Testing in GMM Models Without Truncation

Timothy J. Vogelsang*

Departments of Economics and Statistical Science, Cornell University

First Version August, 2000; This Version June, 2001

Abstract

This paper proposes a new approach to testing in the generalized method of moments (GMM) framework. The new tests are constructed using heteroskedasticity autocorrelation (HAC) robust standard errors computed using nonparametric spectral density estimators without truncation. While such standard errors are not consistent, a new asymptotic theory shows that they lead to valid tests nonetheless. In an over-identified linear instrumental variables model, simulations suggest that the new tests and the associated limiting distribution theory provide a more accurate first order asymptotic null approximation than standard HAC robust tests. Finite sample power of the new tests is shown to be comparable to standard tests. Because use of a truncation lag equal to the sample requires no additional choices for practitioners, the new approach could potentially lead to a standard of practice (which does not currently exist) for the computation of HAC robust standard errors in GMM models.

Keywords: HAC estimators, inference, autocorrelation, covariance matrix estimation, kernel, truncation lag, prewhitening.

1 Introduction

The generalized method of moments (GMM) estimation method has now become one of the standard methodologies in econometrics since it was first introduced to the econometrics literature by the influential paper Hansen (1982). GMM is widely used in empirical macroeconomics and finance. GMM is appealing because it can deliver consistent estimates of parameters in models where likelihood functions are either hard or impossible to write down. The class of GMM models is large and includes such special cases as linear regression, nonlinear regression, instrumental variables (IV), and maximum likelihood.

In macroeconomic and finance applications, heteroskedasticity and/or autocorrelation of unknown form is usually an important specification issue. In many cases, GMM estimators are consistent in spite of heteroskedasticity and/or autocorrelation and can have certain optimality properties. Therefore, heteroskedasticity and/or serial correlation is not a problem, per say, for estimation, but it does affect inference in that standard errors robust to heteroskedasticity and/or serial correlation are required. Such standard errors are often called HAC robust standard errors because they are computed using heteroskedasticity autocorrelation consistent covariance matrix estimators.

A key component in constructing HAC robust standard errors is the estimation of the spectral density at frequency zero of the random process that underlies the moment conditions that define a GMM estimator. The HAC robust standard error literature in econometrics has grown from and extended the spectral density estimation literature in time series statistics. Usually, nonparametric estimators have been proposed although parametric estimators have been receiving some attention recently. Some key contributions to the nonparametric approach include White (1984), Newey and West (1987), Gallant (1987), Gallant and White (1988), Andrews (1991), Andrews and Monahan (1992), Hansen (1992) and Robinson (1998). Recent papers by den Haan and Levin (1997,1998) have argued that parametric estimators based on vector autoregression (VAR) approximations can have certain advantages over the nonparametric approach.

A practical problem with both the nonparametric and parametric approach to constructing HAC robust standard errors is the need to choose "tuning parameters" such as a truncation lag or bandwidth in the nonparametric approach or the lag length in the parametric approach. In both cases, asymptotic theory designed to determine conditions under which HAC robust standard errors will be consistent only provides conditions on the rates at which truncation lags or VAR lags must grow as the sample size increases. However, an empirical researcher faced with a finite sample could choose any truncation lag or VAR lag and justify that choice using clever asymptotic

arguments. Thus, no standard has emerged for the computation of HAC robust standard errors in practice. This situation contrasts models with only heteroskedasticity where the approach proposed by White (1980) has become an empirical standard of practice.

The tuning parameter problem has not gone unnoticed in the HAC literature and many data dependent methods have been proposed for choosing the tuning parameters. The practical drawback to the data dependent approach is that the choice of truncation lag or VAR lag length is replaced with choices such as approximating model of serial correlation (nonparametric approach) or maximal VAR lag length (parametric approach). Data dependent methods, while important improvements over the basic case, ultimately require practical choices and do not establish a standard of practice for the computation of HAC robust standard errors.

The goal of this paper is to propose a different approach to the computation of HAC robust standard errors in GMM models that could potentially lead to a standard of practice in empirical applications. The results in this paper build upon and extend to the GMM framework the approach proposed by Kiefer and Vogelsang (2000). The basic idea is to shed the notion that valid standard errors can only be obtained by searching for consistent covariance matrix estimators. Consistency is a sufficient, but not necessary condition, for obtaining valid standard errors. The class of standard errors considered here are those constructed using nonparametric spectral density estimators but without truncation; the truncation lag is chosen to be the sample size. This new approach requires a new asymptotic theory for HAC robust tests. Kiefer and Vogelsang (2000) developed the required distribution theory for linear regression models. It was not obvious that the results obtained by Kiefer and Vogelsang (2000) extend to over-identified GMM models. However, the new asymptotic theory does go through smoothly for over-identified GMM models. This result is the main theoretical contribution of the paper.

An additional advantage of the new approach is a more accurate first order asymptotic approximation of the finite sample null distributions of t and F tests. It has been well documented that tests based on traditional HAC robust standard errors can perform badly in finite samples and are subject to substantial size distortions. See for example, Andrews (1991), Andrews and Monahan (1992), den Haan and Levin (1997), and, with a focus on GMM, a special issue of the *Journal of Business and Economic Statistics* (volume 14, July 1996). Size distortions associated with the new tests are smaller and yet power remains comparable to standard tests.

The remainder of the paper is organized as follows. The next section describes the model and gives assumptions sufficient for the main theoretical results. Section 3 reviews the standard approach to testing in the GMM framework. The new approach is given in Section 4 along with the theoretical results of the paper. Section 5 illustrates the new tests in the special of linear IV

2

estimation. A finite sample simulation study compares and contrasts the new tests with standard tests. Section 6 concludes with proofs collecting in an appendix.

Finally, the following notation is used throughout the paper. The symbol $\Rightarrow$ denotes weak convergence, $B_j(r)$ denotes a j vector of standard brownian motions (Wiener processes) defined on $r \in [0,1]$, $\widetilde{B}_j(r) = B_j(r) - rB_j(1)$ denotes a j vector of standard Brownian bridges, and $[rT]$ denotes the integer part of rT for $r \in [0,1]$.

2 The Model and Assumptions

Suppose we are interesting in estimating the $p \times 1$ vector of parameters, θ, where θ_0 denotes the true value of θ. Let v_t denote a vector of observed data for $t = 1, 2, ..., T$. Assume that q moment conditions hold for the data and that these moment conditions can be written as

$$E[f(v_t, \theta_0)] = \mathbf{0}, \tag{1}$$

where $f(\cdot)$ is a $q \times 1$ vector of functions with $q \geq p$. The moment conditions given by (1) are often derived from economic models. The basic idea of GMM estimation is to find a value for θ that most closely satisfies an empirical analog of the moment conditions (1). Define $g_t(\theta) = T^{-1} \sum_{j=1}^{t} f(v_j, \theta)$ where $g_T(\theta) = T^{-1} \sum_{t=1}^{T} f(v_t, \theta)$ can be viewed as the sample analog to (1). Then, the GMM estimator of θ based on a sample of T observations is defined as

$$\widehat{\theta}_T = \arg \min_{\theta \in \Theta} g_T(\theta)' W_T g_T(\theta)$$

where W_T is a $q \times q$ positive definite weighting matrix. $\widehat{\theta}_T$ can also be defined in terms of the solution to the first order conditions (FOC) of the minimization problem

$$G_T(\widehat{\theta}_T)' W_T g_T(\widehat{\theta}_T) = \mathbf{0}, \tag{2}$$

where $G_t(\theta) = T^{-1} \sum_{j=1}^{t} \partial f(v_j, \theta) / \partial \theta'$. When the model is exactly identified and $q = p$, the weighting matrix becomes irrelevant and $\widehat{\theta}_T$ is defined by the equation $g_T(\widehat{\theta}_T) = \mathbf{0}$.

A function related to $G_t(\theta)$ will be important for technical developments. Application of the mean value theorem implies that

$$g_t(\widehat{\theta}_T) = g_t(\theta_0) + G_t(\widehat{\theta}_T, \theta_0, \lambda_T)(\widehat{\theta}_T - \theta_0) \tag{3}$$

where $G_t(\widehat{\theta}_T, \theta_0, \lambda_T)$ denotes the $q \times p$ matrix whose i^{th} row is the corresponding row of $G_t(\overline{\theta}_T^{(i)})$ where $\overline{\theta}_T^{(i)} = \lambda_{i,T} \theta_0 + (1 - \lambda_{i,T}) \widehat{\theta}_T$ for some $0 \leq \lambda_{i,T} \leq 1$ and λ_T is the $q \times 1$ vector with i^{th} element $\lambda_{i,T}$.

