes of nonlinear transformations: integrable and locally integrable transformations. Our aim is to analyse misspecified models, within the P&P theoretical framework.

White (1981), Domowitz and White (1982) and Bierens (1984) establish convergence to some pseudo-true value, using the Jennrich (1969) approach[2]. Characterising the limit behaviour of the NLS estimator in the context of misspecified models with unit roots, proves to be a more challenging task. In the presence of unit roots, the applicability of existing econometric techniques, for the asymptotic analysis of extremum estimators (e.g. Jennrich, 1969), is limited (see P&P). This is because the NLS objective function involves components of different orders of magnitude. The applicability of these techniques is further restricted, under misspecification, as the fitted model can be of different order of magnitude than the true specification. As in Park and Phillips (2001), we employ a variety of econometric techniques, to obtain asymptotic results. These relate to the work of Jennrich (1969), Wu (1981) and Wooldridge (1994).

Domowitz and White (1982), show that the NLS estimator has a well defined non-stochastic limit, referred to in the econometric literature as "pseudo-true" value. Moreover, the NLS estimator about the pseudo-true value and scaled by $\sqrt{n}$ (n is the sample size) has a Gaussian limit distribution. Hence, for weakly dependent misspecified models, the limit distribution is still Gaussian, and the rate of convergence is unaffected.

We show that when the covariate is a unit root process, things may be substantially different. In our framework the pseudo-true value can be stochastic. In addition, when the true model is of different order of magnitude than the fitted model, the estimator typically converges to boundary points of the parameter space. When the pseudo-true value is on a boundary, techniques that involve a linearisation of the objective function, about the estimator's limit, e.g. Wooldridge (1994), Andrews (1999) are not applicable. We provide convergence rates and limit distribution results, when the pseudo-true value is an interior point. Again the limit behaviour of the NLS estimator is not always analogous to that reported by Domowitz and White (1982). Sometimes the rates of convergence are slower and the limit distribution different than that obtained under correct specification.

As explained earlier, our results are useful for the development of testing procedures in regression models with unit roots. In addition, our analysis provides guidance for the adequacy of empirical models. We have mentioned, that if FFM is committed in the P&P framework, the estimators may diverge or converge to boundary points in the parameter space. Such behaviour would constitute evidence for misspecification. Therefore, inspecting the behaviour of slope parameters, over different parameter spaces can provide useful information about the adequacy of the fitted model.

The rest of this paper is organised as follows: Section 2 specifies the theoretical framework. Section 3 presents our theoretical results, and Section 4 concludes. Before proceeding to the next section, we introduce some notation. For a vector $x = (x_i)$ or a matrix $A = (a_{ij})$, $|x|$ and $|A|$ denote the vector and matrix respectively of the moduli of their elements. The maximum of the moduli is denoted as $\|.\|$. For a matrix A, $A > \mathbf{0}$ denotes positive definiteness. For a function $g : \mathbb{R}^p \to \mathbb{R}$ define the arrays

$$\dot{g} = \left(\frac{\partial g}{\partial a_i}\right), \; \ddot{g} = \left(\frac{\partial^2 g}{\partial a_i \partial a_j}\right), \; \dddot{g} = \left(\frac{\partial^3 g}{\partial a_i \partial a_j \partial a_k}\right),$$

which are vectors arranged by the lexicographic ordering of their indices. Sometimes is more convenient to express the second derivatives of g in matrix form i.e. $\ddot{G} = \partial^2 g / \partial a \partial a'$. The Borel field on a set A is written as $\mathcal{B}(A)$ and $\mathcal{B}$, when $A = \mathbb{R}$. As usual, $\overset{d}{=}$ denotes distributional equality. Finally, $1\{A\}$ is the indicator function of a set A.

2 DEFINITIONS AND PRELIMINARY RESULTS

This section provides a set of definitions, that specify our theoretical framework and, some preliminary results. The models we consider are the same as those discussed in P&P. We assume that the series $\{y_t\}_{t=1}^n$ is generated by the model:

$$y_t = f(x_t) + u_t \tag{1}$$

where f is an unknown function. The variables x_t and u_t are a unit root process and a martingale difference respectively, in some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The fitted model is:

$$y_t = g(x_t, \hat{a}) + \hat{u}_t \tag{2}$$

where $g(.,a)$ is a transformation of the data that is "different from $f(.)$". This is defined precisely later in this section. The fitted model is estimated by the NLS procedure, i.e.:

$$\hat{a} = \arg\min_{a\in A} Q_n(a),\ \ Q_n(a) = \sum_{t=1}^{n}(y_t - g(x_t, a))^2, \tag{3}$$

where A is a compact subset of $\mathbb{R}^p$.

Next, we specify the processes that generate the covariates and the errors of the model. We assume throughout that the sequence $\{x_t\}_{t=1}^n$ is a unit root process generated by

$$x_t = x_{t-1} + v_t,\ \ x_0 = O_p(1).$$

Further, $\{v_t\}_{t=1}^n$ is the linear process:

$$v_t = \psi(L)\eta_t = \sum_{k=0}^{\infty}\psi_k\eta_{t-k},$$

with $\psi(1) \neq 0$ and $\{\eta_t\}_{t=1}^n$ is a sequence of i.i.d. random variables with mean zero. Define the partial sum processes $V_n(r)$ and $U_n(r)$ as:

$$(V_n(r),\ U_n(r)) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}(v_t,\ u_t),\ r \in [0,1].$$

The following assumption is borrowed from P&P and specifies the properties of x_t and u_t in detail.

Assumption 1

(a) $(u_t, \mathcal{F}_{n,t})$ *is a martingale difference sequence with* $\mathbf{E}(u_t^2|\mathcal{F}_{n,t-1}) = \sigma^2$ *a.s. for every* $t = 1, ..., n$ *and* $\sup_{1\leq t\leq n}\mathbf{E}(|u_t|^\nu |\mathcal{F}_{n,t-1}) < \infty$ *a.s. for* $\nu > 2$. *The variable* x_t *is adapted to* $\mathcal{F}_{n,t-1}$ *for every* $t = 1, ..., n$.

(b) $\sum_{k=0}^{\infty} k|\psi_k| < \infty$ *and* $\mathbf{E}|\eta_t|^p < \infty$ *for some* $p > 2$. *The variable* η_t *has characteristic function* $\phi(s)$ *such that* $\int_{-\infty}^{\infty}|\phi(s)|^\nu ds < \infty$, $\nu \geq 1$.

(c) $(U_n, V_n) \xrightarrow{d} (U, V)$, *where* (U, V) *is a vector Brownian motion.*

Assumption 1 yields strong approximation results for the empirical Brownian motions introduced earlier. For instance, under Assumption 1(c) (see P&P, p. 125 and 152), there is a finer probability space $(\Omega, \mathcal{F}, \mathbf{P})^o$ supporting (U, V) and the partial sum processes (U_n^o, V_n^o) such that:

$$\left(U_n^o\left(\frac{k}{n}\right), V_n^o\left(\frac{k}{n}\right),\ k = 1, ..., n\right) \overset{d}{=} \left(U_n\left(\frac{k}{n}\right), V_n\left(\frac{k}{n}\right),\ k = 1, ..., n\right) \tag{4}$$

and

$$\sup_{0\leq r\leq 1}\|(U_n^o(r), V_n^o(r)) - (U(r), V(r))\| = o_{a.s.}(1)$$

In addition, for the purpose of the subsequent analysis, we need to introduce the (chronological) local time process of the Brownian motion V up to time t defined as

$$L(t,s) = \lim_{\epsilon\searrow 0}\frac{1}{2\epsilon}\int_0^t 1\{|V(r) - s| \leq \epsilon\}dr.$$

The reader is referred to Park and Phillips (2000, 2001) for further discussion about the local time process and its relevance to econometrics.

Under some weak conditions, it is possible to establish embedding results for the NLS estimator and to some functionals of the objective function, useful for our asymptotic analysis. Using the embedding results in P&P, we can construct a copy of the objective function, $Q_n^o(a)$ say, on $(\Omega, \mathcal{F}, \mathbf{P})^o$ as follows. Set $z_n = (x_t, u_t,\ t = 1, ..., n)$. It is obvious from (4) that, for each n, there is a random vector $z_n^o = (x_t^o, u_t^o,\ t = 1, ..., n)$ on $(\Omega, \mathcal{F}, \mathbf{P})^o$ such that $z_n \overset{d}{=} z_n^o$. Then define

$$\begin{aligned} Q_n^o(a) &= \sum_{t=1}^{n} (y_t^o - g(x_t^o, a))^2 \text{ with } y_t^o = f(x_t^o) + u_t^o \\ \text{and } \hat{a}^o &= \arg\min_{a \in A} Q_n^o(a). \end{aligned} \tag{5}$$

The objective function $Q_n(a)$ and its copy can be seen as empirical processes on A. The following distributional result holds for any two continuous empirical processes, on some compact space, that have the same finite dimensional distributions.

LEMMA 1. *Suppose that $G_n(a)$ and $G_n^o(a)$ are continuous empirical processes on some compact set $A \subset \mathbb{R}^p$. If $G_n(a)$ and $G_n^o(a)$ have the same finite dimensional distributions, the following hold:*

(i)

$$\inf_{a \in A} G_n(a) \overset{d}{=} \inf_{a \in A} G_n^o(a) \text{ and } \sup_{a \in A} G_n(a) \overset{d}{=} \sup_{a \in A} G_n^o(a).$$

(ii) *Suppose that $\tilde{a}$ and $\tilde{a}^o$ are the unique minimisers of $G_n(a)$ and $G_n^o(a)$ on A, respectively. Then, we have*

$$\tilde{a} \overset{d}{=} \tilde{a}^o.$$

Using the P&P Shorokhod construction, we can show that $Q_n(a)$, and its copy, have the same finite dimensional distributions. Further, under some additional conditions, $Q_n(a)$ and $Q_n^o(a)$ satisfy Lemma 1. These are stated precisely **by** the subsequent lemma.

LEMMA 2. *Suppose that:*

(a) *Assumption 1 holds.*

(b) *The objective function $Q_n(a)$ is given by (3) and $Q_n^o(a)$ given by (5).*

(c) *The function $f(x)$ of (1) is $\mathcal{B}/\mathcal{B}$-measurable. The function $g(x, a)$ of (2) is $\mathcal{B} \times \mathcal{B}(A)/\mathcal{B}$-measurable and continuous in a.*

Then, for each $n \in \mathbb{N}$, the following hold:

(i) *For $d \in \mathbb{N}$ and any $a_1, ..., a_d \in A$, $y_1, ..., y_d \in \mathbb{R}$, and $y_{d+1} \in \mathbb{R}^{2n}$,*

$$\mathbf{P}\left(Q_n(a_i) \le y_i,\ z_n \le y_{d+1},\ i = 1, ..., d\right) = \mathbf{P}^o\left(Q_n^o(a_i) \le y_i,\ z_n^o \le y_{d+1},\ i = 1, ..., d\right).$$

(ii) *For any $y_1 \in \mathbb{R}$ and $y_2 \in \mathbb{R}^{2n}$,*

$$\mathbf{P}(\inf_{a \in A} Q_n(a) \le y_1,\ z_n \le y_2) = \mathbf{P}^o(\inf_{a \in A} Q_n^o(a) \le y_1,\ z_n^o \le y_2).$$

(iii) *Suppose that* $Q_n(a)$ *and* $Q_n^o(a)$ *have unique minimisers on* A. *Then, for any* $y_1 \in \mathbb{R}^p$ *and* $y_2 \in \mathbb{R}^{2n}$,

$$\mathbf{P}(\hat{a} \le y_1,\ z_n \le y_2) = \mathbf{P}^o(\hat{a}^o \le y_1,\ z_n^o \le y_2).$$

Lemma 1(i) postulates that $Q_n(a)$ and $Q_n^o(a)$ have the same finite dimensional distributions. Lemma 2(ii)-(iii) is a direct consequence of Lemma 1. It provides embedding results for the extrema of the NLS objective function and its minimiser. Some of the techniques we employ to establish convergence of the NLS estimator, require limit theory for the extrema of the objective i.e. $\inf_{a \in A} Q_n(a)$. Therefore, if $\inf_{a \in A} Q_n^o(a) \xrightarrow{d} \inf_{a \in A} Q(a)$, we can assert that $\inf_{a \in A} Q_n(a) \xrightarrow{d} \inf_{a \in A} Q(a)$. In addition, we have $\hat{a} \xrightarrow{d} a^*$, if $\hat{a}^o \xrightarrow{d} a^*$, when Lemma 2(iii) holds.

