An Article Submitted to

Studies in Nonlinear Dynamics & Econometrics

Manuscript 1893

Stochastically Weighted Average Conditional Moment Tests of Functional Form

Jonathan Hill*

*University of North Carolina - Chapel Hill, jbhill@email.unc.edu

Stochastically Weighted Average Conditional Moment Tests of Functional Form*

Jonathan Hill

Abstract

We develop a new consistent conditional moment test of functional form based on nuisance parameter indexed sample moments first presented in Bierens (1982, 1990). We reduce the nuisance parameter space to known countable sets, which leads to a weighted average conditional moment test in the spirit of Bierens and Ploberger's (1997) Integrated Conditional Moment test. The weights are possibly stochastic in an arbitrary way, integer-indexed and flexible enough to cover a range of tests from average to higher quantile to maximum tests. The limit distribution under the null and local alternative belong to the same class as the ICM statistic, hence our test is admissible if the errors are Gaussian, and a flat weight leads to the greatest weighted average local power.

*We thank two anonymous referees for their careful reading and feedback, and Editor Bruce Mizrach for helpful recommendations.

1 Introduction

In this paper we propose a consistent Conditional Moment [CM] test of regression model function form in the spirit of the Integrated Conditional Moment [ICM] test of Bierens (1982) and Bierens and Ploberger (1997). We tackle the nuisance parameter problem by reducing the space to integers, effectively reducing the ICM statistic to a weighted average. We show a max-test is a special case, while an optimal version of the test in the sense of Andrews and Ploberger (1994) is based on a flat weight.

Let $\{y_t, \tilde{x}_t\} \in \mathbb{R} \times \mathbb{R}^k$ be iid random variables with non-degenerate continuous marginal distributions and finite variances. Define the regressor set

$$x_t := (1, \tilde{x}_t')',\ \Im_t := \sigma(x_t),\ \Im := \sigma(\cup_{t\in\mathbb{Z}}\Im_t),$$

and let $f(x_t, \phi)$ denote a known response function $f : \mathbb{R}^{k+1} \times \Phi \rightarrow \mathbb{R}$ with Φ a compact Euclidean subset of $\mathbb{R}^{k+1}$. The regression model is

$$y_t = f(x_t, \phi_0) + \epsilon_t \tag{1}$$

where ϕ_0 is the unique solution to

$$\phi_0 = \underset{\phi\in\Phi}{\text{arginf}} E(y_t - f(x_t, \phi))^2, \text{ an interior point of } \Phi.$$

We want to test whether $f(x_t, \phi_0)$ is a version of $E[y_t|x_t]$ against a general alternative:

$$H_0 : P\left(f(x_t, \phi_0) = E[y_t|x_t]\right) = 1 \text{ against } H_1 : \sup_{\phi\in\Phi} P\left(f(x_t, \phi) = E[y_t|x_t]\right) < 1.$$

An extension to weakly dependent data is straightforward. See de Jong (1996) and Hill (2008a,b), and their references.

Parametric CM tests based on a finite number of L_2-orthogonality conditions are known in general not to be consistent against every alternative (e.g. Ramsey 1969, White 1982, and Newey 1985). Consistency in this class was first established by Bierens (1982, 1990) by introducing an exponential weight $\exp(\gamma' x_t)$ indexed by a smooth nuisance parameter $\gamma \in \mathbb{R}^k$ producing uncountably many moment conditions. Bierens' (1990) celebrated Lemma 1, for example, sharpens results in Bierens (1982) by showing if $\tilde{x}_t$ is bounded and H_1 holds, then for $F(u) = \exp\{u\}$, any compact set $\Gamma \subset \mathbb{R}^{k+1}$ with positive Lebesgue measure, and some countable subset $S \subset \Gamma$

$$E\left[\epsilon_t F\left(\gamma' x_t\right)\right] \neq 0 \ \ \forall \gamma \in \Gamma/S. \tag{2}$$

That is, $\exp\{\gamma' x_t\}$ reveals model mis-specification expect possibly on a set $S \subset \Gamma$ with Lebesgue measure zero. If x_t is not bounded then any Borel measurable one-to-one bonded mapping $\Phi : \mathbb{R}^{k+1} \rightarrow \mathbb{R}^{k+1}$ may be used instead (Bierens 1990).

Stinchcombe and White (1998) provide the only detailed explanation for Bierens' seminal result by proving (2) for any real analytic[1] non-polynomial $F(u)$, where $\gamma' x_t$ is affine (i.e. a constant term must be included) and $\tilde{x}_t$ is bounded. Included weights are

[1] Recall a real analytic function is infinitely differentiable and therefore has an infinite order Taylor expansion.

therefore Bierens' $\exp\{u\}$, the logistic $[1+\exp\{u\}]^{-1}$ explored in Hornik et al (1989) and Lee et al (1993), and trigonometric weights. See also White (1989), Bierens (1994), de Jong and Bierens (1994), Boning and Sowell (1999), Dette (1999), Li et al (2003) and Hill (2008a,b) for related methods. Nonparametric and semiparametric techniques are developed in Yatchew (1992), Hong and White (1995), Zheng (1996), Stute (1997), Koul and Stute (1999), Fan and Li (1996) and Stute and Zhu (2005) amongst many others. Hill (2012) delivers a robust asymptotic power one CM test for heavy tailed time series.

The nuisance parameter γ can be handled by randomly selecting it (e.g. Lee et al 1993, Bierens 1990), or computing test functionals like the supremum or average (Davies 1977, Bierens 1990, Hill 2008a,b). Alternatively, Bierens (1982) and Bierens and Ploberger (1997) integrate a scaled sample version of $(E[\epsilon_t F(\gamma' x_t)])^2$ over Γ for the ICM test. See Section 2, below, for the construction. Extensions of the ICM to simulated ICM for conditional distributions is treated in Bierens and Wang (2011) who also treat optimal selection of the parameter space Γ.

Following Bierens (1982) we derive multinomial weights that reveal mis-specification for integer-valued nuisance parameters. Although $\exp\{\gamma' x_t\}$ is real analytic with affine $\gamma' x_t$, an equally useful interpretation follows directly from Bierens' (1982) and forms the basis of our test. Let $G_\xi(u)$ denote a reveal-valued multinomial mapping for $u, \xi \in \mathbb{R}^k$:

$$G_\xi(u) := \prod\nolimits_{i=1}^{k} u_i^{\xi_i} \text{ provided each } u_i^{\xi_i} \text{ is well defined.}$$

Thus if $u_i \in \mathbb{R}$ then we restrict $\xi_i \in \mathbb{N}$, and if $u_i \in (0,\infty)$ then $\xi_i \in \mathbb{R}$. Bierens (1982: Theorem 2) shows if $\tilde{x}_t$ is bounded then under H_1 there exists an integer $m \in \mathbb{N}^k$ such that

$$E\left[\epsilon_t G_m(\tilde{x}_t)\right] = E\left[\epsilon_t \prod\nolimits_{i=1}^{k} \tilde{x}_{i,t}^{m_i}\right] \neq 0. \tag{3}$$

Replace ϵ_t with $\epsilon_t \prod_{i=1}^{k} \tilde{x}_{i,t}^{m_i}$ and iterate to deduce (3) holds for infinitely many $m \in \mathbb{N}^k$. Since any one-to-one bounded Borel function $\Psi(\tilde{x}_t)$ generates the same σ-field as $\tilde{x}_t$, it follows

$$H_1 : E\left[\epsilon_t G_m(\Psi(\tilde{x}_t))\right] \neq 0 \text{ for infinitely many } m \in \mathbb{N}^k. \tag{4}$$

Similarly, if $\Psi_i(u) \in [0,\infty)$ then for any $\delta \in \mathbb{R}^k$ with non-zero components

$$H_1 : E\left[\epsilon_t G_{\delta \circ m}(\Psi(\tilde{x}_t))\right] \neq 0 \text{ for infinitely many } m \in \mathbb{N}^k. \tag{5}$$

Simply put $\Psi_i(\tilde{x}_t) = \exp\{\tilde{x}_{i,t}\}$ to achieve another version of (2): $E[\epsilon_t \exp\{\gamma' x_t\}] \neq 0$ for ucountably infinitely $\gamma = \delta \circ m$, where the revealing $m's$ intrinsically depend on the choice of δ.

Properties (4) and (5) provide an alternative explanation for why (2) works, where infinite differentiability of $\exp\{\gamma' x_t\}$, γ on compact Γ, and affine $\gamma' x_t$ do not play any role. The constant term is immaterial for a multinomial like $G_m(\Psi(\tilde{x}_t))$, $\Psi_i(u)$ need not be smooth, and the set of revealing $m's$ is countable and therefore has Lebesgue measure zero (while Γ/S has positive Lebesgue measure), all contrary to Stinchcombe and White's (1998) classification.

In Section 2 we use $G_m(u)$ to construct a discretized version of the ICM statistic by generating a weighted sum of a sample version of $(E[\epsilon_t G_m(\Psi(\tilde{x}_t))])^2$ over an integer nuisance parameter space. The summation weights are very general allowing for flat weighting, or stochastic weighting with special cases where very large sample moments are favored, hence a max-test is feasible.

Asymptotic theory is dealt with in Section 3, including a deeper treatment of optimal summation weights. The limit distribution of the proposed statistic belongs to the same class as the ICM statistic, so all the properties of the ICM statistic apply including consistency, admissibility if ϵ_t is Gaussian, and critical value upper bounds. A Monte Carlo study follows in Section 4 where we show a variety of mappings Ψ and test summation weights lead to a sharp test.

The ICM test construction and limit properties have become fundamental tools for composing or understanding mis-specification tests. Boning and Sowell (1999) show the ICM with an exponential weight $F(u)$ and uniform measure has the greatest weighted average local power (cf. Andrews and Ploberger 1994). Fan and Li (2000) show nonparametric tests by Härdle and Mammen (1993), Zhang (1996) and Li and Wang (1998) all have ICM representations with exponential weight $F(u)$. In Section 3 we show Boning and Sowell's (1999) results carry over to our weighted average test.

Throughout $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and finite dimensional distributions, and $\Rightarrow$ denotes weak convergence on a metric space. $|x|_p := (\Sigma_{i=1}^m \Sigma_{j=1}^k |x_{i,j}|^p)^{1/p}$ for $x \in \mathbb{R}^{m \times k}$; $|\cdot| = |\cdot|_2$; and for stochastic matrices $||x||_p = (\Sigma_{i,j} E|x_{i,j}|^p)^{1/p}$. I_k is a k-dimensional identity matrix and 1_k a k-vector of ones. $[z]$ is the integer part of z. ρ is always a number in $(0,1)$ whose value may be different in different places.

2 Stochastically Weighted Average Test

The $\sqrt{n}$-local and global alternatives are

$$H_1^L : y_t = f(x_t, \phi_0) + u_t/\sqrt{n} + \epsilon_t \quad \text{and} \quad H_1^G : y_t = f(x_t, \phi_0) + u_t + \epsilon_t,$$

where $E[\epsilon_t | x_t] = 0$, and u_t is $\Im_t$-measurable, independent and governed by a non-degenerate distribution. Under H_0 we have $u_t = 0$ $a.s.$ See Assumption C in Appendix A for all DGP assumptions, including restrictions on u_t. The following asymptotic properties are easily verified from Assumption C.