Because the focus of this paper is on hypothesis testing, it is taken as given that $\widehat{\theta}_T$ is consistent and asymptotically normally distributed. So, rather than focus on well known regularity conditions under which $\widehat{\theta}_T$ is consistent and asymptotically normally distributed (consult Hansen (1982) and Newey and McFadden (1994)), attention is given to high level assumptions that are sufficient for obtaining the main results of the paper. To that end, the following assumptions are used:

Assumption 1 $p\lim\widehat{\theta}_T = \theta_0$.

Assumption 2 $T^{-1/2}\sum_{t=1}^{[rT]} f(v_t, \theta_0) = T^{1/2}g_{[rT]}(\theta_0) \Longrightarrow \Lambda B_q(r)$ where $\Omega = \Lambda\Lambda' = \sum_{j=-\infty}^{\infty}\Gamma_j$, $\Gamma_j = cov[f(v_t, \theta_0), f(v_{t-j}, \theta_0)]$.

Assumption 3 $p\lim G_{[rT]}(\widehat{\theta}_T) = rG_0$ and $p\lim G_{[rT]}(\widehat{\theta}_T, \theta_0, \lambda_T) = rG_0$ uniformly in $r \in [0,1]$ where $G_0 = E[\partial f(v_t, \theta)/\partial\theta']$.

Assumption 4 W_T is positive semi-definite and $p\lim W_T = W_\infty$ where W_∞ is a matrix of constants.

These assumptions are fairly standard with the exception of Assumption 2. Assumption 2 requires that a functional central limit theorem apply to $T^{1/2}g_t(\theta_0)$. Asymptotic normality of $\widehat{\theta}_T$ requires the less stringent assumption that a central limit theorem apply to $T^{1/2}g_T(\theta_0)$. In the standard approach, however, because Ω must be consistently estimated in order to construct asymptotically valid tests regarding θ_0, regularity conditions needed to obtain a consistent estimate of Ω are more than sufficient for Assumption 2 to hold. For example, a typical regularity condition for consistent estimation of Ω using spectral kernel methods is that $f(v_t, \theta_0)$ be a mean zero fourth order stationary process that is $\alpha - mixing$ (see e.g. Andrews (1991)). Assumption 2 holds under the milder assumption that $f(v_t, \theta_0)$ is a mean zero $2 + \delta$ order stationary process (for some $\delta > 0$) that is $\alpha - mixing$ (see e.g. Phillips and Durlauf (1986)). Therefore, the assumptions used in this paper are slightly weaker than the usual assumptions required for asymptotically valid testing in GMM models.

3 Asymptotic Normality and Covariance Matrix Estimation

The typical starting point for inference regarding the parameter vector θ_0 is an asymptotic normality result for $\widehat{\theta}_T$. The following lemma provides the foundation upon which the test statistics proposed in this paper are built. A proof is given in the appendix.

Lemma 1 *Under Assumptions 1 - 4, as $T \to \infty$,*

$$T^{1/2}(\widehat{\theta}_T - \theta_0) \Rightarrow -(G_0'W_\infty G_0)^{-1}G_0'W_\infty\Lambda B_q(1) \equiv -(G_0'W_\infty G_0)^{-1}\Lambda^* B_p(1),$$

4

where $\Lambda^*\Lambda^{*\prime} = G_0'W_\infty\Lambda\Lambda'W_\infty G_0$.

Given that $B_p(1)$ is a p-vector of standard normal random variables, it immediately follows from the lemma that $T^{1/2}(\widehat{\theta}_T - \theta_0) \to^d N(0, V)$ where $V = (G_0'W_\infty G_0)^{-1}\Lambda^*\Lambda^{*\prime}(G_0'W_\infty G_0)^{-1}$. Exploiting the asymptotic normality of $\widehat{\theta}_T$, asymptotically valid test statistics regarding θ_0 can be constructing in the usual way (i.e. t and Wald tests) provided a consistent estimator of the asymptotic covariance matrix, V, is available. The $(G_0'W_\infty G_0)^{-1}$ term of V is easy to consistently estimate using $[G_T'(\widehat{\theta}_T)W_T G_T(\widehat{\theta}_T)]^{-1}$. The middle term of V is more difficult to estimate. If we write $\Lambda^*\Lambda^{*\prime} = G_0'W_\infty\Lambda\Lambda'W_\infty G_0 = G_0'W_\infty\Omega W_\infty G_0$, then the middle term can be consistently estimated using $G_T'(\widehat{\theta}_T)W_T\widehat{\Omega}W_T G_T(\widehat{\theta}_T)$ where $\widehat{\Omega}$ is a consistent estimator of Ω. Therefore, V can be consistently estimated using

$$\widehat{V} = [G_T'(\widehat{\theta}_T)W_T G_T(\widehat{\theta}_T)]^{-1}G_T'(\widehat{\theta}_T)W_T\widehat{\Omega}W_T G_T(\widehat{\theta}_T)[G_T'(\widehat{\theta}_T)W_T G_T(\widehat{\theta}_T)]^{-1}.$$

Recall from Assumption 2 that Ω is the infinite sum of the autocovariances of $f(v_t, \theta_0)$. It is well known that Ω is equal to 2π times the spectral density matrix of $f(v_t, \theta_0)$ evaluated at frequency zero. Therefore, it has become standard in the GMM framework to use spectral density estimators from the time series literature as a way of estimating Ω.

One of the most popular classes of spectral density estimators, the nonparametric class, take the form

$$\widehat{\Omega} = \sum_{j=-(T-1)}^{T-1} k(j/M)\widehat{\Gamma}_j \tag{4}$$

with

$$\widehat{\Gamma}_j = T^{-1}\sum_{t=j+1}^{T} f(v_t, \widehat{\theta}_T)f(v_{t-j}, \widehat{\theta}_T)' \ for \ j \geq 0,$$

$$\widehat{\Gamma}_j = T^{-1}\sum_{t=-j+1}^{T} f(v_{t+j}, \widehat{\theta}_T)f(v_t, \widehat{\theta}_T)' \ for \ j < 0,$$

where $k(x)$ is a kernel function satisfying $k(x) = k(-x)$, $k(0) = 1$, $|k(x)| \leq 1$, $k(x)$ continuous at $x = 0$ and $\int_{-\infty}^{\infty} k^2(x)dx < \infty$. Often $k(x) = 0$ for $|x| > 1$ so M "trims" the sample autocovariances and acts as a truncation lag. For $\widehat{\Omega}$ to be consistent, $M \to \infty$ and $M/T \to 0$ as $T \to \infty$. The technical requirement that M grows at a suitable rate has long been recognized as an important practical limitation to nonparametric spectral density estimators. The problem in practice is that for any finite sample of size T, any choice of $M \leq T$ can be justified and the "suitable rate" can be satisfied by the "fictional promise" that M will be chosen in certain ways should more data become

available. For example, suppose we adopt the rule that $M = CT^{1/3}$ where C is a finite constant. If $T = 100$ and the practitioner decides that $M = 29$ gives a desired result, then this choice of M could be justified on the grounds of consistency of $\widehat{\Omega}$ by using $C = 6.2478$. Then, in theory, the practitioner would have to use this value of C should more data become available. Presumably, the practitioner is already using all the available data: hence the "fictional promise". But, this promise in no way restricts the choice of M in the sample at hand, because for any choice of M there is always a corresponding choice of C that makes $\widehat{\Omega}$ consistent.

The arbitrary nature of the choice of M in finite samples has led to the development of data dependent choices of M most notably by Andrews (1991) and Newey and West (1994) in the context of covariance matrix estimation. The basic idea is to choose M to minimize the mean square error (MSE) of $\widehat{\Omega}$ (or some asymptotic approximation of the MSE). Because the MSE of $\widehat{\Omega}$ depends on the serial correlation structure of $f(v_t, \theta_0)$, the practitioner must choice an "approximate" model for the serial correlation of $f(v_t, \theta_0)$. Once this "approximate" model is chosen, then the choice of M becomes automatic. However, this approach simply replaces the choice of M with the choice of an approximate model for the serial correlation in $f(v_t, \theta_0)$.

