The Bierens Test for Certain Nonstationary Models

Ioannis Kasparis
University of Cyprus

19 December 2009*

Abstract

We adapt the Bierens (*Econometrica*, 1990) test to the I-regular models of Park and Phillips (*Econometrica*, 2001). Bierens (1990) defines the test hypothesis in terms of a conditional moment condition. Under the null hypothesis, the moment condition holds with probability one. The probability measure used is that induced by the variables in the model, that are assumed to be strictly stationary. Our framework is nonstationary and this approach is not always applicable. We show that Lebesgue measure can be used instead in a meaningful way. The resultant test is consistent against all I-regular alternatives.

1 Introduction

A series of consistent specification tests for parametric regression functions has been initiated by H. Bierens (e.g. 1982, 1984, 1987, 1988, 1990). The most appealing one is that of Bierens (1990). Contrary to the other tests mentioned above, the latter test has a tractable limit distribution. In addition, its consistency is not achieved by randomisation of some test parameter. The test was originally developed for i.i.d. data and was adapted to strictly stationary weakly dependent data by de Jong (1996). To the best of our knowledge, there is no fully consistent test for some class of nonstationary models. In this paper we propose a Bierens (1990) kind of test for the I-regular family of Park and Phillips (2001) (P&P hereafter). This family comprises models, where the regression function is some integrable transformation of a unit root process. Kasparis (2004) and Marmer (2007) develop specification tests for I-regular

*Earlier versions of this paper were presented at the University of York, June 2005, and at the conference in the honour of P. Dhrymes in Paphos, June 2007. I would like to thank Jon Levellen for sharing his data on NYSE returns. In addition, I would like to thank two referees for suggestions that have substantially improved the previous version.

models with a single covariate. The tests of the two aforementioned papers are not fully consistent. The test proposed in this paper is a fully consistent test for the I-regular class. In particular, we consider multi-covariate regressions with additively separable I-regular components and exogenous regressors.

The test hypothesis of Bierens (1990) is determined by some probability measure. Actually, the null hypothesis is defined in terms of a conditional moment condition. Under the null hypothesis (correct specification) the moment condition holds with probability one. For stationary models, there is one-to-one correspondence between the truth of the null/alternative hypothesis and the asymptotic behaviour of the sample moment of the Bierens (1990) test statistic.

The equivalence mentioned above does not always hold when the model is I-regular. In the context of I-regular models we show, that if the null hypothesis of the Bierens test is defined in terms of the Lebesgue measure instead, there is one-to-one correspondence between the truth of the null hypothesis and the asymptotic behaviour of the sample moments of the Bierens test statistic. Our test detects misspecification in large samples, if the true response function differs from the fitted regression function on a set of non-zero Lebesgue measure.

The rest of this paper is organised as follows: Section 2 specifies the test hypothesis and provides some preliminary results. In Section 3, our main results are presented. A Monte Carlo experiment is conducted in Section 4, while Section 5 provides some empirical application to the predictability of stock returns. Before proceeding to the next section, we introduce some notation. For a matrix $A = (a_{ij})$, is $|A|$ the matrix of the moduli of its elements. The maximum of the moduli is denoted by $\|.\|$. For a function $g : \mathbb{R}^p \to \mathbb{R}$, define the arrays

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

to be vectors, arranged by the lexicographic ordering of their indices. By $\int f(s)ds$, we denote the Lebesgue integral of the function f. The Lebesgue measure of some Borel set A on $\mathbb{R}^k$ is denoted by $\lambda[A]$. Finally, $1\{. \in A\}$ is the indicator function of a set A.

2 The model, the test hypothesis and preliminary results

In this section we specify the model under consideration and the test hypothesis. Some preliminary results are also provided. We assume that the series $\{y_t\}_{t=1}^n$ is generated by the following additively separable regression model:

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

where x_t is a vector unit root process, c_o is a constant and $f(x) = \sum_{j=1}^{J} f_j(x_j)$ with $f_j(.)$ being an I-regular function. The I-regular class, comprises integrable transformations that are piecewise Lipschitz (see Park and Phillips, 2001, for full definitions). For a statistical analysis of these models, the reader is referred to P&P and Chang et. al. (2001). The variables x_t, u_t satisfy the following assumption:

Assumption A:

(i) *Let* $x_t = x_{t-1} + v_t$ *with* $x_0 = O_p(1)$ *and*

$$v_t = \Psi(L)\eta_t = \sum_{s=1}^{\infty} \Psi_s \eta_{t-s},$$

with $\Psi(I) \neq \mathbf{0}$ *and* $\sum_{s=1}^{\infty} s \|\Psi_s\| < \infty$. *The sequence* η_t *is iid with mean zero and* $\mathbf{E}\|\eta_t\|^r < \infty$ *with* $r > 4$.

(ii) η_t *has distribution absolutely continuous with respect to Lebesgue measure and has characteristic function* $\varphi(\lambda) = o(\|\lambda\|^{-\delta})$ *as* $\lambda \to \infty$, *for some* $\delta > 0$.

(iii) *The random vector* x_t *is adapted to some filtration* $\mathcal{F}_{t-1}$.

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

(v) $\mathbf{E}\left(u_t^2 \mid \mathcal{F}_{t-1}\right) = \sigma^2 < \infty$ *a.s. and* $\sup_{1 \leq t \leq n} \mathbf{E}(|u_t|^{\gamma} \mid \mathcal{F}_{t-1}) < \infty$ *a.s. for some* $\gamma > 2$.

Define the partial sum processes $V_n(r)$ and $U_n(r)$ as:

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

The processes $V_n(r)$ and $U_n(r)$ take values in the set of cadlag functions on the interval $[0, 1]$.

Assumption A yields strong approximation results for the empirical Brownian motions introduced earlier. In particular (see P&P, p. 125 and 152), there is a finer probability space $(\Omega, \mathcal{F}, \mathbf{P})^o$ supporting a $(J+1)$-dimensional Brownian motion (U, V) and a partial sum processes (U_n^o, V_n^o) such that $(U_n, V_n) \overset{d}{=} (U_n^o, V_n^o)$ and

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

To avoid repetitious embedding arguments, we will assume that $(U_n, V_n) = (U_n^o, V_n^o)$, instead of $(U_n, V_n) \overset{d}{=} (U_n^o, V_n^o)$. Due to this convention, subsequent convergence in probability results should be understood as converngence in distribution, unless the limit is non-stochastic.

For the purposes of our analysis, it is convenient to partition $(U(r), V(r))$ as $(U(r), V_1(r), ..., V_J(r))$. In addition, we need to introduce the (chronological) local time process of the Brownian motion V_j up to time t defined as

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

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

We assume that the fitted model is given by:

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

The regression function is additively separable of the form $g(x,a) = \sum_{j=1}^{J} g_j(x_j, a_j)$, with $g_j(x_j, a_j)$ being an I-regular function on a compact set $A_j \subset \mathbb{R}^{k_j}$. Finally, $(\hat{c}, \hat{a})$ is the NLS estimator defined as

$$(\hat{c}, \hat{a}) = \arg \min_{(c,a) \in \mathbb{R} \times A} Q_n(c,a), \ \ A = A_1 \times ... \times A_J,$$

where

$$Q_n(c,a) = \sum_{t=1}^{n} \left(y_t - c - g(x_t, a)\right)^2.$$

We assume that the fitted regression components $g_j(x_j, a_j)$ are possibly "different" than their true counterparts i.e. $f_j(x_j)$. This is explained precisely later.

Before we present our test, we provide a concise review of the Bierens tests proposed for strictly stationary data. Typically, specification tests (e.g. Newey, 1985), under the null hypothesis (correct specification) impose a finite number of moment conditions of the form

$$\mathbf{E}\left[\left(y_t - c - g(x_t, a)\right) w_i(x_t)\right] = 0, \text{ for some } (c,a) \in \mathbb{R} \times A, \tag{3}$$

where $w_i(.)$ $i = 1,..,l$ are weighting functions. In the stationary case, under misspecification, a test statistic based on weighted residuals, $\hat{T}_i$ say, in large samples typically behaves as

$$\hat{T}_i \sim \sqrt{n}\mathbf{E}\left[\left(y_t - c^* - g(x_t, a^*)\right) w_i(x_t)\right],$$

where c^* and a^* are pseudo-true values (i.e. the limit of the NLS estimator under misspecification). The test is consistent as long as the expectation shown above is non-zero for some $i = 1, ..., l$. Clearly, the more weighting functions are used, the more likely is that the condition of (3) will be violated under misspecification. Nonetheless, as pointed out by Bierens (1990), a test that utilises only finite many moment conditions cannot be consistent against all possible alternatives as it will always be possible to find data generating mechanisms such that misspecification cannot be detected. If infinite many moment conditions were tested, the test could be consistent against all possible alternatives.