Define the regression residual $\hat{\epsilon}_t := y_t - f(x_t, \hat{\phi})$ for some plug-in $\hat{\phi}$, and a scaled sample moment

$$\hat{z}(m) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \hat{\epsilon}_t G_m \left(\Psi(\tilde{x}_t)\right).$$

All that follows carries over to $G_{\delta \circ m}(\Psi(\tilde{x}_t))$ for any $\delta \in \mathbb{R}^k$ with non-zero components, provided each $\Psi_i(\tilde{x}_t) \geq 0$ $a.s.$

A variety of plug-ins are possible, including NLLS (Bierens 1990, de Jong 1996, Hill 2008a) and GMM (Hill 2008b). For the sake of simplicity we ignore over-identifying restrictions and use NLLS:

$$\hat{\phi} = \underset{\phi \in \Phi}{\operatorname{argmin}} \left\{ \sum_{t=1}^{n} (y_t - f(x_t, \phi))^2 \right\}.$$

By the mean-value-theorem and the construction of $\hat{\phi}$, standard arguments reveal under H_1^L

$$(6) \qquad \hat{z}(m) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t g(x_t, m) + \frac{1}{n} \sum_{t=1}^{n} u_t g(x_t, m) + o_p(1) = z_n(m) + o_p(1),$$

say, where $z_n(m) = 1/\sqrt{n}\sum_{t=1}^{n} \epsilon_t g(x_t, m)$,

$$g(x_t, m) = G_m\left(\Psi(\tilde{x}_t)\right) - b(m, \phi_0)' A(\phi_0)^{-1} \frac{\partial}{\partial \phi} f(x_t, \phi)|_{\phi=\phi_0}$$

$$A(\phi) = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial \phi} f(x_t, \phi) \frac{\partial}{\partial \phi'} f(x_t, \phi)$$

$$b(m, \phi) = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} G_m\left(\Psi(\tilde{x}_t)\right) \times \frac{\partial}{\partial \phi} f(x_t, \phi).$$

Thus, the appropriate estimator of $E(\hat{z}(m))^2$ under the null is $\hat{\gamma}(m) = 1/n\sum_{t=1}^{n} \hat{\epsilon}_t^2 \hat{g}(x_t, m)^2$, with

$$\hat{g}(x_t, m) = G_m\left(\Psi(\tilde{x}_t)\right) - \hat{b}(m, \hat{\phi}_0)' \hat{A}(\hat{\phi}_0)^{-1} \frac{\partial}{\partial \phi} f(x_t, \hat{\phi}_0)$$

$$\hat{b}(m, \phi) = \frac{1}{n}\sum_{t=1}^{n} G_m\left(\Psi(\tilde{x}_t)\right) \times \frac{\partial}{\partial \phi} f(x_t, \phi)$$

$$\hat{A}(\phi) = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial \phi} f(x_t, \phi) \frac{\partial}{\partial \phi'} f(x_t, \phi).$$

Since $m \in \mathbb{N}^k$ is unbounded we must control for the fact that if $|\Psi_i(\tilde{x}_t)| > 1$ with positive probability for each i then $|G_m\left(\Psi(\tilde{x}_t)\right)| \xrightarrow{p} \infty$ as $|m| \to \infty$ is possible. Consistency, however, requires we compute $\hat{z}(m)$ for infinitely many $m \in \mathbb{N}^k$, while the ratio $\hat{z}(m)/\ \hat{\gamma}(m)^{1/2}$ need not be well-defined as $m \to \infty$. Thus, we do not consider test functionals like $\sup_m\{\hat{z}(m)^2/\hat{\gamma}(m)\}$, cf. Davies (1977) and Bierens (1990). Instead, we operate directly on $\hat{z}(m)$ a la Bierens (1982) and Bierens and Ploberger (1997).

The proposed test is based on the *Stochastically Weighted Average Conditional Moment* [SWACM] statistic:

$$\hat{T}_n = \sum_{\bar{m}=1}^{\mathcal{N}_n} \hat{z}(m)^2 \omega_{n,m} = \sum_{m\in\mathbb{N}^k:\bar{m}=1}^{\mathcal{N}_n} \hat{z}(m)^2 \omega_{n,m},$$

where $\sum_{\bar{m}=1}^{\mathcal{N}_n}$ denotes $\sum_{m\in\mathbb{N}^k:\bar{m}=1}^{\mathcal{N}_n}$, $\{\mathcal{N}_n\}$ is an increasing sequence of integers, the weights $\{\omega_{n,m}\}$ are defined below, and we write

$$\bar{m} := \sum_{j=1}^{k} m_j.$$

Notice the summation starts at integers $m \in \mathbb{N}^k$ with $\bar{m} = 1$: since $\bar{m} = 0$ for positive integers implies $m = 0$, and $G_0\left(\Psi(\tilde{x}_t)\right) = 1$ by construction, the sample moment $\hat{z}(0)$ cannot reveal mis-specification. Thus $\mathcal{N}_n$ is not in general the number of $m's$, but the maximum sum $\bar{m} := \sum_{j=1}^{k} m_j$. The total number of integer vectors m in the SWACM sum is denoted $\mathcal{M}_n$.

The SWACM statistic is a discreet analogue to the ICM statistic:

$$\widehat{\mathcal{I}}_n = \int_{\gamma \in \Gamma} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \hat{\epsilon}_t F\left(\gamma' \Phi(x_t)\right) \right)^2 d\mu(\gamma) \tag{7}$$

where Φ is a bounded one-to-one Borel function on $\mathbb{R}^{k+1}$, and $\mu(\gamma)$ is a non-stochastic absolutely continuous probability measure on compact $\Gamma \subset \mathbb{R}^{k+1}$. Examples of μ are the truncated normal (e.g. Fan and Li 2000) or uniform (Boning and Sowell 1999, Bierens and Wang 2011).

In order to ensure $\hat{T}_n$ has a well defined limit we require a limiting version $z(m)^2\omega_m$ of $\hat{z}(m)^2\omega_{n,m}$ to be mean summable $\sum_{\bar{m}=1}^{\infty} E(z(m)^2)\omega_m < \infty$. In general if $G_m(\Psi(\tilde{x}_t)) = O_p(a(\bar{m}))$ for some non-random $a : \mathbb{N} \to [0, \infty)$, and $\omega_{n,m} \overset{p}{\to} \omega_m$ a non-random limit, then we must have $\sum_{\bar{m}=1}^{\infty} a(\bar{m})^2\omega_m < \infty$. The following two assumptions, while not unique, ensure such summability.

Assumption A. *$G_m(\Psi(\tilde{x}_t)) = O_p(\rho^{\bar{m}})$ for some $\rho \in (0,1)$ and any $m \in \mathbb{N}^k$.*

Remark: $G_m(\Psi(\tilde{x}_t)) = O_p(\rho^{\bar{m}}))$ forces us to restrict $\Psi : \mathbb{R}^k \to (-1,1)^k$. Valid components $\Psi_i(\tilde{x}_t)$ include $sign(x_{i,t}) \times \exp\{-|x_{i,t}|\}$, $[1 + \exp\{x_{i,t}\}]^{-1}$, and $sign(x_{i,t})[1 + |x_{i,t}|]^{-1}$.

Assumption B. *$\{\mathcal{N}_n\}_{n \geq 1}$ denotes a sequence of increasing integers $\mathcal{N}_{n+1} > \mathcal{N}_n \geq 1$. $\{\omega_{n,m}\}_{m \in \mathbb{N}^k}$ is a sequence of possibly stochastic real numbers where $\liminf_{n\to\infty} P(\omega_{n,m} \geq 0) = 1$ for all m, with strict inequality for some m, and*

$$\limsup_{n \geq 1} P\left(\sum_{\bar{m}=1}^{\mathcal{N}_n} \rho^{\bar{m}} \omega_{n,m} < K \right) = 1 \text{ for some } K < \infty.$$

Similarly, $\{\omega_m\}_{m \in \mathbb{N}^k}$ is the unique non-stochastic sequence, $\omega_m \geq 0$ $\forall m$, $\omega_m > 0$ for some m, and $\sum_{\bar{m}=1}^{\infty} \rho^{\bar{m}}\omega_m < \infty$, that satisfies $\omega_{n,m} \overset{p}{\to} \omega_m$ for every m. In particular $\sum_{\bar{m}=1}^{\mathcal{N}_n} \rho^{\bar{m}}$ $\times$ $|\omega_{n,m} - \omega_m| = o_p(1)$.

Remark 1: By using moment condition (4), a consistent test is assured if $\liminf_{n \geq 1} P(\omega_{n,m} > 0) = 1$ for all m, but that is not necessary. We show below that a max-test is a special case of the SWACM $\hat{T}_n$, placing all weight on one m, say $m^* = \arg\max_{m \in \mathbb{N}}\{|E[u_t g(x_t, m)]|\}$, with summation weights $\omega_{m^*} = 1$ and $\omega_m = 0$ $\forall m \neq m^*$.

Remark 2: If all $\liminf_{n \geq 1} P(\omega_{n,m} > 0) = 1$ then $\{\omega_{n,m}, \omega_m\}$ need not be summable, as long as $\{\rho^{\bar{m}}\omega_{n,m}, \rho^{\bar{m}}\omega_m\}$ are summable in probability for any $n \geq 1$.

Valid test weights under Assumptions A and B include

$$\omega_{n,m}^{(u)} = 1 \quad \text{and} \quad \omega_{n,m}^{(g)} = \rho^{\bar{m}} \tag{8}$$

$$\omega_{n,m}^{(s)} = \frac{\rho^{\bar{m}} + |\hat{z}(m)| / \sqrt{n}}{1 + \sum_{\bar{m}=1}^{\mathcal{N}_n} |\hat{z}(m)| / \sqrt{n}}$$

$$\omega_{n,m}^{(h)}(\pi) = \frac{\rho^{\pi_1 \bar{m}} \times I\left(|\hat{z}(m)/\sqrt{n}| < \hat{z}_{(k_n(\pi_2))}\right) + \hat{z}_{(k_n(\pi_2))} \times I\left(|\hat{z}(m)/\sqrt{n}| \geq \hat{z}_{(k_n(\pi_2))}\right)}{1 + |\hat{z}(m)/\sqrt{n}| \times I\left(|\hat{z}(m)/\sqrt{n}| < \hat{z}_{(k_n(\pi_2))}\right)}$$

$$\omega_{n,m}^{(m)}(\pi) = \rho^{\pi_1 \bar{m}} \times I\left(\left|\hat{z}(m)/\sqrt{n}\right| < \hat{z}_{(k_n(\pi_2))}\right) + I\left(\left|\hat{z}(m)/\sqrt{n}\right| \geq \hat{z}_{(k_n(\pi_2))}\right).$$

Note $\omega_{n,m}^{(h)}(\pi)$ and $\omega_{n,m}^{(m)}(\pi)$ contain tuning parameters $\pi_1, \pi_2 \geq 0$; $\hat{z}_{(i)}$ are the order statistics of $|\hat{z}(m)/\sqrt{n}|$ where $\hat{z}_{(1)} \geq \hat{z}_{(2)} \geq ... \geq \hat{z}_{(\mathcal{M}_n)}$; and $\{k_n(\pi_2)\}$ is a sequence of integers, $0 \leq k_n(\pi_2) \leq \mathcal{M}_n$. As long as $\{y_t, \tilde{x}_t\}$ have absolutely continuous marginal distributions we may assume for any n (cf. Assumption C in Appendix A)

$$\hat{z}_{(1)} > \hat{z}_{(2)} > ... > \hat{z}_{(\mathcal{M}_n)} \ a.s.$$

Recall there are $\mathcal{M}_n$ usable $m's$ and hence $\mathcal{M}_n$ possible order statistics $\hat{z}_{(j)}$. By convention we therefore use "$\hat{z}_{(0)}$" to denote any value greater than the largest $|\hat{z}(m)|/\sqrt{n}$, e.g. $\hat{z}_{(1)} + 1$, in order to allow a non-binding threshold: $I(|\hat{z}(m)/n^{1/2}| < \hat{z}_{(0)}) = 1$ for each sample m.

The flat $\omega_{n,m}^{(u)} = 1$ and geometric $\omega_{n,m}^{(g)} = \rho^{\bar{m}}$ are the only ones above comparable to the ICM measure $\mu(\gamma)$. The flat or uniform $\omega_{n,m}^{(u)}$ is not summable, forcing some form of Assumption A to apply to $G_m(\Psi(\tilde{x}_t))$, while the geometric $\omega_{n,m}^{(g)} = \rho^{\bar{m}}$ implies we may relax Assumption A to just $G_m(\Psi(\tilde{x}_t)) = O_p(1)$.

