Econometric Reviews
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lecr20

Consistent GMM Residuals-Based Tests of Functional Form

Jonathan B. Hill [a]
[a] Department of Economics, University of North Carolina-Chapel Hill, Chapel Hill, North Carolina, USA
Accepted author version posted online: 30 Jul 2012.Published online: 25 Oct 2012.

To cite this article: Jonathan B. Hill (2013) Consistent GMM Residuals-Based Tests of Functional Form, Econometric Reviews, 32:3, 361-383, DOI: 10.1080/07474938.2012.690662

To link to this article: http://dx.doi.org/10.1080/07474938.2012.690662

Econometric Reviews, 32(3):361–383, 2013

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474938.2012.690662

CONSISTENT GMM RESIDUALS-BASED TESTS OF FUNCTIONAL FORM

Jonathan B. Hill

Department of Economics, University of North Carolina–Chapel Hill, Chapel Hill, North Carolina, USA

□ *This paper presents a consistent Generalized Method of Moments (GMM) residuals-based test of functional form for time series models. By relating two moments we deliver a vector moment condition in which at least one element must be nonzero if the model is misspecified. The test will never fail to detect misspecification of any form for large samples, and is asymptotically chi-squared under the null, allowing for fast and simple inference. A simulation study reveals randomly selecting the nuisance parameter leads to more power than supremum-tests, and can obtain empirical power nearly equivalent to the most powerful test for even relatively small n.*

Keywords Consistent test; Conditional moment test; GMM; Nonlinear model.

JEL Classification C12; C45; C52.

1. INTRODUCTION

This paper develops a consistent parametric Conditional Moment (CM) test of regression model functional form within the class of Nonlinear ARX models. We interact two moment conditions in a way such that at least one must be nonzero under misspecification. A score test based on sample versions of the moments does more than obtain asymptotic power of one under misspecification, since this permits asymptotically countably infinitely many possibilities for test failure. Rather, it is asymptotically perfect because there is *no* possibility for test failure. Since the test is based on regression residuals the theory is designed with Generalized Method of Moments (GMM) in mind.

Let $\{y_t, x_t\}$ be the data of interest, x_t is a k-vector of regressors including a constant term, and may contain lags of y_t and other observable random

Address correspondence to Jonathan B. Hill, Department of Economics, University of North Carolina–Chapel Hill, Chapel Hill, NC 27599, USA; E-mail: jbhill@email.unc.edu

variables. The information set at time t is defined as the σ-field $\Im_t := \sigma(\{y_\tau, x_{\tau+1}\} : \tau \leq t)$ assumed to be increasing $\Im_{t-1} \subset \Im_t$. We want to test if some $\Im_{t-1}$-measurable parametric function $f_t(\phi)$, $\phi \in \mathbb{R}^p$, $p \geq 1$, is a version of the conditional mean $E[y_t|\Im_{t-1}]$: for some unique ϕ_0,

$$H_0 : y_t = f_t(\phi_0) + \epsilon_t, \text{ where } E[\epsilon_t \mid \Im_{t-1}] = 0 \text{ with probability one.}$$

The alternative H_1 is simply that H_0 is false uniformly on Φ, covering any direction from H_0:

$$H_1 : \sup_{\phi \in \Phi} P\left(\left|f_t(\phi) = E[y_t \mid \Im_{t-1}]\right|\right) < 1.$$

The model $y_t = f_t(\phi_0) + \epsilon_t$ falls within the class of Nonlinear AR with auxiliary variables (ARX), including linear ARX (Baillie, 1980; Bierens, 1987), Threshold AR (Tong, 1990), Smooth Transition ARX (Luukkonen et al., 1988; Hill, 2008; Teräsvirta, 1994), Artificial Neural Networks with a finite number of hidden layers (Hornik et al., 1989; White, 1989), and Markov Switching if the transition probabilities are $\Im_{t-1}$-measurable (e.g., Filardo, 1994). We allow geometric or hyperbolic memory decay (e.g., Fractional AR); bounded forms of nonstationarity like seasonality, and stochastic breaks; and random volatility errors ϵ_t like Generalized Autoregressive Conditional Heteroskedasticity (GARCH).

Although classic specification tests are known not to be consistent (e.g., Ramsey, 1969; Hausman, 1978; White, 1981; Newey, 1985), a variety of parametric and nonparametric methods promote consistency. Parametric CM tests do so by exploiting scalar weight functions $F(\gamma' x_t)$ indexed by a nuisance vector $\gamma \in \mathbb{R}^k$, effectively producing uncountably infinitely many conditions that "reveal" misspecification. If $F(u)$ is infinitely differentiable and nonpolynomial, then under misspecification $E[\epsilon_t \mid \Im_{t-1}] \neq 0$ the moments $E[\epsilon_t F(\gamma' x_t)]$ are *revealing* in the sense $E[\epsilon_t F(\gamma' x_t)] \neq 0$ for uncountably infinitely many γ in any compact subset $\Gamma \subset \mathbb{R}^k$ (Bierens and Ploberger, 1997; Stinchcombe and White, 1998, cf. Bierens (1982, 1990)). Examples include weights popularized in the (ANN) literature like the exponential $F(\gamma' x_t) = \exp\{\gamma' x_t\}$ (Bierens, 1982, 1990; Hornik et al., 1989) and logistic $F(\gamma' x_t) = [1 + \exp\{\gamma' x_t\}]^{-1}$ (Hornik et al., 1989; White, 1989; Lee et al., 1993); and compound weights in the Smooth Transition Autoregression (STAR) literature like $x_t \exp\{-\tau'(x_t - c)^2\}$ and $x_t[1 + \exp\{-\tau'(x_t - c)\}]^{-1}$ where $\tau \in \mathbb{R}^k_+$ and $c \in \mathbb{R}^k$ (Hill, 2008). See Andrews and Ploberger (1994) and Dette (1999) for related methods. Standard methods for handling γ include randomization (Bierens, 1990; Lee et al., 1993), and power-optimizing supremum and average test functionals (Bierens, 1987, 1990; Andrews and Ploberger, 1994; Hill, 2008), and integrated conditional moments (Bierens, 1982; Bierens and Ploberger, 1997).

Despite the revealing properties of such weights $F(\gamma' x_t)$, the set of γ such that $E[\epsilon_t F(\gamma' x_t)]$ fails to reveal misspecification is not necessarily empty. Indeed, there may be countably infinitely many "bad" $\gamma \in \mathbb{R}^k$: orthogonality $E[\epsilon_t F(\gamma' x_t)] = 0$ despite misspecification. We show in Section 2 by a numerical example that $E[\epsilon_t F(\gamma' x_t)] = 0$ under H_1 is not an abstraction, and a score test may fail in a region near such bad γ. The latter suggests that not even a sup- or ave-test, or Integrated Conditional Moment (ICM) test, may solve the problem of bad $\gamma' s$ in small samples, depending on the choice of Γ.

In this paper we relate parametric moments in a way that *always* reveals model misspecification, and hence we do not require a power-optimizing functional transformation. For example, stacking moments based on Bierens' (1990) $\exp\{\gamma' x_t\}$ and Hill's (2008) $x_t \exp\{\xi' x_t\}$ for some estimable $\xi \in \Gamma$ leads to a perfect vector moment condition for any $\gamma \neq 0$. A small scale Monte Carlo study verifies the merits of the proposed test against existing parametric and nonparametric CM tests.

There are now more tests of functional form in the time series literature than can be reasonably cited. Examples include nonconsistent tests that exploit a Volterra expansion (Tsay, 1986) or the bispectral density function (Rao and Gabr, 1980) or entropy (Maasoumi et al., 2004); nonconsistent F-tests of linearity against smooth transition forms (Luukkonen et al., 1988); consistent tests based of variance differentials (Dette, 1999), with mixtures of discrete and continuous data (Hsiao et al., 2007), for out-of-sample predictive power (Corradi and Swanson, 2002), and that exploit bootstrap semi-parametric methods (Li and Wang, 1998). Consistent nonparametric and semiparametric techniques are also widely available, including Yatchew (1992), Fan and Gencay (1995), Fan and Li (1996, 2002), Hong and White (1995), Zheng (1996), Stute (1997), and Koul and Stute (1999) to name a few. The latter cannot fail asymptotically to reveal misspecification, but require a non-parametric plug-in which will in general affect small sample performance.

The remainder of the paper proceeds as follows. In Section 2 we build the intuition behind our test, construct the statistic and characterize its limit. A simulation study is presented in Section 3, and concluding remarks follow.

Throughout, $\stackrel{p}{\rightarrow}$ and $\stackrel{d}{\rightarrow}$ denote convergence in probability and distribution, $\Rightarrow$ denotes weak convergence on a metric space, and $\rightarrow$ denotes convergence in l_1-norm $|\cdot|$. $\|\cdot\|_p$ denotes the L_p-matrix norm: $\|x\|_p := (\sum_{i,j} E|x_{i,j}|^p)^{1/p}$. K is a finite, positive constant whose value may change from line to line.

2. TEST OF FUNCTIONAL FORM

In this section, we flesh out model details, provide the intuition behind relating moments, and construct the test statistic. Our first task is to develop ARX regressors, estimating equations, and the test weight.

2.1. Nonlinear ARX Model

Let $\{y_t, \tilde{x}_t\} = \{y_t, \tilde{x}_t\}_{t\in\mathbb{Z}} \in \mathbb{R} \times \mathbb{R}^{k-1}$, $k \geq 2$, be a stochastic process with nondegenerate continuous marginal distributions and finite variances. The regressor set $x_t := [1, \tilde{x}_t']'$ contains only observable data for brevity, so the Autoregressive Moving Average (ARMA) class is not included here, although the Nonlinear-ARMAX (Autoregressive Moving Average with Auxiliary Variables) class (Bierens, 1987) can be included by a straightforward extension of de Jong (1996). Notice for $F(\gamma' x_t)$ to be revealing, $\gamma' x_t$ must be affine, and there must be at least one stochastic regressor; hence a constant is included and $k = 1$ is ruled out (Bierens, 1990; Stinchcombe and White, 1998). We simplify sample truncation due to lags by assuming $\{x_t\}_{t=1}^n$ is observed by the econometrician.