Bierens (1990) essentially imposes an infinite many moment conditions, by considering the following weighting function:

$$\exp\left(\sum_{j=1}^{J} m_j \Phi(x_{j,t})\right),\ m_j \in M \subset \mathbb{R},$$

where $\Phi(.)$ is some bounded one-to-one transformation. Notice the expression above entails infinite many weight functions, when M is a continuum of real numbers. The key result for the test consistency is the following. Bierens (1990) shows that under certain regularity conditions, $\mathbf{E}\left[(y_t - g(x_t, a^*))\exp\left(m\Phi(x_t)\right)\right]$ equals zero, only when m belongs to a set of Lebesgue measure zero. Therefore, test consistency can be achieved with a suitable choice of m. For instance, if m is chosen from some continuous distribution, the moment shown above will be non-zero $a.s.$ (e.g. Bierens, 1987). Alternatively, a consistent test can be obtained by some appropriate functional of the test statistic. Bierens (1982, 1984) and Bierens and Ploberger (1997) consider the Cramér-Von Misses functional, while the Bierens (1990) test is based on the Kolmogorov-Smirnov functional. The latter approach is followed here. It should be also mentioned, that the choice of exponential function is not of crucial importance. There several families of weighting functions that can deliver consistent tests. Stinchcombe and White (1998) show that any function that admits an infinite series approximation on compact sets, with non-zero series coefficients, could be employed in the place of the exponential function shown above (see also Escanciano (2006) for further examples of families).

In the stationary framework, by virtue of the Law of Large Numbers, the sample moment of the Bierens test statistic converges to some integral with respect to the probability measure generated by the variables of the model. The null/alternative hypothesis of the Bierens (1990) test is also defined in terms of the same probability measure as shown below:

$$\begin{aligned} H_0 &: \mathbf{P}\left[\mathbf{E}\left[(y_t - c - g(x_t, a)) \mid x_t\right] = 0\right] = 1 \text{ for some } (c, a) \in \mathbb{R} \times A. \\ H_1 &: \mathbf{P}\left[\mathbf{E}\left[(y_t - c - g(x_t, a)) \mid x_t\right] = 0\right] < 1 \text{ for all } (c, a) \in \mathbb{R} \times A. \end{aligned} \tag{4}$$

Under stationarity, there is one-to-one correspondence between the truth of the null/alternative hypothesis and the asymptotic behaviour of the sample moment of the Bierens test statistic.

The limit behaviour of I-regular transformations however, is characterised by integrals with respect to the Lebesgue measure. For instance suppose that $J = 1$ and $c_o = 0$. Then under misspecification and Assumption A, a residual based test statistic asymptotically behaves as:

$$\hat{T}_i \sim \sqrt[4]{n} \int_{-\infty}^{\infty} \left[\left(f(s) - g(s, a^*) \right) w_i(s) \right] ds L(1,0), \tag{5}$$

where $L(1,0)$ is the (chronological) local time the Brownian motion V up to time 1 at the origin. The local time at the origin is $L(1,0) > 0$ $a.s.$, therefore the test is consistent as long as the Lebesgue integral in (5) is non-zero. We show that under Assumption A, when the null hypothesis of the Bierens test is defined in terms of the Lebesgue measure, there is one-to-one correspondence between the truth of the null/alternative hypothesis and the asymptotic behaviour of the sample moments of the Bierens test statistic.

Next, we specify the test hypothesis. Under the null hypothesis, all the fitted components $g_j(x_j, a_j)$, $1 \leq j \leq J$ are correctly specified. In particular, we say that a fitted component is correctly specified, if the function $g_j(., a_j)$ differs from its true counterpart (i.e. $f_j(.)$) on a set of Lebesgue measure zero at most. This is formally stated below:

Definition 1.

H_0 : (correct specification) For all $1 \leq j \leq J$,

$$\lambda \left[\{ s : f_j(s) \neq g_j(s, a_j) \} \right] = 0,$$

for a some $a_j \in A_j$.

H_1 : (incorrect specification) For some $1 \leq j \leq J$,

$$\lambda \left[\{ s : f_j(s) \neq g_j(s, a_j) \} \right] > 0,$$

for all $a_j \in A_j$.

Clearly, our formulation of the test hypothesis is in general different than that of Bierens (1990). Nevertheless, if the covariate x_t has certain properties, the two approaches are equivalent. First, notice that under (1), the null hypothesis that appears in (4) can written as

$$H_0 : \mathbf{P} \left[c_o + f(x_t) = c + g(x_t, a) \right] = 1, \text{ for a some } (c, a) \in \mathbb{R} \times A.$$

The following lemma shows that the test formulation shown above is equivalent to that of Definition 1, under certain conditions.

Lemma 1. *Let $q(.) : \mathbb{R}^J \to \mathbb{R}$ be Borel measurable.*

(i) *If the random vector x has absolutely continuous distribution with respect to the Lebesgue measure, then:*

$$\lambda[s \in \mathbb{R}^J : q(s) \neq 0] = 0 \Rightarrow \mathbf{P}[q(x) = 0] = 1.$$

(ii) *If the random vector* x *satisfies* $\mathbf{P}[x \in D] > 0$, *for all Borel sets* $D \subset \mathbb{R}^J$ *of positive Lebesgue measure, then:*

$$\mathbf{P}[q(x) = 0] = 1 \Rightarrow \lambda[s \in \mathbb{R}^J : q(s) \neq 0] = 0.$$

Therefore, when the conditions of Lemma 1 hold and $q(., a) = f(.) - g(., a)$, it can be seen easily that the Bierens test formulation is equivalent to that of Definition 1. Notice that condition (i) of Lemma 1 requires that the covariates are continuously distributed. This is one of the maintained assumptions of the paper (Assumption A above), which is also a standard assumption in the literature e.g. Park and Phillips (1999, 2001), Jeganathan (2003), Pötscher (2004), de Jong (2004), de Jong and Wang (2005). On the other hand, condition (ii) essentially requires that the covariates are unrestricted. Clearly, it rules out bounded random variables. Condition (ii) can be easily checked for random variables that possess almost everywhere positive density functions (see for example Halmos, 1950 p. 104). For purposes of generality and convenience, we formulate the test hypothesis as in Definition 1 without any further reference.

Finally, we present some preliminary results. In order to derive the asymptotic properties of our test, we need to characterise the limit of the NLS estimator both under the null and the alternative hypothesis. Let $\sigma^{-1}W_j$, $1 \leq j \leq J$ independent standard Gaussian. In addition, W_j's are independent of $L_j(0,1)$'s and U. The following lemma demonstrates the limit theory of the NLS under the null hypothesis:

Lemma 2. *Suppose that:*
(a) *Assumption A holds.*
(b) *and* H_0 *holds for some* $a_o = (a_{o,1}, ..., a_{o,J}) \in A$
(c) *For* $1 \leq j \leq J$*:*
(i) $\dot{g}_j$ *and* $\ddot{g}_j$ *are I-regular on* A_j*.*
(ii) $\int_{-\infty}^{\infty} (f_j(s) - g_j(s, a_j))^2 ds > 0$, *for all* $a_j \neq a_{o,j}$ *in* A_j*.*
(iii) $\int_{-\infty}^{\infty} \dot{g}_j(s, a_{o,j}) \dot{g}_j(s, a_{o,j})' ds > \mathbf{0}$.
Then we have

$$\sqrt{n}\,(\hat{c} - c_o) \xrightarrow{d} U(1)$$

and

$$\sqrt[4]{n}\,(\hat{a}_j - a_{o,j}) \xrightarrow{d} \left(L_j(0,1) \int_{-\infty}^{\infty} \dot{g}(s, a_{o,j}) \dot{g}_j(s, a_{o,j})' ds \right)^{-1/2} W_j,$$

as $n \to \infty$*.*

To obtain the limit properties of the tests under the alternative hypothesis, we need to establish that the NLS estimator has a well defined limit. Sufficient conditions for this are provided by Marmer (2007) in the context of single covariate models. The following result holds for multi-covariate models:

Lemma 3. *Suppose that:*

(a) *Assumption A and* H_1 *hold.*

(b) *For* $j = \{1, ..., J\}$, *there are* $a_j^* \in A_j$ *such that*

$$\int_{-\infty}^{\infty} (f_j(s) - g_j(s, a_j))^2 \, ds > \int_{-\infty}^{\infty} \left(f_j(s) - g_j(s, a_j^*)\right)^2 ds,$$

for all $a_j \neq a_j^*$ *in* A_j.

