Econometric Theory, **27**, 2011, 285–311.
doi:10.1017/S0266466610000265

FUNCTIONAL FORM MISSPECIFICATION IN REGRESSIONS WITH A UNIT ROOT

IOANNIS KASPARIS
University of Cyprus

We examine the limit properties of the nonlinear 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 (2001, *Econometrica* 69, 117–161). We show that the limit behavior of the NLS estimator is largely determined by the relative orders 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 theory 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.

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 rely 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 nonlinear least squares (NLS) estimator under functional form misspecification (FFM) have been studied by White (1981) for independent and identically distributed (i.i.d.) data and by Domowitz and White (1982) for heterogeneous weakly dependent data (for similar results, see also Bierens, 1984). The purpose of this paper is to explore the limit behavior of the NLS estimator in misspecified models with strongly dependent nonstationary regressors. In particular, we consider nonlinear regressions with unit root covariates. The results provided here are not only

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. In addition, I thank three referees for suggestions that have substantially improved the previous version. Address correspondence to Ioannis Kasparis, Department of Economics, University of Cyprus, P.O. Box 20537, CY-1678 Nicosia, Cyprus; e-mail: kasparis@ucy.ac.cy.

© Cambridge University Press 2010

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

of theoretical interest but also useful for the development of specification tests. To obtain asymptotic power rates for certain misspecification tests (tests without specific alternative; e.g., Ramsey, 1969; Bierens, 1990) knowledge about the asymptotic behavior of the estimator under misspecification is necessary. In addition, to determine the limit distribution of certain model selection statistics under the null hypothesis (tests with specific alternative; e.g., Cox, 1961, 1962; Davidson and MacKinnon, 1981; Vuong, 1989), the estimator's limit distribution about the pseudo-true value is required.

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 were explored 25 years ago (e.g., Hansen, 1982; 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 transformations and locally integrable transformations that are asymptotically homogeneous. Our aim is to analyze 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] Characterizing the limit behavior 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 may involve 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. Following P&P, 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 nonstochastic 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 linearization of the objective function about the estimator's limit (e.g., Newey and McFadden, 1994; Wooldridge, 1994; Andrews, 1999) are not applicable.

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

We provide convergence rates and limit distribution results when the pseudo-true value is an interior point. Again the limit behavior 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 addition, our analysis provides guidance for the adequacy of estimated models. We have mentioned that if FFM is committed in the P&P framework, the estimators may converge to boundary points of the parameter space. Such behavior would constitute evidence for misspecification. Therefore, inspecting the behavior of slope parameters over different parameter spaces can provide useful information about the adequacy of the fitted model.

The remainder of the paper is organized 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 by $\|\cdot\|$. For a matrix A, $A > \mathbf{0}$ denotes positive definiteness. For a function $g : \mathbb{R}^p \to \mathbb{R}$ define the arrays

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

which are vectors arranged by the lexicographic ordering of their indexes. Sometimes it is more convenient to express the second derivatives of g in matrix form, i.e., $\ddot{G} = \partial^2 g / \partial a \partial a'$. As usual, $\stackrel{d}{=}$ denotes distributional equivalence, and $\Rightarrow$ denotes implication. The notation $\mu - a.e.$ stands for "almost everywhere with respect to the Lebesgue measure." Finally, $1\{\cdot \in A\}$ is the indicator function of a set A.

2. DEFINITIONS AND THEORETICAL FRAMEWORK

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, on 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 $g(\cdot, a)$ is a transformation of the data that is "different from $f(\cdot)$." This is defined precisely later in this section. The fitted model is estimated by the NLS

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

procedure, i.e.,

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

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

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

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

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

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

The following assumptions are similar to those in P&P and specify 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.* (almost surely) for every $t = 1, \ldots, n$ 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 $t = 1, \ldots, n$.

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

Assumption 2. Suppose that Assumption 1 holds. In addition, assume that

(a) $\sum_{k=0}^{\infty} k\,|\psi_k| < \infty$.

(b) The variable η_t is an i.i.d. sequence with mean zero and $\mathbf{E}|\eta_t|^p < \infty$ for some $p > 2$. In addition, η_t has characteristic function $\phi(s)$ such that $\int_{-\infty}^{\infty} |\phi(s)|^\nu \, ds < \infty$, $\nu \geq 1$.

Assumptions 1 and 2 yield strong approximation results for the empirical Brownian motions introduced earlier. For instance, under Assumption 1(b) (see P&P, pp. 125, 152), there is a finer probability space $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P})^o$ supporting (U, V) and the partial sum processes (U_n^o, V_n^o) such that

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

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

and

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

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

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

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 for 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 $z_n = (x_t, u_t,\ t = 1, \ldots, n)$. It is obvious from (4) that, for each n, there is a random vector $z_n^o = \left(x_t^o, u_t^o,\ t = 1, \ldots, n\right)$ on $(\Omega, \mathcal{F}, \mathbf{P})^o$ such that $z_n \stackrel{d}{=} z_n^o$. Then define

$$Q_n^o(a) = \sum_{t=1}^{n} (y_t^o - g(x_t^o, a))^2 \quad \text{with } y_t^o = f(x_t^o) + u_t^o.$$