Next, we specify the regression functions precisely. The transformations we consider are typically functions of two arguments i.e. $g : \mathbb{R} \times A \to \mathbb{R}$. The first argument corresponds to some economic variable and the second to some parameter(s). Following P&P we restrict $f(x)$ and $g(x, a)$ to be members of two families of transformations: *I-regular* and *H-regular* functions. The *I-regular* family ($\mathcal{I}$) of P&P involves integrable transformations (with respect to x). On the other hand the *H-regular* family ($\mathcal{H}$) of P&P involves locally integrable transformations that exhibit certain homogeneity property.

DEFINITION 1 (*I-regular class*). *The function* $g : \mathbb{R} \times A \to \mathbb{R}$ *is I-regular on* A *if the following hold:*

(a) *For each* $a_o \in A$, *there exist a neighborhood* N_o *of* a_o *and* $T : \mathbb{R} \to \mathbb{R}$ *bounded and integrable such that* $\|g(x, a) - g(x, a_o)\| \le \|a - a_o\| T(x)$ *for all* $a \in N_o$ *and* $\sup_{a \in N_o} |g(x, a)|$ *is integrable.*

(b) *For some* $c > 0$ *and* $k > 6/(p - 2)$ *with* $p > 4$ *given in Assumption (1b),* $\|g(x, a) - g(y, a)\| \le c\,|x - y|^k$ *for all* $a \in A$.

Definition 1 requires $g(x, a)$ to be integrable and sufficiently smooth with respect to x and a. Before we introduce the *H-regular* class, we need to define another class of transformations.

DEFINITION 2 (*regular class*). *The function* $T : \mathbb{R} \times A \to \mathbb{R}$ *is regular on* A *if the following hold:*

(a) *For all* $a \in A$, $T(., a)$ *is continuous in a neighborhood of infinity.*

(b) *For any* $a \in A$ and *compact subset* K *of* $\mathbb{R}$ *given, there exist for each* $\epsilon > 0$ *continuous functions* $\underline{T}_\epsilon$, $\overline{T}_\epsilon$, *and* $\delta_\epsilon > 0$ *such that* $\underline{T}_\epsilon(x, a) \le T(y, a) \le \overline{T}_\epsilon(x, a)$ *for all* $|x - y| < \delta_\epsilon$ *on* K, *such that* $\int_K \left(\underline{T}_\epsilon - \overline{T}_\epsilon\right)(x, a)dx \to 0$ *as* $\epsilon \to 0$.

(c) *For all* $x \in \mathbb{R}$, $T(x, .)$ *is equicontinuous in a neighborhood of* x.

The regular class essentially comprises locally integrable functions that are piecewise continuous with respect to the x argument. Functions with integrable poles e.g. $\ln x$ are not regular. Nonetheless, P&P provide limit theory for "clipped" transformations i.e. sequences of regular functions that approximate transformations with integrable poles, in large samples. Whenever we employ a regression function with integrable poles, we implicitly consider its "clipped" version[3].

DEFINITION 3 (*H-regular class*). *The function* $g : \mathbb{R} \times A \rightarrow \mathbb{R}$ *is H-regular on* A *if the following hold:*

(a) $g(\lambda x, a) = k_g(\lambda, a) h_g(x, a) + R_g(x, \lambda, a)$ *with* $h_g(., a)$ *regular on* A,
(b) $|R_g(x, \lambda, a)| \leq a_g(\lambda, a) P_g(x, a)$, *with* $\limsup_{\lambda\to\infty} \sup_{a\in A} \left\| a_g(\lambda, a) k_g^{-1}(\lambda, a) \right\| = 0$ *or*
(c) $|R_g(x, \lambda)| \leq b_g(\lambda) P_g(x, a) Q_g(\lambda x, a)$, *with* $\limsup_{\lambda\to\infty} \sup_{a\in A} \left\| b_g(\lambda, a) k_g^{-1}(\lambda, a) \right\| < \infty$,
where
(i) $\sup_{a\in A} P_g(., a)$ *is locally bounded such that for some* $c > 0$, $\sup_{a\in A} P_g(x, a) = O(e^{c|x|})$ *as* $|x| \rightarrow \infty$,
(ii) $\sup_{a\in A} Q_g(., a)$ *is bounded with* $\sup_{a\in A} Q_g(x, a) = o(|x|)$ *as* $|x| \rightarrow \infty$.

It follows from Definition 3 that an *H-regular* g is homogenous in the limit i.e.

$$g(\lambda s, a) \sim k_g(\lambda, a) h_g(s, a) \text{ for large } \lambda.$$

The functions $k_g(\lambda, a)$ and $h_g(s, a)$ are the asymptotic order and the limit homogeneous function of g respectively. For notational brevity, we write the asymptotic order of g as $k_g(\lambda) = k_g$. When that depends on some parameter, a^* say, we write $k_g(\lambda, a^*) = k_g^*$. If the asymptotic order of g does not depend on a parameter, g is referred to as H_o*-regular* ($\mathcal{H}_o$ denotes the particular family). In addition, for *H-regular* $\dot{g}$ ($\ddot{g}$) the function $\dot{h}_g$ ($\ddot{h}_g$) and $\dot{k}_g$ ($\ddot{k}_g$) are the relevant limit homogeneous functions and asymptotic orders respectively.

We are confined to transformations that are integrable or locally integrable (i.e. *I-regular* and *H-regular*). The asymptotic behaviour of non-locally integrable transformations is as yet unknown (see de Jong and Wang (2002)). de Jong and Wang (2002) provide asymptotic theory for "nearly non-locally integrable" transformations. Further, Park and Phillips (1999) develop asymptotic theory for another class of functions. This class comprises functions that grow with exponential rate (*E-regular*). Although some of our results may be extended to these two families of functions, such development is not attempted here.

Finally, before proceeding to the next section, we introduce some definitions. The following definition clarifies what we mean by correct/incorrect functional form.

DEFINITION 4. (i) *The fitted model is of correct functional form if*

$$f(.) \neq g(., a_o), \textit{ for some } a_o \in A,$$

on a set of Lebesgue measure zero.
(ii) *The fitted model is of incorrect functional form if*

$$f(.) \neq g(., a), \textit{ for every } a \in A,$$

on a set of positive Lebesgue measure.

Remark:
If the functions $f(.)$ and $g(.)$ are equal almost everywhere with respect to Lebesgue measure, then models (1) and (2) are observationally equivalent. In particular, under

Assumption 1(b), the unit root process, x_t, has absolutely continuous distribution with respect to Lebesgue measure[4], which in turn implies that $f(x_t) = g(x_t)$ *a.s.*

As mentioned earlier, the relative asymptotic order of the true and fitted models are of crucial importance for the asymptotic analysis of the NLS estimator. It is obvious from the asymptotic theory of P&P that integrable transformations of unit root processes are of the same order of magnitude (i.e. sample sums of *I-regular* transformations require the same normalisation to become convergent). On the other hand, different *H-regular* transformations can be of different order of magnitude. The following definition introduces a concept of *relative asymptotic order* for two *H-regular* regression functions:

DEFINITION 5. *Suppose that $f(.)$ is H-regular, and is $g(.,a)$ H-regular on A.*

(a) *$f(.)$ and $g(.,a)$ are of the same asymptotic order if $k_f(\lambda) = k_g(\lambda, a)$, for some $a \in A$. This is denoted by $f \sim g$.*

(b) *$f(.)$ and $g(.,a)$ are of different asymptotic order if $k_f(\lambda) \neq k_g(\lambda, a)$, for all $a \in A$. This is denoted by $f \nsim g$.*

(c) *$f(.)$ is of higher (lower) asymptotic order than $g(.,a)$, if $k_f(\lambda) > k_g(\lambda, a)$ ($k_f(\lambda) < k_g(\lambda, a)$) for all $a \in A$. This is denoted by $f \succ g$ ($f \prec g$).*

3 LIMIT THEORY

3.1 CONVERGENCE TO PSEUDO-TRUE VALUE

This section provides sufficient conditions for the convergence of the NLS estimator $\hat{a}$, defined by (3), to some pseudo-true value a^*. These conditions can be easily checked for a variety of *I-regular* and *H-regular* models. The techniques we employ, to establish convergence, are similar to those used by P&P. Some results follow from a Jennrich (1969) type of argument. Jennrich (1969) shows that under certain regularity conditions, the NLS estimator converges to the (unique) value that minimises $Q(a)$, the probability limit of the objective function $Q_n(a)$. It is more convenient to consider the shifted objective function:

$$D_n(a, a^*) = Q_n(a) - Q_n(a^*), \text{with } a^* \in A,$$

For *I-regular* and *H_o-regular* functions, we establish that the NLS estimator converges in distribution to some pseudo-true value by verifying the following condition (CN1):

CN1 (van de Vaart and Wellner, 1996[5]): *Let v_n be a normalising sequence of real numbers. Suppose that:*

(i) $v_n^{-1} \inf_{a \in A} D_n(a, a^*) \xrightarrow{d} \inf_{a \in A} D(a, a^*)$, *as $n \to \infty$.*

(ii) $D(a^*, a^*) < \inf_{a \in G} D(a, a^*)$ *a.s., for every closed set $G \subset A$ that does not contain a^*.*

Then $\hat{a} \xrightarrow{d} a^*$.

Condition CN1 is reminiscent of the Jennrich (1969) technique employed by White (1981). CN1 postulates that the NLS estimator converges weakly to the minimiser of the limit objective function. Although CN1 is applicable to *I-regular* and *H_o-regular* functions, is not alaways applicable to general *H-regular* functions, as these functions have different rates for different values of a. The following condition (CN2) due to Wu (1981) is more relevant, when the model is determined by some general *H-regular* function:

CN2 (Wu, 1981[6]): *Suppose that* $\liminf_{n\to\infty} \inf_{\{a\in A:\|a-a^*\|\geq\delta\}} D_n(a,a^*) > 0$ *a.s. (in prob.), for any* $\delta > 0$, *such that* $\{a \in A : \|a - a^*\| \geq \delta\}$ *is non-empty. Then* $\hat{a} \overset{a.s.}{\to} a^*$ *(in prob.).*

The limit theory for general *H-regular* models follows from CN2.

Our asymptotic results are qualitatively different than those obtained by White (1981) in two ways. First, in our framework the limit objective function is stochastic, and therefore its minimiser can be stochastic as well. Note that this is not true for the correctly specified models of P&P, as the limit objective function is minimised at the true-parameter in that case. Secondly, when $f \nsim g$, the limit objective function is not a complete quadratic form in f and g, for only the dominant terms feature in the limit. The function $Q(a)$ is often monotonic in a, when is not a complete quadratic form, and therefore attains its minimum at some boundary point.