Under H_0 and H_1^L the next two weights $\omega_{n,m}^{(s)}, \omega_{n,m}^{(h)}(\pi) \xrightarrow{p} \rho^{\bar{m}}$ as $n \rightarrow \infty$ for all m due to the easily shown relation under (6):

$$\frac{\hat{z}(m)}{\sqrt{n}} = \frac{1}{n}\sum_{t=1}^{n} \epsilon_t g(x_t, m) + o_p(1/\sqrt{n}) \xrightarrow{p} 0. \tag{9}$$

Both exploit $\hat{z}(m)/\sqrt{n}$ to augment weight placed on large sample moments, while ensuring $|\omega_{n,m} - \omega_m| \xrightarrow{p} 0$ for some bounded, positive non-stochastic sequence $\{\omega_m\}$. By comparison $\omega_{n,m}^{(m)}(\pi)$ favors, but places flat weight, on large values.

The weight $\omega_{n,m}^{(s)}$ augments large sample moments under H_1^G in a simple way since $\hat{z}(m)/\sqrt{n} \xrightarrow{p} E[u_t g(x_t, m)] \neq 0$, hence

$$\omega_{n,m}^{(s)} \xrightarrow{p} \frac{\rho^{\bar{m}} + |E[u_t g(x_t, m)]|}{1 + \sum_{\bar{m}=1}^{\infty} |E[u_t g(x_t, m)]|} \quad \text{under} \quad H_1^G$$

is large when $|E[u_t g(x_t, m)]|$ is large.

The last two weights $\{\omega_{n,m}^{(h)}(\pi), \omega_{n,m}^{(m)}(\pi)\}$ are inspired by Huber's (1977) theory of robust estimation. In that context large observations are given less weight or are trimmed or truncated as a robustness technique. We invert the premise to give more weight to large $|\hat{z}(m)|/\sqrt{n} \geq \hat{z}_{(k_n(\pi_2))}$ which is tuned by π_1 and π_2. A value $\pi_1 \geq 0$ gives flat or depressed weight $\rho^{\pi_1 \bar{m}}$ on small sample moments $|\hat{z}(m)/\sqrt{n}| < \hat{z}_{(k_n(\pi_2))}$, while $\pi_2 \geq 0$ tunes the threshold $\hat{z}_{(k_n(\pi_2))}$ above which sample moments are considered "large". Under H_0 and H_1^L since $\hat{z}(m)/\sqrt{n} \xrightarrow{p} 0$ uniformly in m we have $|\omega_{n,m}^{(h)}(\pi) - \rho^{\pi_1 \bar{m}}| \xrightarrow{p} 0$. Thus $\pi_1 < \infty$ for $\omega_{n,m}^{(h)}(\pi)$ must be satisfied to ensure a well defined test statistic with non-trivial power asymptotically.

Valid choices of sequences $\{k_n(\pi_2)\}$ include central order $[\pi_2 \mathcal{M}_n]$ where $\pi_2 \in (0,1)$; intermediate order like $[\mathcal{M}_n^{\pi_2}]$ with $\pi_2 \in (0,1)$ or $[\pi_2 \mathcal{M}_n / \ln(\mathcal{M}_n)]$ with $\pi_2 \in (0,1]$; and extreme order $[\pi_2]$ with $\pi_2 \geq 1$, cf. Leadbetter et al (1983). Notice if $\pi_1 = 0$ and $k_n(\pi_2) = 0$ then $\omega_{n,m}^{(m)}(\pi) = 1$ is just the flat weight $\omega_{n,m}^{(u)}$.

The above examples have an intuitive appeal based on their relation to the ICM measure μ and how they use information from large $\hat{z}(m)$. Certainly other possibilities exist, including replacing $\rho^{\bar{m}}$ with the hyperbolic $\bar{m}^{-a}$ for some $a > 0$ with an appropriate restriction on $G_m(\Psi(\tilde{x}_t))$ a la Assumption A. We treat the possibility of an optimal weight $\omega_{n,m}$ in Section 3.

The following examples reveal how $\omega_{n,m}^{(u)}$, $\omega_{n,m}^{(h)}(\pi)$ and $\omega_{n,m}^{(m)}(\pi)$ permit great flexibility ranging from average to max-tests. Write

$$\eta(m) := E\left[u_t g(x_t, m)\right].$$

EXAMPLE 1 (average): Trivially $\sum_{\bar{m}=1}^{\mathcal{N}_n} \hat{z}(m)^2 \omega_{n,m}^{(u)} = \sum_{\bar{m}=1}^{\mathcal{N}_n} \hat{z}(m)^2$ mimics a flat weighted Cramèr-von Mises or ICM statistic. The statistic is well defined for weights $G_m(\Psi(\tilde{x}_t)) = O(\rho^{\bar{m}})$ under Assumption A because $\hat{z}(m) = O_p(\rho^{\bar{m}})$, and it achieves the highest weighted average local power as shown in Section 3.2.

EXAMPLE 2 (upper quantile): Consider $\omega_{n,m}^{(h)}(\pi)$ with a central order sequence $k_n(\pi_2) = [\pi_2 \mathcal{N}_n]$, $\pi_2 \in (0,1)$, and let η_{π_2} be the upper π_2-quantile of $|\eta(m)|$. Then sample moments from the largest to a central value are given the greatest weight. Since by (6) we have $\hat{z}_{(k_n(\pi_2))} = \eta_{\pi_2} + O_p(1/\sqrt{n})$ under H_1^G, it follows $\omega_{n,m}^{(h)}(\pi) \xrightarrow{p} \omega_m^{(h)}(\pi)$ where

$$\begin{aligned}
\omega_m^{(h)}(\pi) &= \eta_{\pi_2} \ \text{ if } \ |\eta(m)| \geq \eta_{\pi_2} \\
&= \frac{1}{1 + |\eta(m)|} \times \rho^{\pi_1 \bar{m}} \ \text{ otherwise.}
\end{aligned}$$

Now let $m_{(i)}$ be the unique integer, with probability one, that satisfies $\hat{z}_{(i)} = \hat{z}_{m_{(i)}}$. If, for example, $\pi_1 = 100$ and $\pi_2 = 1/2$ then $\rho^{\pi_1 \bar{m}} \approx 0$ hence

$$\begin{aligned}
\hat{T}_n &= \hat{z}_{([\mathcal{N}_n/2])} \times \sum_{i=1}^{[\mathcal{N}_n/2]} \left(n \times \hat{z}_{(i)}^2\right) + \sum_{i=[\mathcal{N}_n/2]+1}^{\mathcal{N}_n} \left(n \times \hat{z}_{(i)}^2 \times \frac{\rho^{100 \bar{m}_{(i)}}}{1 + \left|\hat{z}(m_{(i)})/\sqrt{n}\right|}\right) \\
&= \hat{z}_{([\mathcal{N}_n/2])} \times \sum_{i=1}^{[\mathcal{N}_n/2]} \left(n \times \hat{z}_{(i)}^2\right) + r_n(\pi_1),
\end{aligned}$$

where $r_n(\pi_1) \approx 0$, placing equal and nearly all weight on those $|\hat{z}(m)|$ above the median $\hat{z}_{([\mathcal{N}_n/2])}^2$. Notice $r_n(\pi_1)$ never vanishes for $0 \leq \pi_1 < \infty$ although it can be made arbitrarily close to zero for any n, in probability, by setting π_1 large.

EXAMPLE 3 (near-max): Consider $\omega_{n,m}^{(h)}(\pi)$ again, this time with the extreme tail order $k_n(\pi_2) = [\pi_2] \geq 1$: the greatest weight is given to the largest $[\pi_2]$ sample moments. If $\pi_2 = 1$ then the maximum moment is given the largest weight, hence under H_1^G

$$\begin{aligned}
\omega_m^{(h)}(\pi) &= \max_{m \in \mathbb{N}^k} |\eta(m)| \ \text{ if } \ |\eta(m)| = \max_{m \in \mathbb{N}^k} |\eta(m)| \\
&= \frac{1}{1 + |\eta(m)|} \times \rho^{\pi_1 \bar{m}} \text{ otherwise.}
\end{aligned}$$

If $\pi_1 = 100$ then $\hat{T}_n$ works like an adaptive "near" max-test: for some $r_n(\pi_1) \approx 0$ for any n, $\hat{T}_n = \max_{1 \leq \bar{m} \leq \mathcal{N}_n} \{|\hat{z}(m)|^3 / n^{1/2}\} + r_n(\pi_1)$.

It is tempting to force $r_n(\pi_1) \rightarrow 0$ $a.s.$ by taking $\pi_1 \rightarrow \infty$, but the resulting statistic $\max_{m:1 \leq \bar{m} \leq \mathcal{N}_n} \{|\hat{z}(m)|^3 / n^{1/2}\}$ is degenerate under H_0 and H_1^L since $\max_{m:1 \leq \bar{m} \leq \mathcal{N}_n} \{|\hat{z}(m)|/n^{1/2}\} \xrightarrow{p} 0$ by (9).

EXAMPLE 4 (max): The remaining weight $\omega^{(m)}_{n,m}(\pi)$, however, places flat weight on large values. Put $\pi_2 = 1$ to obtain

$$\sum_{\bar{m}=1}^{\mathcal{N}_n} \hat{z}(m)^2 \omega^{(m)}_{n,m}(\pi) = n \times \hat{z}^2_{(1)} + \sum_{i=2}^{\mathcal{N}_n} \hat{z}^2_{(i)} n\rho^{\pi_1 \bar{m}_{(i)}} \rightarrow n \times \hat{z}^2_{(1)} = \max_{1\leq \bar{m}\leq \mathcal{N}_n} \left\{ \hat{z}(m)^2 \right\} \; a.s.$$

as $\pi_1 \rightarrow \infty$, thus the SWACM structure allows for a max-test. The ICM statistic, however, places zero weight $\mu(d\gamma)$ on any particular point $\gamma \in \Gamma$ because μ is absolutely continuous. Thus, $\sup_{\gamma\in\Gamma}(n^{-1/2}\sum_{t=1}^{n} \hat{\epsilon}_t F(\gamma'\Phi(x_t)))^2$ cannot arise as a version of $\widehat{\mathcal{I}}_n$.

3 Asymptotic Theory

We present the main asymptotic results, discuss critical value computation, and end this section by characterizing optimal test weights $\omega_{n,m}$.

3.1 Main Results

In order to derive the limit distribution of $\hat{T}_n$ we require the weak limit of $\hat{z}(m) = 1/\sqrt{n}\sum_{t=1}^{n} \hat{\epsilon}_t G_m(\Psi(\tilde{x}_t))$. It is easier to work with a general functional $\hat{z}(\xi) = 1/\sqrt{n}\sum_{t=1}^{n} \hat{\epsilon}_t G_\xi(\Psi(\tilde{x}_t))$ indexed by real-valued $\xi \in [0,\infty)^k$, and therefore with a strictly positive bounded one-to-one mapping $\Psi : \mathbb{R}^k \rightarrow (0,K]^k$ and some $K > 0$. This does not reduce generality since the main results carry over to one-to-one $\text{sign}\{\Psi_i(\tilde{x}_t)\} \times |\Psi_i(\tilde{x}_t)|^{\xi_i}$. We exploit a tightness argument due to Bierens and Ploberger (1997: Lemma A.1) which relies on Lipschitz continuity for $G_\xi(\Psi(\tilde{x}_t))$. Differentiability with respect to $\xi \in \mathbb{R}^k$ ensures Lipschitz continuity, while $(\partial/\partial\xi)G_\xi(\Psi(\tilde{x}_t)) = G_\xi(\Psi(\tilde{x}_t)) \times \ln\Psi(\tilde{x}_t)$ is well defined if $\Psi : \mathbb{R}^k \rightarrow (0,K]^k$.

We need to show $z_n(\xi)$ converges weakly to some Gaussian element $z(\xi)$ of $\mathcal{C}[0,\infty)^k$. This requires convergence with respect to finite dimensional distributions and uniform tightness. Cf. Billingsley (1999), Bickel and Wichura (1971) and Neuhaus (1971). Assumption C is presented in Appendix A, and proofs are relegated to Appendix B.