Because of these practical problems inherent in nonparametric estimation of Ω, den Haan and Levin (1997,1998) recommend using parametric estimators of Ω based on VAR models fit to $f(v_t, \widehat{\theta}_T)$. The parametric approach can attain the same level of generality with regard to the nature of the serial correlation in $f(v_t, \theta_0)$ provided the lag of the VAR increases at a suitable rate as the sample size increases. Therefore, the choice of M is replaced by the choice of lag length. Again, data dependent methods have been proposed to help make this choice. But, in finite samples, data dependent lag length methods require the practitioner to choice a maximal lag length. There is no guide whatsoever on this choice of maximal lag length other than the maximal lag length must increase as T increases but not too fast. So, the choice of lag length is replaced with the choice of maximal lag length. Again, no standard of practice emerges.

4   A New Approach

Following Kiefer and Vogelsang (2000) consider estimating Ω using (4) but with $M = T$. This estimator is denoted by $\widehat{\Omega}_{M=T}$. This is a specific choice of M that uses information about all of the sample autocovariances in the data. It is well known that letting M grow at the same rate as T results in an inconsistent estimate of Ω. Thus, the corresponding estimate of V,

$$\widehat{V}_{M=T} = [G'_T(\widehat{\theta}_T) W_T G_T(\widehat{\theta}_T)]^{-1} G'_T(\widehat{\theta}_T) W_T \widehat{\Omega}_{M=T} W_T G_T(\widehat{\theta}_T) [G'_T(\widehat{\theta}_T) W_T G_T(\widehat{\theta}_T)]^{-1}, \qquad (5)$$

will not be consistent. However, it is important to keep in mind that a consistent estimate of V is only a sufficient condition for valid asymptotic testing, not a necessary condition. This fact is often overlooked, especially in textbook treatments of asymptotic testing. As shown below, use of $\widehat{\Omega}_{M=T}$ delivers an estimate of V that converges to a random variable that is proportional to V in a useful way and otherwise does not depend on unknown nuisance parameters. Therefore, asymptotically valid testing is still possible except that the distribution theory becomes nonstandard; i.e. t-tests will not be asymptotically normal.

In a sense, merely appealing to consistency of an estimator of V leads to a less than satisfactory asymptotic approximation for t-tests based on $\widehat{\theta}_T$. Consistency is best viewed as the minimally required asymptotic property for an estimator. For example, showing that $\widehat{\theta}_T$ is consistent is the natural first step when determining whether it is a useful estimator. But, if one is interested in testing hypotheses about θ_0, then consistency of $\widehat{\theta}_T$ is not enough, and an asymptotic normality result for $T^{1/2}(\widehat{\theta}_T - \theta_0)$ is required. In other words, testing requires at least a first order asymptotic approximation to the finite sample distribution of $\widehat{\theta}_T$. One should perhaps view consistency as a less than first order asymptotic approximation. When we replace V with a consistent estimator and then treat V as known asymptotically, we are using a constant to approximate the finite sampling variability of a random variable. In small samples, this approximation is likely to be inaccurate. In other words, the standard asymptotic approximation for the t-test ignores the finite sample variability of the estimate of V. By using $\widehat{\Omega}_{M=T}$ and a first order asymptotic approximation of its finite sampling behavior, it may be possible to obtain a more accurate asymptotic approximation for t and F tests regarding θ_0.

4.1 Asymptotic Distribution of $\widehat{\Omega}_{M=T}$.

In this subsection the asymptotic distribution of $\widehat{\Omega}_{M=T}$ is derived. When the model is exactly identified and $q = p$, then $\widehat{\Omega}_{M=T}$ will be asymptotically proportional to $\Lambda\Lambda'$ and it easy to show that the resulting estimate of V will be proportional to V. However, when the model is over-identified and $q > p$, then $\widehat{\Omega}_{M=T}$ will no longer be proportional to $\Lambda\Lambda'$ asymptotically[1]. This does not pose any problems because the middle term of the estimate of V, $G'_T(\widehat{\theta}_T)W_T\widehat{\Omega}_{M=T}W_TG_T(\widehat{\theta}_T)$, does have the required asymptotic proportionality to variance nuisance parameters. The details are as follows.

Let $K_{il} = k((i-l)/T)$. Using algebraic arguments similar to those used by Kiefer and Vogelsang

---

[1] Asymptotic proportionality to $\Lambda\Lambda'$ of $\widehat{\Omega}_{M=T}$ could be obtained by demeaning $f(v_t, \widehat{\theta}_T)$ before computing $\widehat{\Omega}_{M=T}$. However, it can be shown that the term $G'_T(\widehat{\theta}_T)W_T\widehat{\Omega}_{M=T}W_TG_T(\widehat{\theta}_T)$ is computationally equivalent whether or not $f(v_t, \widehat{\theta}_T)$ is demeaned before computing $\widehat{\Omega}_{M=T}$. On the other hand, as shown by Hall (2000), demeaning improves the power of tests for over-identified restrictions and is recommended.

(2000), it is straightforward to show that

$$\widehat{\Omega}_{M=T} = T^{-1} \sum_{l=1}^{T-1} \sum_{i=1}^{T-1} [(K_{il} - K_{i,l+1}) - (K_{i+1,l} - K_{i+1,l+1})] T g_i(\widehat{\theta}_T) T g_l(\widehat{\theta}_T)'$$

$$+ T^{-1} \sum_{l=1}^{T-1} (K_{Tl} - K_{T,l+1}) T g_T(\widehat{\theta}_T) T g_l(\widehat{\theta}_T)' + T^{-1} \sum_{i=1}^{T} f(v_i, \widehat{\theta}_T) K_{iT} T g_l(\widehat{\theta}_T)'. \quad (6)$$

Using (6) it directly follows that

$$G_T'(\widehat{\theta}_T) W_T \widehat{\Omega}_{M=T} W_T G_T(\widehat{\theta}_T)$$

$$= T^{-1} \sum_{l=1}^{T-1} \sum_{i=1}^{T-1} [(K_{il} - K_{i,l+1}) - (K_{i+1,l} - K_{i+1,l+1})] G_T'(\widehat{\theta}_T) W_T T g_i(\widehat{\theta}_T) T g_l(\widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T), \quad (7)$$

where the second and third terms of (6) vanish because from (2) it follows that $G_T'(\widehat{\theta}_T) W_T T g_T(\widehat{\theta}_T) = T g_T(\widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T) = 0$. Define the step function $D_T(r)$ for $r \in [0,1]$ as $D_T(r) = D(j)$ for $\frac{j}{T} \le r < \frac{j+1}{T}$, $j = 0, 1, 2, ..., T-1$ where $D(x) = [k((x+1)/T) - k(x/T)] - [k(x/T) - k((x-1)/T)]$. Simple algebra applied to (7) gives

$$G_T'(\widehat{\theta}_T) W_T \widehat{\Omega}_{M=T} W_T G_T(\widehat{\theta}_T)$$

$$= T^{-1} \sum_{l=1}^{T-1} T^{-1} \sum_{i=1}^{T-1} -T^2 D_T((i-l)/T) G_T'(\widehat{\theta}_T) W_T T^{1/2} g_i(\widehat{\theta}_T) T^{1/2} g_l(\widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T)$$

$$= \int_0^1 \int_0^1 -T^2 D_T(r-s) G_T'(\widehat{\theta}_T) W_T T^{1/2} g_{[rT]}(\widehat{\theta}_T) T^{1/2} g_{[sT]}(\widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T) dr ds. \quad (8)$$

When $k''(x)$ exists and is continuous on the interval $x \in (-1,1)$, it is easy to show that $T^2 D_T(r) \to k''(r)$ uniformly in r. Therefore, given (8), all that is needed to derive the asymptotic distribution of $G_T'(\widehat{\theta}_T) W_T \widehat{\Omega}_{M=T} W_T G_T(\widehat{\theta}_T)$ is the limit of $G_T'(\widehat{\theta}_T) W_T T^{1/2} g_{[rT]}(\widehat{\theta}_T)$ which is given by the following lemma that is proved in the appendix.