We first present limit results for *I-regular* and *H_o-regular* models. The limit objective function takes different forms depending on the relative asymptotic order of f and g. Therefore, different conditions are required to guarantee that $Q(a)$ has a unique minimum. In Theorems 1-3 below, assumption (c) is an identification requirement. It ensures that the limit objective function has a unique minimum, and in view of CN1 is sufficient for the convergence of the NLS estimator to some pseudo-true value. We start with $f,\, g \in \mathcal{I}$.

THEOREM 1: ($f,\, g \in \mathcal{I}$) *Suppose that:*

(a) *Assumption 1(a,b) holds and A is compact.*

(b) $f,\, g \in \mathcal{I}$,

(c) *There is an $a^* \in A$ such that*

$$\int_{-\infty}^{\infty} [f(s) - g(s,a)]^2 ds > \int_{-\infty}^{\infty} [f(s) - g(s,a^*)]^2 ds,$$

for all $a \in A : a \neq a^*$.

Then,

$$\hat{a} \overset{p}{\to} a^*.$$

In particular we have

$$D(a,a^*) = \int_{-\infty}^{\infty} \left\{ [f(s) - g(s,a)]^2 - [f(s) - g(s,a^*)]^2 \right\} ds L(1,0)$$

with $v_n = \sqrt[4]{n}$.

Notice that although $D(a,a^*)$ in Theorem 1 is stochastic, its minimiser is deterministic. Hence, $\hat{a}$ converges in probability to some pseudo-true value.

Next, we consider *H_o-regular* models of the same order. We have the following theorem:

THEOREM 2: ($f,\, g \in \mathcal{H}_o,\ f \sim g$) *Suppose that:*

(a) *Assumption 1(a,c) holds and A is compact.*

(b) $f,\, g \in \mathcal{H}_o$ *with* $k_f = k_g$,

(c) *There is an* $a^* \in A$ *such that*

$$\int_{-\infty}^{\infty} [h_f(s) - h_g(s,a)]^2 L(1,s)ds > \int_{-\infty}^{\infty} [h_f(s) - h_g(s,a^*)]^2 L(1,s)ds \ a.s.,$$

for all $a \in A : a \neq a^*$.
Then,

$$\hat{a} \xrightarrow{d} a^*.$$

In particular we have

$$D(a,a^*) = \int_{-\infty}^{\infty} \left\{ [h_f(s) - h_g(s,a)]^2 - [h_f(s) - h_g(s,a^*)]^2 \right\} L(1,s)ds$$

with $v_n = nk_f(\sqrt{n})^2$.

The pseudo-true value in Theorem 2, can be stochastic. This is demonstrated by the examples given below:

EXAMPLE 1. (a) Let $f(s) = \theta_o s^2 1\{s > 0\}$. The fitted specification is similar to the Michaelis-Menten model (see Bates and Watts, 1988): $g(s,a) = s^3(1+as)^{-1}1\{s > 0\}$ with $\theta_o^{-1} \in A \subset \mathbb{R}_+$. Then, $D(a,a^*) = \left[\left(\theta_o - \frac{1}{a}\right)^2 - \left(\theta_o - \frac{1}{a^*}\right)^2\right] \int_0^\infty s^2 L(1,s)ds$. Hence $a^* = \theta_o^{-1}$.

(b) Let $f(s) = 1\{s > 0\}$ and $g(s,a) = a\,(2 \times 1\{s > 0\} + 1\{s < 0\})$ with $[0,2] \subseteq A$. It can be easily checked that

$$a^* = 2\int_{-\infty}^{\infty} 1\{s > 0\} L(1,s)ds \left(\int_{-\infty}^{\infty} [2 \times 1\{s > 0\} + 1\{s < 0\}]\, L(1,s)ds\right)^{-1}.$$

Finally, we provide sufficient conditions for CN1, when f and g are H_o-regular of different orders. In particular, we consider $f \succ g$. We have the following result:

THEOREM 3: (f, $g \in \mathcal{H}_o$ *with* $f \succ g$) *Suppose that:*

(a) *Assumption 1(a,c) holds and* A *is compact.*

(b) f, $g \in \mathcal{H}_o$ *with* $(k_f k_g^{-1})(\lambda) \to \infty$ *as* $\lambda \to \infty$.

(c) $\int_{-\infty}^{\infty} h_f(s)h_g(s,a)L(1,s)ds < \int_{-\infty}^{\infty} h_f(s)h_g(s,a^*)L(1,s)ds$ *a.s. for all* $a \in A : a \neq a^*$.
Then

$$\hat{a} \xrightarrow{d} a^*.$$

In particular we have

$$D(a,a^*) = 2\int_{-\infty}^{\infty} \{h_f(s)h_g(s,a^*) - h_f(s)h_g(s,a)\}\, L(1,s)ds$$

with $v_n = nk_f(\sqrt{n})k_g(\sqrt{n})$.

Alike Theorems 1 and 2, in Theorem 3 $D(a,a^*)$ is not a complete quadratic form, as the lower order terms in the objective function vanish in the limit. The subsequent example shows that in this case, the limit objective function can be strictly monotonic over the parameter space. As a result, a^* is a boundary point.

EXAMPLE 2. (a) Let $f(s) = s^2$ and $g(s,a) = s^2(1+a|s|)^{-1}$ with $A \subset \mathbb{R}_+$. Then, $D_n(a,a^*) = 2\left(\frac{1}{a^*} - \frac{1}{a}\right)\int_{-\infty}^{\infty} |s|^3 L(1,s)ds$. Hence, a^* is the lower boundary point of A.

(b) Suppose now that f is as before, but g is signed i.e. $g(s,a) = sign(s)s^2(1+a|s|)^{-1}$ with $A \subset \mathbb{R}_+$. Then, $D_n(a,a^*) = 2\left(\frac{1}{a^*} - \frac{1}{a}\right)\int_{-\infty}^{\infty} sign(s)|s|^3 L(1,s)ds$. Notice that a^* is stochastic. It alternates between the upper and lower boundary points of A depending on the realisation of the local time paths. Figure 2 shows the simulated density of $\hat{a}$ for two different choices for the parameter space. In both cases the density is bimodal with picks at the boundary of the parameter space.

Theorems 1-3 are not exhaustive. When $f \prec g$ or when one of the models is *I-regular* and the other *H-regular*, the limit results are analogous to those given above and will not to be discussed here.

Next, we consider general *H-regular* functions. The convenient characterisation of the pseudo-true value as the minimiser of some limit criterion function is not possible for this case, as the Jennrich approach is not applicable. Nonetheless, it is still possible to show that the NLS estimator has a well defined limit. As in P&P, we provide sufficient conditions for CN2, when the parameter appears as an exponent in the model. Again, there is no single general result. Our conditions vary with the relative orders of f and g. We consider three cases: $g \sim f$, $f \prec g$ and $f \succ g$. The limit behaviour of the NLS estimator is comparable to that shown for H_o*-regular* models. For $f \nsim g$, we get convergence to boundary points. We start with $f \sim g$. In Theorem 4 below, the fitted model is correctly specified up some lower order *H-regular* component q.

THEOREM 4. *($f \in \mathcal{H}_o$, $g \in \mathcal{H}$, $f \sim g$) Suppose that:*

(a) *Assumption 1(a,c) holds and A is a compact subset of $\mathbb{R}$.*

(b) *$f(x) \in \mathcal{H}_o$, $g(x,a) \in \mathcal{H}$, with $k_f(\lambda) = k_g(\lambda, a^*)$, for some $a^* \in A$.*

(c) *$f(x) - g(x,a^*) = q(x,a^*)$ with $q(x,a) \in \mathcal{H}$ such that $k_q(\lambda,a^*)k_g(\lambda,a^*)^{-1} \overset{\lambda\to\infty}{\to} 0$.*
Then CN2 holds if:

(i) *for any $a_+ \neq a^*$ and $\bar{c}, \bar{d} > 0$, there exist $\varepsilon > 0$ and a neighborhood N of a_+ such that as $\lambda \to \infty$*

$$k_q(\lambda,a^*)^{-1} \times \inf_{\substack{|c-\bar{c}|<\varepsilon \\ |d-\bar{d}|<\varepsilon}} \inf_{a\in N} |ck_g(\lambda,a^*) - dk_g(\lambda,a)| \to \infty;$$

(ii) *for all $a \in A$ and $\delta > 0$, $\int_{|s|\leq\delta} h_g(s,a)^2 ds > 0$.*

Condition (c) specifies the properties of the true and fitted models, (i) is a regularity condition similar to that of P&P (Theorem 4.3), while (ii) is an identification requirement. An example is provided next:

EXAMPLE 3. Suppose that the true specification is given by $f(s) = s^{\theta_o}(1+s)^{-1}1\{s>0\}$. The fitted model is $g(s,a) = s^a 1\{s>0\}$ with $\theta_o - 1 \in A \subset \mathbb{R}_+$. Theorem 4 holds with $a^* = \theta_o - 1$.

The subsequent result is for $f \prec g$. The NLS estimator converges to the value that minimises the descrepancy between the asymptotic orders of f and g, which is the value that minimises the asymptotic order of g i.e. $k_g(\lambda, a)$. Typically, this corresponds to the lower boundary point of A.

THEOREM 5. *($f \in \mathcal{H}_o$, $g \in \mathcal{H}$, $f \prec g$) Suppose that:*
(a) *Assumption 1(a,c) holds and A is a compact subset of $\mathbb{R}$.*
(b) *$f(x) \in \mathcal{H}_o$, $g(x,a) \in \mathcal{H}$.*
(c) *There is an $a^* \in A$, such that $k_g(\lambda, a^*)k_g(\lambda, a)^{-1} \overset{\lambda\to\infty}{\to} 0$ for all $a \in A : a \neq a^*$.*
Then CN2 holds if:

(i) *for any $a_+ \neq a^*$ and $\bar{c}, \bar{d} > 0$, there exist $\varepsilon > 0$ and a neighborhood N of a_+ such that as $\lambda \to \infty$*

$$k_g(\lambda, a^*)^{-1} \times \inf_{\substack{|c-\bar{c}|<\varepsilon \\ |d-\bar{d}|<\varepsilon}} \inf_{a\in N} |ck_g(\lambda, a^*) - dk_g(\lambda, a)| \to \infty;$$

(ii) *for all $a \in A$ and $\delta > 0$, $\int_{|s|\leq\delta} h_g(s,a)^2 ds > 0$.*

EXAMPLE 4. Theorem 5 holds for $f(s) = \ln(s)1\{s > 0\}$ and $g(s,a) = s^a 1\{s > 0\}$ with $a \in A \subset \mathbb{R}_+$. In this instance the NLS estimator converges to the lower boundary point of the parameter space. This is confirmed by Figure 3 that shows the simulated density of $\hat{a}$ for two different choices for the parameter space.

Finally, we consider $f \succ g$. This is the opposite scenario to that of Theorem 5. Therefore, one would expect that the estimator converges to the value that maximises the asymptotic order of g. Alike Theorem 4 however, in this case the leading term of the objective function is signed in general. The pseudo-true value converges to the value that minimises or maximises k_g, depending on the sign of the dominant term of $Q_n(a)$.