LEMMA 3.1. *Let $\Psi : \mathbb{R}^k \rightarrow (0,K]^k$. Under Assumption C and H_1^L there exists a Gaussian law $z(\xi)$ with mean function $\eta(\xi) := \text{plim}_{n\rightarrow\infty} 1/n\sum_{t=1}^{n} u_t g(x_t,\xi)$ and covariance function $\Gamma(\xi_1,\xi_2) := \text{plim}_{n\rightarrow\infty} 1/n\sum_{t=1}^{n} \epsilon_t^2 g(x_t,\xi_1)g(x_t,\xi_2)$ such that $z_n(\xi) \stackrel{d}{\rightarrow} z(\xi)$ in finite dimensional distributions. Further $\eta(\xi) = O(\rho^{\bar{\xi}})$ and $\Gamma(\xi_1,\xi_2) = O(\rho^{(\bar{\xi}_1+\bar{\xi}_2)})$ under Assumption A.*

LEMMA 3.2. *Let $\Psi : \mathbb{R}^k \rightarrow (0,K]^k$. Under Assumption C $\{z_n(\xi)\}$ is uniformly tight on $[0,\infty)^k$.*

Use the limit process $\{z(\xi)\}$ from Lemma 3.1 to define the distribution

$$T_1 := \sum_{\bar{m}=1}^{\infty} z(m)^2 \omega_m.$$

THEOREM 3.3. *Let $\Psi : \mathbb{R}^k \rightarrow (0,K]^k$. Under H_1^L and Assumptions A-C $\hat{T}_n \Rightarrow T_1$.*

Characterizing the limiting distribution T_1 closely follows Bierens and Ploberger (1997). Let $\{\lambda_i\}$ be the eigenvalues of $\Gamma(m_1,m_2)$ defined in Lemma 3.1, let $\{\vartheta_i(m)\}_{i=1}^{\infty}$

denote an orthonormal sequence

$$\sum_{\bar{m}=1}^{\infty} \vartheta_i(m)\vartheta_j(m)\omega_m = 0 \text{ or } 1 \text{ if } i \neq j \text{ or } i = j,$$

and define

$$\eta_i := \sum_{\bar{m}=1}^{\infty} \eta(m)\vartheta_i(m)\omega_m \text{ where } \eta(m) := \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} u_t g(x_t, m).$$

THEOREM 3.4. *Let H_1^L, $\Psi : \mathbb{R}^k \to (0, K]^k$ and Assumptions A-C hold. There exists a sequence $\{\zeta_i\}_{i=1}^{\infty}$ of iid standard normal random variables, and an orthonormal sequence $\{\vartheta_i(m)\}_{i=1}^{\infty}$ that solves the eigenvalue problem*

$$\sum_{\bar{m}_2=1}^{\infty} \Gamma(m_1, m_2)\vartheta_i(m_2)\omega_{m_2} = \lambda_i\vartheta_i(m_1), \ \ \forall m_1 \in \mathbb{N}^k, \ i = 1, 2, ...,$$

such that $T_1 = \sum_{i=1}^{\infty}(\zeta_i\lambda_i^{1/2} + \eta_i)^2$.

Remark: The limit distribution $T_1 = \sum_{i=1}^{\infty}(\zeta_i\lambda_i^{1/2} + \eta_i)^2$ under H_1^L is identical in form to the limit distribution of Bierens and Ploberger's (1997: Theorem 3) ICM statistic $\widehat{\mathcal{I}}_n$. All of the implied properties of $\widehat{\mathcal{I}}_n$ therefore carry over to $\hat{T}_n$. This includes convergence under null

$$H_0 : \hat{T}_n \xrightarrow{d} T_0 = \sum_{i=1}^{\infty} \zeta_i^2\lambda_i \text{ where } E[T_0] = \sum_{i=1}^{\infty} \lambda_i.$$

If all $\omega_m > 0$ then we achieve consistency under the global alternative since by (4) infinitely many $\eta(m) \neq 0$, hence

$$H_1^G : \hat{T}_n \to \infty \text{ with probability one.}$$

Similarly $\hat{T}_n$ is consistent under "large" local alternatives, and asymptotically admissible for normally distributed ϵ_t.

Although the limit laws of $\hat{T}_n$ and $\widehat{\mathcal{I}}_n$ belong to the same class T_1, the SWACM statistic has a flexibility advantage because $\omega_{n,m}$ are possibly stochastic, discretely indexed, may be zero or positive asymptotically, and need not be summable as long as the limiting terms $E(z(m)^2)\omega_m$ are. The following corollaries explore two cases although many more are feasible. Throughout $\Psi : \mathbb{R}^k \to (0, K]^k$.

Consider testing for linearity in $y_t = \phi_1 + \phi_2 x_t + \epsilon_t$ for scalar iid $x_t \in [0, K]$ *a.s.*, $E[x_t^2] = 1$, and $E[\epsilon_t^2] = 1$. Use a flat summation weight, least squares $\hat{\phi}$ and exponential test weight:

$$\hat{T}_n = \sum_{\bar{m}=1}^{\mathcal{N}_n} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(y_t - \hat{\phi}_1 - \hat{\phi}_2 x_t \right) \exp\{-m x_t\} \right)^2.$$

Assumptions A-C are easily verified. Since the following is a simple case of Theorem 3.4 we omit the proof.

COROLLARY 3.5 (uniform, exponential). *The null distribution $\sum_{i=1}^{\infty}\zeta_i^2\lambda_i$ has eigenvalues λ_i that satisfy $\sum_{\bar{m}_2=1}^{\infty} E[g(x_t, m_1)g(x_t, m_2)]\vartheta_i(m_2) = \lambda_i\vartheta_i(m_1)$ where $g(x_t, m)$*

$= \exp\{-mx_t\} - E[x_t \exp\{-mx_t\}] \times x_t$. *Under* H_1^L *it follows* $\hat{T}_n \xrightarrow{d} \sum_{i=1}^{\infty}(\zeta_i\lambda_i^{1/2} + \sum_{m=1}^{\infty} E[\epsilon_t \exp\{-mx_t\}]\vartheta_i(m))^2$ *where infinitely many* $E[\epsilon_t \exp\{-mx_t\}] \neq 0$.

Now consider the max-test in Example 4. Since $\max_{1\leq\bar{m}\leq\mathcal{N}_n}\{\hat{z}(m)^2\}$ has a SWACM representation a straightforward alteration of Theorem 3.4 applies.

COROLLARY 3.6 (max-test). *Define* $m^* := \arg\max\{m \in \mathbb{N}^k : |E[u_t g(x_t, m)]|\}$. *Under the conditions of Theorem 3.4 if* H_1^L *is true then* $\max_{1\leq\bar{m}\leq\mathcal{N}_n}\{\hat{z}(m)^2\} \xrightarrow{d} T_1 := (\zeta_1\Gamma(m^*, m^*)^{1/2} + \eta(m^*))^2$ *where* $\zeta_1 \overset{iid}{\sim} N(0,1)$, *hence* $T_1/\Gamma(m^*, m^*)$ *is noncentral* $\chi^2(1)$ *with noncentrality* $\eta(m^*)^2/\Gamma(m^*, m^*)$. *Further, if* H_1^G *is true then* $\max_{1\leq\bar{m}\leq\mathcal{N}_n}\{\hat{z}(m)^2\} \xrightarrow{p} \infty$.

Remark: Under H_0 we therefore have $\max_{1\leq\bar{m}\leq\mathcal{N}_n}\{\hat{z}(m)^2\} \xrightarrow{d} \chi^2(1) \times \Gamma(m^*, m^*)$. Recall we cannot in general work with $\max_{1\leq\bar{m}\leq\mathcal{N}_n}\{\hat{z}(m)^2/\hat{\gamma}(m)\}$ because it may be degenerate asymptotically.

Hansen's (1996) bootstrap method applies for computing the p-value for $\hat{T}_n$. Alternately, Bierens and Ploberger (1997) deliver case-independent upper bounds on the asymptotic critical values of the ICM test. Their argument carries over in its entirety to the SWACM test by virtue of their versatile Lemma 7 and Theorem 7, and the fact that the null limit distributions of both $\hat{T}_n$ and $\widehat{\mathcal{I}}_n$ belong to the same class $T_0 = \sum_{i=1}^{\infty} \zeta_i^2\lambda_i$ under the null.

In particular, under the conditions of Theorem 3.4 and H_0 (Bierens and Ploberger 1997: eq. (40))

$$\lim_{n\to\infty} P\left(\hat{T}_n > [3.23,\ 4.26,\ 6.81]' \times \sum_{\bar{m}=1}^{\mathcal{N}_n} \hat{\gamma}(m)\omega_{n,m}\right) \leq [.10,\ .05,\ .01]', \tag{10}$$

where $\hat{\gamma}(m) = 1/n\sum_{t=1}^{n} \hat{\epsilon}_t^2 g(x_t, m)^2$. For example, we reject the null hypothesis at a maximum asymptotic 5% level if $\hat{T}_n/\sum_{\bar{m}=1}^{\mathcal{N}_n} \hat{\gamma}(m)\omega_{n,m} > 4.26$.

3.2 Optimal Weights

Boning and Sowell [BS] (1999) construct a weighted average test statistic $\mathcal{CT}_n := \sum_{i=1}^{\infty} \lambda_i^2 \mathcal{T}_{n,i}^2$ where λ_i are eigenvalues of a covariance function, and $\mathcal{T}_{n,i}$ are smoothed versions of Bierens' (1990) scaled sample moment $1/n^{1/2}\sum_{t=1}^{n} \hat{\epsilon}_t \exp\{\gamma'\Phi(x_t)\}$ over Γ. See below for details. BS (1999) prove $\mathcal{CT}_n$ is optimal in the sense that it has the greatest weighted average local power (cf. Andrews and Ploberger 1994), and is identically the ICM with exponential test weight and flat measure $\int_{\gamma\in\Gamma}(n^{-1/2}\sum_{t=1}^{n} \hat{\epsilon}_t \exp\{\gamma'\Phi(x_t)\})^2 d\gamma$.

It is easy to extend BS's environment with revealing weight $\exp\{\gamma'\Phi(x_t)\}$ on $L_2(\Gamma)$ to ours with multinomial $G_m(\Psi(\tilde{x}_t))$ on $L_2(\mathbb{N}^k)$. As above let $\{\vartheta_i(m)\}_{i=1}^{\infty}$ be an orthonormal basis of $L_2(\mathbb{N}^k)$, the space of square integrable functions on $\mathbb{N}^k$ with covariance function $\Gamma(m_1, m_2)$ defined in Lemma 3.1, and inner product

$$\langle x, y\rangle_\Gamma := \sum_{\bar{m}_1,\bar{m}_2=1}^{\infty} x(m_1)\Gamma(m_1, m_2)\, y(m_2) \text{ where } x, y \in L_2(\mathbb{N}^k).$$

By construction every $x \in L_2(\mathbb{N}^k)$ can be written $x(m) = \sum_{i=1}^{\infty} \langle x, \vartheta_i\rangle_\Gamma\, \vartheta_i(m)$. Define as in BS (1999) a test statistic

$$\mathcal{T}_{n,i} := \langle \hat{z}_n, \vartheta_i \rangle_{\Gamma^{-1}} = \sum_{\bar{m}_1,\bar{m}_2=1}^{\infty} \hat{z}(m_1)\Gamma(m_1,m_2)^{-1}\vartheta_i(m_2).$$

The eigenvalues λ_i of the covariance function $\Gamma(m_1, m_2)$ satisfy $\lambda_i = \langle \vartheta_i, \vartheta_i \rangle_\Gamma$. Now define a single test statistic $\widehat{\mathcal{CT}}_n := \sum_{i=1}^{\infty} \lambda_i^2 \mathcal{T}_{n,i}^2$. The statistic is well defined since by a direct replication of BS's (1999: p. 713) argument

$$\begin{aligned} \sum_{i=1}^{\infty} \lambda_i^2 \mathcal{T}_{n,i}^2 &= \sum_{i=1}^{\infty} \lambda_i^2 \left(\lambda_i^{-1} \sum_{\bar{m}_1,\bar{m}_2=1}^{\infty} \lambda_i \times \hat{z}(m_1)\Gamma(m_1,m_2)^{-1}\vartheta_i(m_2) \right)^2 \\ &= \sum_{i=1}^{\infty} \lambda_i^2 \left(\lambda_i^{-1} \sum_{\bar{m}=1}^{\infty} \hat{z}(m)\vartheta_i(m) \right)^2 = \sum_{\bar{m}=1}^{\infty} \hat{z}(m)^2, \end{aligned}$$

which is identically the flat weight SWACM. Note that the second equality follows by the construction of the orthonomal basis $\{\vartheta_i(m)\}_{i=1}^{\infty}$, and the third by Parseval's Theorem. Since $\sum_{\bar{m}=1}^{\infty} \hat{z}(m)^2$ is well defined under Assumptions A and C, so is $\widehat{\mathcal{CT}}_n$.