We require $\{y_t, \tilde{x}_t\}$ to be L_2-Near Epoch Dependent (L_2-NED) on a strong mixing base $\{\epsilon_t\}$ in order to exploit weak limit theory in Hill (2008). Mixing errors allow for GARCH and stochastic volatility errors (e.g., Carrasco and Chen, 2002), any strong mixing $\{y_t, \tilde{x}_t\}$ is automatically L_2-NED on itself, and in general NED captures a broad array of linear and nonlinear time series with geometric or hyperbolic memory decay, including threshold-like models (e.g., An and Huang, 1996) and a variety of nonlinear Autoregression with Generalized Autoregressive Conditional Heteroskedasticity (AR-GARCH) (e.g., Meitz and Saikkonen, 2008). See Carrasco and Chen (1988), Nze and Doukhan (2004), Hill (2008), and their references. Since all assumptions required for the asymptotic theory are identical or similar to those in de Jong (1996) and Hill (2008), we detail them under Assumption 3 in the supplementary appendix Hill (2011).

The known stochastic response function $f_t : \Phi \to \mathbb{R}$ is almost surely twice continuously differentiable on Φ, a compact Euclidean subset of $\mathbb{R}^p$. The proposed test is based on regression residuals, so we use to estimate ϕ with $\Im_t$-measurable estimating equations $m_t(\phi) \in \mathbb{R}^q$, $q \geq p$. We assume for convenience $m_t(\phi)$ has the form

$$m_t(\phi) = (y_t - f_t(\phi)) \times c_t(\phi), \qquad c_t : \Phi \to \mathbb{R}^q, \quad q \geq p, \tag{1}$$

where $c_t(\phi)$ is $\Im_{t-1}$-measurable and almost surely differentiable on Φ, covering Nonlinear Least Squares (NLLS) while allowing for over-identifying restrictions. A more general GMM framework is a simple extension. The global identification condition

$$E[m_t(\phi)] = 0 \quad \text{if and only if } \phi = \phi_0, \quad \text{a unique interior point of } \Phi,$$

imposes weak orthogonality on the regression error ϵ_t under either hypothesis, but allows for misspecified models under H_1.

The proposed test exploits the general weight class used in Bierens and Ploberger (1997) and Stinchcombe and White (1998). Write $F'(u) := (\partial/\partial u)F(u)$.

Assumption 1. *Assume* $F : \mathbb{R} \to \mathbb{R}$ *is analytic and nonpolynomial on some open interval* $R_0 \subseteq \mathbb{R}$ *containing* 0. *In particular,* $(\partial/\partial u)^i F(u)|_{u=0} = 0$ *for only finitely many* $i \in \mathbb{N}$, $F(0'x_t) = c$ *and* $F'(0'x_t) = d$ *for finite nonstochastic* c *and* d.

Assumption 2. *The stochastic regressors* $\tilde{x}_t \in \mathbb{R}^{k-1}$, $k \geq 2$, *are bounded with probability one.*

Remark 1. Examples of weights F under Assumption 1 include the exponential $\exp\{u\}$, logistic $[1 + \exp\{u\}]^{-1}$, and trigonometric functions like $\sin(u)$ and $\cos(u)$ (Stinchcombe and White, 1998).

Remark 2. The stochastic regressors $\tilde{x}_t$ must be bounded to ensure $\{F(\gamma' x_t) : \gamma \in \Gamma\}$ is weakly dense on the space on which x_t lies (Stinchcombe and White, 1998). If $\tilde{x}_t$ is not bounded, we may use $\Psi(x_t)$ for any bounded, one-to-one measurable mapping $\Psi : \mathbb{R}^k \to \mathbb{R}^k$ (cf. Bierens, 1982, 1990).

2.2. Interacting Moment Condition Intuition

Consider any weight function F satisfying Assumption 1, and define the moment function

$$\mu(\gamma) := E[\epsilon_t F(\gamma' x_t)].$$

Assume $\mu(\gamma)$ is differentiable on an arbitrary compact subset Γ, and assume misspecification H_1 applies. By inspecting $\mu(\gamma)$ and the largest value of $(\partial/\partial\gamma_i)\mu(\gamma)$ in each ith-direction we can always correctly deduce H_1 holds. We only discuss the intuition here. In Appendix A, we formally state the result, and since the proof follows easily from extant theory we relegate it to Hill (2011). The following discussion freely exploits theory developed in Bierens (1990), Bierens and Ploberger (1997), and Stinchcombe and White (1998).

Suppose $\mu(\gamma_0) = 0$ for some $\gamma_0 \in \Gamma$. Then by the revealing nature of $F(\gamma' x_t)$ there exists a neighborhood $N(\gamma_0)$ of γ_0 such that $\mu(\gamma) \neq 0$ *for all* $\gamma \in N(\gamma_0)/\gamma_0$. But this means $\mu(\gamma)$ at γ_0 has a nonzero sth derivative $(\partial/\partial\gamma)^s \mu(\gamma)|_{\gamma=\gamma_0} \neq 0$ for some $s \in \mathbb{N}$. Intuitively, $\mu(\gamma)$ cannot be a constant function on Γ if any one $\mu(\gamma_0) = 0$ since otherwise $F(\gamma' x_t)$ cannot be revealing. Similarly, $\mu(\gamma)$ fails at the origin by construction $\mu(0) = K \times E[\epsilon_t] = 0$. By the same reasoning, the moment function must therefore move in some direction from the origin.

In summary, the moment $\mu(\gamma) = 0$ at $\gamma = 0$ and all "bad" γ_0, and $\mu(\gamma)$ is not a constant function in neighborhoods of 0 and such γ_0. This simply implies $(\partial/\partial\gamma)\mu(\gamma) \neq 0$ somewhere on Γ, even though $(\partial/\partial\gamma)\mu(\gamma) = 0$ is possible at $\gamma = 0$ or γ_0. Indeed, since $\mu(0) = 0$ and $\mu(\gamma)$ is not a constant function near bad γ_0, as long as $\mu(\gamma_0) = 0$ for any other point $\gamma_0 \neq 0$ it must be the case that $(\partial/\partial\gamma_i)\mu(\gamma) > 0$ *and* $(\partial/\partial\gamma_i)\mu(\gamma) < 0$ somewhere on Γ for some ith direction. Therefore, $\sup_{\gamma\in\Gamma}(\partial/\partial\gamma_i)\mu(\gamma) > 0$ *and* $\inf_{\gamma\in\Gamma}(\partial/\partial\gamma_i)\mu(\gamma) < 0$.

Now let $\xi^{(+)}_{(i)}$ optimize the derivative in each ith-direction $\xi^{(+)}_{(i)} = \arg\sup_{\gamma\in\Gamma}(\partial/\partial\gamma_i)\mu(\gamma) \in \mathbb{R}^k$, and suppose in all directions $(\partial/\partial\gamma_i)\mu(\gamma)|_{\gamma=\xi^{(+)}_{(i)}} = 0$. In this case, the moment function $\mu(\gamma)$ is flat or decreasing. Since at the origin $\mu(0) = 0$, and $\mu(\gamma)$ moves away from zero in some direction from the origin, it must be the case that $\mu(\gamma)$ *never* returns to zero. It follows $\mu(\gamma) \neq 0$ for every nonzero $\gamma \in \Gamma$.

There are three key observations from the preceding. First, by working with revealing $F(\gamma' x_t)$, under H_1, we are guaranteed $\mu(\gamma) \neq 0$ for all nonzero $\gamma \in \Gamma$ and/or $\sup_{\gamma\in\Gamma}(\partial/\partial\gamma_i)\mu(\gamma) \neq 0$ in some direction $i \in \{1, \ldots, k\}$.

Second, the relationship does not necessarily hold if we inspect the *absolute magnitude* $|(\partial/\partial\gamma_i)\mu(\gamma)|$. Consider Bierens' $F(\gamma' x_t) = \exp\{\gamma' x_t\}$. Then $\sup_{\gamma\in\Gamma}|(\partial/\partial\gamma_i)\mu(\gamma)| = \sup_{\gamma\in\Gamma}|E[\epsilon_t x_{j,t}\exp\{\gamma' x_t\}]| > 0$ is easy to prove under H_1 (Bierens, 1990). The above argument, however, intimately exploits whether $\sup_{\gamma\in\Gamma}(\partial/\partial\gamma_i)\mu(\gamma) = 0$ since this tells us if the moment function $\mu(\gamma)$ can return to zero over $\Gamma/0$.

Third, the argument is not unique since the same conclusion follows if we use $\xi^{(-)}_{(i)} = \arg\inf_{\gamma\in\Gamma}(\partial/\partial\gamma_i)\mu(\gamma)$, or both $(\partial/\partial\gamma_i)\mu(\gamma)|_{\gamma=\xi^{(+)}_{(i)}}$ and $\partial/\partial\gamma_i)\mu(\gamma)|_{\gamma=\xi^{(-)}_{(i)}}$ simultaneously. Thus if $(\partial/\partial\gamma_i)\mu(\gamma)|_{\gamma=\xi^{(+)}_{(i)}} = 0$ *or* $(\partial/\partial\gamma_i)\mu(\gamma)|_{\gamma=\xi^{(-)}_{(i)}} = 0$ for each i, we know $\mu(\gamma) \neq 0$ for every $\gamma \in \Gamma$ under H_1, and if $\mu(\gamma_0) = 0$ for some $\gamma_0 \neq 0$, then $(\partial/\partial\gamma_i)\mu(\gamma)|_{\gamma=\xi^{(+)}_{(i)}} \neq 0$ *and* $(\partial/\partial\gamma_i)\mu(\gamma)|_{\gamma=\xi^{(-)}_{(i)}} \neq 0$ for some direction i.

Finally, observe by the revealing nature of $F(\gamma' x_t)$ it cannot be the case $(\partial/\partial\gamma_i)\mu(\gamma)|_{\gamma=\xi^{(+)}_{(i)}} = 0$ *and* $(\partial/\partial\gamma_i)\mu(\gamma)|_{\gamma=\xi^{(-)}_{(i)}} = 0$ for each i under H_1 since then $\mu(\gamma)$ is a constant on Γ.

2.3. Example: LSTAR and Misspecification

The relationship between the moment $\mu(\gamma)$ and its curvature $(\partial/\partial\gamma)\mu(\gamma)$ is easily seen by Monte Carlo experiment. See Section 3 for a complete study. We simulate $10,000$ series $\{y_t\}_{t=1}^n$ of size $n = 1{,}000$ generated from a Logistic Smooth Transaction AR (STAR):

$$y_t = 1 + (.8y_{t-1} - .3y_{t-2}) \times \frac{1}{1 + \exp\{-2y_{t-1}\}} + u_t,$$

where u_t is independent and identically distributed (iid) truncated standard normal, with support $[-10^{10}, 10^{10}]$. We fit each series by least squares to an AR(3) $y_t = \phi' x_t + \epsilon_t$, $x_t = [1, y_{t-1}, y_{t-2}, y_{t-3}]'$, compute the residuals $\hat{\epsilon}_t$ and a CM test statistic with logistic $F(\gamma' x_t) = (1 + \exp\{\gamma' x_t\})^{-1}$:

$$\mathscr{W}_n(\gamma) := n \times \hat{\mu}(\gamma)^2 / \hat{s}^2(\gamma) \qquad \text{for } \gamma \in \Gamma = [-2, 2]^4,$$

where $\hat{\mu}(\gamma) := 1/n \sum_{t=1}^{n} \hat{\epsilon}_t F(\gamma' x_t)$, $\hat{s}^2(\gamma) := 1/n \sum_{t=1}^{n} \hat{\epsilon}_t^2 \times (F(\gamma' x_t) - \hat{b}(\gamma)' \widehat{A} x_t)^2$, $\hat{b}(\gamma) := 1/n \sum_{t=1}^{n} x_t F(\gamma' x_t)$, and $\widehat{A} := 1/n \sum_{t=1}^{n} x_t x_t'$ (Lee et al., 1993).