THEOREM 6: *($f \in \mathcal{H}_o$, $g \in \mathcal{H}$, $f \succ g$) Suppose that:*
(a) *Assumption 1(a,c) holds and A is a compact subset of $\mathbb{R}$.*
(b) *$f(x) \in \mathcal{H}_o$, $g(x,a) \in \mathcal{H}$.*
(c) *There are $\underline{a}, \overline{a} \in A$ such that $k_g(\lambda, a)k_g(\lambda, \underline{a})^{-1}$, $k_g(\lambda, \overline{a})k_g(\lambda, a)^{-1} \overset{\lambda\to\infty}{\to} \infty$ for all $a \in A : a \neq \underline{a}, \overline{a}$.*
(d) *Define a^* by*

$$a^* = \begin{cases} \overline{a}, & \text{if } \int_{-\infty}^{\infty} h_g(s,a)h_f(s)L(1,s)ds > 0, \text{ a.s.} \\ \underline{a}, & \text{if } \int_{-\infty}^{\infty} h_g(s,a)h_f(s)L(1,s)ds < 0, \text{ a.s.} \end{cases}$$

Then CN2 holds if:

(i) *for any $a_+ \neq a^*$ and $\bar{c}, \bar{d} \in \mathbb{R}$ such that $\bar{c}\bar{d} > 0$, there exist $\varepsilon > 0$ and a neighborhood N of a_+ such that as $\lambda \to \infty$*

$$k_f(\lambda)k_g(\lambda, \overline{a})^{-2} \times \inf_{\substack{|c-\bar{c}|<\varepsilon \\ |d-\bar{d}|<\varepsilon}} \inf_{a\in N} (ck_g(\lambda, a^*) - dk_g(\lambda, a)) \to \infty;$$

(ii) *for all $a \in A$ and $\delta > 0$, $\int_{|s|\leq\delta} h_g(s,a)h_f(s)ds > 0$ or $\int_{|s|\leq\delta} h_g(s,a)h_f(s)ds < 0$.*

Condition (ii) of Theorem 6 is a regularity condition. It requires that the dominant term in $Q_n(a)$ is either positive or negative in the limit. As shown by the subsequent example, a^* corresponds to the upper or lower boundary point of A, depending on the sign of the integral term in (d).

EXAMPLE 5. (a) Let $f(s) = |s|^{\theta_o}$ and $g(s,a) = |s|^a$ with $a \in A = [0, \theta_o/3]$. Then we have, $a^* = \theta_o/3$.
(b) Let $f(s) = -|s|^{\theta_o}$ and $g(s,a) = |s|^a$ with $a \in A = [0, \theta_o/3]$. Then we have, $a^* = 0$.

3.2 LIMIT DISTRIBUTION ABOUT THE PSEUDO-TRUE VALUE

This section provides limit distribution theory about some pseudo-true value. To obtain limit distribution theory, we need to impose stronger conditions than those of the previous section. In particular, we assume that the fitted response function is differentiable with respect to a. We have seen earlier, that for $f \nsim g$, the NLS estimator may converge to boundary points. Andrews (1999) develops techniques that yield limit distribution results, when the parameter is on a boundary. The Andrews (1999) approach and other methods that rely on linearisation of the objective function (e.g. Wooldridge, 1994) are not applicable in our case. When the pseudo-true value is on a boundary, the limit objective function is not minimised at a turning point. As a result, $\dot{Q}_n(a^*)$ and $\ddot{Q}_n(a^*)$ are of the same order of magnitude. There is no obvious way of obtaining limit distribution theory or convergence rates in this case. Hereafter, we rule out boundary points, by focusing on f, $g \in \mathcal{I}$ and, f, $g \in \mathcal{H}$ with $f \sim g$.

The asymptotic theory developed by P&P is not always sufficient to yield limit distribution results under FFM. For certain kinds of misspecification, second order limit theory is required. Jeganathan (2003) provides second order limit theory for integrable transformations that is utilised here. Second order limit theory for locally integrable transformations is yet undeveloped. For this reason, the limit results we provide for the *H-regular* class, are limited. Suppose that f, $g \in \mathcal{H}$, with $f \sim g$ and set $q = f - g$. Then, there are two possibilities:

P1: The fitted model is correctly specified up to some lower order term, i.e.:

$$k_q(\lambda, a^*)/k_f(\lambda) \to 0, \text{ as } \lambda \to \infty, \text{ for some } a^* \in A,$$

P2: The functions f and g do not agree at all i.e.

$$k_q(\lambda, a)/k_f(\lambda) \nrightarrow 0 \text{ as } \lambda \to \infty, \text{ for all } a \in A.$$

The latter may happen, if the models involve covariates normalised by the sample size. Consider for example $f_n(x) = (1 + \exp(x/\sqrt{n}))^{-1}$ and $g_n(x,a) = a1\{x/\sqrt{n} > 1\}$. These kind of functions are proposed by Saikkonen and Choi (2004) for modelling transition effects in regressions with unit roots. Only the first scenario is considered here, as for the latter, second order limit theory for *H-regular* transformations is required.

To obtain limit distribution results for *I-regular* models, we follow the P&P approach. We utilise the convergence result of Theorem 1 and then linearise the objective function. Limit distribution and convergence rates are the same as those reported by P&P, for correctly specified integrable models. In this respect, the limit theory for the particular class of models is analogous to that for stationary misspecified models.

For *H-regular* models, we obtain limit distribution results following the Wooldridge (1994) approach. The aforementioned method is also based on a linearisation of the objective function. In particular, under suitable conditions we get that:

$$s_n^{-1} v_n' (\hat{a} - a^*) \xrightarrow{d} -\ddot{Q}^{-1} \dot{Q}, \tag{6}$$

where, v_n is sequence of normalising matrices, s_n a normalising sequence of real numbers, and $\dot{Q}$, $\ddot{Q}$ the limits of $\dot{Q}_n$, $\ddot{Q}_n$ respectively. By Wooldridge (1994, Theorem 10.1) and de Jong and Hu (2006, Theorem 1), conditions $C1 - C5a$ below are sufficient for (6).

C1 : $\left((s_n v_n)^{-1} \dot{Q}_n(a^*),\ v_n^{-1} \ddot{Q}_n(a^*) v_n'^{-1} \right) \xrightarrow{d} \left(\dot{Q},\ \ddot{Q} \right)$, *as* $n \to \infty$.
C2 : $\ddot{Q} > \mathbf{0}$ *a.s.*
C3 : *There is a sequence* μ_n *such that* $\mu_n v_n^{-1} \to \mathbf{0}$ *as* $n \to \infty$ *and*
$\sup_{a \in N_n} \left\| \mu_n^{-1} \left(\ddot{Q}_n(a) - \ddot{Q}_n(a^*) \right) \mu_n'^{-1} \right\| = o_p(1)$, *where* $N_n = \left\{ a : \| s_n^{-1} \mu_n (a - a^*) \| \leq 1 \right\}$.
C4 : *(i)* $s_n^{-1} \mu_n (\hat{a} - a^*) = o_p(1)$ *or*
(ii) $Q_n(a)$ *is globally convex.*
C5a : $s_n = 1$.
C5b : $s_n \to \infty$, *with* $\| s_n v_n^{-1} \| \to 0$, *as* $n \to \infty$.

Condition $C5a$ requires the score being of smaller order than the Hessian. For most problems s_n equals one, leading to the familiar v_n*-consistency* for extremum estimators. If FFM is committed under nonstationarity, s_n can be divergent. Under misspecification, the score typically is of higher order of magnitude, than what is under correct specification, and as result the convergence rates are slower. Theorem 10.1 of Wooldridge (1994) and Theorem 1 of de Jong and Hu (2006), can be trivially extended under $C5b$:

LEMMA 3: *Conditions* $C1 - C4$ *and* $C5b$ *are sufficient for (4).*

We utilise Lemma 3, to obtain limit distribution results for misspecified *H-regular* models.

P&P show that limit distribution for *I-regular* models, under correct FFM, is mixed Gaussian and the convergence rate is $n^{1/4}$. Under FFM the convergence rate is the same as that attained under correct specification. The limit distribution is mixed Gaussian, but with two Gaussian components rather than one, which is the case for correctly specified models. The actual result is given by Theorem 7 next:

THEOREM 7: $(f,\, g \in \mathcal{I})$ *Suppose that:*
(a) $f,\ g,\ \dot{g},\ \ddot{g} \in \mathcal{I}$, *and the conditions of Theorem 1 hold.*
(b) a^* *is interior in* A.
(c) $\int_{-\infty}^{\infty} |s z(s, a^*)|\, ds < \infty$, *where* $z(s, a^*) = \dot{g}(s, a^*) \left(f(s) - g(s, a^*) \right)$.
(d) *Let* $\hat{z}$ *be the Fourier transform of* z. *The characteristic function of* η_t *satisfies:*
(i) $|\phi(s)| \leq C_1 |s|^{-\beta}$, *as* $s \to \infty$, *for some* C_1, $\beta > 0$ *and*
(ii) $|\hat{z}(s, a^*) \phi(s)| \leq C_2 |s|^{-(2+\gamma)}$, *as* $s \to \infty$, *for some* $C_2 > 0$, $0 < \gamma < 1$.
(e) $\int_{-\infty}^{\infty} \dot{g}(s, a^*) \dot{g}(s, a^*)' ds - \int_{-\infty}^{\infty} \ddot{G}(s, a^*) \left(f(s, \theta_o) - g(s, a^*) \right) ds > \mathbf{0}$.

Then as $n \to \infty$,

$$n^{1/4}(\hat{a} - a^*) \xrightarrow{d} \left[\int_{-\infty}^{\infty} \dot{g}(s,a^*)\dot{g}(s,a^*)' ds - \int_{-\infty}^{\infty} \ddot{G}(s,a^*)\left(f(s,\theta_o) - g(s,a^*)\right) ds\right]^{-1}$$
$$\times L(1,0)^{-1/2}\left[b^{1/2}W_1(1) + \left(\int_{-\infty}^{\infty} \dot{g}(s,a^*)\dot{g}(s,a^*)' ds\right)^{1/2} W_2(1)\right],$$

where $(W_1(1), W_2(1)/\sigma)$ is standard bivariate Gaussian independent of $L(1,0)$, and b is the constant in Theorem 1 of Jeganathan (2003)[7].

Conditions (c) and (d) of Theorem 7 are required for second order limit theory of integrable transformations (see Jeganathan, 2003). Condition (e) is an identification requirement.

Next, we present limit theory for f, $g \in \mathcal{H}$. Sometimes it is difficult to check Condition $C4(ii)$. To establish our results, we provide sufficient conditions for $C4(i)$ instead. We assume that the fitted model is corectly specified up to the lower order component, q (i.e. **P1** holds). P&P show that for correctly specified *H-regular* models, the convergence rate is $\sqrt{n}k_{\dot{f}}(\sqrt{n})$ and the limit distribution determined by stochastic integrals. For the kind of misspecification under consideration, convergence is slower because the score involves additional components. In particular, the convergence rate is $k_{\dot{g}}(\sqrt{n}, a^*)/k_q(\sqrt{n}, a^*)$. Further, the limit distribution involves functionals of Brownian motion, relating to the q term, but not stochastic integrals. Before proceeding to the next result, we introduce some notation. Let $\varepsilon > 0$ and π_n, be a sequence of real numbers. Define a neighborhood of $a^* \in A \subset \mathbb{R}$ by

$$N(\varepsilon, \pi_n, a^*) = \left\{a : |\pi_n(a - a^*)| \le n^{\varepsilon/3}\right\}.$$

We have the following theorem.