Some of the techniques we employ to establish convergence of the NLS estimator require limit distribution theory for extrema[3] of Q_n of the form $\sup_{a\in F} Q_n(a)$ and $\inf_{a\in F} Q_n(a)$ with $F \subseteq A$ closed. It follows from P&P, e.g., that $\inf_{a\in F} Q_n^o(a) \stackrel{d}{\to} \inf_{a\in F} Q(a)$, where $Q(a)$ is some limit process. From (4) and standard arguments, we get that $Q_n(a)$ and $Q_n^o(a)$ have the same finite dimensional distributions on A. Moreover, it can be shown (see, e.g., Kasparis, 2008b) that the equality of the finite-dimensional distributions together with the continuity of $Q_n(.)$ and the compactness of F implies that

$$\inf_{a\in F} Q_n(a) \stackrel{d}{=} \inf_{a\in F} Q_n^o(a) \quad \text{and} \quad \sup_{a\in F} Q_n(a) \stackrel{d}{=} \sup_{a\in F} Q_n^o(a). \tag{5}$$

We can therefore assert that $\inf_{a\in F} Q_n(a) \stackrel{d}{\to} \inf_{a\in F} Q(a)$, if $\inf_{a\in F} Q_n^o(a) \stackrel{d}{\to} \inf_{a\in F} Q(a)$.

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 properties.

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

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

(i) 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 $\|g(x,a) - g(x,a_o)\| \leq \|a - a_o\| T(x)$ for all $a \in N_o$, and $\sup_{a \in N_o} |g(x,a)|$ is integrable.
(ii) For some $c > 0$ and $k > 6/(p-2)$ with $p > 4$ given in Assumption 2b), $\|g(x,a) - g(y,a)\| \leq c\,|x-y|^k$ for all $a \in A$.

Definition 1 requires $g(x,a)$ to be integrable and sufficiently smooth with respect to x and a. Jeganathan (2004) and Wang and Phillips (2009) provide limit theory for integrable transformations without imposing smoothness conditions on the integrable function $g(x)$. The results of the two aforementioned papers are not readily available for parameterized regression functions. Extensions to parameterized models are possible, but that would divert attention from the main purpose of the paper. Before we introduce the *H-regular* class, we need to define another class of transformations, introduced by P&P.

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

(i) For all $a \in A$, $T(.,a)$ is continuous in a neighborhood of infinity.
(ii) 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 $\underline{T}_\epsilon(x,a) \leq T(y,a) \leq \overline{T}_\epsilon(x,a)$ 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 $\epsilon \to 0$.
(iii) 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.[4]

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

(i) $g(\lambda x, a) = k_g(\lambda, a) h_g(x,a) + R_g(x,\lambda,a)$ with $h_g(.,a)$ regular on A,
(ii) $\left|R_g(x,\lambda,a)\right| \leq a_g(\lambda,a) P_g(x,a)$, with $\limsup_{\lambda\to\infty} \sup_{a\in A} \left\|a_g(\lambda,a)\, k_g^{-1}(\lambda,a)\right\| = 0$, or
(iii) $\left|R_g(x,\lambda)\right| \leq b_g(\lambda) P_g(x,a) Q_g(\lambda x, a)$, with $\limsup_{\lambda\to\infty} \sup_{a\in A} \left\|b_g(\lambda,a)\, k_g^{-1}(\lambda,a)\right\| < \infty$,

where

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

(iv) $\sup_{a\in A} P_g(.,a)$ *is locally bounded such that for some* $c > 0$, $\sup_{a\in A} P_g(x,a) = O(e^{c|x|})$ *as* $|x| \to \infty$,

(v) $\sup_{a\in A} Q_g(.,a)$ *is bounded with* $\sup_{a\in A} Q_g(x,a) = o(|x|)$ *as* $|x| \to \infty$.

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

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

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 functions $\dot{h}_g$ ($\ddot{h}_g$) 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 behavior of nonlocally integrable transformations is as yet unknown (see de Jong and Wang, 2002). de Jong and Wang (2002) provide asymptotic theory for "nearly nonlocally integrable" transformations. Further, Park and Phillips (1999) develop asymptotic theory for another class of functions. This class comprises locally integrable 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

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

a.e. with respect to the Lebesgue measure.

(ii) The fitted model is of incorrect functional form if

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

on a set of positive Lebesgue measure.

Remark 1. If the functions $f(.)$ and $g(.)$ are equal a.e. with respect to Lebesgue measure, then models (1) and (2) are observationally equivalent. In particular, under Assumption 2(b), the unit root process, x_t, has absolutely continuous distribution with respect to Lebesgue measure,[5] which in turn implies that $f(x_t) = g(x_t)$ *a.s.*

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

As mentioned earlier, the relative asymptotic orders 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 normalization 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.*