Then, as $n \to \infty$, *we have*

$$\hat{c} \xrightarrow{p} c_o \text{ and } \hat{a} \xrightarrow{p} a^*,$$

where $a^* = (a_1^{*\prime}, ..., a_J^{*\prime})'$.

Lemmas 2 and 3 are essential for the subsequent analysis as they establish that $\hat{a}$ has a well defined limit both under H_0 and H_1.

Remark: (a) Notice that estimator for the intercept is consistent, even if some integrable component is misspecified.

(b) If the component $g_j(s, a_j)$ is misspecified, the NLS estimator $\hat{a}_j$ converges to a well defined quantity a_j^* that is in general different than the true parameter $a_{o,j}$. It is obvious from Lemma 3 however, that if a $g_j(s, a_j)$ is correctly specified, $\hat{a}_j$ converges to the true parameter, even if there are other misspecified or omitted components. This phenomenon is due to the fact that integrable transformations of different unit root processes, are asymptotically orthogonal.

3 The test

As discussed earlier, the null hypothesis of our test is determined by the Lebesgue measure rather some probability measure. Some basic results due to Bierens (1982, 1990) extend naturally to our framework, when the probability measure is replaced by the Lebesgue measure. The handling of asymptotics however is different in our case, as it relies largely on the asymptotic theory of Park and Phillips (2001) and Chang et. al. (2001). In the remaining of this section we present our test statistic, and derive its limit properties under the null and the alternative hypothesis.

The test statistic under consideration is based on the following sample moment:

$$n^{-1/4} \sum_{t=1}^{n} (y_t - \hat{c} - g(x_t, \hat{a})) \, \mathbf{W}(x_t, m), \ \ m \in M$$

with M a compact subset of $\mathbb{R}^J$, and

$$\mathbf{W}(x_t, m) = \sum_{j=1}^{J} w_j(x_{j,t}) \exp\left(m_j \Phi(x_{j,t})\right).$$

The function $w_j(.) > 0$ is I-regular, $\Phi(.)$ is bijective and bounded such that $w_j(.) \exp(m_j \Phi(.))$ is I-regular. Just like Bierens (1990), the weight function employed is based in the exponential transformation. Nonetheless, there are two noticeable differences between the Bierens weighting function and $\mathbf{W}(x_t, m)$. First, we utilise an additional weighting function, $w_j(.)$, that is chosen to be I-regular. Bierens (1990) points out, that an additional weighting might be employed to improve power against certain alternatives. Under the current framework, the use of an integrable weighting function is necessary. The aim of residual based specification tests is to detect abnormal fluctuation in the residuals that typically arises when the model is misspecified. Integrable transformations of unit root process however, exhibit very weak signal. In particular, the intensity of $g(x_t, a)$ is weaker than that of the error term[1] (u_t). As a result, the functional part of the model is "obscured" by the error term. The employment of some integrable weighting function resolves this problem, because $w_j(x_{j,t})$ and u_t are asymptotically orthogonal. The second difference between $\mathbf{W}(x_t, m)$ and the Bierens (1990) weighting function is that the former is additively separable in the regression variables. Clearly, this conforms with the structure of the models under consideration. Whether a non-additive separable weighing function could form the basis of a consistent test, is an open question, as no limit theory exists for multivariate integrable functions.

Our test statistic is a functional of:

$$\hat{B}(m) = \frac{\left[\sum_{t=1}^{n} (y_t - \hat{c} - g(x_t, \hat{a})) \mathbf{W}(x_t, m)\right]^2}{\hat{\varsigma}^2(m)}, \; m \in M, \tag{6}$$

where:

$\hat{\varsigma}^2(m) = (n^{-1} \sum_{t=1}^{n} \hat{u}_t^2) \sum_{t=1}^{n} \left[A_n(\hat{a}, m) C_n^{-1}(\hat{a}) \dot{g}(x_t, \hat{a}) - \mathbf{W}(x_t, m)\right]^2$ and $\varsigma^2(m)$ its distribution limit.

$A_n(a, m) = n^{-1/2} \sum_{t=1}^{n} \dot{g}'(x_t, a) \mathbf{W}(x_t, m)$ and $A(a, m)$ its distribution limit.

$C_n(a) = n^{-1/2} \sum_{t=1}^{n} \dot{g}(x_t, a) \dot{g}'(x_t, a)$ and $C(a)$ is its distribution limit.

$C_n^{-1}(a)$ and $C^{-1}(a)$ are the inverses of $C_n(a)$ and $C(a)$, when they exist.

The following result is analogous to Theorem 1 of Bierens (1990) and is essential for the development of a fully consistent test for the I-regular family.

Theorem 1: *Let $q : \mathbb{R} \to \mathbb{R}$ integrable with $q(s) \neq 0$ on a set of positive Lebesgue measure. Assume $\Phi : \mathbb{R} \to C$ is bijective and continuously differentiable with C being an open and bounded subset of $\mathbb{R}$. Then, the set*

$$\mathcal{M} = \left\{ m \in \mathbb{R} : \int_{\mathbb{R}} q(s) \exp(m \Phi(s))\, ds = 0 \right\}$$

has Lebesgue measure zero and is non-dense in $\mathbb{R}$.

[1]Notice, for instance, that $\sum_{t=1}^{n} g^2(x_t) = O_p(\sqrt{n})$ while $\sum_{t=1}^{n} u_t^2 = O_p(n)$.

The integral of Theorem 1 corresponds to the numerator of the $\hat{B}(m)$ statistic, in the limit. Under H_1, the numerator of $\hat{B}(m)$ can be zero only on a set of Lebesgue measure zero, in large samples. In fact, $\mathcal{M}$ is a set of isolated points of the real line.

The subsequent assumption is similar to one of the regularity conditions of Bierens (1990). Its purpose is to ensure that the denominator of $\hat{B}(m)$ is non-zero asymptotically.

Assumption B: *There are integrable and Borel measurable functions $\mu_j(s)$, $1 \leq j \leq J$ on $\mathbb{R}$, such that the matrix $\int_{\mathbb{R}} k_j(s)k_j(s)'ds$ is non-singular, with $k_j(s) = \left[\mu_j(s), \dot{g}_j(s)\right]'$.*

The following lemma shows that the $\hat{B}(m)$ statistic has a well defined limit unless m belongs in a null set.

Lemma 4: *Under Assumption B, the set $\mathcal{M}^* = \left\{m \in \mathbb{R}^J : \varsigma(m)^2 = 0\right\}$ has Lebesgue measure zero a.s.*

Theorem 2 next, demonstrates the limit properties of the $\hat{B}(m)$ test statistic under the null and the alternative hypothesis.

Theorem 2: *Suppose that Assumptions A-B hold. Then, for almost all $m \in \mathbb{R}^J$, as $n \to \infty$:*

(i) *Under H_0:*

$$\hat{B}(m) \xrightarrow{d} \chi^2_1,$$

(ii) *Under H_1:*

$$\hat{B}(m)/\sqrt{n} \xrightarrow{d} c(m),$$

with $c(m) > 0$ a.s.

In view of Theorem 2, the function $c(m)$ can be zero only on sets of Lebesgue measure zero. Therefore, consistency can be achieved by choosing m from a continuous distribution. A consistent test of functional form based on randomised m is proposed by Bierens (1987). Alternatively, a consistent test can be based on an appropriate functional of $\hat{B}(m)$. By virtue of Theorem 1, in the limit the numerator of the $\hat{B}(m)$ can be zero only on null sets. Hence, any compact subset of $\mathbb{R}^J$ of positive Lebesgue measure contains some m^* such that $c(m^*) > 0$. An obvious choice for m^* is the maximiser of $\hat{B}(m)$ over a compact interval. This is exactly the approach advocated by Bierens (1990). Following Bierens (1990), we consider the Kolmogorov-Smirnov functional of $\hat{B}(m)$:

$$\sup_{m \in M} \hat{B}(m), \tag{6}$$

where M is a compact subset of $\mathbb{R}^J$.

Next, the limit properties of the sup-statistic are explored. Assumption B ensures that a test statistic based on randomised m is well defined in the limit. Nonetheless, to ensure that the test statistic of (6) is well defined asymptotically, a stronger assumption is required:

Assumption B′: $\inf_{m\in M} \varsigma(m)^2 > 0$ *a.s.*

Theorem 2 essentially follows from the asymptotic theory of Park and Phillips (2001). To obtain the limit distribution of the sup-statistic however, further limit results are required. First, we need some additional assumption about the covariates of the model:

Assumption C: *The process, $t^{-1/2}x_{j,t}$, $1 \leq j \leq J$ has density function $d_{j,t}(x)$ that is uniformly bounded*[2] *i.e.* $\sup_{t\geq 1} \sup_x d_{j,t}(x) < \infty$.