Lemma 2 *Under Assumptions 1-4, as $T \to \infty$,*

$$G_T'(\widehat{\theta}_T) W_T T^{1/2} g_{[rT]}(\widehat{\theta}_T) \Rightarrow G_0' W_\infty \left[\Lambda B_q(r) - r G_0 (G_0' W_\infty G_0)^{-1} G_0' W_\infty \Lambda B_q(1) \right]$$

$$\equiv G_0' W_\infty \Lambda \left[B_q(r) - r B_q(1) \right]$$

$$\equiv \Lambda^* \left[B_p(r) - r B_p(1) \right]$$

$$= \Lambda^* \widetilde{B}_p(r).$$

Using $T^2 D_T(r) \to k''(r)$ and applying Lemma 2 to (8) it follows from the continuous mapping theorem that

$$G_T'(\widehat{\theta}_T) W_T \widehat{\Omega}_{M=T} W_T G_T(\widehat{\theta}_T) \Rightarrow \Lambda^* \int_0^1 \int_0^1 -k''(r-s) \widetilde{B}_p(r) \widetilde{B}_p(s)' dr ds \Lambda^{*'}. \quad (9)$$

8

The reason that $G_T'(\widehat{\theta}_T) W_T \widehat{\Omega}_{M=T} W_T G_T(\widehat{\theta}_T)$ can be used for valid testing is because it is asymptotically proportional to $\Lambda^* \Lambda^{*\prime}$ and the random variable $\int_0^1 \int_0^1 -k''(r-s)\widetilde{B}_p(r)\widetilde{B}_p(s)' dr ds$ does not depend on unknown nuisance parameters.

4.2 Inference Without Truncation

In this section it is shown how asymptotically valid (pivotal) t and F statistics can be constructed using the inconsistent estimate of V based on $\widehat{\Omega}_{M=T}$. Suppose the hypothesis of interest can be written as

$$H_0 : r(\theta_0) = 0$$
$$H_1 : r(\theta_0) \neq 0,$$

where $r(\theta_0)$ is a $m \times 1$ vector ($m \leq p$) of continuously differentiable functions with derivative matrix $R(\theta_0) = \partial r(\theta_0)/\partial \theta'$. Application of the Delta method gives

$$T^{1/2} r(\widehat{\theta}_T) \Rightarrow -R(\theta_0) V^{1/2} B_p(1) \equiv N(\mathbf{0}, V_R)$$

where $V_R = R(\theta_0) V R(\theta_0)'$. Suppose we use the inconsistent estimator of V_R, $R(\widehat{\theta}_T)\widehat{V}_{M=T}R(\widehat{\theta}_T)'$, where $\widehat{V}_{M=T}$ is given by (5) to construct the F statistic

$$F^* = Tr(\widehat{\theta}_T)' \left(R(\widehat{\theta}_T)\widehat{V}_{M=T}R(\widehat{\theta}_T)' \right)^{-1} r(\widehat{\theta}_T)/m.$$

The only difference between F^* and a conventional F test is that $\widehat{\Omega}_{M=T}$ is used instead of a consistent estimator of Ω. In the case where $m = 1$ so that only one restriction is being testing, we can construct a t statistic given by

$$t^* = \frac{T^{1/2} r(\widehat{\theta}_T)}{\sqrt{R(\widehat{\theta}_T)\widehat{V}_{M=T}R(\widehat{\theta}_T)'}}.$$

A practically relevant t test is the test of significance of the individual parameters

$$H_0 : \theta_i = 0$$
$$H_1 : \theta_i \neq 0.$$

In this case the t^* statistic can be written as

$$t^* = \frac{\widehat{\theta}_i}{se(\widehat{\theta}_i)},$$

where $se(\widehat{\theta}_i) = \sqrt{T^{-1}\widehat{V}_{M=T}^{ii}}$ and $\widehat{V}_{M=T}^{ii}$ is the ith diagonal element of the $\widehat{V}_{M=T}$ matrix. A theorem which is proved in the appendix establishes the limiting null distributions of t^* and F^*.

9

Theorem 1 *Under Assumptions 1-4, as $T \to \infty$,*

$$F^* \Rightarrow B_m(1)' \left(\int_0^1 \int_0^1 -k''(r-s)\widetilde{B}_m(r)\widetilde{B}_m(s)'drds \right)^{-1} B_m(1)/m,$$

$$t^* \Rightarrow \frac{B_1(1)}{\sqrt{\int_0^1 \int_0^1 -k''(r-s)\widetilde{B}_1(r)\widetilde{B}_1(s)'drds}}.$$

According to the theorem, the asymptotic distributions are free of nuisance parameters although the distributions are nonstandard and depend on the kernel through $k''(x)$. Given the kernel, critical values are easily obtained using simulation methods. Therefore, using the truncation lag of $M = T$ when estimating the asymptotic variance delivers an asymptotically valid class of tests even though the asymptotic variance is not consistently estimated. The practical advantage of using $M = T$ is that this choice of truncation lag is not arbitrary in finite samples. What remains, though, is the choice of kernel.

4.3   The Bartlett Kernel

The dependence of the asymptotic distributions of F^* and t^* on the kernel can be used to guide the choice of kernel in practice. Kiefer and Vogelsang (2000) showed that among a large group of popular kernels, the Bartlett kernel delivers the highest power for the t^* statistic[2], and they recommend the use of the Bartlett kernel when using $M = T$. Their power analysis easily extends to the GMM framework given Theorem 1. Therefore, it is instructive to analyze the F^* and t^* statistics for the case of the Bartlett kernel.

The Bartlett kernel is given by $k(x) = 1 - |x|$ for $|x| \leq 1$ and $k(x) = 0$ otherwise. Note that $k''(x) = 0$ for $x \neq 0$, but $k''(0)$ does not exist. Kiefer and Vogelsang (2000) showed that heuristically we can define $k''(0) = -2$. Then, when the Bartlett kernel is used, the limiting distributions of F^* and t^* simply to

$$F^* \Rightarrow B_m(1)' \left(2 \int_0^1 \widetilde{B}_m(r)\widetilde{B}_m(r)'dr \right)^{-1} B_m(1)/m,$$

$$t^* \Rightarrow \frac{B_1(1)}{\sqrt{2 \int_0^1 \widetilde{B}_1(r)^2 dr}}.$$

A formal proof of these results follows directly from results in Kiefer and Vogelsang (2001). The distribution of t^* is mixture normal. The density of $B_1(1)/\sqrt{\int_0^1 \widetilde{B}_1(r)^2 dr}$ was derived by Abadir

---

[2]It is interesting to note that the testing approach proposed by Kiefer, Vogelsang and Bunzel (2000) is exactly equivalent to using $\frac{1}{2}\widehat{\Omega}_{M=T}$ with the Bartlett kernel.

and Paruolo (1997) so that analytical critical values are available for t^*. Critical values for t^* are tabulated in Table 1 and were computed by scaling the critical values of $B_1(1)/\sqrt{\int_0^1 \widetilde{B}_1(r)^2 dr}$ by $1/\sqrt{2}$. Analytical representations of the density of F^* have not been derived; however, critical values are easy to obtain using simulation methods. Critical values for F^* for $m = 1, 2, ..., 30$ are tabulated in Table 2.

5   Linear IV Models: Choice of W_T and Finite Sample Performance

In this section attention is focused on the special case of linear instrumental variables (IV) models. Beyond serving as an example, this section has two additional goals. The first goal is to explore the choice of the weighting matrix, W_T, which is greatly simplified in the linear case because the formula for $\widehat{\theta}_T$ can be written in a closed form. The second goal is to illustrate the finite sample performance of the new tests, and this is also simplified (in terms of computation complexity) in the linear model.