(i) $f(.)$ *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* $f \sim g$.
(ii) $f(.)$ *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* $f \nsim g$.
(iii) $f(.)$ *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 minimizes $Q(a)$, some limit version of the objective function $Q_n(a)$. The asymptotic behavior of the NLS estimator is determined by the components of $Q_n(a)$ that depend on a. It is convenient to consider some shifted version of the objective function of the form

$$D_n(a) = Q_n(a) - \mathbf{s}_n. \tag{6}$$

The shifting sequence $\mathbf{s}_n$ is employed to remove (higher order) "nuisance" terms that do not depend on 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 der Vaart and Wellner, 1996).[6] *Consider the objective function of (6) and let* v_n *be a normalizing sequence of real numbers. Suppose that for some random element* $a^* \in A$ *and each closed set* $F \subseteq A$

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

(i) $v_n^{-1}[\inf_{a\in F} D_n(a),\ \inf_{a\in A} D_n(a)] \xrightarrow{d} [\inf_{a\in F} D(a),\ \inf_{a\in A} D(a)]$, *as* $n \to \infty$, *where* $D(a)$ *is a stochastic process on* A;

(ii) $\mathbf{P}(a^* \notin F \Rightarrow D(a^*) < \inf_{a\in F} D(a)) = 1$.

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

Remark 2.

(a) Note that CN1(i) is a stronger requirement than $D_n(a) \xrightarrow{d} D(a)$ for each $a \in A$. Indeed, pointwise weak convergence of $D_n(a)$ follows from CN1(i) if we choose the set F to be the singleton $F = \{a\}$.

(b) Under our assumptions, the process $D_n(a)$ and its limit are continuous on A (see P&P, Lem. 8). By the continuity of $D(a)$ and the compactness of A, it can be shown that the following condition is sufficient for CN1(ii):

There is a random $a^* \in A$ *such that for all random* $a \in A$,

$$\mathbf{P}\left(a^* \neq a \Rightarrow D(a^*) < D(a)\right) = 1. \tag{7}$$

For many examples of interest, CN1(ii) can be established directly. In some cases however, it is convenient to show that CN1(ii) holds via (7).

(c) Note that for deterministic a^*, (7) simplifies to the usual identification requirement (e.g., Jennrich, 1969):

There is an $a^* \in A$ *such that for all* $a \in A$,

$$a^* \neq a \Rightarrow \mathbf{P}\left(D(a^*) < D(a)\right) = 1. \tag{8}$$

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

CN2 (Wu, 1981).[7] *Let* a^* *be a deterministic element of* A. *Suppose that* $D_n(a) = Q_n(a) - Q_n(a^*)$ *and*

$$\liminf_{n\to\infty} \inf_{\{a\in A:\|a-a^*\|\geq\delta\}} D_n(a) > 0 \quad a.s.\ (\textit{in prob.},)$$

for any $\delta > 0$, *such that* $\{a \in A : \|a - a^*\| \geq \delta\}$ *is nonempty. Then* $\hat{a} \xrightarrow{a.s.} a^*$ (*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. Therefore, its minimizer can be stochastic also. Note that this is not true for the

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

correctly specified models of P&P, as the limit objective function is minimized at the true parameter in that case. Second, 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. For scalar a, $Q(a)$ can be monotonic when it is not a complete quadratic form. Therefore, it attains its minimum at some boundary point of A.

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 in this section, assumption (iii) is an identification requirement sufficient for (7). We start with $f, g \in \mathcal{I}$.

THEOREM 1 ($f, g \in \mathcal{I}$). *Let* $D_n(a) = Q_n(a) - \sum_{t=1}^n u_t^2$. *Suppose that*

(i) Assumption 2 holds and A is compact.
(ii) $f, g \in \mathcal{I}$.
(iii) 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

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

In particular we have

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

with $v_n = \sqrt{n}$.

Remark 3. Notice that although the limit objective function $D(a)$ in Theorem 1 is stochastic, it has a deterministic minimizer. Condition (iii) of Theorem 1 ensures that (8) holds. For this case, $\xrightarrow{p}$ is the relevant mode of convergence.

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

THEOREM 2 ($f \in \mathcal{H}, g \in \mathcal{H}_o, f \sim g$). *Let* $D_n(a) = Q_n(a) - \sum_{t=1}^n u_t^2$. *Suppose that*

(i) Assumption 1 holds and A is compact.
(ii) $f, g \in \mathcal{H}_o$ *with* $k_f = k_g$.
(iii) There is a random $a^* \in A$ *such that for all random* $a \in A$

$$\mathbf{P}\left(a \neq a^* \Rightarrow \int_{-\infty}^{\infty} [h_f(s) - h_g(s,a)]^2 L(1,s)\, ds > \int_{-\infty}^{\infty} [h_f(s) - h_g(s,a^*)]^2 L(1,s)\, ds \right) = 1.$$

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

Then

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

In particular we have

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

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

Condition (iii) of Theorem 2 is an identification requirement sufficient for (7). The pseudo-true value of Theorem 2 can be stochastic. This is demonstrated by the examples that follow.

Example 1

(a) Let $f(s) = \theta_o s^2 1\{s > 0\}$ with $\theta_o \in \mathbb{R}$. 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, $D(a) = (\theta_o - (1/a))^2 \int_0^{\infty} s^2 L(1,s)\,ds$. It can be easily seen that CN1(ii) holds with $a^* = \theta_o^{-1}$.