In addition, define $z_n(m)$ as

$$z_n(m) = \frac{n^{-1/4} \sum_{t=1}^{n} \left[A(\hat{a}, m)C^{-1}(\hat{a})\dot{g}(x_t, \hat{a}) - \mathbf{W}(x_t, m)\right] u_t}{\sqrt{s^2(m)}}.$$

The process $z_n(m)$ is asymptotically equal to $\hat{B}(m)$. Moreover, by virtue of Assumption C, $z_n(m)$ is tight. This is formally stated in the subsequent lemma:

Lemma 5: *Suppose that H_0 holds.*
(i) *Under Assumptions A and B′, we have* $\sup_{m\in M} \left|\hat{B}(m) - z_n^2(m)\right| = o_p(1)$,
(ii) *Under Assumptions A and C, z_n is tight.*

Next, we report our main result. Let $C(M)$ be the space of all continuous functions on M equipped with the metric $\rho(c_1, c_2) = \sup_{m\in M} |c_1(m) - c_2(m)|$. The following theorem shows the limit distribution of our test statistic under the null hypothesis. Moreover, it establishes that the sup-statistic diverges in probability against any I-regular alternative.

Theorem 3: *Let Assumptions A, B′ and C hold. Then, as $n \to \infty$, we have:*

(i) *Under H_0, $\hat{B}(m)$ converges to $z(m)^2$, where $z(m)$ is a mixed Gaussian element of $C(M)$ with covariance function*

$$\Gamma(m_1, m_2) = \mathbf{E}\left[\frac{\sum_{j=1}^{J} \int_{-\infty}^{\infty} G_j(m_1, s)G_j(m_2, s)ds}{\sqrt{s^2(m_1)s^2(m_2)}}\right],$$

[2] By Lemma 3.1 in Pötscher (2004), the following requirement is sufficient for Assumption C: $\eta_{j,t}$ has characteristic function $\varphi_j(r)$ such that $\lim_{r\to\infty} |r|^\delta \varphi_j(r) = 0$, for some $\delta > 1$.

where $m_1, m_2 \in M$ *and for* $i = \{1,2\}$,
$G_j(m_i, s) = L_j^{1/2}(1,0)\left[A_j(a_{o,j}, m_{i,j})C_j^{-1}(a_{o,j})\dot{g}_j(s, a_{o,j}) - w_j(s)\exp\left(m_{i,j}\Phi(s)\right)\right]$ *and*
$s^2(m_i) = \sum_{j=1}^{J} L_j(1,0)\int_{-\infty}^{\infty}\left[A_j(a_{o,j}, m_{i,j})C_j^{-1}(a_{o,j})\dot{g}_j(s, a_{o,j}) - w_j(s)\exp\left(m_{i,j}\Phi(s)\right)\right]^2 ds.$

In addition,

$$\sup_{m\in M}\hat{B}(m) \xrightarrow{d} \sup_{m\in M} z(m)^2.$$

(ii) *Under* H_1,

$$\sup_{m\in M}\hat{B}(m)/\sqrt{n} \xrightarrow{d} \sup_{m\in M} c(m),$$

with $\sup_{m\in M} c(m) > 0$ *a.s.*

For stationary models (e.g. Bierens (1990), de Jong (1996)), the limit distribution of $\hat{B}(m)$ under H_0 is Gaussian. In our case it is mixed Gaussian. In addition, under stationarity the limit distribution of the sup-statistic is data dependent (see Bierens (1990) and de Jong (1996)). This is true under the present framework as long as the fitted model involves multiple covariates. The limit distribution depends on the local times $L_j(1,0)$ that relate to the covariates of the model. Notice, however, that no local time features in the limit, when the empirical model involves a single covariate. In this instance the limit distribution is Gaussian rather than mixed Gaussian and the covariance function $\Gamma(m_1, m_2)$ depends only on the regression function and the weighting employed. Under H_1 the test is consistent. In particular, the test statistic diverges with rate $\sqrt{n}$ which is slower than the rate attained for stationary data (n).

Theorem 3 suggests that the truth of H_0 implies certain properties for the asymptotic moments of the sup-statistic. The following result is analogous to Corollary 1 of Bierens (1984). It demonstrates that there is one-to-one correspondence between the truth of H_0 and the asymptotic behaviour of the sup-statistic.

Lemma 6: H_1 *holds if and only if* $\sup_{m\in M} c(m) > 0$ *a.s.*

The limit distribution of the sup-statistic is not pivotal. Bierens (1990) suggests a modification of the sup-statistic in order to obtain a tractable limit distribution under the null hypothesis. The modified test statistic has a chi-square limit distribution. This approach is applicable to our models as well.

Lemma 7: *Let Assumptions A-B hold. Choose independently of the data generating process* $\gamma > 0$, $\rho \in (0, 1/2)$ *and some* $m_o \in M$. *Let* $\hat{m} = \arg\max_{m\in M}\hat{B}(m)$ *and let*

$$\tilde{m} = m_o \text{ if } \hat{B}(\hat{m}) - \hat{B}(m_o) \leq \gamma n^{\rho} \text{ and } \tilde{m} = \hat{m} \text{ if } \hat{B}(\hat{m}) - \hat{B}(m_o) > \gamma n^{\rho}.$$

Then as $n \to \infty$ *we have:*
(i) *Under* H_0, $\hat{B}(\tilde{m}) \xrightarrow{d} \chi_1^2$,
(ii) *Under* H_1, $\hat{B}(\tilde{m})/\sqrt{n} \xrightarrow{d} \sup_{m\in M} c(m)$.

In view of Lemma 6, the modified statistic $\hat{B}(\tilde{m})$ has a pivotal limit distribution and yields a consistent test. It is reasonable to expect that the penalty term γn^{ρ} affects the properties of the test in finite samples. In fact, a small penalty term should result in size larger than the nominal one. The sensitivity of the test on the choice of the penalty term is explored in the simulation experiment of Section 4.

It follows from Bierens (1990) that any test based on conditional moment conditions can be converted into a fully consistent (under strict stationarity). This is true for I-regular family as well. Marmer (2007) proposes a RESET type of test for I-regular models. In particular, the regression residuals are regressed on integrable polynomials (basis functions) of a unit root covariate. The aim of the polynomials is to approximate neglected nonlinear components. The significance of the basis functions is checked with the aid of an F-test. Marmer's test is not a fully consistent one, but it can be converted into a fully consistent test. Under misspecification Marmer's test has power as long as the inner product of f (f is a neglected I-regular term) and at least one of the basis functions (ψ_k's) employed is non-zero i.e.

$$\int_{-\infty}^{\infty} f(s)\psi_k(s)ds \neq 0,$$

Clearly, a Bierens weighting function can be added in Marmer's test statistic to ensure that the above is true i.e.

$$\int_{-\infty}^{\infty} f(s)\psi_k(s)\exp(m\Phi(s))ds \neq 0.$$

4 Simulation Evidence

Next, we assess the finite sample properties of Bierens tests presented in the previous section. Our simulation experiment is based on 5000 replications. The data is generated by the following I-regular model:

$$\begin{aligned} y_t &= f(x_t) + u_t,\ x_t = x_{t-1} + v_t \text{ with} \\ \begin{pmatrix} v_{t-1} \\ u_t \end{pmatrix} &\sim i.i.d.\ N\left(\mathbf{0}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right). \end{aligned}$$

and f is chosen as follows:

$f_1(x) = 1\{0 < x < 1\}$
$f_2(x) = (1 - 0.5x)\,1\{0 < x < 2\}$
$f_3(x) = x^2 1\left\{0 < x < 3^{1/3}\right\}$
$f_4(x) = 2\phi(x + 0.25) - \phi(x - 0.75)$ (ϕ is the standard Gaussian density)
$f_5(x) = x$
$f_6(x) = \ln(0.1 + |x|)$
$f_7(x) = \exp(x)(1 + \exp(x))^{-1}$

The functions f_1-f_4 are the same as those used in the simulation experiment of Marmer (2007). In addition, we consider f_5-f_7 in order to investigate the power properties of our tests against non-integrable alternatives. The fitted specification is:

$$y_t = \hat{c} + \hat{u}_t \tag{7}$$

Our test statistic utilises the following weight functions: $w(s) = (1 + s^2)^{-1}$ and $\Phi(s) = \tan^{-1}(s/10)$. The set M is the interval $[-15, 15]$. Finally, m_o (Lemma 7) is chosen from a uniform distribution over M.