Figure 1(a) plots the simulation average $\mathscr{W}_n(\gamma)$ and sample moment $\hat{\mu}(\gamma)$ over one plane of $\Gamma = [-2, 2]^4$ by fixing $\gamma_1 = \gamma_3 = \gamma_4 = 1$, and discretizing $\gamma_2 \in \{-2, -1.96, \ldots, 2\}$, a set of 100 elements. We now write compactly $\mathscr{W}_n(\gamma_2)$ and $\hat{\mu}(\gamma_2)$ for clarity.

There are four notable outcomes: (i) there are two breakdown points where the sample moment $\hat{\mu}(\gamma_2) \approx 0$, corresponding to $\gamma_2 = -0.52$ and $\gamma_2 = 0.32$; (ii) $(\partial/\partial_2)\hat{\mu}(\gamma_2) \approx 0$ at neither breakdown point; (iii) at the breakdown points the simulation mean $\mathscr{W}_n(\gamma_2) \approx 1$; and (iv) in a fairly broad neighborhood connecting the two breakdown points γ_2 the simulation mean $\mathscr{W}_n(\gamma_2) \approx 1$.

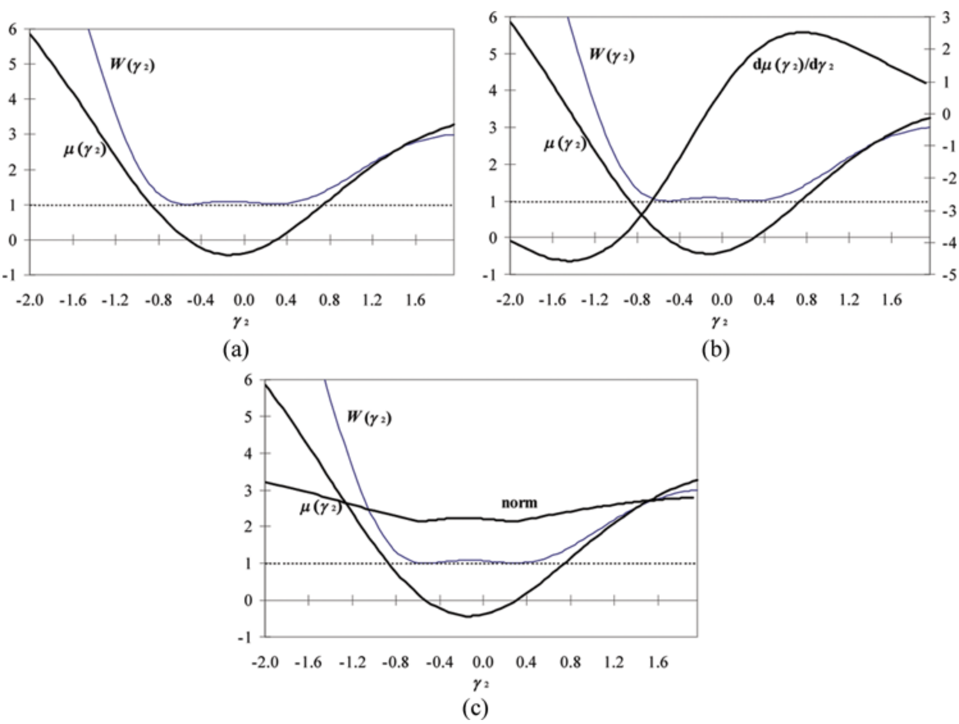

FIGURE 1 (a) $\mathscr{W}_n(\gamma_2)$ and Sample Mean $\hat{\mu}(\gamma_2)$; (b) $\mathscr{W}_n(\gamma_2)$, $\hat{\mu}(\gamma_2)$ and $(\partial/\partial\gamma_2)\hat{\mu}(\gamma_2)$. The right Y-axis represents $(\partial/\partial\gamma_2)\hat{\mu}(\gamma_2)$; (c) $\mathscr{W}_n(\gamma_2)$, $\hat{\mu}(\gamma_2)$ and $\|[\hat{\mu}(\gamma_2), [\sup\{(\partial/\partial\gamma_i)\hat{\mu}(\gamma)\}_{i=1}^4]]\|$. $\mathscr{W}_n(\gamma_2) = n\hat{\mu}(\gamma_2)^2/\hat{s}^2(\gamma_2)$ is Lee et al.'s (1993) test statistic evaluated at fixed $\gamma_1 = \gamma_3 = \gamma_4 = 1 \cdot \hat{\mu}(\gamma)$ is the sample moment $1/n \sum_{t=1}^{n} \hat{\epsilon}_t F(\gamma' x_t)$ with logistic $F(\gamma' x_t) = (1 + \exp\{\gamma' x_t\})^{-1}$. (Figure available in color online.)

Finding (ii) is logical in lieu of breakdown (i) since in a neighborhood of a bad γ_2 the function $\mu(\gamma_2)$ cannot be constant. See Fig. 1(b) for a plot of the simulation means of $\mathscr{W}_n(\gamma_2)$, $\hat{\mu}(\gamma_2)$, and $(\partial/\partial\gamma_2)\hat{\mu}(\gamma_2)$, and Fig. 1(c) for plots of $\mathscr{W}_n(\gamma_2)$, $\hat{\mu}(\gamma_2)$ and the Euclidean norm of the stacked vector $[\hat{\mu}(\gamma_2), [\sup_{\gamma\in\Gamma}\{(\partial/\partial\gamma_i')\hat{\mu}(\gamma)\}]_{i=1}^4]'$. The norm appears never to fail to reveal mis-specification since it contains an element that is nowhere zero. Obviously the norm must be scaled to control for its dispersion and to ensure a non-degenerate distribution limit. Our approach is precisely to use this scaled stacked vector because for each nuisance parameter value some element will be non-zero.

Finding (iii) is also logical since at true breakdown points $\mu(\gamma_2) = 0$ the CM statistic satisfies $\mathscr{W}_n(\gamma_2) \xrightarrow{d} \chi^2(1)$; hence $E[\mathscr{W}_n(\gamma_2)] \to 1$ by the Helly–Bray theorem.

Finding (iv), therefore, suggests $\mathscr{W}_n(\gamma_2)$ exhibits weak empirical power *near* breakdown points. Although $\mu(\gamma_2) \neq 0$ in a neighborhood of bad γ_2, recall the function $\mu(\gamma_2)$ is continuous. In small samples $\hat{\mu}(\gamma_2)$ may, therefore, insignificantly differ from zero near bad γ_2. If bad γ_2 are close together, then $\hat{\mu}(\gamma_2) \approx 0$ for a range of γ_2 is possible, which is precisely what we see here. This simply provides evidence that even power-optimizing average, supremum, and integrated transforms may suffer from weak power, depending on Γ.

Finally, we repeat the experiment by generating a new set of 10,000 samples $\{y_t\}_{t=1}^n$, and use a new interval $\Gamma = [0, 2]^4$ that contains the origin, with fixed $\gamma_1 = \gamma_3 = \gamma_4 = 0$. We plot each statistic over $\gamma_2 \in \{0, .02, \ldots, 2\}$ in Fig. 2. The results again perfectly mimic the intuition given in Section 2.2. The sample moment fails at the origin $\hat{\mu}(0) = 0$ and the derivative is nowhere zero. In particular, there are no breakdown points in $\hat{\mu}(\gamma_2)$ and, therefore, in the norm of $[\hat{\mu}(\gamma_2), [\sup_{\gamma\in\Gamma}\{(\partial/\partial\gamma_i')\hat{\mu}(\gamma)\}]_{i=1}^4]'$, and the mean $E[\mathscr{W}_n(\gamma_2)] > 1$ for all points γ_2.

Notice $\mathscr{W}_n(\gamma)$ spikes at the origin $\gamma = 0$. This occurs because we include a constant term: $\hat{\mu}(\gamma) = 0$ by construction, hence $E[\hat{\mu}(\gamma)^2] = 0$ so $\mathscr{W}_n(\gamma_2)$ is degenerate at $\gamma = 0$ (Bierens, 1990).

2.4. Score Test

Let Υ be a positive definite matrix in $\mathbb{R}^{q\times q}$, and $\hat{\Upsilon}_n$ a sample version that satisfies $\hat{\Upsilon}_n \xrightarrow{p} \Upsilon$. Define the GMM estimator

$$\hat{\phi} = \underset{\phi\in\Phi}{\operatorname{argmin}} \left\{ \bar{m}(\phi)' \times \hat{\Upsilon} \times \bar{m}(\phi) \right\} \qquad \text{where } \bar{m}(\phi) = \frac{1}{n}\sum_{t=1}^{n} m_t(\phi),$$