(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 CN1(ii) is satisfied with

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

Remark 4. Suppose that the true response function is $f(x) = \theta_o e^x (1+e^x)^{-1}$ with $\theta_o \in A$. The fitted regression function is $g(x,a) = a1\{x \geq 0\}$. In this case, $f \sim g$. In particular, $f(x) = g(x,\theta_o) + q(x,\theta_o)$, with q being a lower order component (integrable). It follows from Theorem 2 that the LS estimator $\hat{a} \xrightarrow{p} \theta_o$. Nevertheless, θ_o has to be interpreted with caution in this case. When a model is misspecified, the notion of "correct parameter" may be ill defined. A parameter has a meaningful interpretation when it is associated with some regression function. In fact, a regression parameter is the unknown part of some regression function. Note that

(a) The elasticity of y_t with respect to x_t is determined by the derivative $\partial f(x_t)/\partial x_t$. The regression function $g(x,\theta_o)$, given previously, yields wrong elasticity.

(b) The optimal mean square error predictor of y_t is $f(x_t)$. The predictor $g(x_t,\theta_o)$ is suboptimal (see, e.g., Hamilton, 1994, pp. 72–73).

(c) When the model is misspecified, LS estimator may have different limit distribution and, in some situations, a slower convergence rate, as shown in the subsequent section.

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

Finally, we provide sufficient conditions for CN1 when $f \succ g$. We have the following result.

THEOREM 3 ($f \in \mathcal{H}$, $g \in \mathcal{H}_o$, $f \succ g$). *Let* $D_n(a) = Q_n(a) - \sum_{t=1}^{n} y_t^2$. *Suppose that*

(i) Assumption 1 holds and A is compact.
(ii) $f, g \in \mathcal{H}_o$ *with* $(k_f k_g^{-1})(\lambda) \to \infty$ *as* $\lambda \to \infty$.
(iii) There is a random $a^* \in A$ *such that for all random* $a \in A$

$$\mathbf{P}\left(a \neq a^* \Rightarrow \int_{-\infty}^{\infty} h_f(s) h_g(s, a) L(1, s)\, ds < \int_{-\infty}^{\infty} h_f(s) h_g(s, a^*) L(1, s)\, ds\right) = 1.$$

Then

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

In particular we have

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

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

Unlike Theorems 1 and 2, in Theorem 3 $D(a)$ is not a complete quadratic form as the lower order terms of 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 $D(a) = -(2/a) \int_{-\infty}^{\infty} |s|^3 L(1, s)\, ds$. Hence, for all $a \neq a^*$ we have $D(a) > D(a^*)$ *a.s.*, whenever a^* is the lower boundary point of A.

(b) Suppose now that f is as before but g is signed, i.e., $g(s, a) = \text{sign}\,(s) s^2 (1 + a|s|)^{-1}$ with $A \subset \mathbb{R}_+$. Then, $D(a) = -(2/a) \int_{-\infty}^{\infty} s^3 L(1, s)\, ds$. Notice that a^* is stochastic. It alternates between the upper and lower boundary points of A depending on the realization of the local time paths. In particular, CN1(ii) is satisfied with

$$a^* = \begin{cases} \underline{a}, & \int_{-\infty}^{\infty} s^3 L(1, s)\, ds > 0 \\ \overline{a}, & \int_{-\infty}^{\infty} s^3 L(1, s)\, ds < 0, \end{cases}$$

where $\overline{a}$, $\underline{a}$ are the upper and lower boundary points of A, respectively.

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

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 earlier and will not be discussed here.

The convenient characterization of the pseudo-true value as the minimizer of some limit criterion function is not possible for general *H-regular* functions, 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 behavior of the NLS estimator is comparable to that shown for H_o*-regular* models. For $f \asymp g$, we get convergence to boundary points. We start with $f \sim g$. In Theorem 4, which follows, the fitted model is correctly specified up to some lower order *H-regular* component q.

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

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

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

(iii) $f(x) - g(x,a^*) = q(x,a^*)$ *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

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

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

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

Condition (iii) specifies the properties of the true and fitted models, (iv) is a regularity condition similar to that of P&P (Thm. 4.3), and (v) 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 $a^* = \theta_o - 1$.

For $f \asymp g$, the NLS estimator converges to the value that minimizes the discrepancy between the asymptotic orders of f and g. In particular, when $f \prec g$, a^* is the value that minimizes the asymptotic order of g, i.e., $k_g(\lambda,a)$. Typically, this corresponds to the lower boundary point of A.

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

THEOREM 5 ($f, g \in \mathcal{H}$, $f \prec g$). *Suppose that*

(i) Assumption 1 holds and A is a compact subset of $\mathbb{R}$.
(ii) $f(i) \in \mathcal{H}_o$, $g(x,a) \in \mathcal{H}$.
(iii) There is a deterministic $a^* \in A$ 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

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

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

(v) 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.

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 maximizes the asymptotic order of g. Unlike 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 minimizes or maximizes k_g, depending on the sign of the dominant term of $Q_n(a)$.