Table 1 shows the finite sample properties of the Bierens modified test ($\hat{B}(\tilde{m})$) and the Bierens randomised test ($\hat{B}(m_o)$). For the construction of the $\hat{B}(\tilde{m})$ statistic we have used three different penalty terms. For $n = 500$, the randomised test has size very close to the nominal one. The modified test has good size in most cases, but overrejects the null hypothesis, when the penalty term is small. Clearly, the size properties of the tests improve, when sample size increases. Both Bierens tests have good power properties. In many cases however, the modified test has significantly superior power than that of the randomised test. Table 2 shows the finite sample properties of Marmer's RESET test. This test has good finite sample properties as well. It is not our objective to make close power comparisons between the Bierens and the RESET tests[3]. It seems however, that the Bierens modified test has better power than the RESET test, for the particular choice of weighting and basis functions[4]. The RESET test performs better, when $f = f_4$. Notice that in this case, the basis function used resembles the neglected component and as a result it provides a good approximation. It should be stressed again that the RESET test can be converted into a fully consistent test. We expect that the employment of a Bierens weighting function, in the RESET test statistic, could improve power.

It follows from our simulation experiment, that all the tests under consideration have reasonable power in the presence of a neglected locally integrable component. Hong and Phillips (2005) and Kasparis (2008) develop specification tests that have power against locally integrable alternatives but have no power, when there is some neglected integrable term. Contrary to the tests proposed in the two aforementioned papers, our Bierens tests (as well as the Marmer's test) provide valid testing procedures for both families of transformations.

5 Application

The question whether certain financial ratios (i.e. dividend yield (DY), book-to-market (BM) and equity-to-price ratio (EP)) can predict stock returns has received much attention over the years. A substantial body of applied work in this area focuses

[3] Clearly, for a close comparison, we should consider a variaty of basis and weighting functions.
[4] The basis functions utilised for the RESET test are $\psi_k(x) = x^k \phi(x)$, $k = \{1, 2, 3\}$.

on the following linear model (see for example Levellen (2004) and the references therein):

$$r_t = c_o + a_o x_{t-1} + u_t,$$

where r_t is stock returns and x_t is some financial ratio. Stambaugh (1999), Levellen (2004), Goyal and Welch (2003) among others, explore the predictability of NYSE returns by testing the significance of the slope parameter a_o.

A simple inspection of the NYSE data, reveals that the returns and the financial ratios series have very different properties. Namely, the NYSE returns series exhibits mean reversion, constant variance and little autocorrelation. On the other hand financial ratios, exhibit no mean reversion, strong persistence and time varying variance. Actually, the financial ratio series are reminiscent of integrated processes. The fact that two sets of series exhibit very different characteristics give support to a possible non-linear relationship between returns and financial ratios. A non-linear transformation applied to some trending series may attenuate its intensity and bound its variance.

Marmer (2007) studies the relationship between NYSE stock returns and the dividend yield. Marmer proposes the following non-linear model:

$$r_t = c_o + f(x_{t-1}) + u_t,$$

with f being some integrable transformation and x_t a unit root process. An integrable transformation applied to an integrated process attenuates the intensity of the process. Integrable transformations of unit root processes tend to be non-zero, when the process is in the vicinity of some spatial point and zero, when the unit root is away from that point. Therefore, $f(x_t)$ takes non-zero values rarely, due to the null-recurrent behaviour of x_t. An integrable transformation added to some stationary sequence produces a seemingly stationary process. In view of this, Marmer (2007) points out that the above model provides a small deviation from the martingale difference hypothesis. It should be mentioned that locally integrable transformations (see Park and Phillips (2001)) may also reduce the intensity of some integrated process, but to a lesser extend. Therefore, integrable transformations produce smaller deviations from the martingale difference hypothesis. Marmer tests for predictability of returns by regressing returns on an intercept term and then applies his RESET test to check for some neglected integrable function of the DY variable. The RESET test provides evidence for a neglected non-linear component relating to the DY.

In this section we employ the Bierens tests to investigate whether returns are predictable. Apart from DY, we also consider BM and EP as possible predictors. Our data are the same as those used by Levellen (2004). A series of stationarity tests suggest that DY, BM and EP are integrated series (see Table 3). In addition, the stationarity tests suggest that the returns series a stationary one. This is not inconsistent with the I-regular formulation of returns. The Dickey-Fuller (DF) test favours the stationary alternative, when performed on an I-regural series (see Park and Phillips, 1998). Partial sum tests like the KPSS and the CUSUM, also favour stationarity (see Kasparis, 2008).

We apply the Bierens test on the residuals of a model that involves an intercept term only. Table 4 provides results for the modified Bierens statistic ($\hat{B}(\tilde{m})$). We also provide results for Marmer's RESET test[5]. The modified statistic has been estimated for $M = [-15, 15]$. Instead of using a randomised m_o, we have chosen $m_o = 1$, 15, in order to enable the reader to verify our results. We consider two weighting functions: $\mathbf{W}_i(s) = (1 + s^2)^{-1} \exp(m\Phi_i(s))$, $i = 1, 2$, with $\Phi_1(s) = \tan^{-1}(s/10)$ and $\Phi_2(s) = \tan^{-1}(s/2)$. In addition, we consider two penalty terms, for the modified Bierens statistic.

Both the Bierens and the RESET tests indicate that the three financial ratios have predictive power. The Bierens tests reject then null hypothesis at 5%, when $\mathbf{W}_2$ is employed for all three variables. In addition the null is rejected at 1% level in most cases for the particular weighting function. The null hypothesis is also rejected at 5% level in most occasions, when $\mathbf{W}_1$ is employed. It is obvious from our simulation experiment that, when a small penalty term is used, the $\hat{B}(\tilde{m})$ statistic tends to be significantly larger $\hat{B}(m_o)$. In this instance the modified Bierens test may suffer from overrejection of the null hypothesis. Notice however, that in Table (4) $\hat{B}(\tilde{m})$ equals $\hat{B}(m_o)$ in most cases, even when a small penalty term is employed.

6 Appendix

Lemma A. *Let $q(s) : \mathbb{R} \to \mathbb{R}$ be Borel measurable. The function $\Phi(s) : \mathbb{R} \to C$* is *bijective and continuously differentiable with C an open and bounded subset of $\mathbb{R}$. If $q(.)$ is integrable, then*

$$\lambda\left[\{s : q(s) \neq 0\}\right] > 0 \textit{ if and only if } \int_{\mathbb{R}} q(s)e^{m\Phi(s)}ds \neq 0$$

for $m \in \mathbb{R}$ in an arbitrarily small neighborhood of zero.

Proof of Lemma A: The "if" part is trivial. We show the "only if" part. Consider the Borel measurable functions

$$q_1(s) = \max\{q(s), 0\} \text{ and } q_2(s) = \max\{-q(s), 0\}$$

and notice that $q = q_1 - q_2$. Assume that

$$c_1 = \int_{\mathbb{R}} q_1(s)ds > 0 \text{ and } c_2 = \int_{\mathbb{R}} q_2(s)ds > 0.$$

[5] We have used a single basis function (ϕ) for the RESET statistic.

Define the probability measures[6] F_i, $i = \{1, 2\}$ on the Borel field restricted on C ($\mathcal{B}_C$) as

$$F_i(B) = \frac{1}{c_i} \int_B q_i(\Phi^{-1}(s)) \left|\dot{\Phi}^{-1}(s)\right| ds, \; B \in \mathcal{B}_C.$$

Then

$$\begin{aligned} \int_{\mathbb{R}} q(s)e^{m\Phi(s)}ds &= \int_{\mathbb{R}} q_1(s)e^{m\Phi(s)}ds - \int_{\mathbb{R}} q_2(s)e^{m\Phi(s)}ds \\ &= \int_C q_1(\Phi^{-1}(s))e^{ms} \left|\dot{\Phi}^{-1}(s)\right| ds - \int_C q_2(\Phi^{-1}(s))e^{ms} \left|\dot{\Phi}^{-1}(s)\right| ds \\ &= c_1 \int_C e^{mr} dF_1(r) - c_2 \int_C e^{mr} dF_2(r) \\ &= c_1\eta_1(m) - c_2\eta_2(m), \end{aligned}$$

where the second equality above is due to Billingsley (1979, Theorem 17.2). Notice that $\eta_i(m)$ is the moment generating function of F_i, $i = \{1, 2\}$. In view of this, and using the same arguments as Bierens (1982, Theorem 1(I)) it can be shown that $\lambda[\{s : q(s) \neq 0\}] > 0$ implies $\int_{\mathbb{R}} q(s)e^{m\Phi(s)}ds \neq 0$ for some $m \in \mathbb{R}$. In addition, because $\Phi(s)$ is bounded, the requisite result follows along the lines of Bierens (1982, Theorem 1(II)).■