5.1   Linear IV Model

Let x_t denote a $p \times 1$ vector of regressors. Consider the linear regression

$$y_t = x_t'\theta_0 + u_t,$$

where u_t is a mean zero error term and x_t and u_t could be correlated. Suppose that a $q \times 1$ vector of instruments, z_t, are available that satisfy the moment conditions $E(z_t u_t) = \mathbf{0}$. For the instruments to be valid we also need the assumption that $cov(x_t, z_t) \neq \mathbf{0}$. In the general notation from above, we have $v_t = (y_t, x_t', z_t')'$ and $f(v_t, \theta_0) = z_t u_t = z_t(y_t - x_t'\theta_0)$. Then it follows that $g_T(\theta) = T^{-1} \sum_{t=1}^{T} z_t(y_t - x_t'\theta)$. Given a weighting matrix, W_T, it is easy to show that

$$\widehat{\theta}_T = \left[\left(\sum_{t=1}^{T} z_t x_t' \right)' W_T \left(\sum_{t=1}^{T} z_t x_t' \right) \right]^{-1} \left[\left(\sum_{t=1}^{T} z_t x_t' \right)' W_T \left(\sum_{t=1}^{T} z_t y_t \right) \right]. \tag{10}$$

The asymptotic variance of $\widehat{\theta}_T$ is

$$V = (Q_{zx}'W_\infty Q_{zx})^{-1} Q_{zx}'W_\infty \Omega W_\infty Q_{zx}(Q_{zx}'W_\infty Q_{zx})^{-1},$$

where $Q_{zx} = p\lim(T^{-1} \sum_{t=1}^{T} z_t x_t')$ and $\Omega = \lim_{T \to \infty} \mathrm{VAR}(T^{-1/2} \sum_{t=1}^{T} z_t u_t)$. The asymptotic variance matrix can be consistently estimated using

$$\widehat{V} = (\widehat{Q}_{zx}'W_T\widehat{Q}_{zx})^{-1}\widehat{Q}_{zx}'W_T\widehat{\Omega}W_T\widehat{Q}_{zx}(\widehat{Q}_{zx}'W_T\widehat{Q}_{zx})^{-1},$$

where $\widehat{Q}_{zx} = T^{-1} \sum_{t=1}^{T} z_t x_t'$ and $\widehat{\Omega}$ is given by (4) using $f(v_t, \widehat{\theta}_T) = z_t \widehat{u}_t = z_t(y_t - x_t'\widehat{\theta}_T)$.

11

As shown by Hansen (1982), the weighting matrix that gives the optimal (minimum asymptotic variance) GMM estimator is $W_T = \widehat{\Omega}^{-1}$. In this case, the asymptotic variance of $\widehat{\theta}_T$ simplifies to $V_{opt} = (Q'_{zx}\Omega^{-1}Q_{zx})^{-1}$ and the corresponding estimator simplifies to $\widehat{V}_{opt} = (\widehat{Q}'_{zx}\widehat{\Omega}^{-1}\widehat{Q}_{zx})^{-1}$. Because $\widehat{\Omega}$ is constructed using $\widehat{\theta}_T$, there can be improvements in the precision of $\widehat{\Omega}$ through iteration. Let $\widehat{\theta}^0_T$ denote an initial estimate of θ obtained with an arbitrary weighting matrix (e.g. $W_T = I$). Let $\widehat{\Omega}(\widehat{\theta}^0_T)$ denote the estimate of Ω obtained using $\widehat{\theta}^0_T$. Now, using the weighting matrix, $W_T = \widehat{\Omega}(\widehat{\theta}^0_T)^{-1}$ we obtain the updated estimate of θ, $\widehat{\theta}^1_T$. Using $\widehat{\theta}^1_T$ we can then construct a new estimate of Ω, $\widehat{\Omega}(\widehat{\theta}^1_T)$ and a new weighting matrix $W_T = \widehat{\Omega}(\widehat{\theta}^1_T)^{-1}$. This algorithm is iterated until the estimate of θ converges. In what follow, iteration is always used when using $W_T = \widehat{\Omega}^{-1}$ and the corresponding estimate of θ is denoted by $\widehat{\theta}^{opt}_T$. Note that for $\widehat{\theta}_T$ to be consistent, the weighting matrix must converge to a constant matrix as T increases. Thus, truncation must be used given that $W_T = \widehat{\Omega}^{-1}_{M=T}$ converges to a random matrix.

If one would like to completely avoid the choice of truncation lag when doing inference, a weighting matrix other than $W_T = \widehat{\Omega}^{-1}$ is required. In the case where the errors and instruments are *i.i.d.*, the optimal weighting matrix simplifies to $W_T = \left(T^{-1}\sum_{t=1}^{T} z_t z'_t\right)^{-1}$. Use of this weighting matrix generates the well known generalized IV estimator (GIVE) which is denoted by $\widehat{\theta}^{GIVE}_T$.

5.2   Finite Sample Results

Monte Carlo simulations are used to compare and contrast the performance of standard GMM tests with the new tests. The role of the weighting matrix is highlighted in this exercise. To keep matters simple, attention is focused on a simple regression model with $AR(2)$ errors:

$$y_t = \theta_1 + \theta_2 x_t + u_t \tag{11}$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \xi_t,$$

where $\xi_t \sim i.i.d.\ N(0,1)$, $u_{-1} = u_0 = 0$, x_t is a scalar $AR(1)$ process given by $x_t = 0.5x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.\ N(0,1)$, $x_0 = 0$. There are two instruments, $z_t = (z_{1t}, z_{2t})'$ available for x_t given by $z_{it} = x_t + \eta_{it}$ where $\eta_{it} \sim i.i.d.\ N(0,1)$, $i = 1,2$. All the errors are assumed to be uncorrelated with each other (although ξ_t and ε_t could be modeled as correlated given that IV estimation is used). The vector of parameters $\theta = (\theta_1, \theta_2)'$ is estimated using (10) with the weighting matrices $W_{1T} = \widehat{\Omega}^{-1}$ (with iteration) and $W_{2T} = \left(T^{-1}\sum_{t=1}^{T} z_t z'_t\right)^{-1}$. The focus is on θ_2 and the respective estimates denoted by $\widehat{\theta}^{opt}_2$ and $\widehat{\theta}^{GIVE}_2$. The null hypothesis of interest is $H_0 : \theta_2 \leq 0$, $H_1 : \theta_2 > 0$. For each estimate of θ_2 both the standard t statistic and t^* were computed. The standard t statistic is denoted by t_{HAC}. For t_{HAC}, Ω was estimated using the quadratic spectral kernel with the truncation lag chosen using the data dependent method proposed by

12

Andrews (1991) using the $VAR(1)$ plug-in method. Following the recommendation of Kiefer and Vogelsang (2000) t^* was implemented using the Bartlett kernel (without truncation). Following Andrews and Monahan (1992), t_{HAC} and t^* were also implemented using $VAR(1)$ prewhitening. See Andrews and Monahan (1992) for computational details

To assess the accuracy of the limiting null distributions, data was generated according to (11) with $\theta_2 = 0$ (θ_1 was set to zero without loss of generality). Empirical null rejection probabilities were computed using 2,000 replications for the sample sizes $T = 25, 50, 100, 200$. Asymptotic 5% right tail critical values were used so that the nominal level is 0.05. Results were obtained for $AR(1)$ errors with $\rho_2 = 0$ and $\rho_1 = -0.5, -0.3, 0.0, 0.3, 0.5, 0.7, 0.9, 0.95, 0.99$. Results were also obtained for $AR(2)$ models with $\rho_1 = 1.3$ and $\rho_2 = -0.8, -0.6, -0.4, -0.35$. Serial correlation parameter values relevant for macroeconomic data include the $AR(1)$ model with $\rho_1 \geq 0.7$ and the $AR(2)$ model with $\rho_2 \geq -0.6$.

The results are reported in Table III. Several interesting patterns emerge from the table. First, in nearly all cases empirical null rejection probabilities are closer to 0.05 for t^* than for t_{HAC}. In some cases the differences are quite large. For example, with $T = 25$ and $\rho_1 = \rho_2 = 0$, the probability of rejection is 0.089 for t_{HAC} and 0.063 for t^*. Second, models with more persistent serial correlation have greater distortions. This fact is well known in the HAC literature. Third, whereas prewhitening usually improves accuracy for t^* (even when the errors are $AR(2)$), prewhitening sometimes decreases accuracy in the asymptotic approximation for t_{HAC}. Although for highly persistent errors ($\rho_1 \geq 0.7$), prewhitening does improve the asymptotic approximation. Fourth, for both t_{HAC} and t^* the asymptotic approximation is more accurate for and $\widehat{\theta}_2^{GIVE}$ than for $\widehat{\theta}_2^{opt}$. It is not obvious why this should be the case, and the pattern holds regardless of sample size or error model. Finally, when the sample size is relatively large ($T = 200$), empirical rejection probabilities are remarkably close to 0.05 for prewhitened t^* using $\widehat{\theta}_2^{GIVE}$.