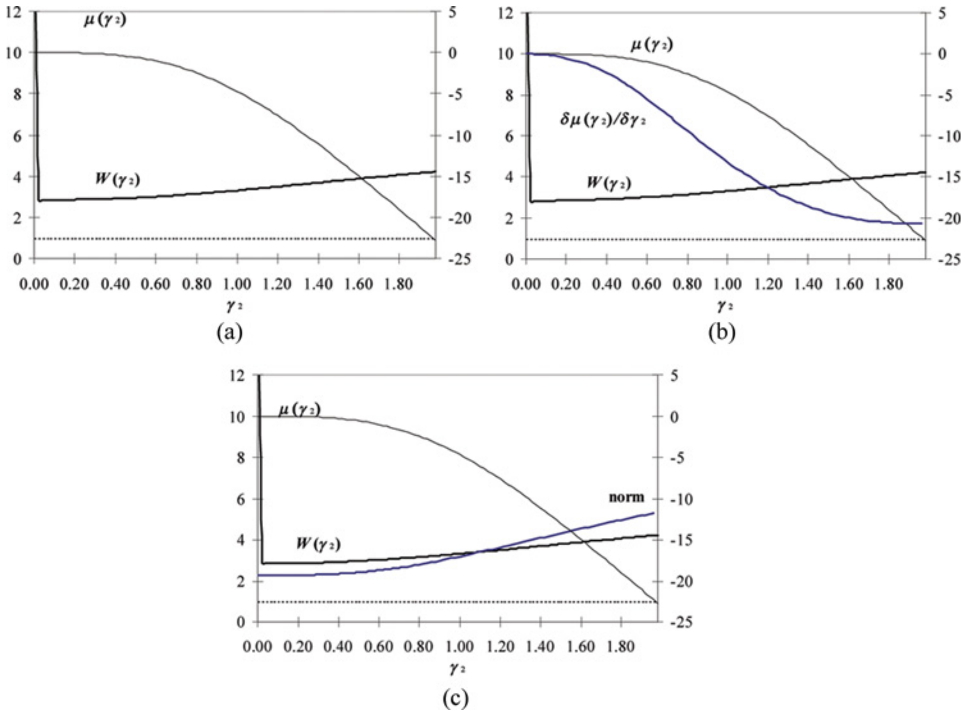

FIGURE 2 (a) $\mathscr{W}_n(\gamma_2)$ and Sample Mean $\hat{\mu}(\gamma_2)$; (b) $\mathscr{W}_n(\gamma_2)$, $\hat{\mu}(\gamma_2)$ and $(\partial/\partial\gamma_2)\hat{\mu}(\gamma_2)$. The left Y-axis is $\mathscr{W}_n(\gamma_2)$; (c) $\mathscr{W}_n(\gamma_2)$, $\hat{\mu}(\gamma_2)$ and $\|[\hat{\mu}(\gamma_2), [\sup\{(\partial/\partial\gamma_i)\hat{\mu}(\gamma)\}_{i=1}^4]]\|$. The left Y-axis is $\mathscr{W}_n(\gamma_2)$. (Figure available in color online.)

and define

$$\hat{\epsilon}_t := y_t - f_t(\hat{\phi}) \qquad \text{and} \qquad \hat{\mu}(\gamma) := \frac{1}{n}\sum_{t=1}^{n} \hat{\epsilon}_t F(\gamma' x_t)$$

$$\widehat{\mathfrak{Z}}_{n,i}^{(+)} := \left\{ \underset{\gamma\in\Gamma}{\text{argsup}} \left\{ \frac{\partial}{\partial\gamma_i}\hat{\mu}(\gamma) \right\} \right\} \qquad \text{and} \qquad \widehat{\mathfrak{Z}}_{n,i}^{(-)} := \left\{ \underset{\gamma\in\Gamma}{\text{arginf}} \left\{ \frac{\partial}{\partial\gamma_i}\hat{\mu}(\gamma) \right\} \right\}.$$

In order to reduce notation, we only discuss $\widehat{\mathfrak{Z}}_{n,i} := \widehat{\mathfrak{Z}}_{n,i}^{(+)}$ since the following carries over to $\widehat{\mathfrak{Z}}_{n,i}^{(-)}$.

In practice, $\arg\sup_{\gamma\in\Gamma}(\partial/\partial\gamma_i)\mu(\gamma)$ is both unknown and possibly not unique; hence $\widehat{\mathfrak{Z}}_{n,i}$ is a set with possibly more than one element for any n, and asymptotically. For example, asymptotically $\widehat{\mathfrak{Z}}_{n,i}$ is identically Γ under the null, so it has positive Lebesgue measure. We therefore propose a sample plug-in matrix

$$\hat{\xi} = \left[\hat{\xi}_{(1)}, \ldots, \hat{\xi}_{(k)}\right] \in \Gamma^* := \Gamma \times \cdots \times \Gamma \subset \mathbb{R}^{k\times k},$$

where $\hat{\xi}_{(i)} \in \mathbb{R}^k$ is any unique element of $\widehat{\mathfrak{Z}}_{n,i}$, as long as the same selection criterion is used for each n. Examples include the first element of the set $\widehat{\mathfrak{Z}}_{n,i}$ ordinally ranked according to some criterion; the first element of an ordinally ranked subset of $\widehat{\mathfrak{Z}}_{n,i}$ with the minimum, median, mean, or maximum Euclidean norm $\|\hat{\xi}_{(i)}\|$; and so on. We simply assume there exists a unique element $\xi^*_{(i)}$ of the set $\mathfrak{Z}_i := \{\text{argsup}_{\gamma \in \Gamma}(\partial/\partial\gamma_i)E[\epsilon_t F(\gamma' x_t)]\} = \{\xi_{(i)}\}$ satisfying

$$\xi^*_{(i)} := \underset{n\to\infty}{\text{plim}}\, \hat{\xi}_{(i)}.$$

In lieu of the key observations of Section 2.2, we stack weight-types popularized in the ANN and STAR literatures:

$$w_t(\gamma, \hat{\xi}) = \left[F(\gamma' x_t); \left\{x_{i,t}F'\left(\hat{\xi}'_{(i)}x_t\right)\right\}_{i=1}^{k}\right]' \in \mathbb{R}^{k+1}.$$

Now define the $\sqrt{n}$-scaled sample moment

$$\hat{z}_n(\gamma) := \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \hat{\epsilon}_t \times w_t(\gamma, \hat{\xi}) \in \mathbb{R}^{k+1},$$

and covariance matrix $\widehat{\Sigma}_n(\gamma) := 1/n \sum_{t=1}^{n}\left\{\hat{\epsilon}_t^2 \times \hat{g}_t(\gamma, \hat{\xi}) \times \hat{g}_t(\gamma, \hat{\xi})'\right\} \in \mathbb{R}^{(k+1)\times(k+1)}$, where

$$\hat{g}_t(\gamma, \xi) =: w_t(\gamma, \xi) + \widehat{A}(\hat{\phi}, \gamma, \xi) \times c_t(\hat{\phi}) \in \mathbb{R}^{k+1}$$

$$\widehat{A}(\phi, \gamma, \xi) := -\left(\frac{1}{n}\sum_{t=1}^{n} w_t(\gamma, \xi) \times \frac{\partial}{\partial \phi'} f_t(\phi)\right) \times \left(\hat{H}(\phi)^{-1}\frac{\partial}{\partial \phi'}\bar{m}(\phi) \times \hat{\Upsilon}\right) \in \mathbb{R}^{(k+1)\times q}$$

$$\hat{H}(\phi) := -\frac{\partial}{\partial \phi'}\bar{m}(\phi) \times \hat{\Upsilon} \times \frac{\partial}{\partial \phi}\bar{m}(\phi) \in \mathbb{R}^{p\times p}$$

and $c_t(\phi)$ is defined in (1). The test statistic has a quadratic form

$$\mathcal{T}_n(\gamma) = \hat{z}_n(\gamma)'\widehat{\Sigma}_n(\gamma)^{-1}\hat{z}_n(\gamma).$$

We analyze $\mathcal{T}_n(\gamma)$ under a $\sqrt{n}$-local alternative

$$H_1^L : y_t = f_t(\phi_0) + u_t/\sqrt{n} + \epsilon_t,$$

where $\epsilon_t = y_t - E[y_t|\Im_{t-1}]$, and u_t is an $\Im_{t-1}$-measurable random variables. Measurability implies u_t may be a function of lags of y_t and other random variables contained in x_t, hence trivially $E[\epsilon_t u_t] = 0$ by the martingale difference property of ϵ_t. It also implies switching or break models are not included under H_1^L if the switch or break mechanism is not $\Im_{t-1}$-measurable (e.g., Hamilton, 1994).[1]

The null hypothesis is $u_t = 0$ $a.s.$, and a global alternative is simply

$$H_1^G : y_t = f_t(\phi_0) + u_t + \epsilon_t.$$

Finally, let Γ_ι be the subset of Γ with bounded $|\gamma| > \iota$ for arbitrary $\iota > 0$. The main result of this paper follows. See Appendix A for a sketch of the proof. As shown in Section 2.2 the agument for a perfectly revealing parametric moment condition follows from results in Bierens (1982, 1990), Bierens and Ploberger (1997), and Stinchcombe and White (1998). See Hill (2011) for data generating process specifics under Assumption 3.

Theorem 2.1. *Let Assumptions* 1–3 *hold. Under* H_1^L, $\mathcal{T}_n(\gamma) \Rightarrow z(\gamma)^2$ *on* $\mathbb{C}[\Gamma_\iota]$ *where* $z^2(\gamma)$ *is an almost everywhere continuous chi-squared process. In particular,* $\mathcal{T}_n(\gamma) \xrightarrow{d} \chi^2(k+1)$ *under the null* H_0 *pointwise on* Γ_ι *and under the distant alternative* H_1^G, $\mathcal{T}_n(\gamma) \xrightarrow{p} \infty$ *with probability one for every* $\gamma \in \Gamma_\iota$.

Remark 3. Bounding $|\gamma| > \iota$ provides a mathematical safety net to ensure the test statistic is non-degenerate for γ or $\xi_{(i)}$ very close to zero (cf. Hill, 2008).

Remark 4. If the regression model $y_t = f_t(\phi_0) + \epsilon_t$ is misspecified, then by simultaneously using sample versions of $E[\epsilon_t F(\gamma' x_t)]$ and $E[\epsilon_t x_{i,t} F'(\xi'_{(i)} x_t)]$ our test statistic diverges $\mathcal{T}_n(\gamma) \xrightarrow{p} \infty$ for any nuisance parameter value $\gamma \neq 0$.

Remark 5. Theorem 2.1 ensures continuous test functionals $h(\mathcal{T}_n(\gamma))$ have well defined limits, including supremum and average functionals (Davies, 1977; Bierens, 1990). Since $\mathcal{T}_n(\gamma)$ is consistent for any $\gamma \neq 0$ a test functional is not required, but may be inspected by case for small sample performance as we do below.

Remark 6. If the response is linear $f_t(\phi) = \phi' x_t$ and the plug-in is least squares computed as exactly identified GMM (i.e., $c_t(\phi) = x_t$), then the test scale $\widehat{\Sigma}_n(\gamma) := 1/n \sum_{t=1}^n \{\hat{\epsilon}_t^2 \hat{g}_t(\gamma, \hat{\xi}) \hat{g}_t(\gamma, \hat{\xi})'\}$ is computed with

$$\hat{g}_t(\gamma, \xi) = w_t(\gamma, \xi) + \left(\frac{1}{n}\sum_{t=1}^n w_t(\gamma, \xi) x_t'\right) \times \left(\frac{1}{n}\sum_{t=1}^n x_t x_t'\right)^{-1} \times x_t. \qquad (2)$$

[1]Test procedures for Markov Switching and stochastic break models are studied elsewhere. See Bai and Perron (1998) and Carrasco et al. (2009) and their references.