Proof of Lemma 1: The proof for part (i) is trivial (see for example Halmos (1950) p. 104 and 128). For (ii) set $D = \left\{s \in \mathbb{R}^J : q(s) \neq 0\right\}$ and suppose that $\lambda[D] > 0$. Then by assumption, $\mathbf{P}[x \in D] > 0$. Clearly this implies that $\mathbf{P}[q(x) = 0] < 1$, which is a contradiction. Therefore, $\lambda[D] = 0$.■

Proof of Lemma 2: We first prove consistency. As in Marmer (2007), we consider the concentrated objective function:

$$Q_n(a) = \sum_{t=1}^{n} (y_t - \hat{c}(a) - g(x_t, a))^2 ,$$

where $\hat{c}(a) = n^{-1}\sum_{t=1}^{n}(y_t - g(x_t, a)) = n^{-1}\sum_{t=1}^{n}(f(x_t) - g(x_t, a)) + n^{-1}\sum_{t=1}^{n} u_t + c_o$. Next,

$$\sup_{a\in A} |\hat{c}(a) - c_o| \leq \sup_{a\in A} \left| n^{-1} \sum_{t=1}^{n} (f(x_t) - g(x_t, a)) \right| + \left| n^{-1} \sum_{t=1}^{n} u_t \right| = o_p(1),$$

where the last equality above follows from Park and Phillips (2001, Theorem 3.2) and Chang et. al. (2001, Lemma 3.1). This establishes consistency of $\hat{c}$. For $\hat{a}$ notice that if $x_{j,t}$ has absolutely continuous distribution with respect to Lebesgue measure,

[6] Notice that $\int_C q_i(\Phi^{-1}(s)) \left|\dot{\Phi}^{-1}(s)\right| ds = \int_{\mathbb{R}} q_i(s)ds = c_i$.

then $f_j(x_{j,t}) = g_j(x_{jt}, a_{o,j})$ $a.s.$ Given this, Park and Phillips (2001, Theorem 3.2) and Chang et. al. (2001, Lemma 3.1), it is easy to show that:

$$n^{-1/2}\left(Q_n(a) - Q_n(a_o)\right) \xrightarrow{p} \sum_{j=1}^{J} L_j(0,1) \int_{-\infty}^{\infty} \left(g_j(s, a_{o,j}) - g_j(s, a_j)\right)^2 ds$$

uniformly in a. In view of the above, condition c(ii) of Lemma 2 and condition CN1 of Park and Phillips (2001, p. 133) we get $\hat{a} \xrightarrow{p} a_o$.

Given the consistency of the LS estimator, the limit distribution result follows easily along the lines of Park and Phillips (2001, Theorem 5.1).■

Proof of Lemma 3: The consistency of $\hat{c}$ can be established as above. Also, using the same arguments as those in the previous proof, we get

$$n^{-1/2}\left(Q_n(a) - Q_n(a^*)\right) \xrightarrow{p} \sum_{j=1}^{J} L_j(0,1) \int_{-\infty}^{\infty} \left[\left(f_j(s) - g_j(s, a_j)\right)^2 - \left(f_j(s) - g_j(s, a_j^*)\right)^2\right] ds,$$

uniformly in a and the result follows. ■

Proof of Theorem 1: The proof is similar to that of Bierens (1990). Suppose m_o is such that $\int_{\mathbb{R}} q(s) e^{m_o \Phi(s)} ds = 0$. Notice that

$$\lambda\left[\left\{s : q(s) e^{m_o \Phi(s)} \neq 0\right\}\right] = \lambda\left[\left\{s : q(s) \neq 0\right\}\right] > 0.$$

By Lemma A, there is $\delta > 0$ such that $\int_{\mathbb{R}} \left[q(s) e^{m_o \Phi(s)}\right] e^{m \Phi(s)} ds \neq 0$, for $0 < |m| < \delta$. Therefore, $\int_{\mathbb{R}} q(s) e^{m \Phi(s)} ds \neq 0$, for $0 < |m - m_o| < \delta$. Hence, $\inf_{m \in \mathcal{M}, m \neq m_o} |m - m_o| > 0$. This implies that $\mathcal{M}$ is a set of isolated points of the real line, and therefore is countable. It is also straightforward to show that is non-dense in $\mathbb{R}$.■

Proof of Lemma 4: By Lemma 3.1 of Chang et. al. (2001) the result can be proved along the lines of Lemma 2 in Bierens (1990).■

Proof of Theorem 2: We start with the limit result under H_0. By Lemma 3.1 of Chang et. al. (2001), the terms $A_n(a, m)$ and $C_n(a)$ have well defined limits, $A(a, m)$, $C(a)$ say. Set $D(a, m) = A'(a, m) C^{-1}(a, m)$ and notice that $D(a, m)$ can be partitioned as $D(a, m) = [D_1(a_1, m_1), ..., D_J(a_J, m_J)]$. Let $\bar{a}_n$, $\tilde{a}_n$ be mean values of $\hat{a}$ and a_o. By Lemma 2, Lemma 3.1 of Chang et. al. (2001) and the mean value theorem, the numerator of $\hat{B}(m)$ rescaled by $n^{-1/2}$ is

$$\left[n^{-1/4} \sum_{t=1}^{n} \left(y_t - \hat{c} - g(x_t, \hat{a})\right) \mathbf{W}(x_t, m)\right]^2$$

$$= \left[n^{-1/4} \left\{\sum_{t=1}^{n} \left[A_n'(\bar{a}_n, m) C_n^{-1}(\tilde{a}_n, m) \dot{g}(x_t, a_o) - \mathbf{W}(x_t, m)\right] u_t - (\hat{c} - c_o) \sum_{t=1}^{n} \mathbf{W}(x_t, m)\right\}\right]^2$$

$$= \left[\sum_{j=1}^{J} n^{-1/4} \sum_{t=1}^{n} \left[D_j(m_j, a_{o,j}) \dot{g}_p(x_{j,t}, a_{o,j}) - w_j(x_{j,t}) \exp\left(m_j \Phi(x_{j,t}) \right) \right] u_t + O_p(n^{-1/2}) \right]^2 \xrightarrow{d}$$

$$\left[\sum_{j=1}^{J} \left\{ L_j(1,0) \int_{-\infty}^{\infty} \left[D_j(m_j, a_{o,j}) \dot{g}_j(s, a_{o,j}) - w_j(s) \exp\left(m_j \Phi(s) \right) \right]^2 ds \right\}^{1/2} W \right]^2, \tag{A1}$$

where $W \sim N(0, \sigma^2)$ is independent of $L_j(1,0)$'s. In addition, by Lemma 3.1 of Chang et. al. (2001), the denominator of $\hat{B}(m)$ rescaled by $n^{-1/2}$ is

$$\begin{aligned} & n^{-1/2} \hat{\varsigma}^2(m) \xrightarrow{p} \varsigma^2(m, a_o) \\ \equiv \; & \sigma^2 \sum_{j=1}^{J} \left\{ L_j(1,0) \int_{-\infty}^{\infty} \left[D_j(m_j, a_{o,j}) \dot{g}_j(s, a_{o,j}) - w_j(s) \exp\left(m_j \Phi(s) \right) \right]^2 ds \right\} \end{aligned} \tag{A2}$$

The result follows from (A1) and (A2).