THEOREM 8. ($f \in \mathcal{H}_o$, $g \in \mathcal{H}$, $f \sim g$) *Suppose that:*

(a) *Assumption 1(a,c) holds with A compact subset of $\mathbb{R}$.*

(b) *$f \in \mathcal{H}_o$ and g, $\dot{g}$, $\ddot{g} \in \mathcal{H}$ on A.*

(c) *There exists an interior point of A, a^* such that $f(s) - g(s,a^*) = q(s,a^*)$ with q, $\dot{g} \in \mathcal{H}$, and $q \prec g$, $\dot{g}$.*

(d) *There are functions $\overline{h}$, $\underline{h} \in \mathcal{H}_o$ with asymptotic orders $\overline{k}$ and $\underline{k}$ respectively such that $\pi_n \equiv \left(\underline{k}/\overline{k}k_q(a^*)\right)(\sqrt{n}) \to \infty$ as $n \to \infty$, and $\overline{h} \ge \sup_{a\in A}|\dot{g}|$, $\underline{h} \le \inf_{a\in A}\dot{g}^2$.*

(e) *$\int_{|s|\le\delta} \dot{h}_g(s,a^*)^2 ds > 0$ for all $\delta > 0$,*

(f) *For any $\bar{s} > 0$, there exists $\varepsilon > 0$ such that as $n \to \infty$,*

$$\dot{k}_g(\sqrt{n}, a^*)^{-2}\left(\sup_{|s|\le\bar{s}}\left|\ddot{g}(\sqrt{n}s, a^*)q(\sqrt{n}s, a^*)\right|\right) \to 0, \tag{7}$$

$$n^{\varepsilon}\pi_n^{-2}k_q(\sqrt{n}, a^*)^{-1}\left(\sup_{|s|\le\bar{s}}\sup_{a\in N(\varepsilon,\pi_n,a^*)}\left|\ddot{g}(\sqrt{n}s, a)\right|\right) \to 0, \tag{8}$$

$$n^{\varepsilon}\pi_n^{-2}k_q(\sqrt{n}, a^*)^{-1}\dot{k}_g(\sqrt{n}, a^*)^{-1}\left(\sup_{|s|\le\bar{s}}\sup_{a\in N(\varepsilon,\pi_n,a^*)}\left|\dddot{g}(\sqrt{n}s, a)q(\sqrt{n}s, a^*)\right|\right) \to 0; \tag{9}$$

Then as $n \to \infty$,

$$\frac{\dot{k}_g(\sqrt{n},a^*)}{k_q(\sqrt{n},a^*)}(\hat{a}-a^*) \xrightarrow{d} \left[\int_0^1 \dot{h}_g(V(r),a^*)^2 dr\right]^{-1} \int_0^1 \dot{h}_g h_q(V(r),a^*)dr.$$

Theorem 8 provides sufficient conditions for Lemma 3. Condition (c) specifies the kind of misspecification under consideration. Condition (d) is sufficient for $C4(i)$, (e) is an identification requirement and (f) is sufficient for $C3$. Condition (d) is somewhat restrictive for general *H-regular* models. It can be checked however for several H_o*-regular* specifications[8]. Note that for general *H-regular* models, convergence rates depend on the pseudo-true value a^*. Some examples are provided next.

EXAMPLE 7. (a) Let $f(s) = \theta_o s^2$ and $g(s,a) = |s|^3 (1 + a|s|)^{-1}$ with $A \subset \mathbb{R}_+$. It follows from Theorem 8 that $a^* = \theta_o^{-1}$ and

$$\sqrt{n}(\hat{a}-a^*) \xrightarrow{d} \left[\int_0^1 |V(r)|^4 dr\right]^{-1} \int_0^1 |V(r)|^3 dr.$$

(b) Let $f(s) = |s|^{\theta_{1o}} (1 + |s|^{\theta_{2o}})^{-1}$ with $\theta_{1o} > \theta_{2o} > 0$ and $\theta_{1o} > (3\theta_{2o} - 1)/2$ (local integrability requirement). The empirical model is determined by $g(s,a) = |s|^a$ with $A = [\underline{a}, \overline{a}] \subset \mathbb{R}_+$. Suppose that $2\underline{a} > (3\theta_{1o} - 5\theta_{2o})/4 + \overline{a} + \delta$ for some $\delta > 0$ (sufficient for condition (c)). It follows from Theorem 8 that $a^* = \theta_{1o} - \theta_{2o}$ and

$$n^{\theta_{2o}/2} \ln\left(\sqrt{n}\right)(\hat{a}-a^*) \xrightarrow{d} -\left[\int_0^1 |V(r)|^{2a^*} dr\right]^{-1} \int_0^1 |V(r)|^{2a^*-\theta_{2o}} dr.$$

4 CONCLUSION

Accepting that any empirical model is a mere approximation rather a true data generating mechanism, it is important to know the estimators' limit behaviour under functional form misspecification. This is exactly the problem we have addressed in this paper. We have explored the limit behaviour of the NLS estimator, when the true and fitted models involve a unit root covariate. For nonstationary misspecified models the behaviour of the NLS estimator depends on the nature of the true and fitted models. We have shown that, when the fitted regression function is of different order of magnitude than the true model, the estimator may converge to boundary points of the parameter space. White (1981) and Domowitz and White (1982) show that, for stationary models the convergence rates and the limit distribution theory under misspecification are the same as those obtained under correct specification. This is not always the case, when the covariate is a unit root process. We have demonstrated that if FFM is committed, convergence rates can be slower and limit distribution different than that obtained for correctly specified models.

Our analysis provides some guidance for the adequacy of estimated models. For example, estimates that are close to boundary points constitute evidence for misspecification. In addition, our results are useful for the development of testing procedures for regression models with integrated regressors. To obtain power rates for certain specification tests (tests without specific alternative), it is necessary to characterise the limit of the NLS estimator under misspecification. Further, the implementation of some model selection procedures (tests with specific alternative) requires limit distribution theory about the pseudo-true value. Specification testing procedures that exploit the results presented here are under development by the author.

Our analysis is not exhaustive. To obtain the limit distribution of the least squares estimator, under certain type of misspecification, second order asymptotic theory for H-regular transformations is required. In addition, following P&P we have considered models with a single covariate. Some results for multi-covariate models linear in parameters are provided by Kasparis (2005). This work shows that analysing single covariate models is in itself a complicated problem. Therefore, extensions to multi-covariate models nonlinear in parameters may prove to be a challenging task.

NOTES

1. See also de Jong (2004), Pötscher (2004), Jeganathan (2003, 2004) for some further developments.
2. Jennrich (1969) considers nonlinear regressions with fixed covariates. For extensions of this approach see for example Domowitz and White (1982) and the references there in.
3. de Jong (2004) extends the asymptotic theory of P&P for *regular* functions to a more general class of transformations. This class comprises *locally integrable* functions with finitely many poles that are continuous and monotone between them. Pötscher (2004) generalises the limit theory further, making it applicable to all locally integrable functions under more restrictive assumptions about the errors that drive the unit root processes. The results of de Jong (2004) and Pötscher (2004) are not readily available for parameterised regression functions. Extensions to models non-linear in parameter is possible, but that would divert attention from the main purpose of the paper.
4. See Lemma 3.1 in Pötscher (2004).
5. See Lemma 3.2.1 in van de Vaart and Wellner (1996).
6. See Lemma 1 in Wu (1981).
7. I would like to thank Peter Phillips for pointing out the relevance of Jeganathan's work for the proof of this result.
8. When $g \in \mathcal{H}_o$, the sequence π_n simplifies to $\dot{k}_g(\sqrt{n})/k_q(\sqrt{n})$.

REFERENCES

Andrews, D.W.K. (1999) Estimation when a parameter is on a boundary, *Econometrica,* 67, 1341-1383.

Bates D.M. and D.G. Watts (1988) Nonlinear regression analysis and its applications. Wiley.

Bierens, H.J. (1984) Model specification testing of time series regressions. *Journal of Econometrics* 26, 323-353.

———— (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.

Cox, D.R. (1961) Tests for separate families of hypotheses, *Journal of the Royal Statistical Society,* Series B, 32, 323-344.

———— (1962) Further results on tests for separate families of Hypotheses, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability,* Vol 1. Berkeley: University of California Press.

Davidson, R. and J.G. McKinnon (1981) Several tests for model specification in the presence of alternative hypotheses, *Econometrica,* 49, 781-793.

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

de Jong, R.M. and L. Hu (2006) Nonlinear models with integrated regressors and convergence order results, Working Paper, Department of Economics, Ohio State University.

de Jong, R.M. and C-H. Wang (2002) Further results on the asymptotics for nonlinear transformations of integrated time series, *Econometric Theory* 21, p. 413-430.

Domowitz, I. and H. White (1982) Misspecified models with dependent observations *Journal of Econometrics*, 20, 35-58.

Hansen, L.P. (1981) Large sample properties of generalized method of moments estimators," *Econometrica,* 50, 1029-1054.

Jeganathan, P. (2003) Second order limits of functionals of sums of linear processes that converge to fractional stable motions," Working Paper, Indian Statistical Institute.

——— (2004) Limits of functionals of sums of linear processes that converge to fractional stable motions, *The Annals of Probability* 32, 1771–1795.

Jennrich, R.I. (1969) Asymptotic properties of non-linear least squares estimation, *The Annals of Mathematical Statistics* 40, 633-643.

Kasparis, I. (2005) Detection of functional form misspecification in cointegrating relations, ESWC 2005.

Newey, W.K. and D.M. McFadden (1994) Large sample estimation and hypothesis testing, in *Handbook of Econometrics,* Vol. IV, ed. by R.F. Engle and D.L. McFadden. Amsterdam: Elsevier, 2213-2145.

Park, J.Y. and P.C.B. Phillips (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.

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.

Voung, Q. (1989) Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica,* 57, 307-333.

White, H. (1981) Consequences and detection of misspecified nonlinear regression models, *Journal of the American Statistical Association,* 76, 419-433.

White, H and I. Domowitz (1984) Nonlinear regression with dependent observations *Econometrica,* 52, 143-162.

Wooldridge, J.M. (1994) Estimation and inference for dependent processes, in *Handbook of Econometrics,* Vol. IV, ed. by R.F. Engle and D.L. McFadden. Amsterdam: Elsevier, 2639-2738.

Wu, C.F. (1981) Asymptotic theory for non-linear least squares estimation, *The Annals of Statistics,* 9, 501-513.

5 Appendix A (technical results)

For a function $f : X \to Y$, $f^{-1}[A]$ denotes its inverse image under the set $A \subset Y$. In addition, $\bar{A}$ is the closure of the set A.

LEMMA A1. *Suppose that the random vector (X,Y) is defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and the random vector $(X,Y)^o$ on the probability space $(\Omega, \mathcal{F}, \mathbf{P})^o$. If $(X,Y) \stackrel{d}{=} (X,Y)^o$, the following hold:*

(a) $(X, Y, X+Y) \stackrel{d}{=} (X, Y, X+Y)^o$.

(b) $(X, Y, XY) \stackrel{d}{=} (X, Y, XY)^o$.

(c) $(X, Y, f(X)) \stackrel{d}{=} (X, Y, f(X))^o$*, for $f(.)$ $\mathcal{B}/\mathcal{B}$-measurable.*

Proof of Lemma A1: (a) Suppose that $F(x,y)$ is the distribution function of (X,Y) and, $F^o(x,y)$ is the distribution function of $(X,Y)^o$. Then, for $\bar{x}$, $\bar{y}$, $\bar{z} \in \mathbb{R}$ we have

$$\begin{aligned}\mathbf{P}\left(X \le \bar{x},\, Y \le \bar{y},\, X+Y \le \bar{z}\right) &= \int_{x=-\infty}^{\bar{x}} \int_{y=-\infty}^{\min(\bar{y},\, \bar{z}-x)} dF(x,y) \\ &= \mathbf{P}^o\left(X^o \le \bar{x},\, Y^o \le \bar{y},\, X^o+Y^o \le \bar{z}\right).\end{aligned}$$

The last equality above follows from the fact that $F(x,y) = F^o(x,y)$.