TABLE 1 DGP

AR(2)	$y_t = 0.8y_{t-1} - 0.4y_{t-2} + \epsilon_t$
SETAR	$y_t = 0.8y_{t-1} - 1.2y_{t-1}I\left(y_{t-1} \geq 0\right) + \epsilon_t$
ESTAR	$y_t = 0.8y_{t-1} - 1.2y_{t-2} \times \exp\{-1.5y_{t-1}^2\} + \epsilon_t$
LSTAR	$y_t = 0.8y_{t-1} - 1.2y_{t-1} \times [1 + \exp\{-1.5y_{t-1}\}]^{-1} + \epsilon_t$
BILIN	$y_t = 0.9y_{t-1}\left\|\epsilon_{t-1}\right\|^{1.5} + \epsilon_t$

3. MONTE CARLO STUDY

We simulate 10,000 samples of linear and nonlinear AR processes, with sample sizes $n \in \{200, 500, 2000\}$. In all cases we test if the true data generating process (DGP) is linear AR. The nonlinear processes are Self-Exciting AR (SETAR), Exponential and Logistic Smooth Transition AR (ESTAR and LSTAR), and Bilinear (BILIN), each with an error $\epsilon_t \overset{iid}{\sim} N(0,1)$. See Table 1 for details.

Now write $x_t = [1, \tilde{x}_t']' = [1, y_{t-1}, \ldots, y_{t-5}]'$ and $z_t = [1, y_{t-1}, \ldots, y_{t-10}]'$. We estimate an AR(5) with a constant term $y_t = \phi_0' x_t + \epsilon_t$ by GMM, with equations $m_t(\phi) = (y_t - \phi' x_t) z_t$ and naïve weight $\hat{\Upsilon} = I_{11}$. In simulations not reported here there is essentially no difference between $\hat{\Upsilon} = I_{11}$ and the asymptotically efficient weight $\hat{\Upsilon} = (1/n \sum_{t=1}^n [m_t(\hat{\phi}) m_t(\hat{\phi})'])^{-1}$ based on a first step naïve GMM or Ordinary Least Squares (OLS) plug-in $\hat{\phi}$.

3.1. Construction of $\mathcal{T}_n(\gamma)$

We test the GMM residuals $\hat{\epsilon}_t = y_t - \hat{\phi}' x_t$ for omitted nonlinearity at the 5%-level using $\sup_{\gamma\in\Gamma_n}\{\mathcal{T}_n(\gamma)\}$, randomized $\mathcal{T}_n(\gamma_*)$, and $\mathcal{T}_n(\gamma_0)$ with a fixed γ_0. Throughout, we use a discretized set $\Gamma_n = [\gamma_1, \ldots, \gamma_{J_n}]$ where for each sample γ_i are uniformly randomized on $\Gamma = [0.5, 10]^6$, and $J_n = [n/4]$. The scale $\widehat{\Sigma}_n(\gamma)$ is computed using $\hat{g}_t(\gamma, \xi)$ in (2).

We use test weights $w_t(\gamma, \xi)$ based on $\hat{\xi} = \hat{\xi}^{(+)} = [\hat{\xi}_{(1)}^{(+)}, \ldots, \hat{\xi}_{(k)}^{(+)}]$, and exponential $F(u) = \exp\{u\}$ or logistic $F(u) = (1 + \exp\{u\})^{-1}$ functions. For example, the exponential-based weight is

$$w_t(\gamma, \hat{\xi}) = \left[\exp\{\gamma' \Psi(x_t)\}, \left[\Psi_i(x_t) \exp\{\hat{\xi}_{(i)}^{(+)\prime} \Psi(x_t)\}\right]_{i=1}^{6}\right]' \in \mathbb{R}^7$$

where the bounded argument $\Psi(x_t) = [\Psi_i(x_t)]_{i=1}^6$ is defined below. In this study the set of solutions $\widehat{\mathfrak{Z}}_{n,i} = \{\hat{\xi}_{(i)}\}$ was a singleton for every simulated series $\{y_t\}$ under any hypothesis: for the given design $\hat{\mu}(\gamma)$ always exhibited sufficient curvature for a unique solution $\hat{\xi}_{(i)} := \text{argsup}_{\gamma\in\Gamma_n}\{(\partial/\partial\gamma_i)\hat{\mu}(\gamma)\}$.[2]

[2]The same sufficient curvature arose for exceptionally large samples not reported here (e.g. $n = 100,000$).

Test results where only $\hat{\xi}_{(i)}^{(-)}$ is used are essentially identical, and when both $[\hat{\xi}_{(i)}^{(-)}, \hat{\xi}_{(i)}^{(+)}]$ are used there is a slight reduction in empirical power due to the increased weight dimension.

Now consider $\Psi(x_t)$, let $[\varsigma]$ round to the nearest integer; for any random scalar ς_t write $\varsigma_t^{(a)} := |\varsigma_t|$; and let $\varsigma_{(i)}^{(a)}$ denote an order statistic: $\varsigma_{(1)}^{(a)} \geq \varsigma_{(2)}^{(a)} \geq \cdots \geq \varsigma_{(n)}^{(a)}$. We use a one-to-one argument $\Psi(x_t) = [1, \tilde{\Psi}(y_{t-1}), \ldots, \tilde{\Psi}(y_{t-5})]'$ that imposes a smooth weight at a high threshold:

$$\tilde{\Psi}\left(y_{t-i}\right) = y_{t-i} \qquad \text{if } \left|y_{t-i}\right| \leq y_{([0.001\times n])}^{(a)}$$

$$\tilde{\Psi}\left(y_{t-i}\right) = sign\left\{y_{t-i}\right\} \times y_{([0.001\times n])}^{(a)}\left(2 - \exp\left\{y_{([0.001\times n])}^{(a)} - \left|y_{t-i}\right|\right\}\right)$$
$$\text{if } \left|y_{t-i}\right| > y_{([0.001\times n])}^{(a)}.$$

If q_λ denotes the $\lambda \in (0,1)$ two-tailed quantile $P(|y_t| \leq q_\lambda) = \lambda$, then as $n \to \infty$ by construction $|\tilde{\Psi}\left(y_{t-i}\right)| \leq 2q_\lambda$ for $\lambda = 0.999$, in particular $|\tilde{\Psi}\left(y_{t-i}\right)| \in [q_\lambda, 2q_\lambda]$ if $|y_{t-i}| \geq q_\lambda$. Other arguments lead to roughly similar but suboptimal results, including sign $\{y_{t-i}\} \times \exp\left\{-\left|y_{t-i}\right|\right\}$ proposed in Hill (2008), or $\tan^{-1}(y_{t-i})$ proposed in Bierens (1982, 1990) where y_{t-i} is standardized. We choose such a high percentile λ since in small samples a reduction in power occurs for smaller values like $\lambda = 0.99$.[3] There does not exist a practicum for selecting Ψ, and any choice will impact test results; hence all subsequent results should be interpreted with some caution.

Each process is stationary geometrically strong mixing and $\|y_t\|_{4+\iota} < \infty$ (An and Huang, 1996). It is then straightforward to verify each condition under Assumption 3. Further, under stationary geometric strong mixing the central order statistic $y_{([\lambda n])}^{(a)}$ of $|y_t|$ is consistent for the quantile q_λ (e.g., de Haan et al., 1991).

In the case of $\sup_{\gamma\in\Gamma_n}\{\mathcal{T}_n(\gamma)\}$, the p-value is computed by Hansen's (1996) boostrap/Monte Carlo method based on 1,000 simulated samples of iid standard normals, drawn for each sample $\{y_t\}_{t=1}^n$. The present framework is nested within the context developed in Hill (2008), hence the asymptotic p-value is consistent (Hill, 2008). An average test functional, by comparison, is dominated by the supremum because the alternatives here are not local or infinitesimally "close" to the null (see Andrews and Ploberger, 1994).

In randomized tests γ_* is a uniform draw from Γ for *each sample*, and in the fixed case we use the set mid-point $\gamma_0 = [5.25, \ldots, 5.25]$. Under the AR(5) null both $\mathcal{T}_n(\gamma_*)$ and $\mathcal{T}_n(\gamma_0)$ are asymptotically $\chi^2(7)$ from which critical values are taken.

[3]Despite the necessity for boundedness of Ψ based on arguments in Stinchcombe and White (1998), within the present simulation design we find the best argument is simply x_t itself. Our chosen $\Psi(x_t)$ with $\lambda = 0.999$ works very well since there is no, or nearly no, difference between $\Psi(x_t)$ and x_t for small and large n.

3.2. Most Powerful (MP) Tests

In order to gauge the power of the proposed test, we compute most powerful (MP) tests against each alternative. Since $\epsilon_t \overset{iid}{\sim} N(0,1)$ and ϕ are known within the simulation, each model can be written as $y_t(\phi_1) = \phi_2 \varsigma_t(\gamma) + \epsilon_t$ for some scalar $\varsigma_t(\gamma)$ and $y_t(\phi_1) = y_t - \phi_1 y_{t-1}$. The regressors are: SETAR $\varsigma_t(\gamma) = y_{t-1} I(y_{t-1} > 0)$; ESTAR $\varsigma_t(\gamma) = y_{t-1}\exp\{-\gamma y_{t-1}^2\}$; LSTAR $\varsigma_t(\gamma) = y_{t-1}(1+\exp\{-\gamma y_{t-1}\})^{-1}$; and BILIN $\varsigma_t(\gamma) = y_{t-1}|\epsilon_{t-1}|^{1.5}$. The errors are known to be iid standard normal so the MP test statistic reduces to $M_n(\gamma) = y(\phi_1)'\varsigma(\gamma)\left[\varsigma(\gamma)'\varsigma(\gamma)\right]^{-1}\varsigma(\gamma)'y(\phi_1)$.

3.3. Additional Tests

We also compute Bierens' (1990) and Lee et al.'s (1993) statistics $\mathscr{W}_n(\gamma) := (1/\sqrt{n}\sum_{t=1}^n \hat{\epsilon}_t F(\gamma'\Psi(x_t)))^2/\hat{s}^2(\gamma)$ for exponential and logistic $F(u)$, where $\hat{s}^2(\gamma) = 1/n\sum_{t=1}^n \hat{\epsilon}_t^2 \times (F(\gamma'\Psi(x_t)) - \hat{b}(\gamma)'\widehat{A}x_t)^2$, $\hat{b}(\gamma) = 1/n\sum_{t=1}^n x_t \exp\{\gamma'\Psi(x_t)\}$ and $\widehat{A} = 1/n\sum_{t=1}^n x_t x_t'$. We compute $\mathscr{W}_n(\gamma_*)$, $\mathscr{W}_n(\gamma_0)$ and $\sup_{\gamma\in\Gamma_n}\{\mathscr{W}_n(\gamma)\}$ and use Hansen's (1996) p-value method for the latter.