THEOREM 6 ($f, g \in \mathcal{H}$, $f \succ g$). *Suppose that*

(i) Assumption 1 holds and A is a compact subset of $\mathbb{R}$.
(ii) $f(x)$, $g(x,a) \in \mathcal{H}$.
(iii) 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 $a \in A : a \neq \underline{a}, \overline{a}$.
(iv) For all $a \in A$, $h_g(s,a)h_f(s) > 0$ $\mu-a.e.$ or $h_g(s,a)h_f(s) < 0$ $\mu-a.e.$
(v) The term a^* is deterministic defined by

$$a^* = \begin{cases} \overline{a}, \text{ if for all } a \in A,\ h_g(s,a)h_f(s) > 0\ \mu-a.e. \\ \underline{a}, \text{ if for all } a \in A,\ h_g(s,a)h_f(s) < 0\ \mu-a.e. \end{cases}$$

Then CN2 holds if

(vi) 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 $\lambda \to \infty$

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

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

Condition (iv) of Theorem 6 is a regularity condition. It requires that the dominant term in $Q_n(a)$ is not vanishing asymptotically. In particular, it postulates that, for all $a \in A$, the integral term $\int_{-\infty}^{\infty} h_g(s,a)h_f(s)L(1,s)\,ds$ is either positive a.s. or negative a.s. (for more details, see Kasparis, 2008b). 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 (iv).

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 the pseudo-true value. To obtain limit distribution theory, we need to impose stronger conditions than those of the previous section. In particular, the fitted response function is assumed to be 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 linearization 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 minimized 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 $f \sim g$. In particular, for the *H-regular* case, we assume that the fitted model is correctly specified up to some lower order *H-regular* component q, i.e., $q = f - g$ with

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

For *I-regular* models, we follow the P&P approach. We utilize the convergence result of Theorem 1 and then linearize 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 the asymptotic analysis of *I-regular* models, under FFM, limit theory additional to that of P&P is needed. For this class of models, second-order limit theory is required. Jeganathan (2003) provides second-order asymptotic theory for integrable transformations that is utilized in this paper to obtain limit distribution results and convergence rates.

For *H-regular* models, we obtain limit distribution results following the Wooldridge (1994) approach. The aforementioned method is also based on a

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

linearization of the objective function. In particular, under suitable conditions we get

$$s_n^{-1} v_n'(\hat{a} - a^*) \xrightarrow{d} -\ddot{Q}^{-1}\dot{Q}, \tag{9}$$

where v_n is a sequence of normalizing matrices, s_n a normalizing sequence of real numbers, and $\dot{Q}$, $\ddot{Q}$ the limits of $\dot{Q}_n$, $\ddot{Q}_n$, respectively. By Wooldridge (1994, Thm. 10.1) and de Jong and Hu (2006, Thm. 1), conditions C1–C5a, which follow, are sufficient for (9).

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* $N_n = \left\{a : \left\| s_n^{-1}\mu_n(a - a^*) \right\| \le 1\right\}$.
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* $\left\| s_n v_n^{-1} \right\| \to 0$, *as* $n \to \infty$.

Condition C5a requires the score to be 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 is typically of higher order of magnitude than what is under correct specification, and as a 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 1. *Conditions* C1−C4 *and* C5b *are sufficient for (9).*

We utilize Lemma 1 to obtain limit distribution results for misspecified *H-regular* models.

P&P show that the 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*

(i) $f, g, \dot{g}, \ddot{g} \in \mathcal{I}$, *and the conditions of Theorem 1 hold.*
(ii) a^* *is deterministic interior in* A.
(iii) $\int_{-\infty}^{\infty} |sz(s, a^*)|\, ds < \infty$, *where* $z(s, a^*) = \dot{g}(s, a^*)\left(f(s) - g(s, a^*)\right)$.
(iv) Let $\hat{z}$ *be the Fourier transform of* z. *The characteristic function of* η_t *satisfies*

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

(a) $|\phi(s)| \leq C_1 |s|^{-\beta}$, *as* $s \to \infty$, *for some* C_1, $\beta > 0$ *and*
(b) $\left|\hat{z}(s, a^*)\phi(s)\right| \leq C_2 |s|^{-(2+\gamma)}$, *as* $s \to \infty$, *for some* $C_2 > 0$, $0 < \gamma < 1$.
(c) $\int_{-\infty}^{\infty} \dot{g}(s, a^*)\dot{g}(s, a^*)' ds - \int_{-\infty}^{\infty} \ddot{G}(s, a^*)\,(f(s) - g(s, a^*))\, ds > \mathbf{0}$.

Then as $n \to \infty$

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

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).*[8]

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

Next, we present limit theory for $f, g \in \mathcal{H}$. Sometimes it is difficult to check Condition C4(ii). Sufficient conditions for C4(i) are provided instead. We assume that the fitted model is correctly specified up to the lower order component, q. 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

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

We have the following theorem.