The proof of (b) is similar to that of (a). We show (c). For f measurable and B_1, B_2, $B_3 \in \mathcal{B}$ we have

$$\begin{aligned}\mathbf{P}\left(Y \in B_1,\, X \in B_2 \cap f^{-1}[B_3]\right) &= \mathbf{P}^o\left(Y^o \in B_1,\, X^o \in B_2 \cap f^{-1}[B_3]\right) \Leftrightarrow \\ \mathbf{P}\left(Y \in B_1,\, X \in B_2,\, f(X) \in B_3\right) &= \mathbf{P}^o\left(Y^o \in B_1,\, X^o \in B_2,\, f(X^o) \in B_3\right),\end{aligned}$$

as required.■

LEMMA A2. *Let $G : A \times X \to \mathbb{R}$, where A a compact subset of $\mathbb{R}^p$ and X a measurable space. $G(.,x)$ continuous for each $x \in X$. For each $a \in A$, $G(a,.)$ is a measurable function. Suppose that $\{A_k\}$ is an increasing sequence of finite subsets of A, whose limit is dense in A. Then,*

$$\lim_{k\to\infty} \inf_{a\in A_k} G(a,x) = \inf_{a\in A} G(a,x),$$

everywhere on X.

Proof of Lemma A2: By the compactness of A, there is $a_*(x) \in A$ (which is measurable e.g. Jennrich 1969) such that

$$G(a_*(x), x) = \inf_{a\in A} G(a,x)$$

We can find a measurable sequence $\{a_k(x)\}$ that satisfies:

$$\|a_*(x) - a_k(x)\| = \inf_{a\in A_k} \|a_*(x) - a\|.$$

Since $A = \overline{\cup_{k\in\mathbb{N}} A_k}$, for each $\epsilon > 0$,

$$\|a_*(x) - a(x)\| < \epsilon,$$

for some $a(x) \in \cup_{k\in\mathbb{N}} A_k$. Clearly, there is some $N(\epsilon, x) \in \mathbb{N}$ such that $a(x) \in A_{N(\epsilon,x)}$. In addition, because $\{A_k\}$ is increasing, we have $a(x) \in A_k$ for $k \geq N(\epsilon, x)$. Hence,

$$\|a_*(x) - a_k(x)\| < \epsilon, \text{ for } k \geq N(\epsilon, x).$$

Therefore,

$$\lim_{k\to\infty} a_k(x) = a_*(x),$$

everywhere on X. In view of this and the continuity of $G(., x)$ we get

$$G(a_*(x), x) = \lim_{k\to\infty} G(a_k(x), x) \geq \lim_{k\to\infty} \inf_{a\in A_k} G(a, x) \geq G(a_*(x), x),$$

as required.■

6 Appendix B (main results)

Proof of Lemma 1. (i) Suppose that $G_n(a)$ is defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and $G_n^o(a)$ on some probability space $(\Omega, \mathcal{F}, \mathbf{P})^o$. Let $\{A_k\}$ be an increasing sequence of finite subsets of A, whose limit is dense in A. Define the sets $C(A_k)$ and $C^o(A_k)$ as

$$C(A_k) = \left\{\omega \in \Omega : \inf_{a\in A_k} G_n(a) > y\right\} \text{ and } C^o(A_k) = \left\{\omega \in \Omega^o : \inf_{a\in A_k} G_n^o(a) > y\right\}$$

with $y \in \mathbb{R}$. Due to the equality of the finite dimensional distributions we have

$$\mathbf{P}(C(A_k)) = \mathbf{P}^o(C^o(A_k)) \tag{10}$$

for all $k \in \mathbb{N}$. In view of (10) and Lemma A2

$$\mathbf{P}(C(A)) = \lim_{k\to\infty} \mathbf{P}(C(A_k)) = \lim_{k\to\infty} \mathbf{P}^o(C^o(A_k)) = \mathbf{P}^o(C^o(A))$$

as required.

(ii) Next, we show part (ii). Let $\{A_k\} \subset A$ be an increasing sequence of finite sets, whose limit is dense in A. Define $\tilde{a}_k$ and $\tilde{a}_k^o$ **as**

$$\tilde{a}_k = \arg\min_{a\in A_k} G_n(a) \text{ and } \tilde{a}_k^o = \arg\min_{a\in A_k} G_n^o(a)$$

Then

$$G_n(\tilde{a}) = \inf_{a\in A} G_n(a) = \lim_{k\to\infty} \inf_{a\in A_k} G_n(a) = \lim_{k\to\infty} G_n(\tilde{a}_k), \tag{11}$$

where the second equality above is due to Lemma A2. Fix $\delta > 0$. Then, by the uniqueness of $\tilde{a}$, the compactness of A and the continuity of $G_n(.)$ we have

$$\epsilon := \inf_{\{a\in A:\|a-\tilde{a}\|\geq\delta\}} G_n(a) - G_n(\tilde{a}) > 0.$$

Therefore, by (11) for k large enough, we have

$$G_n(\tilde{a}_k) - G_n(\tilde{a}) < \epsilon \Rightarrow G_n(\tilde{a}_k) < \inf_{\{a\in A:\|a-\tilde{a}\|\geq\delta\}} G_n(a),$$

which in turn implies that

$$\lim_{k\to\infty} \tilde{a}_k = \tilde{a}. \tag{12}$$

Similarly,

$$\lim_{k\to\infty} \tilde{a}_k^o = \tilde{a}^o. \tag{13}$$

Choose $\tilde{a}_k$ as follows. Denote by $m(k)$ the number of points in the set A_k and write $A_k = \{a_j,\ j = 1, ..., m(k)\}$. Then define

$$\tilde{a}_k(\omega) = \sum_{i=1}^{m(k)} a_i \inf_{1\le j\le m(k)} 1\{\omega \in E_{ij}\},\ E_{ij} = \left\{ \begin{array}{l} \{\omega \in \Omega : G_n(a_i) \le G_n(a_j)\}\,,\ i \le j \\ \{\omega \in \Omega : G_n(a_i) < G_n(a_j)\}\,,\ i > j \end{array} \right.$$

and

$$\tilde{a}_k(\omega) = \sum_{i=1}^{m(k)} a_i \inf_{1\le j\le m(k)} 1\{\omega \in E_{ij}^o\},\ E_{ij}^o = \left\{ \begin{array}{l} \{\omega \in \Omega^o : G_n^o(a_i) \le G_n^o(a_j)\}\,,\ i \le j \\ \{\omega \in \Omega^o : G_n^o(a_i) < G_n^o(a_j)\}\,,\ i > j \end{array} \right.$$

It is easy to check that $\tilde{a}_k$ and $\tilde{a}_k^o$ are measurable minimisers of $Q_n(a)$ and $Q_n^o(a)$ on A_k, respectively. Notice, that for all $y \in \mathbb{R}^p$,

$$\mathbf{P}\left(\tilde{a}_k \le y\right) = \sum_{i=1}^{m(k)} \mathbf{P}\left(\{a_i \le y\}\,, \cap_{1\le j\le m(k)} E_{ij}\right) \tag{14}$$

and

$$\mathbf{P}\left(\tilde{a}_k^o \le y\right) = \sum_{i=1}^{m(k)} \mathbf{P}^o\left(\{a_i \le y\}\,, \cap_{1\le j\le m(k)} E_{ij}^o\right). \tag{15}$$

In view of (14) and (15) and by the equality of the finite dimensional distributions, we have $\tilde{a}_k \stackrel{d}{=} \tilde{a}_k^o$. Therefore, (12) and (13) give

$$\mathbf{P}\left(\tilde{a} \le y\right) = \lim_{k\to\infty} \mathbf{P}\left(\tilde{a}_k \le y\right) = \lim_{k\to\infty} \mathbf{P}^o\left(\tilde{a}_k^o \le y\right) = \mathbf{P}^o\left(\tilde{a}^o \le y\right),$$

as required.■

Proof of Lemma 2. Part (i) follows from repeated application of Lemma A1. Part (ii) and (iii) follow from Lemma 1. We shall provide an alternative proof for parts (ii) and (iii).

We start with the proof of part (ii). Write $Q_n\left(a\right) = \bar{Q}_n(z_n, a)$ and $Q_n^o\left(a\right) = \bar{Q}_n(z_n^o, a)$ where $\bar{Q}_n(z, a)$ is a sequence of functions $\bar{Q}_n : Z_n \times A \to \mathbb{R}$, with Z_n the measurable space $(\mathbb{R}^{2n}, \mathcal{B}(\mathbb{R}^{2n}))$. Now, by the compactness of A and the continuity of $\bar{Q}_n(z, .)$ there is a measurable function $a_n(z)$ (c.f. Jennrich, 1969) such that

$$\inf_{a\in A} \bar{Q}_n(z, a) = \bar{Q}_n(z, a_n(z))$$

Notice that $a_n(z_n) \stackrel{d}{=} a_n(z_n^o)$. Therefore,

$$\bar{Q}_n(z_n, a_n(z_n)) \stackrel{d}{=} \bar{Q}_n(z_n^o, a_n(z_n^o)),$$

as required.

For part (iii) notice that if $\bar{Q}_n(z,.)$ has a unique minimum for each $z \in Z_n$, then $\hat{a} = a_n(z_n)$ and $\hat{a}^o = a_n(z_n^o)$ and therefore $\hat{a} \overset{d}{=} \hat{a}^o$.■

Proof of Theorem 1. First we check condition (i) of CN1. By Theorem 3.2 and Lemma 7 in P&P,

$$\sup_{a\in A} |D_n^o(a,a^*) - D(a,a^*)| = o_p(1)$$

where $D(a,a^*) = \int_{-\infty}^{\infty} \{[f(s,\theta_o) - g(s,a)]^2 - [f(s,\theta_o) - g(s,a^*)]^2\} dsL(1,0)$. Therefore, $\inf_{a\in A} D_n^o(a,a^*) \overset{d}{\to} \inf_{a\in A} D(a,a^*)$ and by Lemma 2(ii), $\inf_{a\in A} D_n(a,a^*) \overset{d}{\to} \inf_{a\in A} D(a,a^*)$.

Next, we check condition (ii). Let G be a closed subset of A that does not contain a^*. By the continuity of $D(a,a^*)$ (Lemma 8b in P&P) and the compactness of G, $D(a,a^*)$ attains a minimum on G. Therefore, it follows from condition (b) of Theorem 1 that $\inf_{a\in G} D(a,a^*) > D(a^*,a^*)$ as required.■

Proof of Theorems 2 and 3. Similar to the proof of Theorem 1.■

Proof of Theorem 4. Define

$$m(\sqrt{n},a)^2 = \frac{1}{nk_g(\sqrt{n},a)^2} \sum_{t=1}^n g(x_t^o,a)^2,$$
$$m(a)^2 = \int_{-\infty}^{\infty} h_g(s,a)^2 L(1,s) ds.$$

It follows from Lemma A6(c) and Theorem 3.3 of P&P that $\sup_{a\in A} |m(\sqrt{n},a) - m(a)| = o_{a.s.}(1)$. Moreover, $m(a)$ continuous $a.s.$ by Lemma 8(a) of P&P and greater than zero $a.s.$ due to condition (ii).