We also perform the following classic tests of regression model misspecification: Ramsey's (1969) RESET LM-statistic based on the auxiliary regression $\hat{\epsilon}_t = \beta_0' x_t + \sum_{i=2}^3\sum_{j=2}^3 \beta_{i,j} x_{t,j}^i + u_t$; the McLeod and Li (1983) statistic

$$\sum_{t=1}^n \left(\hat{\epsilon}_t^2 - \hat{\sigma}^2\right)\left(\hat{\epsilon}_{t-h}^2 - \hat{\sigma}^2\right) \Big/ \sum_{t=1}^n \left(\hat{\epsilon}_t^2 - \hat{\sigma}^2\right)^2, \quad h = 1, 2, 3,$$

where $\hat{\sigma}^2 = \hat{\epsilon}_t'\hat{\epsilon}_t/n$, and an asymptotic version F_n of Tsay's (1986) F-statistic. The latter is based on first regressing nonredundant stochastic regressor cross-products vech$[\tilde{x}_t\tilde{x}_t']$ on all regressors x_t, vech$[\tilde{x}_t\tilde{x}_t'] = \vartheta' x_t + u_t$; and then computing $F_n = \sum_{t=1}^n(\hat{\epsilon}_t\hat{u}_t) \times [\sum_{t=1}^n \hat{u}_t\hat{u}_t']^{-1} \times \sum_{t=1}^n(\hat{\epsilon}_t\hat{u}_t')$. Under the null for the given design standard arguments reveal $F_n \overset{d}{\to} \chi^2(p(p+1)/2)$ where $p = 5$.

Finally, we compute Hong and White's (1995) nonparametric test statistic

$$\frac{1}{(2\ln n)^{1/2}}\left(\frac{1}{1/n\sum_{t=1}^n \hat{\epsilon}_t^2}\sum_{t=1}^n \hat{\epsilon}_t\left(\hat{f}_t - \hat{\phi}' x_t\right) - \ln n\right),$$

with Gallant's (1981) Flexible Fourier Form estimator $\hat{f}_t = \sum_{i=1}^{[\ln(n)]}\beta_i \exp\{\zeta_i' x_t\}$ of $E[y_t|x_t]$. The parameters ζ_i are for each sample uniformly randomly selected from Γ, and β is estimated by least squares. The nonparametric estimator $\hat{f}_t$ is consistent by Corollary 1 of Bierens (1990), and our series length $\ln(n)$ choice is defended by Theorem 3.2 of Hong and White (1995). Under regularity conditions covered by our DGP, if the null is true, then $\widehat{M}_n \xrightarrow{d} N(0,1)$, else $\widehat{M}_n \xrightarrow{p} \infty$, and hence a one-sided test is performed.

3.4. Simulation Results

See Tables 2 and 3 for all randomized and sup-CM test results, and Table 4 for the remaining test results. We do not report fixed CM test results here since $\mathcal{T}_n(\gamma_0)$ is qualitatively similar to $\mathcal{T}_n(\gamma_*)$. See Hill (2011).

We now summarize the most important findings.

TABLE 2 Randomized CM test rejection frequencies

	AR	LSTAR	ESTAR	SETAR	BILIN
			$n = 200$		
L-$\mathcal{T}_n(\gamma_*)$	0.043[b]	0.256	0.750	0.453	0.058
E-$\mathcal{T}_n(\gamma_*)$	0.052	0.426	0.497	0.302	0.035
L-$\mathcal{W}_n(\gamma_*)$[c]	0.048	0.119	0.347	0.213	0.078
E-$\mathcal{W}_n(\gamma_*)$	0.064	0.121	0.277	0.183	0.034
L-$\mathcal{H}_n(\gamma_*)$[d]	0.047	0.134	0.362	0.196	0.090
E-$\mathcal{H}_n(\gamma_*)$	0.050	0.133	0.292	0.167	0.033
UMP(γ_*)[e]	–	0.956	1.00	1.00	0.314
			$n = 500$		
L-$\mathcal{T}_n(\gamma_*)$	0.048	0.587	0.998	0.910	0.142
E-$\mathcal{T}_n(\gamma_*)$	0.053	0.754	0.966	0.835	0.036
L-$\mathcal{W}_n(\gamma_*)$	0.049	0.304	0.661	0.444	0.099
E-$\mathcal{W}_n(\gamma_*)$	0.052	0.318	0.566	0.425	0.013
L-$\mathcal{H}_n(\gamma_*)$	0.043	0.306	0.655	0.449	0.092
E-$\mathcal{H}_n(\gamma_*)$	0.062	0.330	0.547	0.400	0.017
UMP(γ_*)	–	1.00	1.00	1.00	0.484
			$n = 2000$		
L-$\mathcal{T}_n(\gamma_*)$	0.048	0.993	1.00	1.00	0.465
E-$\mathcal{T}_n(\gamma_*)$	0.047	1.00	1.00	1.00	0.039
L-$\mathcal{W}_n(\gamma_*)$	0.061	0.752	0.923	0.837	0.159
E-$\mathcal{W}_n(\gamma_*)$	0.063	0.778	0.895	0.821	0.007
L-$\mathcal{H}_n(\gamma_*)$	0.037	0.723	0.933	0.858	0.159
E-$\mathcal{H}_n(\gamma_*)$	0.044	0.751	0.905	0.855	0.001
UMP(γ_*)	–	1.00	1.00	1.00	0.711

[a]The nuisance parameter is randomized on $\Gamma = [0.5, 10]^{p+1}$ for each sample.
[b]L = logistic; E = exponential. Values are rejection frequency at the 5% level.
[c]Bierens' (1990) and Lee et al.'s (1996) CM test with logistic or exponential weight $F(\gamma'\psi_t)$.
[d]Hill's (2008) STAR test with logistic or exponential weight $x_t F(\gamma'\psi_t)$.
[e]The Uniformly Most Powerful test. Each test statistic is designed to be UMP for the particular H_1.
In simulations not reported here, each statistic obtains empirical size roughly equal to nominal size.

TABLE 3 Sup-CM Test Rejection Frequencies[a]

	AR	LSTAR	ESTAR	SETAR	BILIN
			$n = 200$		
L-sup $\mathcal{T}_n(\gamma)$	0.056[b]	0.189	0.772	0.390	0.037
E-sup $\mathcal{T}_n(\gamma)$	0.061	0.333	0.510	0.224	0.021
L-sup $\mathcal{W}_n(\gamma)$[c]	0.051	0.086	0.352	0.111	0.050
E-sup $\mathcal{W}_n(\gamma)$	0.068	0.093	0.279	0.109	0.018
L-sup $\mathcal{H}_n(\gamma)$[d]	049	0.091	0.399	0.112	0.067
E-sup $\mathcal{H}_n(\gamma)$	0.054	0.102	0.314	0.085	0.018
supUMP(γ)[e]	–	0.852	1.00	0.991	0.254
			$n = 500$		
L-sup $\mathcal{T}_n(\gamma)$	0.051	0.512	1.00	0.889	0.129
E-sup $\mathcal{T}_n(\gamma)$	0.054	0.691	1.00	0.771	0.027
L-sup $\mathcal{W}_n(\gamma)$	0.047	0.258	0.690	0.402	0.082
E-sup $\mathcal{W}_n(\gamma)$	0.050	0.229	0.602	0.383	0.009
L-sup $\mathcal{H}_n(\gamma)$	0.047	0.244	0.679	0.368	0.073
E-sup $\mathcal{H}_n(\gamma)$	0.058	0.300	0.582	0.361	0.003
supUMP(γ)	–	0.912	1.00	0.937	0.483
			$n = 2000$		
L-sup $\mathcal{T}_n(\gamma)$	0.045	0.948	1.00	0.979	0.383
E-sup $\mathcal{T}_n(\gamma)$	0.039	0.952	1.00	0.940	0.028
L-sup $\mathcal{W}_n(\gamma)$	0.052	0.742	0.942	0.804	0.111
E-sup $\mathcal{W}_n(\gamma)$	0.062	0.689	0.912	0.763	0.009
L-sup $\mathcal{H}_n(\gamma)$	0.035	0.706	0.955	0.791	0.134
E-sup $\mathcal{H}_n(\gamma)$	0.028	0.666	0.931	0.772	0.001
supUMP(γ)	–	0.919	1.00	0.949	0.647

[a]The statistic is optimized on $\Gamma = [0.5, 10]^{p+1}$, and the p-value is computed by bootstrap.
[b]L = logistic; E = exponential. Values are rejection frequency at the 5% level.
[c]Bierens' (1990) and Lee et al.'s (1996) CM test with logistic or exponential weight $F(\gamma'\psi_t)$.
[d]Hill's (2008) STAR test with logistic or exponential weight $x_t F(\gamma'\psi_t)$.
e. The Uniformly Most Powerful test. Each test statistic is designed to be UMP for the particular H_1.
In simulations not reported here, each statistic obtains empirical size roughly equal to nominal size.

i. Constructing a sup-test functional $\sup_{\gamma\in\Gamma_n}\{\mathcal{T}_n(\gamma)\}$ in order to improve small sample power is utterly ineffective in the majority of cases. Indeed, randomized $\mathcal{T}_n(\gamma_*)$ and fixed $\mathcal{T}_n(\gamma_0)$ exhibit non-negligible power improvements over the sup-test in nearly every case. The reason for the improvement is likely two-fold: $\mathcal{T}_n(\gamma_*)$ is asymptotically perfect so $\sup_{\gamma\in\Gamma_n}\{\mathcal{T}_n(\gamma)\}$ need not have more power; and the sup-test requires a bootstrapped p-value which is subject to additional sampling error that vanishes as $n \to \infty$.

ii. There is essentially no difference between a randomized test and a test based on our choice of fixed γ_0. Simulations not reported here reveal a wide range of fixed γ_0 render similar results.