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

(i) *Assumption 1 holds with A a compact subset of* $\mathbb{R}$.
(ii) $f \in \mathcal{H}_o$ *and* $g, \dot{g}, \ddot{g} \in \mathcal{H}$ *on A.*
(iii) *There exists a deterministic interior point of A,* a^*, *such that* $f(s) - g(s, a^*) = q(s, a^*)$ *with* $q, \dot{g} \in \mathcal{H}$, *and* $q \prec g, \dot{g}$.
(iv) *There are functions* $\overline{h}, \underline{h} \in \mathcal{H}_o$ *with asymptotic orders* $\overline{k}$ *and* $\underline{k}$, *respectively, such that* $\pi_n \equiv \left(\underline{k}/\overline{k} k_q(a^*)\right)\left(\sqrt{n}\right) \to \infty$ *as* $n \to \infty$ *and* $\overline{h} \geq \sup_{a \in A} |\dot{g}|$, $\underline{h} \leq \inf_{a \in A} \dot{g}^2$.
(v) $\int_{|s| \leq \delta} \dot{h}_g(s, a^*)^2 ds > 0$ *for all* $\delta > 0$.

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

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

$$\dot{k}_g(\sqrt{n}, a^*)^{-2} \left(\sup_{|s| \leq \bar{s}} \left| \ddot{g}(\sqrt{n}s, a^*) q(\sqrt{n}s, a^*) \right| \right) \to 0, \tag{10}$$

$$n^{\varepsilon} \pi_n^{-2} k_q(\sqrt{n}, a^*)^{-1} \left(\sup_{|s| \leq \bar{s}} \sup_{a \in N(\varepsilon, \pi_n, a^*)} \left| \ddot{g}(\sqrt{n}s, a) \right| \right) \to 0, \tag{11}$$

$$n^{\varepsilon} \pi_n^{-2} k_q(\sqrt{n}, a^*)^{-1} \dot{k}_g(\sqrt{n}, a^*)^{-1} \times \left(\sup_{|s| \leq \bar{s}} \sup_{a \in N(\varepsilon, \pi_n, a^*)} \left| \dddot{g}(\sqrt{n}s, a) q(\sqrt{n}s, a^*) \right| \right) \to 0. \tag{12}$$

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 1. Condition (iii) specifies the kind of misspecification under consideration. Condition (iv) is sufficient for C4(i), and (v) is an identification requirement. Moreover, (vi) provides some useful technical conditions similar to those considered by P&P (Thm. 5.3). In particular, (10) is useful for establishing C1, whereas (11) and (12) are sufficient for C3. Condition (iv) is somewhat restrictive for general *H-regular* models. It can be checked however for several *H_o-regular* specifications.[9] Note that for general *H-regular* models, convergence rates depend on the pseudo-true value a^*. Some examples are provided next.

Example 6

(a) Let $f(s) = \theta_o s^2$ with $\theta_o \in \mathbb{R}$ 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 (iv)). It follows from Theorem 8 that $a^* = \theta_{1o} - \theta_{2o}$ and

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

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

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 behavior under FFM. This is exactly the problem we have addressed in this paper. We have explored the limit behavior of the NLS estimator when the true and fitted models involve a unit root covariate. For nonstationary misspecified models, the behavior of the NLS estimator depends on the nature of the true and fitted regression functions. 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) shows 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, it is necessary to characterize the limit of the NLS estimator under misspecification. Further, the implementation of some model selection procedures 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. Following P&P we have considered models with a single covariate. Some results for multicovariate models linear in parameters are provided by Kasparis (2008a). This work shows that analyzing single covariate models is in itself a complicated problem. Therefore, extensions to multicovariate models nonlinear in parameters may prove to be a challenging task.

NOTES

1. See also de Jong (2004), Pötscher (2004), and Jeganathan (2003, 2004) for some further developments.

2. Jennrich (1969) considers nonlinear regressions with fixed covariates. For a more general result, see, e.g., Domowitz and White (1982) and the references therein.

3. Under weak conditions, it can be shown that the minimizers of $Q_n(a)$ and $Q_n^o(a)$ agree in distribution also. (For more details, see Kasparis, 2008b.)

4. 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) generalizes 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 parameterized regression functions. Extensions to models nonlinear in parameter are possible, but that would divert attention from the main purpose of the paper.

5. See Lemma 3.1 in Pötscher (2004).

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

6. See Lemma 3.2.1 in van der Vaart and Wellner (1996).
7. See Lemma 1 in Wu (1981).
8. I thank Peter Phillips for pointing out the relevance of Jeganathan's work for the proof of this result.
9. 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, 1341–383.
Bates, D.M. & 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.
Bierens, H.J. (1990) A consistent test of functional form. *Econometrica* 58, 1443–1458.
Cox, D.R. (1961) Tests for separate families of hypotheses. *Journal of the Royal Statistical Society, Series B* 32, 323–344.
Cox, D.R. (1962) Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society Series B* 24, 406–424.
Davidson, R. & 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. & L. Hu (2006) Nonlinear Models with Integrated Regressors and Convergence Order Results. Working paper, Department of Economics, Ohio State University.
de Jong, R.M. & C.-H. Wang (2002) Further results on the asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 21, 413–430.
Domowitz, I. & H. White (1982) Misspecified models with dependent observations. *Journal of Econometrics* 20, 35–58.
Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press.
Hansen, L.P. (1982) 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.
Jeganathan, P. (2004) Limits of functionals of sums of linear processes that converge to fractional stable motions. *Annals of Probability* 32, 1771–1795.
Jennrich, R.I. (1969) Asymptotic properties of non-linear least squares estimation. *Annals of Mathematical Statistics* 40, 633–643.
Kasparis I. (2008a) Detection of functional form misspecification in cointegrating relations. *Econometric Theory* 24, 1373–1403.
Kasparis I. (2008b) Functional Form Misspecification in Regressions with a Unit Root. Working paper 08-02, Department of Economics, University of Cyprus.
Newey, W.K. & D.M. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics,* vol. 4, pp. 2213–2145. Elsevier.
Park, J.Y. & P.C.B. Phillips (1999) Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 15, 269–298.
Park , J.Y. & P.C.B. Phillips (2000) Nonstationary binary choice. *Econometrica* 68, 1249–1280.
Park, J.Y. & P.C.B. Phillips (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.