By Lemma 3.1 of Chang et. al. (2001) and Lemma 3, under H_1

$$\begin{aligned} & n^{-1/2} \hat{B}(m) \xrightarrow{p} c(m) \\ \equiv \; & \frac{\left\{ \sum_{j=1}^{J} L_j(1,0) \int_{-\infty}^{\infty} \left[\left(f_j(s) - g_j(s, a_j^*) \right) w_j(s) \exp\left(m_j \Phi(s) \right) \right] ds \right\}^2}{\varsigma^2(m, a^*)} \\ \equiv \; & \frac{\left\{ \sum_{j=1}^{J} L_j(1,0) q_j(m_j) \right\}^2}{\varsigma^2(m, a^*)}. \end{aligned}$$

Set $L = (L_1(1,0), ..., L_J(1,0))'$ and consider any non-zero deterministic vector $\mathbf{x}$ ($J \times 1$). The inner product $L'\mathbf{x}$ is a continuously distributed random variable. Hence $L'\mathbf{x} \neq 0$ $a.s.$ In addition, by Theorem 1, each $q_j(m_j)$ is non-zero for almost every $m_j \in \mathbb{R}$. In view of this and Lemma 4 the result follows. ■

Proof of Lemma 5:

(i) Let $K_n(a, m) = n^{-1/4} \sum_{t=1}^{n} \left(y_t - g(x_t, a) \right) \mathbf{W}(x_t, m)$. Then, the mean value theorem gives

$$K_n(\hat{a}, m) - K_n(a_o, m) = n^{-1/2} \dot{K}_n(\bar{a}(m), m) n^{1/4} \left(\hat{a} - a_o \right), \tag{A3}$$

where $\sup_{m \in M} \|\bar{a}(m) - a_o\| \leq \|\hat{a} - a_o\| = o_p(1)$. Also let

$$A(a_o, m) = \sum_{j=1}^{J} L_j(1,0) \int_{-\infty}^{\infty} \dot{g}_j(s, a_{o,j}) w_j(s) \exp\left(m_j \Phi(s) \right) ds.$$

Next

$$\sup_{m \in M} \left\| n^{-1/4} \dot{K}_n(\bar{a}(m), m) - A(a_o, m) \right\| = o_p(1), \tag{A4}$$

by Theorem 3.2 of Park and Phillips (2001). By (A3) and (A4) we therefore have

$$\sup_{m\in M}\left\|K_n(\hat{a},m)-K_n(a_o,m)+A(a_o,m)n^{1/4}\left(\hat{a}-a_o\right)\right\|=o_p(1). \tag{A5}$$

Also note that

$$\sup_{m\in M}\left\|n^{1/4}\left(\hat{a}-a_o\right)-C(a_o)^{-1}n^{-1/4}\sum_{t=1}^{n}\dot{g}(x_t,a_o)u_t\right\|=o_p(1). \tag{A6}$$

Now (A5) and (A6) give

$$\sup_{m\in M}\left|K_n(\hat{a},m)-z_n(m)\sqrt{\varsigma^2(m)}\right|=o_p(1). \tag{A7}$$

Next note that

$$\sup_{m\in M}\left|\hat{\varsigma}^2(m)-\varsigma^2(m)\right|=o_p(1), \tag{A8}$$

by Theorem 3.2 of Park and Phillips (2001). Now the result follows by (A7) and (A8) and the assumption that $\inf_{m\in M}\varsigma^2(m)>0$.

(ii) By Theorem 8.2 of Billingsley (1968), the following conditions are sufficient for the requisite result:

C1: For any $\delta>0$ and some $m_o\in M$ there is an $\epsilon>0$ such that

$$\sup_n \mathbf{P}\left(z_n(m_o)>\epsilon\right)\leq\delta.$$

C2: For any $\delta>0$ and $\epsilon>0$ there is a $\xi>0$ such that

$$\sup_n \mathbf{P}\left(\sup_{\|m_1-m_2\|<\xi}|z_n(m_1)-z_n(m_2)|>\epsilon\right)\leq\delta.$$

Verifying C1 and is C2 is what we set out to do. Note that C1 holds trivially as $z_n(m_o)=O_p(1)$ by Theorem 3.2 of Park and Phillips (2001). Next choose $\xi<1$ and let $H_j\left(m_{1,j},m_{2,j},x_{j,t}\right)=w_j(x_{j,t})\left[\exp\left(m_1\Phi(x_{j,t})\right)-\exp\left(m_2\Phi(x_{j,t})\right)\right]$. Condition C2 follows from the continuity of $A(a_o,m)$ and the fact that

$$\mathbf{E}\left\{\sup_{\|m_1-m_2\|<\xi}n^{-1/4}\left|\sum_{t=1}^{n}\sum_{j=1}^{J}H_j\left(m_{1,j},m_{2,j},x_{j,t}\right)u_t\right|\right\}\leq\xi\bar{C}, \tag{A9}$$

for some $\bar{C} < \infty$. The last result can be established as follows:

$$\mathbf{E}\left\{\sup_{\|m_1-m_2\|<\xi} n^{-1/4}\left|\sum_{t=1}^{n}\sum_{j=1}^{J} H_j\left(m_{1,j}, m_{2,j}, x_{j,t}\right) u_t\right|\right\} \tag{A10}$$

$$\leq \xi \exp\left\{\left(1+\sup_{m\in M}\|m\|\right)\left(\sup_s|\Phi(s)|\right)\right\}\sum_{j=1}^{J}\mathbf{E}\left\{n^{-1/4}\left|\sum_{t=1}^{n} w_j(x_{j,t})u_t\right|\right\}$$

$$\leq \xi \exp\left\{\left(1+\sup_{m\in M}\|m\|\right)\left(\sup_s|\Phi(s)|\right)\right\}\sum_{j=1}^{J}\left\{\mathbf{E}\left(n^{-1/4}\sum_{t=1}^{n} w_j(x_{j,t})u_t\right)^2\right\}^{1/2}$$

$$\leq \xi \exp\left\{\left(1+\sup_{m\in M}\|m\|\right)\left(\sup_s|\Phi(s)|\right)\right\}\sum_{j=1}^{J}\left\{\mathbf{E}\left(n^{-1/2}\sum_{t=1}^{n} w_j(x_{j,t})^2\mathbf{E}\left(u_t^2 \mid \mathcal{F}_{t-1}\right)\right)\right\}^{1/2}$$

$$\leq \sigma\xi \exp\left\{\left(1+\sup_{m\in M}\|m\|\right)\left(\sup_s|\Phi(s)|\right)\right\}\sum_{j=1}^{J}\left\{\mathbf{E}\left(n^{-1/2}\sum_{t=1}^{n} w_j(x_{j,t})^2\right)\right\}^{1/2},$$

where the first inequality is due to equation (B9) in Bierens (1990), and second one is due to Liapunov's inequality. The last term above is:

$$\begin{aligned}\sum_{j=1}^{J}\mathbf{E}\left(n^{-1/2}\sum_{t=1}^{n} w_j(x_{j,t})^2\right) &= \sum_{j=1}^{J} n^{-1/2}\sum_{t=1}^{n}\int_{-\infty}^{\infty} w_j^2(t^{1/2}x)d_{j,t}(x)dx \\ &= \sum_{j=1}^{J} n^{-1/2}\sum_{t=1}^{n} t^{-1/2}\int_{-\infty}^{\infty} w^2(s)d_{j,t}(s/t^{1/2})ds \\ &\leq 2\sum_{j=1}^{J}\sup_{t\geq 1}\sup_x\|d_{j,t}(x)\|\int_{-\infty}^{\infty} w_j^2(s)ds + o(1) < \infty,\end{aligned}$$

where the last inequality is due to Assumption C. Therefore, $\sum_{j=1}^{J}\mathbf{E}\left(n^{-1/2}\sum_{t=1}^{n} w_j(x_{j,t})^2\right)$ is bounded. This together with (A10) establishes (A9). ■

Proof of Theorem 3: The result under H_0 follows easily from Lemma 4 and Lemma 3.1 of Chang et. al. (2001). The result under H_1 follows directly from Theorem 3.2 Park and Phillips (2001) and Lemma 3.1 of Chang et. al. (2001).■

Proof of Theorem 4: (i) Note that $[z_n(m_1), ..., z_n(m_d)] \xrightarrow{d} [z(m_1), ..., z(m_d)]$ by Theorem 3.2 of Park and Phillips (2001), for any finite d. In view of Lemma 4(ii) this ensures that z_n converges in distribution to z. Moreover, because $\sup_{m\in M}(.)^2$ is a continuous mapping from $C(M)$ on $\mathbb{R}$, the result follows.

(ii) The result under H_1 follows directly from Theorem 3.2 of Park and Phillips (2001). ■

Proof of Lemma 6: The result follows easily from Lemma A.■

Proof of Lemma 7: Same as the proof of Theorem 4 of Bierens (1990).■

7 References

Bierens, H.J., 1982. Consistent model specification testing. Journal of Econometrics 20, 3105-134.

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

Bierens, H.J., 1987. ARMAX model specification testing, with an application to unemployment in Netherlands. Journal of Econometrics 35, 161-190.

Bierens, H.J., 1988. ARMA memory index modeling of economic time series (with discussion). Econometric Theory, 4, 35-59.

Bierens, H.J., 1990. A Consistent Conditional Moment Test of Functional Form. Econometrica 67, 1341-1383.

Bierens, H.J. and W. Ploberger, 1997. Asymptotic theory of conditional moment integrated test. Econometrica 65, 1129-1151.

Billinglsey P., 1968. Convergence of Probability Measures, Wiley.

Billinglsey P., 1979. Probability and Measure, Wiley.

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.

de Jong R. M., 1996. The Bierens Test Under Data Dependence. Journal of Econometrics 72, 1-32.

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 Wang, C-H., 2005. Further results on the asymptotics for nonlinear transformations of integrated time series. Econometric Theory 21, 413-430.