Fix $\delta > 0$ such that the set $A_o = \{|a - a^*| \geq \delta\} \subset A$, is non-empty. Let a_+ be an arbitrary point in A_o. Set $\bar{c} = m(a^*)$, $\bar{d} = m(a_+)$ and notice that $\bar{c}, \bar{d} > 0$ $a.s.$, since $m > 0$ $a.s.$ Fix $\epsilon > 0$. Then, by the continuity of $m(.)$ and P&P (Theorem 3.3), there is a neighborhood of a_+, N say, such that

$$|m(\sqrt{n},a^*) - \bar{c}|, \sup_{a\in N} \left|m(\sqrt{n},a) - \bar{d}\right| < \epsilon \ a.s.,$$

for n large enough.

Next, let

$$A_n(a) = \frac{1}{n} \sum_{t=1}^n (g(x_t^o,a) - g(x_t^o,a^*))^2,$$
$$B_n(a) = \frac{1}{n} \sum_{t=1}^n (g(x_t^o,a) - g(x_t^o,a^*)\, u_t^o,$$
$$C_n(a) = \frac{1}{n} \sum_{t=1}^n (g(x_t^o,a) - g(x_t^o,a^*)) q(x_t^o,a^*).$$

Due to condition (c), the objective function $D_n^o(a,a^*)$ can be written as

$$n^{-1} D_n^o(a,a^*) = A_n(a) - 2B_n(a) - 2C_n(a).$$

By the backward triangle inequality[1], we get

$$A_n(a)^{1/2} \geq \left|k_g(\sqrt{n},a^*) m(\sqrt{n},a^*) - k_g(\sqrt{n},a) m(\sqrt{n},a)\right|$$

[1] i.e. $\|x - y\| \geq |\|x\| - \|y\||$.

In view of this and condition (i) we have

$$k_q(\sqrt{n},a^*)^{-2}\inf_{a\in N}A_n(a)\overset{a.s.}{\to}\infty \text{ and } k_q(\sqrt{n},a^*)\sup_{a\in N}A_n(a)^{-1/2}\overset{a.s.}{\to}0 \tag{16}$$

Further, since

$$\begin{array}{l}\frac{1}{n}\sum_{t=1}^n (u_t^o)^2 = O_p(1),\\ \frac{1}{n}\sum_{t=1}^n q(x_t^o,a^*)^2 = O_p\left(k_q(\sqrt{n},a^*)^2\right),\end{array}$$

it follows from the Cauchy-Schwarz inequality that

$$\begin{array}{c}\sup_{a\in N}A_n(a)^{-1}\left|B_n(a)\right| \le \sup_{a\in N}A_n(a)^{-1/2}O_p(1) = o_p(1),\\ \sup_{a\in N}A_n(a)^{-1}\left|C_n(a)\right| \le k_q(\sqrt{n},a^*)\sup_{a\in N}A_n(a)^{-1/2}O_p(1) = o_p(1).\end{array} \tag{17}$$

Now from (16) and (17) we have

$$\begin{aligned} n^{-1}\inf_{a\in N}D_n^o(a,a^*) &\ge \inf_{a\in N}\left(A_n(a)\right)\left(1-2\sup_{a\in N}\left(A_n(a)^{-1}\left|B_n(a)\right|\right)-2\sup_{a\in N}\left(A_n(a)^{-1}\left|C_n(a)\right|\right)\right)\\ &= \inf_{a\in N}A_n(a)(1+o_p(1))\overset{p}{\to}\infty, \end{aligned} \tag{18}$$

Because A_o is compact, we can find a finite number of open balls, $\{N_i\}_{1\le i\le k\in\mathbb{N}}$ say, with centres in A_o, that cover A_o. In view of this and (18)

$$n^{-1}\inf_{a\in A_o}D_n^o(a,a^*)\ge n^{-1}\min_{1\le i\le k}\inf_{a\in N_i}D_n^o(a,a^*)\overset{p}{\to}\infty \tag{19}$$

Therefore by (19) and Lemma 2(ii), for any $\zeta,\eta>0$

$$\mathbf{P}\left(n^{-1}\inf_{a\in A_o}D_n(a,a^*)>\zeta\right)=\mathbf{P}^o\left(n^{-1}\inf_{a\in A_o}D_n^o(a,a^*)>\zeta\right)>1-\eta,$$

when n is large enough, and the result follows.■

Proof of Theorem 5. Fix $\delta>0$ such that the set $A_o=\{|a-a^*|\ge\delta\}\subset A$, is non-empty. Let a_+ be an arbitrary point in A_o and let N be a neighborhood of a_+ given in condition (i). Next, define

$$\begin{array}{l} A_n(a)=\frac{1}{n}\sum_{t=1}^n\left(g(x_t^o,a)-g(x_t^o,a^*)\right)^2,\\ B_n(a)=\frac{1}{n}\sum_{t=1}^n\left(g(x_t^o,a)-g(x_t^o,a^*)\right)u_t^o,\\ C_n(a)=\frac{1}{n}\sum_{t=1}^n\left(g(x_t^o,a)-g(x_t^o,a^*)\right)f(x_t^o),\\ E_n(a)=-\frac{1}{n}\sum_{t=1}^n\left(g(x_t^o,a)-g(x_t^o,a^*)\right)g(x_t^o,a^*).\end{array}$$

and note that the objective function $D_n^o(a,a^*)$ is

$$n^{-1}D_n^o(a,a^*)=A_n(a)-2B_n(a)-2C_n(a)-2E_n(a). \tag{20}$$

Next, using similar arguments as those in the previous proof we have

$$k_g(\sqrt{n},a^*)^{-2}\inf_{a\in N}A_n(a)\overset{a.s.}{\to}\infty \text{ and } k_g(\sqrt{n},a^*)\sup_{a\in N}A_n(a)^{-1/2}\overset{a.s.}{\to}0 \tag{21}$$

Moreover, since

$$\begin{array}{l}\frac{1}{n}\sum_{t=1}^{n}\left(u_t^o\right)^2 = O_p(1),\\ \frac{1}{n}\sum_{t=1}^{n}f(x_t^o)^2 = O_p\left(k_f(\sqrt{n})^2\right),\\ \frac{1}{n}\sum_{t=1}^{n}g(x_t^o,a^*)^2 = O_p\left(k_g(\sqrt{n},a^*)^2\right),\end{array}$$

by the Cauchy-Schwarz inequality and (21) we get

$$\begin{aligned}\sup_{a\in N} A_n(a)^{-1}\left|B_n(a)\right| &\leq \sup_{a\in N} A_n(a)^{-1/2}O_p(1) = o_p(1), \\ \sup_{a\in N} A_n(a)^{-1}\left|C_n(a)\right| &\leq k_f(\sqrt{n})\sup_{a\in N} A_n(a)^{-1/2}O_p(1) = o_p(1), \\ \sup_{a\in N} A_n(a)^{-1}\left|E_n(a)\right| &\leq k_g(\sqrt{n},a^*)\sup_{a\in N} A_n(a)^{-1/2}O_p(1) = o_p(1),\end{aligned} \tag{22}$$

Now from (21) and (22) we have

$$\begin{aligned} n^{-1}\inf_{a\in N} D_n^o(a,a^*) &\geq \inf_{a\in N} A_n(a)\times \\ &\quad \left(1-2\sup_{a\in N}\left(A_n^{-1}(a)\left|B_n(a)\right|\right)-2\sup_{a\in N}\left(A_n^{-1}(a)\left|C_n(a)\right|\right)-2\sup_{a\in N}\left(A_n^{-1}(a)\left|E_n(a)\right|\right)\right) \\ &= \inf_{a\in N} A_n(a)(1+o_p(1)) \xrightarrow{p} \infty.\end{aligned}$$

Using the same arguments as those in the previous proof, we get:

$$n^{-1}\inf_{a\in A_o} D_n(a,a^*) \xrightarrow{p} \infty$$

and the result follows.■

Proof of Theorem 6. First note that $D_n^o(a,a^*)$ is determined by (20). Next, let

$$\begin{array}{l} m(\sqrt{n},a) = \frac{1}{nk_f(\sqrt{n})k_g(\sqrt{n},a)}\sum_{t=1}^{n} g(x_t^o,a)f(x_t^o),\\ m(a) = \int_{-\infty}^{\infty} h_g(s,a)h_f(s)L(1,s)ds.\end{array}$$

It follows from Lemma A6(c) and Theorem 3.3 of P&P that $\sup_{a\in A}\left|m(\sqrt{n},a)-m(a)\right| = o_{a.s.}(1)$. Moreover, $m(a)$ continuous $a.s.$ by Lemma 8(a) of P&P and greater than zero $a.s.$ due to condition (ii).

Fix $\delta > 0$ such that the set $A_o = \{|a-a^*| \geq \delta\} \subset A$, is non-empty. Let a_+ be an arbitrary point in A_o. Set $\bar{c} = m(a^*)$, $\bar{d} = m(a_+)$ and notice that $\bar{c}\bar{d} > 0$ $a.s.$ by condition (ii). Fix $\epsilon > 0$. Then, by Theorem 3.3. of P&P and the continuity of m (Lemma A8 in P&P), there is a neighborhood of a_+, N say, such that

$$\left|m(\sqrt{n},a^*)-\bar{c}\right|, \sup_{a\in N}\left|m(\sqrt{n},a)-\bar{d}\right| < \epsilon \ a.s.,$$

Therefore, for n large enough,

$$\begin{aligned} & k_f(\lambda)\times \inf_{\substack{|c-\bar{c}|<\varepsilon \\ |d-\bar{d}|<\varepsilon}}\inf_{a\in N}\left(ck_g(\lambda,a^*)-dk_g(\lambda,a)\right) \\ \leq\ & k_f(\lambda)\times\left(k_g(\lambda,a^*)m(\sqrt{n},a^*)-k_g(\lambda,a)m(\sqrt{n},a)\right) = -C_n(a),\end{aligned}$$

(C_n is as in the previous proof) and in view of condition (i),

$$k_g(\sqrt{n}, \bar{a})^{-2} \inf_{a \in N} (-C_n(a)) \overset{a.s.}{\to} \infty \text{ and } k_g(\sqrt{n}, \bar{a})^2 \sup_{a \in N} \left(-C_n(a)^{-1}\right) \overset{a.s.}{\to} 0. \qquad (23)$$

Therefore, by condition (i) and (23),

$$\begin{aligned} &\inf\nolimits_{a \in N} \left(-C_n(a)^{-1}\right) A_n(a) \leq \\ &k_g(\sqrt{n}, \bar{a})^2 \inf\nolimits_{a \in N} (-C_n(a))^{-1} \left(k_g(\sqrt{n}, \bar{a})^{-2} \sup\nolimits_{a \in N} A_n(a)\right) = o_p(1) O_p(1) = o_p(1), \end{aligned}$$

$$\begin{aligned} &\sup\nolimits_{a \in N} \left(-C_n(a)^{-1}\right) |B_n(a)| \leq \\ &k_g(\sqrt{n}, \bar{a})^2 \sup\nolimits_{a \in N} (-C_n(a))^{-1} k_g(\sqrt{n}, \bar{a})^{-2} \sup\nolimits_{a \in N} B_n(a) = o_p(1) o_p(1) = o_p(1), \end{aligned}$$

$$\begin{aligned} &\sup\nolimits_{a \in N} \left(-C_n(a)^{-1}\right) |E_n(a)| \leq \\ &k_g(\sqrt{n}, \bar{a})^{-2} \sup\nolimits_{a \in N} (-C_n(a))^{-1} k_g(\sqrt{n}, \bar{a})^{-2} \sup\nolimits_{a \in N} E_n(a) = o_p(1) o_p(1) = o_p(1). \end{aligned}$$