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes: With Applications in Statistics*. Springer Series in Statistics. Springer-Verlag.

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

Wang, Q. & P.C.B. Phillips (2009) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710–738.

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

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

Wooldridge, J.M. (1994) Estimation and inference for dependent processes. In R.F. Eagle & D.L. McFadden (eds.) *Handbook of Econometrics,* vol. 4, pp. 2639–2738, Elsevier.

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

APPENDIX

Proof of Theorem 1. We shall check that condition CN1 holds. By Theorem 3.2 and Lemma 7 in P&P, for any closed $F \subseteq A$, we get

$$\sup_{a\in F}\left|v_n^{-1}D_n^o(a) - D(a)\right| = o_p(1),$$

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

$$v_n^{-1}\left[\inf_{a\in F} D_n^o(a),\ \inf_{a\in A} D_n^o(a)\right] \xrightarrow{p} \left[\inf_{a\in F} D(a),\ \inf_{a\in A} D(a)\right],$$

and by (5), we get

$$v_n^{-1}\left[\inf_{a\in F} D_n(a),\ \inf_{a\in A} D_n(a)\right] \xrightarrow{d} \left[\inf_{a\in F} D(a),\ \inf_{a\in A} D(a)\right].$$

In view of condition (iii) of Theorem 1 and Remark 2(c), the result follows. ■

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 $\sup_{a\in A}\left|m(\sqrt{n},a) - m(a)\right| = o_{a.s.}(1)$. Moreover, $m(a)$ is continuous a.s. by Lemma 8(a) of P&P and greater than zero a.s. as a result of condition (v).

Fix $\delta > 0$ such that the set $A_o = \{\left|a - a^*\right| \geq \delta\} \subset A$ is nonempty. 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.* because $m > 0$ *a.s.*

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

Fix $\varepsilon > 0$. Then, by the continuity of $m(.)$ and P&P (Thm. 3.3), there is a neighborhood of a_+, N say, such that

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

for n large enough.

Next, let

$$A_n(a) = \frac{1}{n} \sum_{t=1}^{n} \left(g(x_t^o, a) - g(x_t^o, a^*)\right)^2,$$

$$B_n(a) = \frac{1}{n} \sum_{t=1}^{n} \left(g(x_t^o, a) - g(x_t^o, a^*)\right) u_t^o,$$

$$C_n(a) = \frac{1}{n} \sum_{t=1}^{n} \left(g(x_t^o, a) - g(x_t^o, a^*)\right) q(x_t^o, a^*).$$

Because of condition (iii), the objective function $D_n^o(a)$ can be written as

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

By the backward triangle inequality, (i.e., $\|x - y\| \geq |\|x\| - \|y\||$), we get

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

In view of this and condition (iv) we have

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

Further, because

$$\frac{1}{n} \sum_{t=1}^{n} \left(u_t^o\right)^2 = O_p(1),$$

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

it follows from the Cauchy–Schwarz inequality that

$$\sup_{a \in N} A_n(a)^{-1} |B_n(a)| \leq \sup_{a \in N} A_n(a)^{-1/2} O_p(1) = o_p(1),$$

$$\sup_{a \in N} A_n(a)^{-1} |C_n(a)| \leq k_q(\sqrt{n}, a^*) \sup_{a \in N} A_n(a)^{-1/2} O_p(1) = o_p(1). \tag{A.2}$$

Now from (A.1) and (A.2) we have

$$\begin{aligned} n^{-1} \inf_{a \in N} D_n^o(a) &\geq \inf_{a \in N} (A_n(a)) \left(1 - 2 \sup_{a \in N} \left(A_n(a)^{-1} |B_n(a)|\right)\right. \\ &\qquad \left. -2 \sup_{a \in N} \left(A_n(a)^{-1} |C_n(a)|\right)\right) \\ &= \inf_{a \in N} A_n(a)(1 + o_p(1)) \xrightarrow{p} \infty. \end{aligned} \tag{A.3}$$

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

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 centers in A_o that cover A_o. In view of this and (A.3)

$$n^{-1}\inf_{a\in A_o} D_n^o(a) \ge n^{-1}\min_{1\le i\le k}\inf_{a\in N_i} D_n^o(a) \xrightarrow{p} \infty. \tag{A.4}$$