Escanciano, J.C., 2006. Goodness-of-fit statistics for linear and nonlinear time series models. Journal of the American Statistical Association 101, 531-541.

Goyal, A, and Welch, I., 2003. A note on predicting returns with financial ratios. mimeo.

Halmos P. R., 1950. Measure Theory, Springer-Verlag.

Hong, S. H. and P.C.B. Phillips, 2005. Testing linearity in cointegrating relations with an application to PPP. Cowles Foundation Discussion Paper 1541.

Kasparis I., 2004. Testing for functional form under non-stationarity. Unpublished Manuscript, University of Cyprus.

Kasparis I., 2008. Detection of functional form misspecification in cointegrating relations. Econometric Theory, 24 1373-1403.

Marmer V., 2007. Nonlinearity, nonstationarity and spurious forecasts. Journal of Econometrics, 142. 1-27.

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

Park, J.Y. and P.C.B. Phillips, 1998. Unit roots in nonlinear transformations of integrated time series. Unpublished Manuscript, Yale University.

Park, J.Y. and P.C.B. Phillips, 1999. Asymptotics for nonlinear transformations of integrated time series. Econometric Theory 15, 269-298.

Park, J.Y. and P.C.B. Phillips, 2000. Nonstationary binary choice. Econometrica 68, 1249-1280.

Park, J.Y. and 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.

Jeganathan, P., 2003. Second order limits of functionals of sums of linear processes that converge to fractional stable motions. Unpublished Manuscript, Indian Statistical Insitute.

Lewellen J., 2004. Predicting returns with financial ratios. Journal of Financial Economics. 74, 209-235.

Stambaugh, R., 1999. Predicting regressions. Journal of Financial Economics 54, 375-421.

Stinchcombe, M.B., H., White H., 1998. Consistent specification testing with nuisance parameters present only under the alternative. Econometric Theory 14, 295-325.

Table 1. Empirical size and power of the Bierens tests (5% level).

	$n = 200$				$n = 500$			
Test	$\hat{B}(\tilde{m})$	$\hat{B}(\tilde{m})$	$\hat{B}(\tilde{m})$	$\hat{B}(m_o)$	$\hat{B}(\tilde{m})$	$\hat{B}(\tilde{m})$	$\hat{B}(\tilde{m})$	$\hat{B}(m_o)$
(γ, ρ)	$(1.5, 0.2)$	$(1.5, 0.3)$	$(2, 0.3)$	–	$(1.5, 0.2)$	$(1.5, 0.3)$	$(2, 0.3)$	–
$f = 0$	0.1154	0.0551	0.0457	0.0434	0.0964	0.0512	0.0484	0.0476
$f = f_1$	0.7920	0.7495	0.7235	0.5240	0.8550	0.8205	0.8075	0.6480
$f = f_2$	1.0000	1.0000	1.0000	0.9555	1.0000	1.0000	1.0000	0.9610
$f = f_3$	1.0000	1.0000	1.0000	0.9545	1.0000	1.0000	1.0000	0.9695
$f = f_4$	0.6920	0.5825	0.5270	0.4145	0.8235	0.7475	0.7100	0.5540
$f = f_5$	1.0000	1.0000	1.0000	0.9560	1.0000	1.0000	1.0000	0.9607
$f = f_6$	1.0000	1.0000	1.0000	0.9400	1.0000	1.0000	1.0000	0.9613
$f = f_7$	0.7490	0.7070	0.6787	0.5103	0.8470	0.8227	0.8123	0.6660

Table 2. Empirical size and power of Marmer's RESET test (5% level).

	$n = 200$			$n = 500$		
	$k = 1$	$k = 2$	$k = 3$	$k = 1$	$k = 2$	$k = 3$
$f = 0$	0.0405	0.0469	0.0510	0.0439	0.0468	0.0486
$f = f_1$	0.2940	0.6920	0.6935	0.4310	0.8045	0.8015
$f = f_2$	0.8940	0.9300	0.9305	0.9065	0.9270	0.9245
$f = f_3$	0.9945	0.9940	0.9960	0.9965	0.9950	0.9965
$f = f_4$	0.6970	0.6595	0.6325	0.8330	0.8015	0.7875
$f = f_5$	0.8943	0.9383	0.9377	0.9103	0.9310	0.9300
$f = f_6$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$f = f_7$	0.1870	0.4347	0.4507	0.3263	0.6233	0.6350

Table 3. Stationarity tests

	Returns	DY	BM	EP
ADF	-24.172	-0.846	-0.710	-0.584
Phillips-Perron	-24.613	-0.858	-0.693	-0.554
KPSS	0.050	2.894	2.175	1.302

Table 4. Tests for the predictability of value weighted NYSE returns.

Predictor: DY

$(\gamma,\rho)=(1.5,0.3)$	$\mathbf{W}_1$	$\mathbf{W}_2$
$m_o=1$	$\hat{B}(\tilde{m})=4.992^{**}$ $\hat{B}(m_o)=4.992^{**}$	$\hat{B}(\tilde{m})=5.139^{**}$ $\hat{B}(m_o)=5.139^{**}$
$m_o=15$	$\hat{B}(\tilde{m})=6.266^{**}$ $\hat{B}(m_o)=6.266^{**}$	$\hat{B}(\tilde{m})=6.798^{***}$ $\hat{B}(m_o)=6.798^{***}$
$(\gamma,\rho)=(0.5,0.1)$	$\mathbf{W}_1$	$\mathbf{W}_2$
$m_o=1$	$\hat{B}(\tilde{m})=6.862^{***}$ $\hat{B}(m_o)=4.992^{**}$	$\hat{B}(\tilde{m})=6.879^{***}$ $\hat{B}(m_o)=5.139^{**}$
$m_o=15$	$\hat{B}(\tilde{m})=6.266^{**}$ $\hat{B}(m_o)=6.266^{**}$	$\hat{B}(\tilde{m})=6.798^{***}$ $\hat{B}(m_o)=6.798^{***}$
$RESET=6.856^{***}$		

Predictor: BM

$(\gamma,\rho)=(1.5,0.3)$	$\mathbf{W}_1$	$\mathbf{W}_2$
$m_o=1$	$\hat{B}(\tilde{m})=3.632^{*}$ $\hat{B}(m_o)=3.632^{*}$	$\hat{B}(\tilde{m})=4.353^{**}$ $\hat{B}(m_o)=4.353^{**}$
$m_o=15$	$\hat{B}(\tilde{m})=3.490^{*}$ $\hat{B}(m_o)=3.490^{*}$	$\hat{B}(\tilde{m})=8.583^{***}$ $\hat{B}(m_o)=8.583^{***}$
$(\gamma,\rho)=(0.5,0.1)$	$\mathbf{W}_1$	$\mathbf{W}_2$
$m_o=1$	$\hat{B}(\tilde{m})=4.938^{**}$ $\hat{B}(m_o)=3.632^{*}$	$\hat{B}(\tilde{m})=8.583^{***}$ $\hat{B}(m_o)=4.353^{**}$
$m_o=15$	$\hat{B}(\tilde{m})=4.938^{**}$ $\hat{B}(m_o)=3.490^{*}$	$\hat{B}(\tilde{m})=8.583^{***}$ $\hat{B}(m_o)=8.583^{***}$
$RESET=4.271^{**}$		

Predictor: EP

$(\gamma,\rho)=(1.5,0.3)$	$\mathbf{W}_1$	$\mathbf{W}_2$
$m_o=1$	$\hat{B}(\tilde{m})=1.580$ $\hat{B}(m_o)=1.580$	$\hat{B}(\tilde{m})=6.815^{***}$ $\hat{B}(m_o)=6.815^{***}$
$m_o=15$	$\hat{B}(\tilde{m})=7.518^{***}$ $\hat{B}(m_o)=7.518^{***}$	$\hat{B}(\tilde{m})=5.949^{**}$ $\hat{B}(m_o)=5.949^{**}$
$(\gamma,\rho)=(0.5,0.1)$	$\mathbf{W}_1$	$\mathbf{W}_2$
$m_o=1$	$\hat{B}(\tilde{m})=1.580$ $\hat{B}(m_o)=1.580$	$\hat{B}(\tilde{m})=6.815^{***}$ $\hat{B}(m_o)=6.815^{***}$
$m_o=15$	$\hat{B}(\tilde{m})=7.518^{***}$ $\hat{B}(m_o)=7.518^{***}$	$\hat{B}(\tilde{m})=7.437^{***}$ $\hat{B}(m_o)=5.949^{**}$
$RESET=6.881^{***}$		

* significant at 10% level. ** significant at 5% level. *** significant at 1% level.