TABLE 4 Nonlinearity test rejection frequencies

	AR	LSTAR	ESTAR	SETAR	BILIN
			$n = 200$		
RESET[a]	0.025[b]	0.180	0.661	0.461	0.013
TSAY[c]	0.058	0.440	0.994	0.649	0.728
ML[d]	0.038	0.103	0.613	0.184	0.648
HW[e]	0.017	0.002	0.410	0.048	0.096
			$n = 500$		
RESET	0.029	0.816	0.992	0.981	0.043
TSAY	0.044	0.926	1.00	0.992	0.863
ML	0.043	0.308	0.905	0.491	0.973
HW	0.025	0.086	0.998	0.164	0.438
			$n = 2000$		
RESET	0.024	1.00	1.00	1.00	0.197
TSAY	0.045	1.00	1.00	1.00	0.968
ML	0.041	0.886	0.998	0.976	0.998
HW	0.079	0.460	1.00	0.729	0.957

[a]Ramsey's RESET test.
[b]Rejection frequency at the 5% level.
[c]Asymptotic version of Tsay's F-test.
[d]McLeod–Li test with 3 lags.
[e]Hong and White's (1996) nonparametric test with Gallant's (1981) FFF plug-in.

iii. In the majority of cases, $\mathcal{T}_n(\gamma_*)$ offers a non-negligible power improvement over Bierens' (1990) and Lee et al.'s (1993) test statistic $\mathcal{W}_n(\gamma_*)$. There are, however, multiple interpretations of this outcome. Hill (2008) obtains related results when STAR components $x_t F(\gamma'\Psi(x_t))$ are used in a vector-valued CM test, and since x_t contains a constant the weight nests an ANN weight $F(\gamma'\Psi(x_t))$. The present statistic $\mathcal{T}_n(\gamma_*)$ is similarly constructed from ANN and STAR-like terms, whereas $\mathcal{W}_n(\gamma_*)$ uses only an ANN weight $F(\gamma'\Psi(x_t))$. If $F(u)$ is exponential, for example, we use $\exp\{\gamma'\Psi(x_t)\}$ and $\Psi_i(x_t)\exp\{\xi_{(i)}^{(+)\prime}\Psi(x_t)\}$, similar to Hill's (2008) $x_t \exp\{\gamma'\Psi(x_t)\}$; except $\Psi_i(x_t)\exp\{\xi_{(i)}^{(+)\prime}\Psi(x_t)\}$ cleverly ensures a perfect test asymptotically.

Hill (2008), however, cannot provide a theoretical reason for why his test statistic *should* have greater power against a general alternative, although his test is consistent. We compute Hill's (2008) test statistic $\mathcal{H}_n(\gamma)$ for comparison, which is simply $\mathcal{W}_n(\gamma)$ with $F(\gamma'\Psi(x_t))$ replaced by $x_t F(\gamma'\Psi(x_t))$. See Table 2. Based on the present design the simplest statistic $\mathcal{W}_n(\gamma_*)$ achieves the lowest power. Then $\mathcal{H}_n(\gamma_*)$ as expected, while $\mathcal{T}_n(\gamma_*)$ exhibits a non-negligible improvement over $\mathcal{H}_n(\gamma_*)$. The construction

$[F(\gamma'\Psi(x_t)), [\Psi_i(x_t)F'(\xi_{(i)}^{(+)\prime}\Psi(x_t))]_{i=1}^{6}]'$ appears to have a non-negligible impact on empirical power.

iv. In the case of ESTAR and SETAR data $\mathcal{T}_n(\gamma_*)$ exhibits empirical power nearly identical to the Uniformly Most Powerful (UMP) test for $n \geq 500$. In no case does $\mathcal{W}_n(\gamma_*)$ or $\mathcal{H}_n(\gamma_*)$ exhibit comparable power with a sample size $n = 500$.

v. Each remaining test works well in some cases. Both McLeod–Li and Hong–White tests, however, suffer from low power for SETAR.

3.5. Asymptotic Power

Finally, we investigate power over sample size n by computing $\mathcal{T}_n(\gamma_*)$ and $\mathcal{W}_n(\gamma_*)$ for ESTAR, LSTAR SETAR, and BILIN data, and the corresponding UMP statistic, for each $n \in \{50, 60, \ldots, 990, 1000\}$. See Fig. 3 for power plots. The statistic $\mathcal{T}_n(\gamma_*)$ has power near the UMP test for $n \geq 250$ in the case of ESTAR, and for $n \geq 500$ for SETAR, while $\mathcal{W}_n(\gamma_*)$ lags behind in all cases at each n, often substantially.

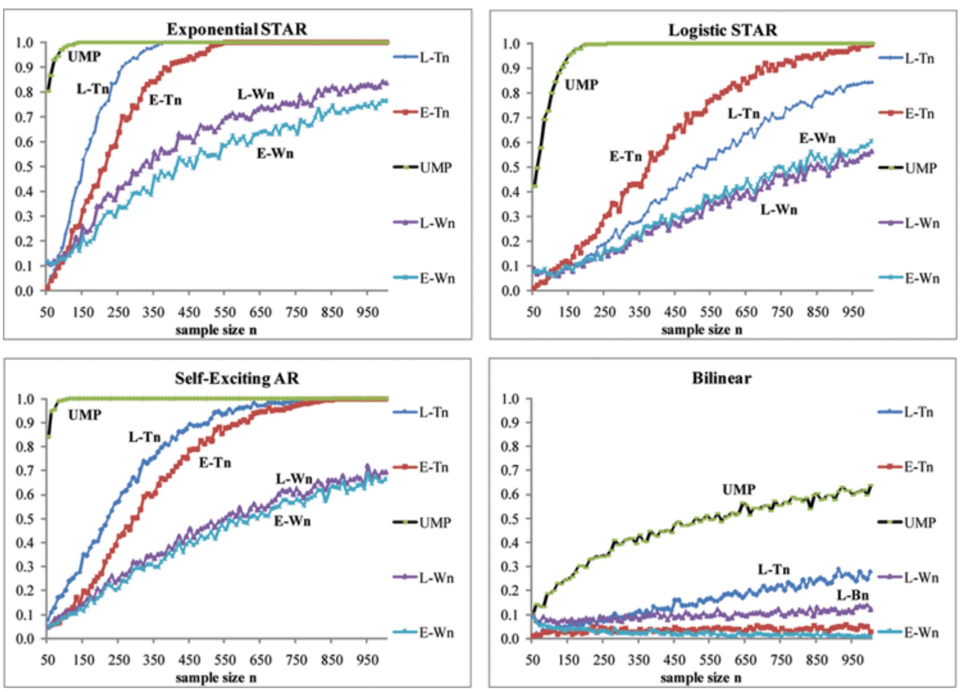

FIGURE 3 Empirical power against nonlinear AR alternatives. L-Tn = $\mathcal{T}_n(\gamma)$ with a logistic-based weight. L-Wn = $\mathcal{W}_n(\gamma)$ with a logistic weight. UMP = uniformly most powerful test. Each statistic uses a randomized nuisance parameter. (Figure available in color online.)

4. CONCLUSION

We exploit the CM test structure developed in Bierens (1982, 1990) and elsewhere in order to optimize CM test performance. The resulting test exploits two sets of interacting moment conditions that ensure any form of misspecification is detected for any nuisance parameter value $\gamma \neq 0$. The interacting moment conditions are based on those popularly used in the ANN/STAR literatures. A simulation experiment reveals a score statistic with a randomized nuisance parameter can obtain power nearly equal to the associated most powerful test, even for relatively small samples, while the combination of moment conditions provides a non-negligible increase in empirical power relative to tests with only ANN or STAR interpretations. Both results provide small sample evidence of our test's asymptotic capabilities.

APPENDIX A: PROOF OF THEOREM 2.1

Theorem 2.1 is easily verified using extensions of results developed elsewhere. Although intuition is presented in Section 2.2, the fact that the interacting moment condition is perfect, a la Lemma A.1, below, follows from theory developed in Bierens (1982, 1990), and then broadened in Bierens and Ploberger (1997) and especially Stinchcombe and White (1998), while the weak limit theory is identical to Bierens (1991), de Jong (1996), and Hill (2008). In view of this, we present below only an overview of the main steps in the proof of Theorem 2.1. A complete proof can be found in Hill (2011), an unpublished supplementary appendix.

In order to prove Theorem 2.1, we must first formally demonstrate the proposed interacting moment condition reveals model misspecification for any non-zero nuisance parameter value. We then present asymptotic theory for the weak limit in Theorem 2.1. Finally, we prove Theorem 2.1.

A.1. Interacting Moment Conditions

Recall the argument in Section 2.2 crucially exploits moment condition failure $\mu(0) = E[\epsilon_t F(0'x_t)] = 0$ at the origin. We therefore require $0 \in \Gamma$, but sets Γ *not* containing zero may be considered in practice. Indeed, Theorem 2.1 implicitly suggests the analyst ensures $0 \notin \Gamma$ precisely because $\mu(0) = 0$. Consider $\xi_{(i)}^{(+)}$ since all arguments extend to $\xi_{(i)}^{(-)}$, and drop "(+)" everywhere.

Construct the set $\mathfrak{Z} = \mathfrak{Z}_1 \otimes \cdots \otimes \mathfrak{Z}_k$ of matrices $\xi = [\xi_{(1)}, \ldots, \xi_{(k)}] \in \Gamma^*$, $\xi_{(i)} = [\xi_{(i),j}]_{j=1}^k$, from

$$\xi_{(i)} \in \mathfrak{Z}_i := \left\{ \underset{\gamma\in\Gamma}{\operatorname{argsup}} \left\{ \frac{\partial}{\partial \gamma_i} \mu(\gamma) \right\} \right\} = \left\{ \underset{\gamma\in\Gamma}{\operatorname{argsup}} \left\{ E\left[\epsilon_t \frac{\partial}{\partial \gamma_i} F(\gamma' x_t) \right] \right\} \right\}.$$

In general $\mathfrak{Z}$ may contain more than one element under either hypothesis, and may have zero or positive Lebesgue measure. Under the null hypothesis, for example,

$$\frac{\partial}{\partial \gamma} \mu(\gamma) = E\left[\epsilon_t x_t F'(\gamma' x_t)\right] = 0 \qquad \text{for every } \gamma \in \Gamma;$$

hence $\mathfrak{Z} = \Gamma^*$ contains uncountably infinitely many elements. Similarly, under misspecification H_1, if $\mu(\gamma)$ is nonmonotonic, then $\arg\sup_{\gamma\in\Gamma} \{E[\epsilon_t x_t F'(\gamma' x_t)]\}$ need not be unique.

Now construct a vector weight function (note $x_{1,t} = 1$):

$$w_t(\gamma, \xi) = \left[F(\gamma' x_t); x_{1,t} F'\left(\xi'_{(1)} x_t\right), \ldots,\ x_{k,t} F'\left(\xi'_{(k)} x_t\right)\right]' \in \mathbb{R}^{k+1}.$$

The weight $w_t(\gamma, \xi)$ is perfectly revealing in the sense that the set $S = \cap_{i=1}^k \{\gamma \in \Gamma : E[\epsilon_t w_t(\gamma, \xi)] = 0, \text{ and } P(\gamma' x_t \in R_0) = 1\}$ contains only $\gamma = 0$, or is empty.