Given that the asymptotic null approximation appears better for t^*, the obvious question to ask is: does t^* have any power? The answer is yes as the next simulation experiment illustrates. To simulate finite sample power of the tests, data was generated according to (11) for $\theta_2 = 0.0, 0.05, 0.1, ..., 0.95, 1.0$ (again, $\theta_1 = 0$). Results are reported for $T = 100$ for two error models. The first model has $\rho_1 = 0.7, \rho_2 = 0$ while the second model has $\rho_1 = 1.3, \rho_2 = -0.6$. Power was calculated using finite sample critical values, i.e. power is size adjusted. This was done so that the empirical null rejection probabilities are the same for all statistics thus making power comparisons more meaningful in a theoretical sense. Of course, such size correction cannot be done in practice on actual data, but it serves a useful purpose here for comparing the performance of different tests statistics. As before, 2,000 replications were used. Results are only reported for

the case of prewhitening. Similar results were obtained without prewhitening.

The results are given by Figures 1 and 2. Two patterns emerge. First, power is similar for all the tests with power of t_{HAC} slightly higher than t^*. Second, for both statistics, power is higher for $\widehat{\theta}_2^{GIVE}$ than for $\widehat{\theta}_2^{opt}$. This result is surprising and counter-intuitive given that $\widehat{\theta}_2^{opt}$ has a smaller asymptotic variance than $\widehat{\theta}_2^{GIVE}$ for the two error models considered here. It should be noted that in unreported simulations, power was often greater for $\widehat{\theta}_2^{opt}$ than for $\widehat{\theta}_2^{GIVE}$ when the errors are more persistent. But when errors are less persistent, but still serially correlated, the counter-intuitive rankings consistently emerged.

The finite sample results suggest that t^* is a good alternative to t_{HAC} in practice. t^* has a more accurate asymptotic null approximation and power is competitive. The asymptotic approximation is most accurate and power can be higher when t^* is implemented with $W_{2T} = \left(T^{-1} \sum_{t=1}^{T} z_t z_t' \right)^{-1}$. When implemented in this way, t^* has, in a sense, a fully automatic bandwidth. Reinforcing the recommendation of Andrews and Monahan (1992), the simulations also show that $VAR(1)$ prewhitening improves the asymptotic approximation even if the errors have higher order serial correlation. An obvious topic for future research is to see whether these finite sample results in the linear IV model continue to hold in nonlinear models estimated by GMM.

6 Conclusions

This paper proposes a new approach to testing in the GMM framework. The new tests are constructed using HAC robust standard errors computed using nonparametric spectral density estimators without truncation. While such standard errors are not consistent, a new asymptotic theory shows that they lead to valid tests nonetheless. In the context of the linear IV model, the new tests and the associated limiting distribution theory are shown to have better finite sample size, i.e. a more accurate first order asymptotic null approximation, than standard HAC robust tests. Finite sample power of the new tests is shown to be comparable to standard tests. Because use of a truncation lag equal to the sample size is specific recommendation, it is fully automatic. This approach could potentially lead to a standard of practice (which does not currently exist) for the computation of HAC robust standard errors in GMM models.

7 Appendix

Proof of Lemma 1: Setting $t = T$, multiplying both sides of (3) by $G_T'(\widehat{\theta}_T)W_T$, and using the first order condition $G_T'(\widehat{\theta}_T)W_T g_T(\widehat{\theta}_T) = 0$ gives

$$0 = G_T'(\widehat{\theta}_T)W_T g_T(\theta_0) + G_T'(\widehat{\theta}_T)W_T G_T(\widehat{\theta}_T, \theta_0, \lambda_T)(\widehat{\theta}_T - \theta_0). \qquad (12)$$

Solving (12) for $(\widehat{\theta}_T - \theta_0)$ and scaling by $T^{1/2}$ gives

$$T^{1/2}(\widehat{\theta}_T - \theta_0) = -[G_T'(\widehat{\theta}_T)W_T G_T(\widehat{\theta}_T, \theta_0, \lambda_T)]^{-1} G_T'(\widehat{\theta}_T)W_T T^{1/2} g_T(\theta_0).$$

The lemma follows because $p\lim G_T'(\widehat{\theta}_T)W_T G_T(\widehat{\theta}_T, \theta_0, \lambda_T) = G_0' W_\infty G_0$ by assumptions 3 and 4 and because $G_T'(\widehat{\theta}_T)W_T T^{1/2} g_T(\theta_0) \Rightarrow G_0' W_\infty \Lambda B_q(1)$ by assumptions 2, 3, and 4.

Proof of Lemma 2: Setting $t = [rT]$ and multiplying both sides of (3) by $T^{1/2}$ gives

$$T^{1/2} g_{[rT]}(\widehat{\theta}_T) = T^{1/2} g_{[rT]}(\theta_0) + G_{[rT]}(\widehat{\theta}_T, \theta_0, \lambda_T) T^{1/2}(\widehat{\theta}_T - \theta_0). \tag{13}$$

Using assumptions 2, 3 and 4 and Lemma 1 it follows that

$$T^{1/2} g_{[rT]}(\widehat{\theta}_T) \Rightarrow \Lambda B_q(r) - r G_0 (G_0' W_\infty G_0)^{-1} G_0' W_\infty \Lambda B_q(1). \tag{14}$$

Because $p\lim G_T'(\widehat{\theta}_T)W_T = G_0' W_\infty$ by assumptions 3 and 4, it follows from (13) and (14) that

$$G_T'(\widehat{\theta}_T)W_T T^{1/2} g_{[rT]}(\widehat{\theta}_T) \Rightarrow G_0' W_\infty \left(\Lambda B_q(r) - r G_0 (G_0' W_\infty G_0)^{-1} G_0' W_\infty \Lambda B_q(1)\right),$$

which completes the proof.

Proof of Theorem 1: Once the result for F^* is established, the result for t^* trivially follows when $m = 1$. Applying the delta method to the result in Lemma 1 and using the fact that $B_q(1)$ is a vector of independent standard normal random variables gives

$$\begin{aligned}
T^{1/2} r(\widehat{\theta}_T) &\Rightarrow -R(\theta_0)\left(G_0' W_\infty G_0\right)^{-1} G_0' W_\infty \Lambda B_q(1) \\
&\equiv -R(\theta_0)\left(G_0' W_\infty G_0\right)^{-1} \Lambda^* B_p(1) \\
&\equiv \Lambda^{**} B_m(1),
\end{aligned} \tag{15}$$

where Λ^{**} is the matrix square root of $R(\theta_0)\left(G_0' W_\infty G_0\right)^{-1} \Lambda^* \Lambda^{*\prime} \left(G_0' W_\infty G_0\right)^{-1} R(\theta_0)'$. Using (9) it directly follows that

$$R(\widehat{\theta}_T)\widehat{V}_{M=T} R(\widehat{\theta}_T)'$$
$$= R(\widehat{\theta}_T)[G_T'(\widehat{\theta}_T)W_T G_T(\widehat{\theta}_T, \theta_0, \lambda_T)]^{-1} G_T'(\widehat{\theta}_T)W_T \widehat{\Omega}_{M=T} W_T G_T(\widehat{\theta}_T)[G_T'(\widehat{\theta}_T)W_T G_T(\widehat{\theta}_T, \theta_0, \lambda_T)]^{-1} R(\widehat{\theta}_T)'$$
$$\tag{16}$$

$$\Rightarrow R(\theta_0)\left(G_0' W_\infty G_0\right)^{-1} \Lambda^* \int_0^1 \int_0^1 -k''(r-s)\widetilde{B}_p(r)\widetilde{B}_p(s)' dr ds \Lambda^{*\prime} \left(G_0' W_\infty G_0\right)^{-1} R(\theta_0)'$$

$$\equiv \Lambda^{**} \int_0^1 \int_0^1 -k''(r-s)\widetilde{B}_m(r)\widetilde{B}_m(s)' dr ds \Lambda^{**\prime}, \tag{17}$$

where we use the fact that

$$R(\theta_0)\left(G_0'W_\infty G_0\right)^{-1}\Lambda^*\widetilde{B}_p(r) = R(\theta_0)\left(G_0'W_\infty G_0\right)^{-1}\Lambda^*\left(B_p(r) - rB_p(1)\right)$$
$$\equiv \Lambda^{**}\left(B_m(r) - rB_m(1)\right)$$
$$= \Lambda^{**}\widetilde{B}_m(r).$$