Hence,

$$\begin{aligned} n^{-1} D_n^o(a, a^*) &\geq -C_n(a) \left(2 + 2C_n(a)^{-1} |B_n(a)| + 2C_n(a)^{-1} |E_n(a)| - C_n(a)^{-1} A_n(a)\right) \\ &\geq \inf_{a \in N} (-C_n(a)) \times \\ &\quad \left(2 - 2 \sup_{a \in N} \left(-C_n(a)^{-1} |B_n(a)|\right) - 2 \sup_{a \in N} \left(-C_n(a)^{-1} |E_n(a)|\right) + \inf_{a \in N} \left(-C_n(a)^{-1} A_n(a)\right)\right) \\ &= \inf_{a \in N} (-C_n(a)) (1 + o_p(1)) \overset{p}{\to} \infty. \end{aligned}$$

In view of the above and, the fact that a_+ has been chosen arbitrarily, the result follows. ■

Proof of Lemma 3. The proof is the same as the proof of Theorems 8.1, 10.1 in Wooldridge (1994) and the proof Theorem 1 of de Jong and Hu (2006).■

Proof of Theorem 7. Notice that $\int_{-\infty}^{\infty} z(s, a^*) ds = 0$, for a^* is interior in A and $\sup_{a \in A} |z(s, a)|$ integrable (by the definition of *I-regularity*). In view of this, and Theorem 1 of Jeganathan (2003), it follows that

$$n^{-1/4} \dot{Q}_n(a^*) \overset{d}{\to} L(1, 0)^{1/2} \left[b^{1/2} W_1(1) + \left(\int_{-\infty}^{\infty} \dot{g}(s, a^*) \dot{g}(s, a^*)' ds \right)^{1/2} W_2(1) \right].$$

The rest of the proof follows easily.■

Proof of Theorem 8. Set $s_n = n^{1/2} k_{n,q}^*$ and $v_n = n^{1/2} \dot{k}_{n,g}^*$. In view of (5), conditions $C1 - C2$ can be established easily along the lines of Theorem 5.3 in P&P. Consider the following modifications of $C3$ and $C4(i)$:

$C3'$: $\sup_{a \in N_n} \left\| z_n s_n^{-1} \mu_n^{-1} \left(\ddot{Q}_n(a) - \ddot{Q}_n(a^*) \right) \mu_n'^{-1} \right\| = o_p(1)$, *where* $z_n : \|z_n^{-1} \mu_n\| \to \infty$ *as* $n \to \infty$ *and* $N_n = \{a : \|z_n^{-1} \mu_n (a - a^*)\| \leq 1\}$.

$C4(i)' : z_n^{-1}\mu_n(\hat{a} - a^*) = o_p(1).$

It is easy to show that Lemma 1 still holds, when $C3$ and $C4(i)$ are replaced by $C3'$ and $C4(i)'$. We first check $C3'$. Define the sequence $z_n = n^{1/2}\dot{k}_{n,g}^* \pi_n^{-1}$. Fix δ such that $0 < \delta < \varepsilon/3$, and Set $\mu_n = n^{1/2-\delta}\dot{k}_{n,g}^*$ with δ such that $0 < \delta < \varepsilon/3$, and $\|z_n^{-1}\mu_n\| \to \infty$ as $n \to \infty$. Next, write

$$\ddot{Q}_n(a) - \ddot{Q}_n(a^*) = \left(\ddot{D}_{1n}(a) + \ddot{D}_{1n}(a)'\right) + \ddot{D}_{2n}(a) + \ddot{D}_{3n}(a) + \ddot{D}_{4n}(a) + \ddot{D}_{5n}(a^*),$$

where

$$\ddot{D}_{1n}(a) = \sum_{t=1}^{n} \dot{g}(x_t, a^*)(\dot{g}(x_t, (a) - \dot{g}(x_t, a^*))', \; \ddot{D}_{2n}(a) = \sum_{t=1}^{n} (\dot{g}(x_t, a) - \dot{g}(x_t, a^*))(\dot{g}(x_t, a) - \dot{g}(x_t, a^*))',$$

$$\ddot{D}_{3n}(a) = \sum_{t=1}^{n} \ddot{G}(x_t, a)(g(x_t, a) - g(x_t, a^*)), \; \ddot{D}_{4n}(a) = \sum_{t=1}^{n} \left(\ddot{G}(x_t, a^*) - \ddot{G}(x_t, a)\right) q(x_t, a^*).$$

$$\ddot{D}_{5n}(a^*) = -\sum_{t=1}^{n} \left(\ddot{G}(x_t, a) - \ddot{G}(x_t, a^*)\right) u_t,$$

and define

$$\bar{\omega}_{in}^2(a) = \left| z_n s_n^{-1} \mu_n^{-2} \ddot{D}_{in}(a) \right|, \; i = 1, ..., 6.$$

Denote by $\bar{\omega}_{in}^2(a)^o$ the copies of $\bar{\omega}_{in}^2(a)$ on the expanded probability space. Notice that $z_n s_n^{-1} = \dot{k}_{n,g}^* \left(k_{n,q}^* \pi_n\right)^{-1}$. Now using similar arguments as those in P&P, it can be shown that (6)-(7) are sufficient for $\bar{\omega}_{in}^2(a)^o = o_{a.s.}(1)$, $i = 1, ..., 6$, uniformly in N_n. Hence, in view of Lemma 2 we $\bar{\omega}_{in}^2(a) = o_p(1)$, $i = 1, ..., 6$, uniformly in N_n. This establishes $C3'$. Finally, we check $C4(i)'$. Our exposition is similar to that of de Jong and Hu (2006). Set

$$A_n = \sum_{t=1}^{n} \underline{h}(x_t), \text{ and } B_n = \sum_{t=1}^{n} \overline{h}(x_t) \left| (q(x_t, a^*) + u_t) \right|.$$

Fix $\epsilon > 0$ and choose some $K_\epsilon > 0$, such that

$$\mathbf{P}\left(B_n / \left(\pi_n^{-1} A_n\right) > K_\epsilon\right) < \epsilon,$$

for n large enough. Next,

$$\begin{aligned} \mathbf{P}\left(|\hat{a} - a^*| > \pi_n^{-1} K_\epsilon\right) &\leq \mathbf{P}\left(\inf_{\left\{a \in A : |a - a^*| > K\pi_n^{-1}\right\}} Q_n(a) \leq Q_n(a^*)\right) \\ &\leq \mathbf{P}\left(\inf_{\left\{a \in A : |a - a^*| > K\pi_n^{-1}\right\}} \left\{- |a - a^*| B_n + |a - a^*|^2 A_n\right\} \leq 0\right) \end{aligned}$$

The last expression is minimal for $|a - a^*| = K_\epsilon \pi_n^{-1}$, when $K_\epsilon \pi_n^{-1} \geq B_n / A_n$. In view of this the rest of the proof follows along the lines of de Jong and Hu (2006).■

FIGURE 1

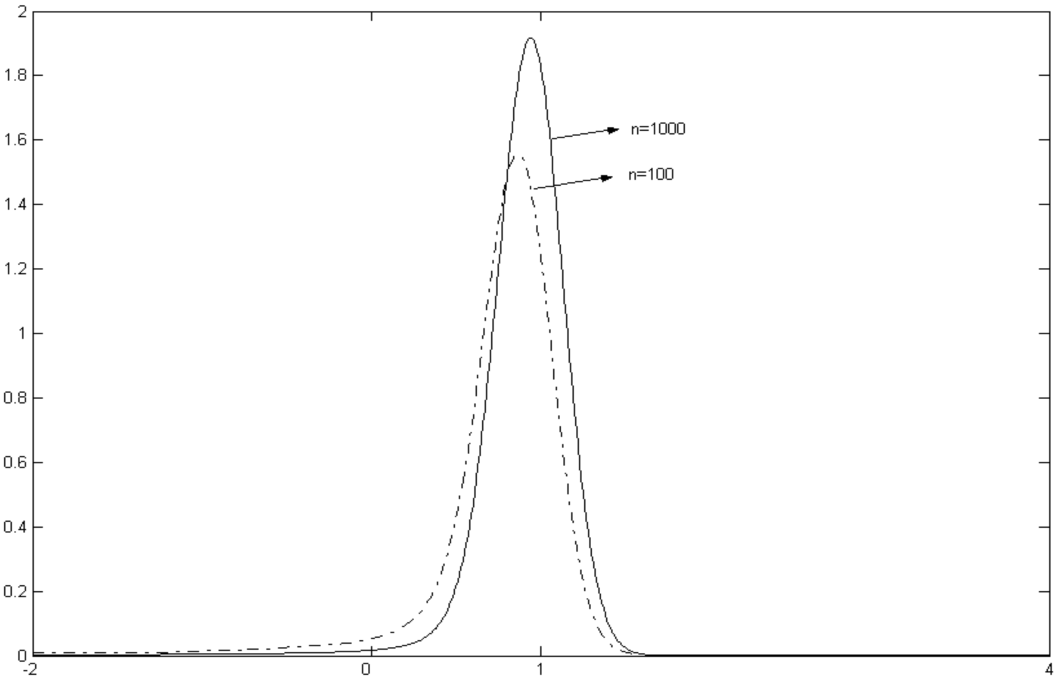

FIGURE 2

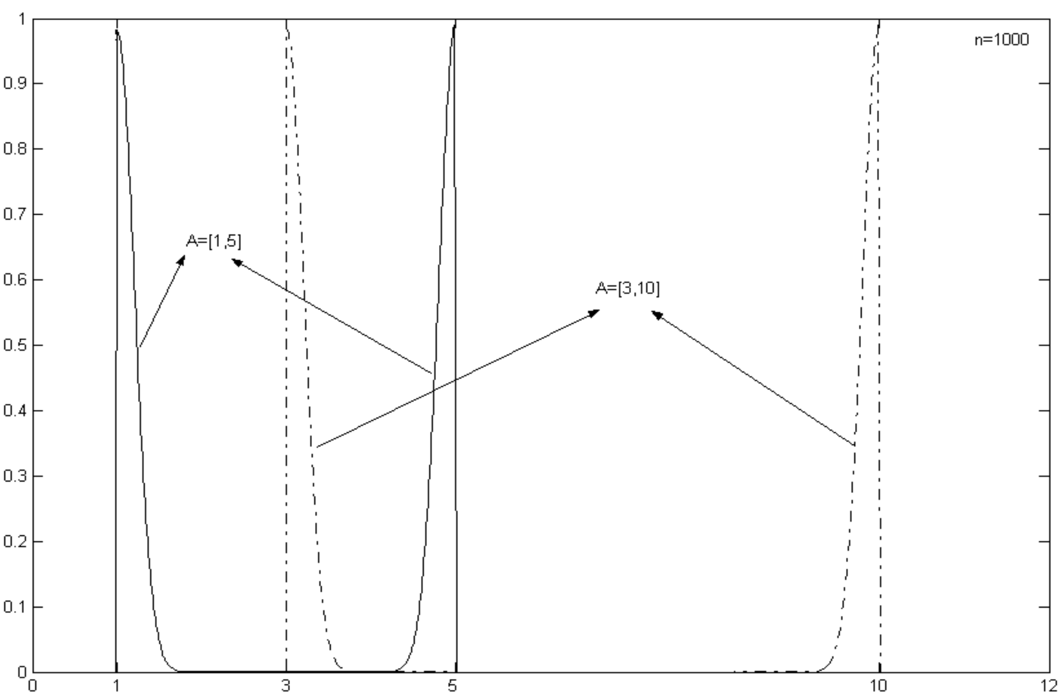

FIGURE 3