Theorem A.1. *Assume Assumptions* 1 *and* 2, *and* $P(E[\epsilon_t|\Im_{t-1}] = 0) < 1$ *hold. Then* $S = \{0\}$ *if and only if* ϵ_t *is* L_2*-orthogonal to the closed linear span of* $\{x_{i,t} F'(\xi'_{(i)} x_t)\}_{i=1}^k$, *and otherwise* S *is empty.*

Example 1 (Exponential Weight). Let $F(u) = \exp\{u\}$, and assume x_t is bounded uniformly in t. If $P[E(\epsilon_t|\Im_{t-1}) = 0] < 1$, then $\forall \gamma \neq 0$

$$\left[E\left[\epsilon_t \exp\{\gamma' x_t\}\right], E\left[\epsilon_t x_{1,t} \exp\{\xi'_{(1)} x_t\}\right], \ldots, E\left[\epsilon_t x_{k,t} \exp\{\xi'_{(k)} x_t\}\right]\right]' \neq 0.$$

Example 2 (Linearity). Suppose $\xi = 0$ such that $F'(\xi'_{(i)} x_t) = F'(0' x_t) = d$ by Assumption 1. If $d = 0$, then $0 \in S$, and if $d \neq 0$, then $0 \in S$ *if and only if* $E[\epsilon_t x_t] = 0$.

For example, consider testing linearity $f_t(\phi) = \phi' x_t$. If we estimate ϕ_0 by OLS as exactly identified GMM (i.e., $c_t(\phi) = x_t$), and $\xi = 0$ and $F'(0' x_t) \neq 0$, then $E[\epsilon_t x_t F'(\xi'_{(i)} x_t)] = d \times E[\epsilon_t x_t] = 0$ under the GMM identification condition. Hence $S = \{0\}$ by Theorem A.2.

A.2. Asymptotic Theory

Standard arguments for expanding GMM estimators under smoothness conditions are presented in Newey and McFadden (1994). By the mean-value-theorem and Assumption 3 in Hill (2011), under H_1^L we have the exanpsion

$$\hat{z}_n(\gamma) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t w_t(\gamma,\xi) + A(\phi_0,\gamma,\xi)\frac{1}{\sqrt{n}}\sum_{t=1}^{n} m_t(\phi) + \frac{1}{n}\sum_{t=1}^{n} u_t g_t(\gamma,\xi) + r_n(\gamma) = z_n(\gamma) + r_n(\gamma),$$

say, where the remainder $r_n(\gamma) \in \mathbb{R}^{k+1}$ satisfies $\sup_{\gamma\in\Gamma}|r_n(\gamma)| \xrightarrow{p} 0$ under Assumption 3, $g_t(\gamma,\xi) := w_t(\gamma,\xi) + A(\phi,\gamma,\xi) \times c_t(\phi)$, and $A(\phi,\gamma,\xi) := \text{plim}_{n\to\infty}\widehat{A}(\phi,\gamma,\xi)$ uniformly in $\{\phi,\gamma,\xi\}$. Define

$$\mu(\gamma) := \underset{n\to\infty}{\text{plim}}\frac{1}{n}\sum_{t=1}^{n} u_t g_t(\gamma,\xi), \quad \Sigma(\gamma_1,\gamma_2) := \underset{n\to\infty}{\text{plim}}\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 g_t(\gamma_1,\xi)\, g_t(\gamma_2,\xi)',$$

and hence $\Sigma(\gamma) = \Sigma(\gamma,\gamma)$.

The weak limit under H_1^L in Theorem 2.1 requires a revealing test weight $w_t(\gamma,\xi)$ characterized by Theorem A.1, a pointwise central limit theorem and tightness, and a uniformly consistent variance estimator $\widehat{\Sigma}_n(\gamma)$. The asymptotic results stated below follow from Assumption 3, and theory developed in Hill (2008).

Lemma A.2. *Let $z(\gamma)$ denote a Gaussian element of $\mathbb{C}[\Gamma_\iota]$ with mean function $\Sigma(\gamma)^{-1/2}\mu(\gamma)$ and covariance function $E[z(\gamma_1)z(\gamma_2)'] := \Sigma(\gamma_1)^{-1/2}\Sigma(\gamma_1,\gamma_2)\Sigma(\gamma_2)^{-1/2}$. Under Assumptions* 1 *and* 3 *and H_1^L, $\Sigma(\gamma)^{-1/2}z_n(\gamma) \xrightarrow{d} z(\gamma)$ pointwise in Γ_ι. Moreover, $|\Sigma(\gamma)^{-1/2}z_n(\gamma)/\sqrt{n}| \to \infty$ with probability one under H_1^G for every $\gamma \in \Gamma_\iota$.*

Lemma A.3. *Under Assumptions* 1 *and* 3 *and H_1^L, the sequence $\{\Sigma(\gamma)^{-1/2}(z_n(\gamma) - \mu(\gamma))\}$ is tight on Γ_ι.*

Lemma A.4. *Under Assumptions* 1 *and* 3, $\sup_{\gamma\in\Gamma}|\widehat{\Sigma}_n(\gamma) - \Sigma(\gamma)| \xrightarrow{p} 0$.

Proof of Theorem 2.1. Under H_1^L use Lemmas A.2–A.4 to deduce $\widehat{\Sigma}_n(\gamma)^{-1/2}\hat{z}_n(\gamma) \Rightarrow z(\gamma)$ on $\mathbb{C}[\Gamma_\iota]$ where $z(\gamma)$ is defined in Lemma A.2. Now apply the mapping theorem to prove the claim under H_1^L. The remaining claims follow instantly from Lemma A.2 and the mapping theorem. $\mathcal{QED}$. □

ACKNOWLEDGMENT

This research benefited from the suggestions of an associate editor, two anonymous referees, and editor Esfandiar Maasoumi.

REFERENCES

An, H. Z., Huang, F. C. (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica* 6:943–956.

Andrews, D. W. K., Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 82:1383–1414.

Bai, J., Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* 66:47–78.

Baillie, R. T. (1980). Predictions from ARMAX models. *Journal of Econometrics* 12:365–374.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometric* 20:105–134

Bierens, H. J. (1987). ARMAX Model specification testing, with an application to unemployment in the netherlands. *Journal of Econometrics* 35:161–190.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica* 58:1443–1458.

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.

Bierens, H. J., Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* 65:1129–1151.

Carrasco, M., Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18:17–39.

Carrasco, M., Hu, L., Ploberger, W. (2009). *Optimal Test for Markov Switching Parameters.* Mimeo: University of Montreal.

Corradi, V., Swanson, N. (2002). A consistent test for nonlinear out of sample predictive accuracy. *Journal of Econometrics* 110:353–381.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64:247–254.

de Haan, L., Leadbetter, M. R., Rootzen, H. (1991). *Tail and Quantile Estimation for Strongly Mixing Stationary Sequences.* Technical Report, Department of Statistics. University of North Carolina-Chapel Hill.

de Jong, R. (1996). The bierens test under data dependence. *Journal of Econometrics* 72:1–32.

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Annals of Statistics* 27:1012–1040.

Fan, Y., Gencay, R. (1995). A consistent nonparametric test of symmetry in linear regression models. *Journal of the American Statistical Association* 90:551–557.

Fan, Y., Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica* 64:865–890.

Fan, Y., Li, Q. (2002). A consistent model specification test based on kernel sum of squares of residuals. *Econometric Reviews* 21:337–352.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics* 12:299–308.

Gallant, A. R. (1981). Unbiased determination of production technologies. *Journal of Econometrics* 20:285–323.

Gallant, A. R., White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* New York: Basil Blackwell.

Hamilton, J. D. (1994). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:357–384.

Hansen, B. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64:413–430.

Hausman, J. A. (1978). Specification testing in econometrics. *Econometrica* 46:1251–1271.

Hill, J. B. (2008). Consistent and non-degenerate model specification tests against smooth transition and neural network alternatives. *Annales D'Economie et de Statistique* 90:145–179.

Hill, J. B. (2011). *Consistent GMM Residuals-Based Tests of Functional Form: Appendix B: Omitted Proofs and Simulation Results*. Department of Economics, University of North Carolina – Chapel Hill. The appendix is available upon request, or from www.unc.edu/~jbill/append_super_con_cm.pdf.

Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks* 2:359–366.

Hong, Y., White, H. (1995). Consistent specification testing via nonparametric series regression. *Econometrica* 63:1133–1159.

Hsiao, C., Li, Q., Racine, J. S. (2007). A consistent model specification test with mixed discrete and continuous data. *Journal of Econometrics* 128:802–826.

Koul, H. L., Stute, W. (1999). Nonparametric Model Checks for Time Series. *The Annals of Statistics* 27:204–236.

Lee, T., White, H., Granger, C. W. J. (1993). Testing for neglected nonlinearity in time-series models: a comparison of neural network methods and alternative tests. *Journal of Econometrics* 56:269–290.

Li, Q., Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* 87:145–165.

Luukkonen, R., Saikkonen, P., Terasvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika* 75:491–9.

Maasoumi, E., Racine, J., Granger, C. (2004). A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis* 25:649–699.

McLeod, A. I., Li, W. K. (1983). Diagnostic checking arma time series models using squared residual autocorrelations. *Journal of Time Series Analysis* 4:269–273.

Meitz, M., Saikkonen, P. (2008). Stability of nonlinear AR-GARCH models. *J. Time Ser. Anal.* 29:453–475.

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

Newey, W. K., McFadden, D. L. (1994). Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D. eds., *Handbook of Econometrics*. Vol. 4. Amsterdam: North-Holland.

Nze, P. A., Doukhan, P. (2004). Weak dependence: models and applications to econometrics. *Econometric Theory* 20:995–1045.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression. *Journal of the Royal Statistical Society Series B* 31:350–371.

Rao, T. S., Gabr, M. M. (1980). A test for linearity of time series. *Journal of Time Series Analysis* 1:145–158.

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

Stute, W. (1997). Nonparametric Model Checks for Regression. *The Annals of Statistics* 25:613–641.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89:208–218.

Tong, H. (1990). *Non-Linear Time Series: A Dynamic System Approach*. Oxford: Oxford University Press.

Tsay, R. (1986). Nonlinearity tests for time series.*Biometrika* 73:461–466.

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

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*. Vol. 2. New York, Washington D.C.: IEEE Press.

Yatchew, A. J. (1992). Nonparametric Regression Tests Based on Least Squares. *Econometric Theory* 8:435–451.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75:263–289.