BS (1999: p. 715-717) crucially exploit the revealing properties of $\exp\{\gamma'\Phi(x_t)\}$ to prove optimality of the class $\widehat{\mathcal{CT}}_n := \sum_{i=1}^{\infty} \lambda_i^2 \mathcal{T}_{n,i}^2$. Since other weight functions and other nuisance parameter spaces lead to revealing moments and the appropriate weak limit theory, their argument instantly extends to $G_m(\Psi(\tilde{x}_t))$ on $L_2(\mathbb{N}^k)$. We omit a proof because it simply mimics BS (1999: Sections 3 and 4.3). See especially their equations (3)-(6).

THEOREM 3.7 (optimal SWACM). *Let $\Psi : \mathbb{R}^k \to (0, K]^k$. Under Assumptions A and C the flat weighted SWACM $\sum_{\bar{m}=1}^{\mathcal{N}_n} \hat{z}(m)^2$ obtains asymptotically the highest weighted average power against H_1^L (cf. Andrews and Ploberger 1994).*

Remark: Optimality aligns with the flat weight $\omega_{n,m}^{(u)} = 1$. The more flexible weight $\omega_{n,m}^{(m)}(\pi)$ therefore achieves optimality with $\pi_1 = 0$ and $k_n(\pi_2) = 0$.

4 Monte Carlo Study

In this final section a Monte Carlo study is performed. We draw 10,000 samples $\{\epsilon_t, w_t\}_{t=1}^n$ of iid standard normal random variables ϵ_t and $w_t \in \mathbb{R}^{10}$, for $n \in \{100, 500, 1000\}$. For each sample we use k regressors $x_t = [w_{1,t}, ..., w_{k,t}]'$ to construct the dependent variable y_t, where k is uniformly randomly selected from $\{1, ..., 10\}$. The models of y_t are

$$\begin{aligned} &\text{Linear:} && y_t = \phi_1' x_t + \epsilon_t \\ &\text{Switching:} && y_t = \phi_1' x_t + \phi_2' x_t \times I(x_{1,t} > 0) + \epsilon_t \\ &\text{Bilinear:} && y_t = \phi_1' x_t + \beta x_{1,t} x_{2,t} + \epsilon_t \\ &\text{Quadratic:} && y_t = \phi_1' x_t + \gamma x_{1,t}^2 + \epsilon_t \\ &\text{Logistic:} && y_t = \phi_1' x_t + \beta \left[1 + \exp\{\phi_2' x_t\}\right]^{-1} + \epsilon_t. \end{aligned}$$

We randomly select all parameters for each sample: ϕ_i is uniformly randomly selected from $[-.9, .9]^k$, and β and γ are uniformly randomized on $[0, 10]$ and $[-.9, .9]$ respectively.

We estimate a linear regression model $y_t = \phi' w_t + \epsilon_t$ using all regressors w_t by least squares. We then test the residuals $\hat{\epsilon}_t$ for omitted nonlinearity by computing the SWACM statistic with m taken from a subset of $\mathbb{N}^k$ described below. The test weights $G_m^{(\cdot)} = G_m(\Psi_t^{(\cdot)})$ are based on one of two arguments $\Psi_t^{(\cdot)}$ that satisfy Assumption A:

$$\text{Exponential } \Psi_{t,i}^{(E)} = \exp\{-|w_{t,i}|\} \times sign(w_{t,i})$$
$$\text{Logistic } \Psi_{t,i}^{(L)} = [1 + \exp\{-w_{t,i}\}]^{-1}.$$

We use the five summation weights $\omega_{n,m}^{(\cdot)}$ defined in Section 2. The Huber-type $\omega_{n,m}^{(h)}(\pi)$ is computed with $\pi_1 = 1$ and central order $k_n(\pi_2) = [.05\mathcal{N}_n]$ which favors and places large weight on the top 5^{th}-percentile of $\hat{z}(m)^2$. Similarly $\omega_{n,m}^{(m)}(\pi)$ is computed with $\pi_1 = 1$ and extreme order $\pi_2 = 2$ which favors but places flat weight on the first and second largest $|\hat{z}(m)|$. We find that substantially depressing the weight on the smaller sample moments (e.g. $\pi_1 = 100$) decreases empirical power by damping usable information from those moments. In particular, although we find a max-test performs well based on simulations not reported here, it obtains the least power when compared with other SWACM statistics. This is not surprising since the highest average power is obtained by the flat $\omega_{n,m}^{(u)} = 1\ \forall m$ as we discuss below, whereas a max-test puts zero weight on all but one $|\hat{z}(m)|$.

The statistic $\hat{T}_n = \sum_{m \in \mathfrak{N}_n} \hat{z}(m)^2 \omega_{n,m}$ is computed over an increasing integer set of integers $\mathfrak{N}_n \subset \mathbb{N}^k$ constructed as follows. Let $\{J_n\}$ be a sequence of scalar integers to be defined below, $J_n \to \infty$ as $n \to \infty$, and let $j_i^{(k)}$ be a k-vector with the value j for the i^{th} component and the value k in all other components. For example, $2_3^{(0)} = [0, 0, 2, 0, ..., 0]'$ and $2_1^{(2)} = [2, 2, ..., 2]'$. Let $\tilde{N}_n$ be a set with $[(\ln n)^{1/2}]$ integer vectors randomly selected from $\{[0, ..., 0]', ..., [[J_n^{1/2}], ..., [J_n^{1/2}]']\}$. Let $\check{N}_n$ be the set of all integers in the hypercube $\{[1, ..., 0]'$, ..., $[[J_n^{1/8}], ..., [J_n^{1/8}]]'\}$. Finally, let $\breve{N}_n$ denote the set of all simple integers $\{\{i_j^{(0)}\}_{i=1}^{[J_n^{1/2}]}\}_{j=1}^5$ and $\{1_1^{(1)}, 2_1^{(2)}, ..., [J_n^{1/2}]_1^{[J_n^{1/2}]}\}$. Then

$$\mathfrak{N}_n := \tilde{N}_n \cup \check{N}_n \cup \breve{N}_n.$$

Thus $\mathfrak{N}_n$ contains integer vectors ranging from $[1, 0, ..., 0]$ to $[[J_n^{1/2}], ..., [J_n^{1/2}]]'$ hence $\bar{m} \in \{1, ..., 5 \times [J_n^{1/2}]\}$, and $M_n \to \mathbb{N}^5$.

Theorems 3.3 and 3.4 do not provide a basis for choosing J_n since in theory any $J_n \to \infty$ is allowed. We find, however, that $J_n = o(n)$ leads to superlative test performance where faster convergence rates are associated with poor size. We therefore use $J_n = \ln(n)$.

4.1 ICM and Most Powerful Tests

Finally, we compute ICM and Most Powerful [MP] test statistics. Since the ICM is applied elsewhere (Fan and Li 2000, Bierens and Wang 2011) we limit our scope, in particular since our data are iid Gaussian and therefore do not deviate from previous studies. We compute the ICM $\widehat{\mathcal{I}}_n$ on $\Gamma = [-1, 1]$, with flat measure $\mu(d\gamma) = 1$ in lieu of optimality (Boning and Sowell 1999), weight $F(u) = \exp\{u\}$ or $[1 + \exp\{u\}]^{-1}$, and argument $\Phi_i(x_t) = \arctan((x_{i,t} - 1/n \sum_{t=1}^n x_{i,t})/s_i)$ where $s_i^2 = 1/n \sum_{t=1}^n (x_{i,t} - 1/n \sum_{t=1}^n x_{i,t})^2$ as in Bierens (1990) and Bierens and Wang (2011). Critical value upper bounds are identically (10), cf. Bierens and Ploberger (1997: eq. (40)).

Finally, the MP statistic is easily deduced. Since the parameters ϕ_i, β, and δ and regressor dimension k are known within the simulation, each model can be written as $y_t(\phi_1) = \varsigma' z_t(\tau) + \epsilon_t$, where $y_t(\phi_1) = y_t - \phi_1' x_t$ and $\varsigma' z_t(\tau)$ is some model specific parametric function. In the switching model $\varsigma' z_t(\tau) = \phi_2' x_t I(x_{1,t} > 0)$, $\varsigma = \phi_2$; bilinear $\varsigma' z_t(\tau) = \beta x_{1,t} x_{2,t}$, $\varsigma = \beta$; quadratic $\varsigma' z_t(\tau) = \gamma x_{2,t}^2$, $\varsigma = \gamma$; and logistic $\varsigma' z_t(\tau) = \beta[1 + \exp\{\phi_2' x_t\}]^{-1}$, $\varsigma = \beta$ and $\tau = \phi_2$. Since the errors are iid standard normal, an appeal to a generalization of the Neyman-Pearson lemma leads to the MP statistic $W_n(\tau) = y(\phi_1)' z(\tau) \left[z(\tau)' z(\tau)\right]^{-1} z(\tau)' y(\phi_1)$. It is easy to show $W_n(\tau) \xrightarrow{d} \chi^2(J)$ under H_0 for any point τ, where J is the dimension of ς.

4.2 Simulation Results

Empirical size and power of the SWACM and ICM statistics are presented in Tables 1-4. We only report power for MP tests in Table 5 since empirical size is near the nominal size. In simulation experiments not reported here we find Hansen's (1996) p-value method leads to similar results[2].

Table 1 - SWACM Test Sizes[a]

Weight \ n		100	500	1000
		10%, 5%, 1%	10%, 5%, 1%	10%, 5%, 1%
$\omega_{n,m}^{(u)}$	L[b]	.08, .05, .01[c]	.07, .04, .01	.09, .05, .01
	E	.08, .05, .01	.07, .04, .01	.09, .05, .01
$\omega_{n,m}^{(g)}$	L	.11, .06, .02	.09, .05, .01	.10, .05, .01
	E	.07, .04, .01	.04, .02, .00	.06, .03, .01
$\omega_{n,m}^{(s)}$	L	.08, .06, .03	.11, .06, .02	.09, .04, .01
	E	.05, .03, .01	.04, .02, .01	.03, .02, .01
$\omega_{n,m}^{(h)}$	L	.08, .04, .01	.09, .05, .01	.06, .03, .00
	E	.04, .02, .00	.02, .01, .00	.00, .00, .00
$\omega_{n,m}^{(m)}$	L	.07, .03, .01	.09, .04, .01	.09, .05, .01
	E	.01, .01, .00	.04, .02, .01	.05, .03, .01

a. Critical value upper bounds are taken from (10) in Section 3.1.
b. The SWACM moment condition weight $G_m(\Psi_t)$ is based on a logistic (L) or exponential (E) argument Ψ_t.
c. Rejection frequencies at the 10%, 5% and 1% levels.

The SWACM test generates reasonable empirical size at the nominal 5% level considering the rejection rates have 99% bounds .05 $\pm$ .0195. Thus, Bierens and Ploberger's (1997) critical value bounds work exceptionally well here. Although in all cases power is large once $n \geq 500$, the sharpest size and highest power occurs with the flat weighted SWACM with logistic $G_m^{(L)}$.