Using (15) and (17) it directly follows that

$$F^* = Tr(\widehat{\theta}_T)'\left(R(\widehat{\theta}_T)\widehat{V}_{M=T}R(\widehat{\theta}_T)'\right)^{-1}r(\widehat{\theta}_T)/m$$
$$= T^{1/2}r(\widehat{\theta}_T)'\left(R(\widehat{\theta}_T)\widehat{V}_{M=T}R(\widehat{\theta}_T)'\right)^{-1}T^{1/2}r(\widehat{\theta}_T)/m$$
$$\Rightarrow (\Lambda^{**}B_m(1))'\left(\Lambda^{**}\int_0^1\int_0^1 -k''(r-s)\widetilde{B}_m(r)\widetilde{B}_m(s)'drds\Lambda^{**\prime}\right)^{-1}(\Lambda^{**}B_m(1))/m$$
$$\equiv B_m(1)'\left(\int_0^1\int_0^1 -k''(r-s)\widetilde{B}_m(r)\widetilde{B}_m(s)'drds\right)^{-1}B_m(1)/m,$$

which completes the proof.

References

Abadir, K. M. and Paruolo, P. (1997), Two Mixed Normal Densities from Cointegration Analysis, *Econometrica* **65**, 671–680.

Andrews, D. W. K. (1991), Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* **59**, 817–854.

Andrews, D. W. K. and Monahan, J. C. (1992), An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator, *Econometrica* **60**, 953–966.

den Haan, W. J. and Levin, A. (1997), A Practictioner's Guide to Robust Covariance Matrix Estimation, *in* G. Maddala and C. Rao (eds), *Handbook of Statistics: Robust Inference*, Volume 15, Elsevier, New York.

den Haan, W. J. and Levin, A. (1998), Vector Autoregressive Covariance Matrix Estimation, working paper, International Finance Division, FED Board of Governors.

Gallant, A. (1987), *Nonlinear Statistical Models*, Wiley, New York.

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

Hall, A. (2000), Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test, *Econometrica* **68**, 1517–1528.

Hansen, B. E. (1992), Consistent Covariance Matrix Estimation for Dependent Heterogenous Processes, *Econometrica* **60**, 967–972.

Hansen, L. P. (1982), Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* **50**, 1029–1054.

Kiefer, N. M. and Vogelsang, T. J. (2000), A New Approach to the Asymptotics of Heteroskedasticity-Autocorrelation Robust Testing, mimeo, Department of Economics, Cornell University.

Kiefer, N. M. and Vogelsang, T. J. (2001), Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation, mimeo, Department of Economics, Cornell University.

Kiefer, N. M., Vogelsang, T. J. and Bunzel, H. (2000), Simple Robust Testing of Regression Hypotheses, *Econometrica* **68**, 695–714.

Newey, W. K. and McFadden, D. L. (1994), Large Sample Estimation and Hypothesis Testing, *in* R. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier Science Publishers, Amsterdam, The Netherlands, pp. 2113–2247.

Newey, W. K. and West, K. D. (1987), A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* **55**, 703–708.

Newey, W. K. and West, K. D. (1994), Automatic Lag Selection in Covariance Estimation, *Review of Economic Studies* **61**, 631–654.

Phillips, P. C. B. and Durlauf, S. N. (1986), Multiple Regression with Integrated Processes, *Review of Economic Studies* **53**, 473–496.

Robinson, P. (1998), Inference-Without Smoothing in the Presence of Nonparametric Autocorrelation, *Econometrica* **66**, 1163–1182.

White, H. (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica* **48**, 817–38.

White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, New York.

Table I: Asymptotic Critical Values for t^* Using Bartlett Kernel

1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
-6.090	-4.771	-3.764	-2.740	0.000	2.740	3.764	4.771	6.090

Note: Critical values were obtained by scaling by $1/\sqrt{2}$ the critical values from line 1 of Table I in Abadir and Paruolo (1997).

Table II: Asymptotic Critical Values for F^* Using Bartlett Kernel

$q =$	1	2	3	4	5	6	7	8	9	10
90%	14.28	17.99	21.13	24.24	27.81	30.36	33.39	36.08	38.94	41.71
95%	23.14	26.19	29.08	32.42	35.97	38.81	42.08	45.32	48.14	50.75
97.5%	33.64	35.56	37.88	40.57	44.78	47.94	50.81	54.22	57.47	59.98
99%	51.05	48.74	51.04	52.39	56.92	60.81	62.27	67.14	69.67	72.05

$q =$	11	12	13	14	15	16	17	18	19	20
90%	44.56	47.27	50.32	52.97	55.71	58.14	60.75	63.35	65.81	68.64
95%	53.70	56.70	60.11	62.83	65.74	68.68	70.59	73.76	76.42	79.50
97.5%	63.14	65.98	69.46	72.46	75.51	78.09	80.94	83.63	86.20	89.86
99%	74.74	78.80	82.09	85.12	88.86	91.37	94.08	97.41	99.75	103.2

$q =$	21	22	23	24	25	26	27	28	29	30
90%	70.80	73.41	76.19	78.40	81.21	83.59	85.83	88.11	90.92	93.63
95%	82.00	84.76	87.15	89.67	92.70	95.49	97.57	99.48	102.9	105.8
97.5%	92.32	94.54	98.06	100.4	103.5	106.6	108.8	110.7	114.6	117.5
99%	105.4	108.0	111.8	114.7	117.6	120.8	123.4	124.5	129.6	132.1

Notes: The critical values were calculated via simulation methods using normalized partial sums of 1,000 *i.i.d.* $N(0,1)$ random deviates to approximate the standard Brownian motions in the respective distributions. 50,000 replications.were used. q is the number of restrictions being tested.

Table III: Empirical Null Rejection Probabilities in Linear IV Model

Panel A: $T = 25$

| | | No Prewhitening | | | | With Prewhitening | | | |
| | | $\widehat{\theta}_2^{GIVE}$ | | $\widehat{\theta}_2^{opt}$ | | $\widehat{\theta}_2^{GIVE}$ | | $\widehat{\theta}_2^{opt}$ | |
ρ_1	ρ_2	t_{HAC}	t^*	t_{HAC}	t^*	t_{HAC}	t^*	t_{HAC}	t^*
-0.50	0.00	0.057	0.043	0.074	0.057	0.067	0.040	0.098	0.063
-0.30	0.00	0.068	0.054	0.089	0.066	0.076	0.049	0.115	0.075
0.00	0.00	0.089	0.063	0.115	0.087	0.092	0.060	0.133	0.086
0.30	0.00	0.105	0.073	0.136	0.097	0.107	0.069	0.147	0.095
0.50	0.00	0.115	0.075	0.146	0.102	0.114	0.073	0.154	0.103
0.70	0.00	0.127	0.091	0.154	0.113	0.120	0.078	0.151	0.104
0.90	0.00	0.146	0.100	0.168	0.115	0.132	0.081	0.161	0.108
0.95	0.00	0.149	0.100	0.180	0.120	0.138	0.083	0.170	0.105
0.99	0.00	0.149	0.099	0.178	0.118	0.141	0.080	0.167	0.102
1.30	-0.80	0.108	0.073	0.148	0.099	0.094	0.063	0.138	0.093
1.30	-0.60	0.136	0.094	0.166	0.121	0.116	0.077	0.147	0.110
1.30	-0.40	0.150	0.112	0.172	0.128	0.132	0.089	0.165	0.115
1.30	-0.35	0.158	0.102	0.183	0.125	0.147	0.086	0.172	0.114