Therefore by (5) and (A.4), for any $\zeta,\eta>0$

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

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 nonempty. Let a_+ be an arbitrary point in A_o and let N be a neighborhood of a_+ given in condition (iv). Next, define

$$A_n(a)=\frac{1}{n}\sum_{t=1}^n\left(g(x_t^o,a)-g(x_t^o,a^*)\right)^2,$$

$$B_n(a)=\frac{1}{n}\sum_{t=1}^n\left(g(x_t^o,a)-g(x_t^o,a^*)\right)u_t^o,$$

$$C_n(a)=\frac{1}{n}\sum_{t=1}^n\left(g(x_t^o,a)-g(x_t^o,a^*)\right)f(x_t^o),$$

$$E_n(a)=-\frac{1}{n}\sum_{t=1}^n\left(g(x_t^o,a)-g(x_t^o,a^*)\right)g(x_t^o,a^*)$$

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

$$n^{-1}D_n^o(a)=A_n(a)-2B_n(a)-2C_n(a)-2E_n(a). \tag{A.5}$$

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

$$k_g(\sqrt{n},a^*)^{-2}\inf_{a\in N}A_n(a)\xrightarrow{a.s.}\infty \quad\text{and}\quad k_g(\sqrt{n},a^*)\sup_{a\in N}A_n(a)^{-1/2}\xrightarrow{a.s.}0. \tag{A.6}$$

Moreover, because

$$\frac{1}{n}\sum_{t=1}^n\left(u_t^o\right)^2=O_p(1),$$

$$\frac{1}{n}\sum_{t=1}^n f(x_t^o)^2=O_p\left(k_f(\sqrt{n})^2\right),$$

$$\frac{1}{n}\sum_{t=1}^n g(x_t^o,a^*)^2=O_p\left(k_g(\sqrt{n},a^*)^2\right),$$

by the Cauchy–Schwarz inequality and (A.6) we get

$$\begin{aligned}
&\sup_{a\in N}A_n(a)^{-1}|B_n(a)|\le\sup_{a\in N}A_n(a)^{-1/2}O_p(1)=o_p(1),\\
&\sup_{a\in N}A_n(a)^{-1}|C_n(a)|\le k_f(\sqrt{n})\sup_{a\in N}A_n(a)^{-1/2}O_p(1)=o_p(1),\\
&\sup_{a\in N}A_n(a)^{-1}|E_n(a)|\le k_g(\sqrt{n},a^*)\sup_{a\in N}A_n(a)^{-1/2}O_p(1)=o_p(1).
\end{aligned} \tag{A.7}$$

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

Now from (A.6) and (A.7) we have

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

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

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

and the result follows. ∎

Proof of Theorem 6. First note that $D_n^o(a)$ is determined by (A.5). Next, let

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

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, $m(a)$ is continuous a.s. by Lemma 8(a) of P&P and greater than zero a.s. because of condition (iv).

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

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

for n large enough. Therefore, for large n

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

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

$$k_g(\sqrt{n},\bar{a})^{-2}\inf_{a\in N}(-C_n(a))\xrightarrow{a.s.}\infty\quad\text{and}\quad k_g(\sqrt{n},\bar{a})^2\sup_{a\in N}\left(-C_n(a)^{-1}\right)\xrightarrow{a.s.}0.\qquad\textbf{(A.8)}$$

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

Therefore, by condition (vi) and (A.8),

$$\begin{aligned}
&\inf_{a\in N}\left(-C_n(a)^{-1}\right)A_n(a)\\
&\quad\le k_g(\sqrt{n},\bar{a})^2\inf_{a\in N}(-C_n(a))^{-1}\left(k_g(\sqrt{n},\bar{a})^{-2}\sup_{a\in N}A_n(a)\right)\\
&\quad= o_p(1)O_p(1)=o_p(1),
\end{aligned}$$

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

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

Hence,

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

In view of the preceding results, the proof follows along the lines of the proof of Theorem 4. ■

Proof of Lemma 1. The proof is the same as the proof of Theorems 8.1 and 10.1 in Wooldridge (1994) and the proof of 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

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

The rest of the proof follows easily. ■

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

Proof of Theorem 8. Set $s_n = n^{1/2} k^*_{n,q}$ and $\upsilon_n = n^{1/2} \dot{k}^*_{n,g}$. In view of (10), conditions C1 and 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 : $\left\| z_n^{-1} \mu_n \right\| \to \infty$ *as* $n \to \infty$ *and* $N_n = \left\{ a : \left\| z_n^{-1} \mu_n (a - a^*) \right\| \leq 1 \right\}$.

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 $\left\| z_n^{-1} \mu_n \right\| \to \infty$ 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

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

Denote by $\bar{\omega}^2_{in}(a)^o$ the copies of $\bar{\omega}^2_{in}(a)$ on the expanded probability space $(\Omega, \mathcal{F}, \mathbf{P})^o$. 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 (11) and (12) are sufficient for $\bar{\omega}^2_{in}(a)^o = o_{a.s.}(1)$, $i = 1, ..., 6$, uniformly in N_n. Hence, in view of (5) we get $\bar{\omega}^2_{in}(a) = o_p(1)$, $i = 1, ..., 6$, 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

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

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press

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,

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

The last expression is minimal for $\left|a - a^*\right| = K_\epsilon \pi_n^{-1}$ 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). ■

https://doi.org/10.1017/S0266466610000265 Published online by Cambridge University Press