[2] The data are iid Gaussian and the nuisance parameter space is countable, hence it is easy to show Hansen's approximate p-value is consistent (cf. de Jong 1996, Hill 2008a).

Table 2 - SWACM Test Power (weights $\omega_{n,m}^{(u)}, \omega_{n,m}^{(g)}, \omega_{n,m}^{(s)}$)

Weight	Model \ n		100	500	1000
			10%, 5%, 1%	10%, 5%, 1%	10%, 5%, 1%
$\omega_{n,m}^{(u)}$	Switching	L	.59, .51, .40	.83, .80, .75	.93, .91, .85
		E	.58, .51, .41	.84, .81, .75	.94, .92, .85
	Bilinear	L	.71, .67, .59	.74, .68, .59	.81, .76, .67
		E	.69, .64, .57	.69, .65, .57	.80, .74, .65
	Quadratic	L	.85, .82, .76	.91, .90, .88	.94, .92, .91
		E	.85, .82, .77	.90, .90, .88	.93, .93, .91
	Logistic	L	.56, .51, .46	.87, .85, .80	.93, .89, .85
		E	.66, .51, .48	.80, .79, .76	.92, .88, .85
$\omega_{n,m}^{(g)}$	Switching	L	.56, .46, .41	.81, .73, .66	.86, .84, .77
		E	.43, .32, .17	.69, .63, .54	.81, .77, .70
	Bilnear	L	.44 .44, .32	.47, .45, .36	.52, .52, .47
		E	.64, .58, .46	.78, .74, .59	.89, .85, .71
	Quadratic	L	.82, .79, .71	.94, .90, .84	.97, .96, .94
		E	.74, .68, .64	.86, .82, .80	.96, .95, .93
	Logistic	L	.64, .56, .48	.81, .77, .74	.93, .91, .90
		E	.51, .41, .21	.76, .71, .68	.91, .89, .80
$\omega_{n,m}^{(s)}$	Switching	L	.52, .47, .31	.80, .78, .70	.89, .88, .85
		E	.41, .32, .20	.77, .68, .64	.86, .84, .76
	Bilnear	L	.48, .45, .34	.52, .48, .40	.69, .58, .49
		E	.68, .61,. .44	.72, .64, .50	.81, .71, .62
	Quadratic	L	.80, .76, .73	.93, .91, .89	.95, .95, .94
		E	.75, .70, .64	.90, .89, .87	.94, .92, .89
	Logistic	L	.51, .41, .34	.87, .82, .79	.93, .90, .88
		E	.37, .31, .17	.79, .77, .74	.90, .89, .86

The flat weighted SWACM and ICM with uniform weight perform similarly. In some cases either may perform better, but the improvement is mild and not across cases. This verifies the ICM concept successfully extends to other nuisance parameter spaces and other test weight classes, and shows great promise and flexibility beyond the original scope presented in Bierens (1982) and Bierens and Ploberger (1997).

Finally, as a separate experiment we compute the logistic SWACM and MP statistics for quadratic and logistic models and sample sizes $n = 20, 40, ..., 500$. See Figure 1 in Appendix C for empirical powers at the 5% level. In the quadratic case SWACM power nearly matches MP power for $n \geq 400$, and is nearly *identical* to MP power for *all* $n \geq 100$ in the logistic case.

5 Conclusion

We exploit seminal results in Bierens (1982, 1990) and Bierens and Ploberger (1997) to create a class of multinomial tests weights and orthogonality conditions with an integer nuisance parameter for testing regression model specifications. We use the weights to create discretely spaced sample moments, and a test statistic that is a stochastically weighted average of those sample moments squared. The stochastic weights are flexible enough to cover a range of statistics from average to max statistics, including statistics that favor any subset of ranked sample moments. The statistic has asymptotic properties identical to the ICM, including the null distribution, consistency, admissibility for Gaussian errors, and optimality with a flat weight. In a controlled experiment the flat weighted SWACM obtained the sharpest size and highest power.

Table 3 - SWACM Test Power (weight $\omega_{n,m}^{(m)}, \omega_{n,m}^{(h)}$)

Weight	Model \ n		100	500	1000
			10%, 5%, 1%	10%, 5%, 1%	10%, 5%, 1%
		L	.57, .48, .35	.83, .81, .73	.89, .88, .82
	Switching	E	.40, .28, .15	.80, .72, .68	.88, .86, .79
		L	.63, .57, .50	.70, .65, .61	.76, .70, .74
$\omega_{n,m}^{(m)}$	Bilinear	E	.76, .75, .69	.83, .80, .75	.89, .85, .78
		L	.78, .77, .71	.92, .91, .91	.95, .95, .94
	Quadratic	E	.75, .67, .63	.91, .90, .85	.95, .93, .92
		L	.60, .51, .49	.85, .83, .78	.92, .90, .88
	Logistic	E	.52, .42, .24	.82, .77, .74	.90, .86, .81
		L	.60, .51, .34	.86, .83, .73	.86, .85, .83
	Switching	E	.36, .25, .12	.73, .64, .54	.82, .78, .69
		L	.66, .62, .57	.70, .64, .56	.75, .70, .62
$\omega_{n,m}^{(h)}$	Bilinear	E	.80, .75, .63	.82, .80, .74	.83, .82, .77
		L	.79, .76, .71	.92, .90, .87	.93, .92, .90
	Quadratic	E	.70, .68, .60	.86, .83, .82	.90, .87, .86
		L	.61, .54, .41	.88, .86, .80	.92, .90, .89
	Logistic	E	.41, .33, .18	.75, .72, .67	.84, .81, .79

Table 4 - ICM[a] Test Size and Power[b]

Model \ n		100	500	1000
		10%, 5%, 1%	10%, 5%, 1%	10%, 5%, 1%
	L	.09, .04, .01	.11, .06, .01	.10, .05, .01
Linear	E	.09, .03, .01	.11, .06, .01	.10, .05, .01
	L	.56, .47, .35	.86, .84, .75	.90, .89, .81
Switching	E	.56, .46, .35	.86, .84, .75	.91, .89, .80
	L	.68, .62, .53	.74, .67, .55	.80, .72, .61
Bilinear	E	.62, .57, .50	.63, .59, .51	.74, .68, .59
	L	.78, .75, .68	.92, .90, .88	.95, .93, .91
Quadratic	E	.78, .75, .68	.91, .89, .87	.94, .93, .90
	L	.67, .59, .47	.86, .85, .77	.92, .90, .84
Logistic	E	.67, .59, .47	.87, .85, .77	.91, .89, .84

a. The argument $\Phi_i(x_t)$ is the arctan of standardized $x_{i,t}$; the measure is $\mu(d\gamma) = d\gamma$; and the parameter space is $\Gamma = [-1, 1]$.

b. Critical value upper bounds are taken from (10) in Section 3.1.

Table 5 - Most Powerful Test Power

Model \ n	100	500	1000
	10%, 5%, 1%	10%, 5%, 1%	10%, 5%, 1%
Switching	.97, .96, .96	.99, .99, .99	.99, .99, .99
Bilinear	.98, .98, .98	.99, .99, .99	1.0, 1.0, 1.0
Quadratic	.94, .93, .91	.95, .95, .94	.97, .96, .96
Logistic	.74, .67, .55	.83, .81, .77	.94, .92, .85

Appendix A: Assumption C

Assumption C1: *$\{y_t, \tilde{x}_t\} \in \mathbb{R} \times \mathbb{R}^k$ is an iid process with non-degenerate absolutely continuous marginal distributions and finite variances. The parameter space Φ is a compact subset of $\mathbb{R}^{k+1}$. $\phi_0 = \arg\inf_{\phi\in\Phi} E(y_t - f(x_t, \phi))^2 \in \text{interior}\{\Phi\}$. $f(\cdot, \phi)$ is for each ϕ Borel measurable, and twice continuously differentiable on Φ.*

Assumption C2: *Let $A_n(\phi) = (1/n)\sum_{t=1}^n (\partial/\partial\phi)f(x_t, \phi)(\partial/\partial\phi')f(x_t, \phi)$, then $A_n(\phi) \to A(\phi)$ uniformly on Φ, where $A(\phi)$ is a non-stochastic positive definite matrix. The NLLS estimator $\hat{\phi} = \arg\min_{\phi\in\Phi} \sum_{t=1}^n (y_t - f(x_t, \phi))^2$ satisfies*

$$\sqrt{n}\left(\hat{\phi} - \phi_0\right) = A(\phi_0)^{-1}\left(\sum_{t=1}^n \frac{\epsilon_t}{\sqrt{n}}\frac{\partial}{\partial\phi}f(x_t, \phi_0) + \frac{1}{n}\sum_{t=1}^n u_t\frac{\partial}{\partial\phi}f(x_t, \phi_0)\right) + o_p(1).$$

Assumption C3: *Write $\hat{b}(m, \phi) = (1/n)\sum_{t=1}^n G_m(\Psi(\tilde{x}_t))(\partial/\partial\phi')f(x_t, \phi) \in \mathbb{N}^k \times \Phi$, where $G_m(\Psi(\tilde{x}_t)) = O_p(\rho^{\bar{m}})$. Then $\hat{b}(m, \phi) \to b(m, \phi)$ uniformly on $\mathbb{N}^k \times \Phi$ where $b(m, \phi)$ is a non-stochastic function satisfying $\sup_{\phi\in\Phi, m\in\mathbb{N}^k} |b(m, \phi)| < \infty$.*

Assumption C4:

i. *$(1/n)\sum_{t=1}^n E[\epsilon_t^2(\partial/\partial\phi)f(x_t, \phi)(\partial/\partial\phi')f(x_t, \phi)] \to A_2$, a finite non-stochastic matrix.*

ii. *u_t is governed by a non-generate distribution, and $\text{plim}_{n\to\infty}(1/n)\sum_{t=1}^n u_t^2$ exists and is constant and finite. There exists a mapping $\eta : \mathbb{N}^k \to \mathbb{R}$ such that $\text{plim}_{n\to\infty}(1/n)\sum_{t=1}^n u_t g(x_t, m) = \lim_{n\to\infty}(1/n)\sum_{t=1}^n E[u_t g(x_t, m)] = \eta(m)$ uniformly on $\mathbb{N}^k$. If $G_m(\Psi(\tilde{x}_t)) = O_p(q(m))$ for some $q : \mathbb{N}^k \to \mathbb{R}$ then $\eta(m) = O(q(m))$.*

iii. *There exists a matrix functional $\Gamma(m_1, m_2)$ on $\mathbb{N}^{k\times k}$ such that $(1/n)\sum_{t=1}^n \{E[\epsilon_t^2|x_t] \times g(x_t, m_1)g(x_t, m_2)\} \to \Gamma(m_1, m_2)$, $\text{plim}_{n\to\infty}(1/n)\sum_{t=1}^n \{\epsilon_t^2 g(x_t, m_1)g(x_t, m_2)\} \to \Gamma(m_1, m_2)$ and $(1/n)\sum_{t=1}^n E[\epsilon_t^2 g(x_t, m_1)g(x_t, m_2)] \to \Gamma(m_1, m_2)$ pointwise on $\mathbb{N}^{k\times k}$. If $G_m(\Psi(\tilde{x}_t)) = O_p(q(m))$ for some $q : \mathbb{N}^k \to \mathbb{R}$ then $\Gamma(m_1, m_2) = O(q(m_1) \times q(m_2))$.*

iv. *Let $G_\xi(\Psi) \in H^+_{G(\Psi)}$. For each $\xi \in \mathbb{R}^k$, $\{\epsilon_t^2 g(x_t, \xi)^2\}$ is uniformly integrable and $\liminf_{n\geq 1} 1/n\sum_{t=1}^n \epsilon_t^2 g(x_t, \xi)^2 \geq \iota$ for some $\iota > 0$. Moreover, $\limsup_{n\to\infty}(1/n)\sum_{t=1}^n E[\epsilon_t^2 \times \sup_{\xi\in[0,u]^k} |(\partial/\partial\xi)g(x_t, \xi)|] < \infty$ for each $u \in [0, \infty)$.*

Appendix B: Proofs of Main Results

PROOF OF LEMMA 3.1. Recall from (6) $z_n(m) = 1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t g(x_t, m)$, and define $g(x_t, \xi, \pi) := \sum_{i=1}^{l}\pi_i g(x_t, \xi_i)$ for arbitrary $\pi \in \mathbb{R}^l$, $\pi'\pi = 1$, $\xi_i \in \mathbb{R}^{k+1}$ and $l \geq 1$. Assumption C, Cramér's Theorem, and the Lindeberg central limit theorem guarantee

$$\sum_{i=1}^{l}\pi_i z_n(\xi_i) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t g(x_t, \xi, \pi) + \frac{1}{n}\sum_{t=1}^{n}u_t g(x_t, \xi, \pi)$$
$$\xrightarrow{d} N\left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}u_t g(x_t, \xi, \pi),\ \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 g(x_t, \xi, \pi)^2\right).$$

The claim follows by invoking the Cramér-Wold Theorem. The bounds $\text{plim}_{n\to\infty} 1/n\sum_{t=1}^{n}u_t g(x_t, \xi) = O(\rho^{\bar{\xi}})$ and $\text{plim}_{n\to\infty} 1/n\sum_{t=1}^{n}\epsilon_t^2 g(x_t, \xi_1) \times g(x_t, \xi_2)$ $= O(\rho^{(\bar{\xi}_1+\bar{\xi}_2)})$ follow from Assumptions A and C. $\mathcal{QED}$.

PROOF OF LEMMA 3.2. It suffices to show $\{1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t g(x_t, \xi)\}_{t=1}^{n}$ is uniformly tight on $[0, \varpi]^k$ for each $\varpi \geq 0$ by straightforward extensions of results in Bickel and Wichura (1971) and Neuhaus (1971). We will apply Lemma A.1 of Bierens and Ploberger (1997). The functional $g(x_t, \xi)$ must satisfy a Lipschitz continuity condition on $[0, \varpi]^k$ for arbitrary ϖ:

$$|g(x_t, \xi_1) - g(x_t, \xi_2)| \leq K_t \times |\xi_1 - \xi_2| \ \ a.s.$$

for every $\xi_1, \xi_2 \in [0, \varpi]^k$, and for some K_t measurable with respect to $\Im_t$ that satisfies $\limsup_{n\to\infty} 1/n\sum_{t=1}^{n}E[\epsilon_t^2 K_t^2] < \infty$. Finally, we need

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}E\left[\epsilon_t^2 g(x_t, \xi_0)\right] < \infty,$$

for one arbitrary $\xi_0 \in [0, \varpi]^k$. All requirements are met under Assumption C by choosing K_t $= \sup_{\xi\in[0,\varpi]^k}|(\partial/\partial\xi)g(x_t, \xi)|$. Since $(\partial/\partial\xi)G_m(\Psi(\tilde{x}_t)) = G_m(\Psi(\tilde{x}_t)) \times \ln\Psi(\tilde{x}_t)$ is bounded $\Im$-$a.e.$ by construction of $G_m(\cdot)$ under Assumption A and $\Psi : \mathbb{R}^k \to (0, K]^k$, it follows

$$\begin{aligned}\left|\frac{\partial}{\partial\xi}g(x_t, \xi)\right| &\leq \left(|G_m(\Psi(\tilde{x}_t))| + \left|b(m, \phi_0)'A(\phi_0)^{-1}\partial f(x_t, \phi_0)\right|\right) \times |\ln\Psi(x_t)| \\ &= O_p\left(\rho^{\bar{\xi}} \times |\ln\Psi(\tilde{x}_t)|\right) = O_p(\rho^{\bar{\xi}}).\ \mathcal{QED}.\end{aligned}$$

PROOF OF THEOREM 3.3. Using the non-stochastic limiting sequence $\{\omega_m\}$, apply expansion (6), Lemmas 3.1 and 3.2, and the continuous mapping theorem to verify under Assumption C

$$\sum_{\bar{m}=1}^{\infty}\hat{z}(m)^2\omega_m \Rightarrow \sum_{\bar{m}=1}^{\infty}z(m)^2\omega_m.$$

By (6) it therefore suffices to prove $|\sum_{\bar{m}=1}^{\mathcal{N}_n} z_n(m)^2\omega_{n,m} - \sum_{\bar{m}=1}^{\infty} z_n(m)^2\omega_m| = o_p(1)$. By the triangle inequality and $\omega_m \geq 0$

$$\begin{aligned}\left|\sum_{\bar{m}=1}^{\mathcal{N}_n} z_n(m)^2\omega_{n,m} - \sum_{\bar{m}=1}^{\infty} z_n(m)^2\omega_m\right| &\leq \left|\sum_{\bar{m}=1}^{\mathcal{N}_n} z_n(m)^2\left\{\omega_{n,m} - \omega_m\right\}\right| + \sum_{\bar{m}=\mathcal{N}_n+1}^{\infty} z_n(m)^2\omega_m \\ &= \mathcal{A}_{1,n} + \mathcal{A}_{2,n}.\end{aligned}$$

Assumptions A-C and $\mathcal{N}_n \to \infty$ imply

$$E\left[\mathcal{A}_{2,n}\right] = \sum_{\bar{m}=\mathcal{N}_n+1}^{\infty} E\left[z_n(m)^2\right]\omega_m \leq K\sum_{\bar{m}=\mathcal{N}_n+1}^{\infty} \rho^{2\bar{m}}\omega_m \leq K\rho^{\mathcal{N}_n} \to 0,$$

hence by Chebyshev's inequality $\mathcal{A}_{2,n} \xrightarrow{p} 0$.

For $\mathcal{A}_{1,n}$ observe

$$\left|\sum_{\bar{m}=1}^{\mathcal{N}_n} z_n(m)^2\left\{\omega_{n,m} - \omega_m\right\}\right| \leq \sum_{\bar{m}=1}^{\mathcal{N}_n} \rho^{-\bar{m}} z_n(m)^2 \times \sum_{\bar{m}=1}^{\mathcal{N}_n} \rho^{\bar{m}}\left|\omega_{n,m} - \omega_m\right| = \mathcal{B}_{1,n} \times \mathcal{B}_{2,n}.$$

Assumption B states $\mathcal{B}_{2,n} \xrightarrow{p} 0$. By Lemmas 3.1 and 3.2 $\mathcal{B}_{1,n} \Rightarrow \sum_{\bar{m}=1}^{\infty} \rho^{-\bar{m}} z(m)^2$ where $\sum_{\bar{m}=1}^{\infty} \rho^{-\bar{m}} E[z(m)^2] \leq K\sum_{\bar{m}=1}^{\infty} \rho^{\bar{m}} < \infty$. Therefore $\mathcal{A}_{1,n} \xrightarrow{p} 0$ which completes the proof. $\mathcal{QED}$.

PROOF OF THEOREM 3.4. Denote by $\mathcal{H} = (\mathcal{H}, ||\cdot||_\omega)$ the inner product space of sequences $y = \{y(\xi_i)\}_{i=1}^{\infty}$ of continuous functions $y(\xi) \in C[0,\infty)^k$ with bound $y(\xi) = O(\rho^{\bar{\xi}})$, metrized with $||y||_\omega = (\sum_{\bar{m}=1}^{\infty} y(m)^2\omega_m)^{1/2}$. Let $\langle x, y\rangle_\omega = \sum_{\bar{m}=1}^{\infty} x(m)y(m)\omega_m$ be the supporting inner product. Then $\mathcal{H}$ is separable and complete[3], hence a separable Hilbert space. Separable inner product spaces have countably infinite orthonormal basis, say $\{\vartheta_i(m)\}_{i=1}^{\infty}$, $\sum_{m\in\mathbb{N}^k} \vartheta_i(m)\vartheta_j(m)\omega_m = I_{i=j}$ (e.g. Giles, 2000: Theorem 3.27).

Now define Fourier coefficients

$$w_i := \sum_{\bar{m}=1}^{\infty} z(m)\vartheta_i(m)\omega_m. \tag{11}$$

Then $z \in \mathcal{H}$ admits a coordinate-wise expansion

$$z(m) = \sum_{i=1}^{\infty} \vartheta_i(m)w_i. \tag{12}$$

Because $\Gamma(m_1, m_2) = \text{plim}_{n\to\infty} 1/n\sum_{t=1}^{n} \epsilon_t^2 g(x_t, m_1)g(x_t, m_2)$ is a symmetric positive-semi-definite $O(\rho^{(\bar{m}_1+\bar{m}_2)})$-bounded function under Lemma 3.1, it follows that $\Gamma = (\Gamma(m_1, m_2))_{m_1,m_2\in\mathbb{N}^k}$ is a linear compact self-adjoint operator (Giles, 2000: §15). By the spectral theorem for compact self-adjoint operators on a Hilbert space there exists an orthonormal basis of $\mathcal{H}$ consisting of eigenfunctions of Γ, where each eigenvalue λ_i is real and

[3] It is straightforward to show Davidson's (1994: Theorem 5:15) argument carries over to $\mathcal{H}$ due to boundedness $y(\xi) = O(2^{-\bar{\xi}})$ of every $y \in \mathcal{H}$.

non-negative (Giles, 2000: Theorem 20.4.1). It is immediate that $\{\vartheta_i(m)\}_{i=1}^{\infty}$ denotes the eigenfunctions of Γ,

$$\sum_{\bar{m}_2=1}^{\infty} \Gamma(m_1, m_2)\vartheta_i(m_2)\omega_{m_2} = \lambda_i \vartheta_i(m_1),$$

and $\Gamma(m_1, m_2)$ obtains the series representation $\Gamma(m_1, m_2) = \sum_{i=1}^{\infty} \lambda_i \vartheta_i(m_1)\vartheta_i(m_2)$. Use Parseval's identity, (11) and (12) and orthonormality to get

$$T_1 = \sum_{\bar{m}=1}^{\infty} z(m)^2 \omega_m = \sum_{i=1}^{\infty} w_i^2.$$

Each $z(m)$ under H_1^L is mean zero Gaussian by Lemma 3.1, therefore each Fourier coefficient $w_i = \sum_{\bar{m}=1}^{\infty} z(m)\vartheta_i(m)\omega_m$ is Gaussian, completely characterized by means $\eta_i = E[w_i] = \sum_{\bar{m}=1}^{\infty} \eta(m)\vartheta_i(m)\omega_m$ and pair-wise covariances

$$E\left[\left(\sum_{\bar{m}=1}^{\infty} [z(m) - \eta(m)]\, \vartheta_i(m)\omega_m\right)\left(\sum_{\bar{m}=1}^{\infty} [z(m) - \eta(m)]\, \vartheta_j(m)\omega_m\right)\right]$$
$$= \sum_{\bar{m}_1=1}^{\infty} \sum_{\bar{m}_2=1}^{\infty} \Gamma(m_1, m_2)\vartheta_i(m_1)\vartheta_j(m_2)\omega_{m_1}\omega_{m_2} = \lambda_i I_{i=j}.$$

Thus $w_i \overset{iid}{\sim} N(\eta_i, \lambda_i)$ which completes the proof. $\mathcal{QED}$.

PROOF OF COROLLARY 3.6. Since the argument simply mimics the proof of Theorem 3.4, we only present a sketch. Write $\omega_m = \omega_m^{(m)}(\pi)$.

Exploit $\omega_m = I(|\eta(m)| \geq \eta_{(1)})$ to deduce the following. First, $\sum_{m \in \mathbb{N}^k} \vartheta_i(m)\vartheta_j(m)\omega_m = \vartheta_i(m^*)\vartheta_j(m^*) = 1$ if $i = j$ and 0 otherwise; hence $\vartheta_1(m^*) = 1$ and $\vartheta_i(m^*) = 0\ \forall i \geq 2$; hence the Fourier coefficients $w_i = \sum_{\bar{m}=1}^{\infty} z(m)\vartheta_i(m)\omega_m$ are $w_1 := z(m^*)$ and $w_i = 0$ *a.s.* $\forall i \geq 2$.

Second, $\Gamma(m_1, m^*)\vartheta_i(m^*) = \lambda_i \vartheta_i(m_1)$, hence $\lambda_1 = \Gamma(m^*, m^*)$ and $\lambda_i = 0\ \forall i \geq 2$.

Third, use Parseval's identity and orthonormality to obtain the trivial identity $T_1 = \sum_{\bar{m}=1}^{\infty} z(m)^2 \omega_m = \sum_{i=1}^{\infty} w_i^2 = z(m^*)^2 = \max_{m \in \mathbb{N}^k} \{z(m)^2\}$.

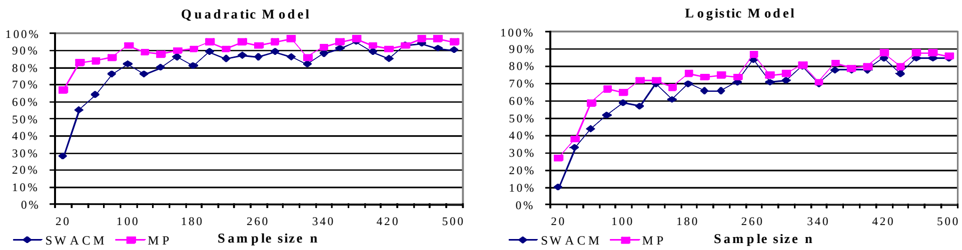
Fourth, the Fourier coefficients $w_i = \sum_{\bar{m}=1}^{\infty} z(m)\vartheta_i(m)\omega_m$ are Gaussian, completely characterized by means $\eta_i = E[w_i] = \sum_{\bar{m}=1}^{\infty} \eta(m)\vartheta_i(m)\omega_m = \eta(m^*)\vartheta_i(m^*) = \eta(m^*)$ if $i = 1$ and 0 $\forall i \geq 2$, and pair-wise covariances that reduce to $\Gamma(m^*, m^*)\vartheta_i(m^*)\vartheta_j(m^*) = 0$ if $i \neq j$, $i = j \neq 1$ and λ_1 otherwise.

Therefore $w_1 \overset{iid}{\sim} N(\eta(m^*), \lambda_1)$ and $w_i = 0$ *a.s.* $\forall i \geq 2$. Coupled with $T_1 = \sum_{i=1}^{\infty} w_i^2$ this completes the proof under H_1^L. Under H_1^G we have $\hat{z}(m)/n^{1/2} \overset{p}{\to} \eta(m)$ hence by (4) $\max_{1 \leq \bar{m} \leq \mathcal{N}_n} \{\hat{z}(m)^2\} \overset{p}{\to} \infty$. $\mathcal{QED}$.

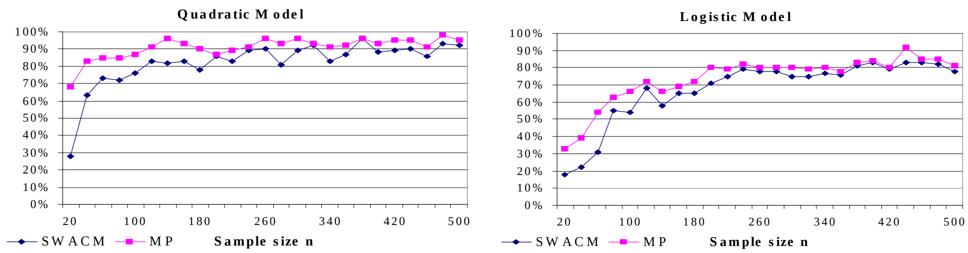
Appendix C

Figure 1: Logistic SWACM Power at 5% level

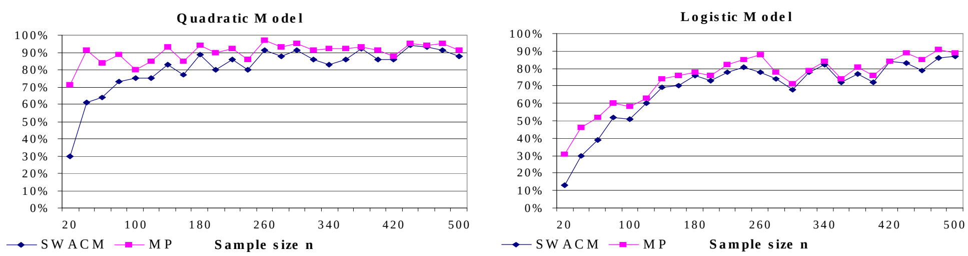
Uniform/Flat $\omega_{n,m}^{(u)}$



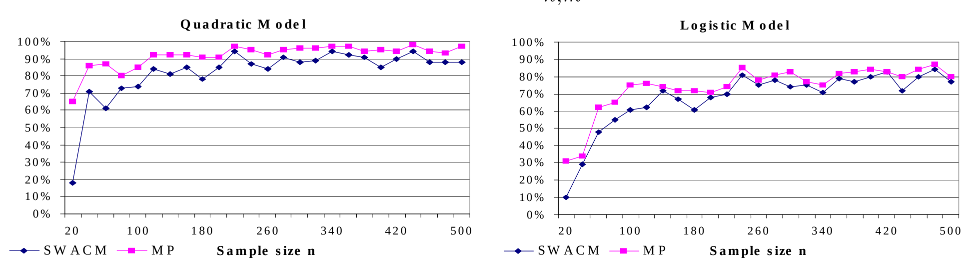
Simple Geometric $\omega_{n,m}^{(g)}$



Upper Quantile $\omega_{n,m}^{(h)}$



Near-Max $\omega_{n,m}^{(m)}$



References

[1] Andrews, D.W.K. and W. Ploberger (1994): "Optimal Tests when a Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 82, 1383-1414.

[2] Bickel, P.J. and M.J. Wichura (1971): "Convergence Criteria for Multiparameter Stochastic Processes and Some Applications", *Annals of Mathematical Statistics*, 42, 1656-1670.

[3] Bierens, H.J. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105-134.

[4] Bierens, H.J. (1984): "Model Specification Testing of Time Series Regressions," *Journal of Econometrics*, 26, 323-353.

[5] Bierens, H.J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443-1458.

[6] Bierens, H.J. (1991): "Least Squares Estimation of Linear and Nonlinear ARMAX Models Under Data Heterogeneity," *Annales d'Economie et de Statistique*, 20/21, 143-169.

[7] Bierens, H.J. (1994): "Comment on *Artificial Neural Networks: An Econometric Perspective*," by C-M. Kuan and H. White, *Econometric Reviews*, 13, 93-97.

[8] Bierens, H.J. and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129-1151.

[9] Bierens, H.J. and L. Wang (2011): "Integrated Conditional Moment Tests for Parametric Conditional Distributions," *Econometric Theory*: forthcoming.

[10] Billingsley, P. (1999): *Convergence of Probability Measures*: John Wiley & Sons: New York.

[11] Blake, A.P. and G. Kapetanios (2007): "Testing for Neglected Nonlinearity in Cointegrating Relationships," *Journal of Time Series Analysis*, 28, 807-826.

[12] Boning, B. Wm. and F. Sowell (1999): "Optimality for the Integrated Conditional Moment Test," *Econometric Theory*, 15, 710-718.

[13] Davies, R.B. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 64, 247-254.

[14] de Jong, R. (1996): "The Bierens Test under Data Dependence," *Journal of Econometrics*, 72, 1-32.

[15] de Jong, R. and H.J. Bierens (1994): "On the Limit Behavior of a Chi-Squared Type Test if the Number of Conditional Moments Tested Approaches Infinity," *Econometric Theory*, 9, 70-90.

[16] Dette, H. (1999): "A Consistent Test for the Functional Form of a Regression Based on a Difference of Variance Estimators," *Annals of Statistics*, 27, 1012-40.

[17] Fan, Y. and Q. Li (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64, 865-890.

[18] Fan, Y. and Q. Li (2000): "Consistent Model Specification Tests: Kernel-Based Tests versus Bierens' ICM Tests," *Econometric Theory*, 16, 1016-1041.

[19] Giacomini, R. and H. White (2006): "Tests of Conditional Predictive Ability," *Econometrica*, 74, 1545-1578.

[20] Giles, J.R. (2000): *Introduction to the Analysis of Normed Linear Spaces*: Cambridge University Press, U.K.

[21] Hansen, B. (1996): "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413-430.

[22] Härdle, W. and E. Mammen (1993): "Comparing Nonparametric Versus Parametric

Regression Fits," *Annals of Statistics*, 21, 1926-1947.

[23] Hill, J.B. (2008a): "Consistent and Non-Degenerate Model Specification Tests Against Smooth Transition and Neural Network Alternatives," *Annale's d'Economie et de Statistique*, 90, 145-179.

[24] Hill, J.B. (2008b): "Consistent GMM Residuals-Based Tests of Functional Form," *Econometric Reviews*: forthcoming.

[25] Hill, J.B. (2012): "Heavy-Tail and Plug-In Robust Consistent Conditional Moment Tests of Functional Form," to appear in X. Chen and N. Swanson (ed.'s), *Festschrift in Honor of Hal White*: Springer, New York.

[26] Holley, A. (1982): "A Remark on Hausman's Specification Test," *Econometrica*, 50, 749-760.

[27] Hong, Y. and H. White (1995): "Consistent Specification Testing via Nonparametric Series Regression," *Econometrica*, 63, 1133-1159.

[28] Hong, Y. and Y-J Lee (2005): "Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroscedasticity of Unknown Form," *Review of Economic Studies*, 72, 499-541.

[29] Hornik, K., M. Stinchcombe and H. White (1989): "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, 2, 359-366.

[30] Huber, P J. (1977): *Robust Statistical Procedures*: Society for Industrial and Applied Mathematics, Philadelphia.

[31] Koul, H.L. and W. Stute (1999): "Nonparametric Model Checks for Time Series," *Annals of Statistics*, 27, 204-236.

[32] Leadbetter, M.R., G. Lindgren and H. Rootzén (1983): *Extremes and Related Properties of Random Sequences and Processes*: Springer-Verlag, New York.

[33] Lee T., H. White H., C.W.J. Granger (1993): "Testing for Neglected Nonlinearity in Time-Series Models: A Comparison of Neural Network Methods and Alternative Tests," *Journal of Econometrics*, 56, 269-290.

[34] Li, Q., C. Hsiao, and J. Zinn (2003): "Consistent Specification Tests for Semiparametric/Nonparametric Models Based on Series Estimation Methods," *Journal of Econometrics*, 112, 295-325.

[35] Neuhaus, G. (1971): "On Weak Convergence of Stochastic Processes with Multidimensional Time Parameter," *Annals of Mathematical Statistics*, 42, 1285-1295.

[36] Newey, W.K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047-1070.

[37] Ramsey, J.B. (1969): "Tests for Specification Errors in a Classical Linear Least-Squares Regression Analysis," *Journal of the Royal Statistical Society, Series B*, 31, 350-371.

[38] Stinchcombe, M.B. and H. White (1998): "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative," *Econometric Theory*, 14, 295-325.

[39] Stute, W. (1997): "Nonparametric Model Checks for Regression," *Annals of Statistics*, 35, 613-641.

[40] Stute, W. and L-X. Zhu (2005): "Nonparametric Checks for Single-Index Models," *Annals of Statistics*, 33, 1048-1083.

[41] Teräsvirta, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition

Autoregressive Models," *Journal of the American Statistical Association*, 89, 208-218.

[42] White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-2.

[43] White, H. (1989): "An Additional Hidden Unit Test of Neglected Nonlinearity in Multilayer Feedforward Networks, in *Proceedings of the International Joint Conference on Neural Networks*, Washington D.C., Vol. 2. IEEE Press, New York.

[44] Yatchew, A.J. (1992): "Nonparametric Regression Tests Based on Least Squares," *Econometric Theory*, 8, 435-451.

[45] Zheng, J. (1996): "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263-289.