Panel B: $T = 50$

| | | No Prewhitening | | | | With Prewhitening | | | |
| | | $\widehat{\theta}_2^{GIVE}$ | | $\widehat{\theta}_2^{opt}$ | | $\widehat{\theta}_2^{GIVE}$ | | $\widehat{\theta}_2^{opt}$ | |
ρ_1	ρ_2	t_{HAC}	t^*	t_{HAC}	t^*	t_{HAC}	t^*	t_{HAC}	t^*
-0.50	0.00	0.061	0.054	0.065	0.057	0.070	0.056	0.077	0.060
-0.30	0.00	0.067	0.061	0.071	0.068	0.075	0.059	0.084	0.069
0.00	0.00	0.068	0.061	0.081	0.070	0.079	0.058	0.092	0.072
0.30	0.00	0.082	0.062	0.103	0.076	0.081	0.057	0.099	0.071
0.50	0.00	0.087	0.064	0.109	0.075	0.081	0.061	0.108	0.075
0.70	0.00	0.093	0.076	0.117	0.087	0.084	0.069	0.105	0.078
0.90	0.00	0.121	0.081	0.131	0.099	0.102	0.068	0.115	0.085
0.95	0.00	0.128	0.089	0.141	0.101	0.105	0.072	0.118	0.084
0.99	0.00	0.123	0.079	0.141	0.092	0.109	0.068	0.113	0.079
1.30	-0.80	0.088	0.060	0.103	0.070	0.070	0.054	0.085	0.065
1.30	-0.60	0.096	0.074	0.119	0.085	0.079	0.059	0.103	0.076
1.30	-0.40	0.119	0.085	0.128	0.099	0.095	0.067	0.107	0.084
1.30	-0.35	0.129	0.091	0.138	0.102	0.109	0.069	0.122	0.085

19

Panel C: $T = 100$

| | | No Prewhitening | | | | With Prewhitening | | | |
| | | $\widehat{\theta}_2^{GIVE}$ | | $\widehat{\theta}_2^{opt}$ | | $\widehat{\theta}_2^{GIVE}$ | | $\widehat{\theta}_2^{opt}$ | |
ρ_1	ρ_2	t_{HAC}	t^*	t_{HAC}	t^*	t_{HAC}	t^*	t_{HAC}	t^*
-0.50	0.00	0.040	0.049	0.043	0.051	0.045	0.049	0.053	0.051
-0.30	0.00	0.044	0.044	0.048	0.050	0.049	0.045	0.057	0.054
0.00	0.00	0.048	0.048	0.054	0.056	0.049	0.047	0.060	0.057
0.30	0.00	0.067	0.056	0.078	0.063	0.065	0.054	0.073	0.062
0.50	0.00	0.079	0.061	0.093	0.069	0.072	0.056	0.082	0.068
0.70	0.00	0.081	0.059	0.093	0.073	0.070	0.055	0.081	0.066
0.90	0.00	0.096	0.070	0.101	0.080	0.071	0.058	0.085	0.068
0.95	0.00	0.091	0.068	0.101	0.079	0.075	0.058	0.079	0.069
0.99	0.00	0.093	0.071	0.107	0.077	0.071	0.056	0.081	0.065
1.30	-0.80	0.063	0.060	0.078	0.064	0.050	0.054	0.061	0.055
1.30	-0.60	0.081	0.064	0.095	0.069	0.058	0.055	0.071	0.063
1.30	-0.40	0.097	0.069	0.109	0.079	0.073	0.052	0.086	0.066
1.30	-0.35	0.095	0.070	0.108	0.079	0.074	0.053	0.079	0.064

Panel D: $T = 200$

| | | No Prewhitening | | | | With Prewhitening | | | |
| | | $\widehat{\theta}_2^{GIVE}$ | | $\widehat{\theta}_2^{opt}$ | | $\widehat{\theta}_2^{GIVE}$ | | $\widehat{\theta}_2^{opt}$ | |
ρ_1	ρ_2	t_{HAC}	t^*	t_{HAC}	t^*	t_{HAC}	t^*	t_{HAC}	t^*
-0.50	0.00	0.043	0.052	0.043	0.055	0.044	0.049	0.049	0.056
-0.30	0.00	0.042	0.048	0.046	0.050	0.049	0.049	0.052	0.052
0.00	0.00	0.048	0.050	0.051	0.052	0.049	0.050	0.051	0.052
0.30	0.00	0.059	0.048	0.065	0.051	0.052	0.048	0.058	0.053
0.50	0.00	0.063	0.055	0.066	0.060	0.057	0.055	0.061	0.059
0.70	0.00	0.068	0.053	0.070	0.059	0.054	0.053	0.062	0.057
0.90	0.00	0.079	0.057	0.081	0.065	0.061	0.053	0.066	0.062
0.95	0.00	0.077	0.060	0.081	0.064	0.064	0.051	0.069	0.060
0.99	0.00	0.082	0.059	0.085	0.064	0.069	0.050	0.072	0.051
1.30	-0.80	0.054	0.048	0.059	0.050	0.040	0.044	0.044	0.047
1.30	-0.60	0.056	0.053	0.065	0.055	0.044	0.048	0.048	0.052
1.30	-0.40	0.082	0.055	0.084	0.057	0.061	0.049	0.064	0.051
1.30	-0.35	0.077	0.059	0.082	0.065	0.063	0.050	0.069	0.058

Notes: The t statistics were computed for model (11) which was estimated by GMM using (10). $\widehat{\theta}_2^{opt}$ is the estimate of θ_2 using the optimal weighting matrix $W_{1T} = \widehat{\Omega}^{-1}$ (with iteration). $\widehat{\theta}_2^{GIVE}$ is the estimate of θ_2 using the weighting matrix $W_{2T} = \left(T^{-1}\sum_{t=1}^{T} z_t z_t'\right)^{-1}$. The null hypothesis of interest is $H_0 : \theta_2 \leq 0$, $H_1 : \theta_2 > 0$, and asymptotic 5% right tail critical values were used. 2,000 replications were used all cases.
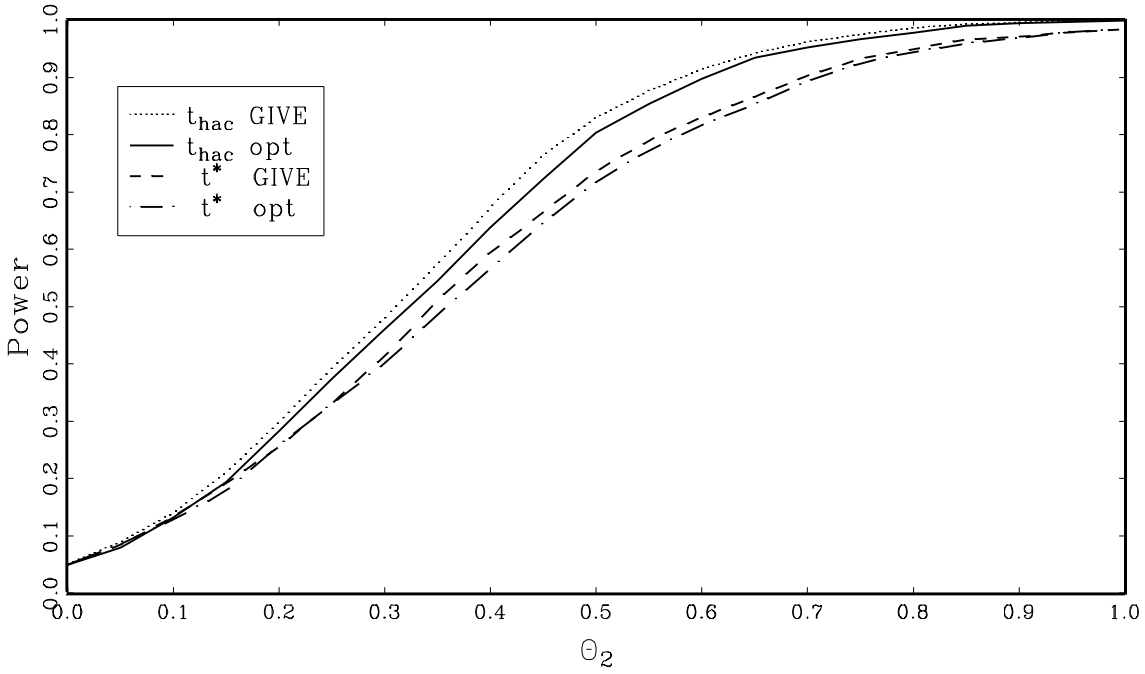
Figure 1: Finite Sample Power (size adjusted, 5% level)
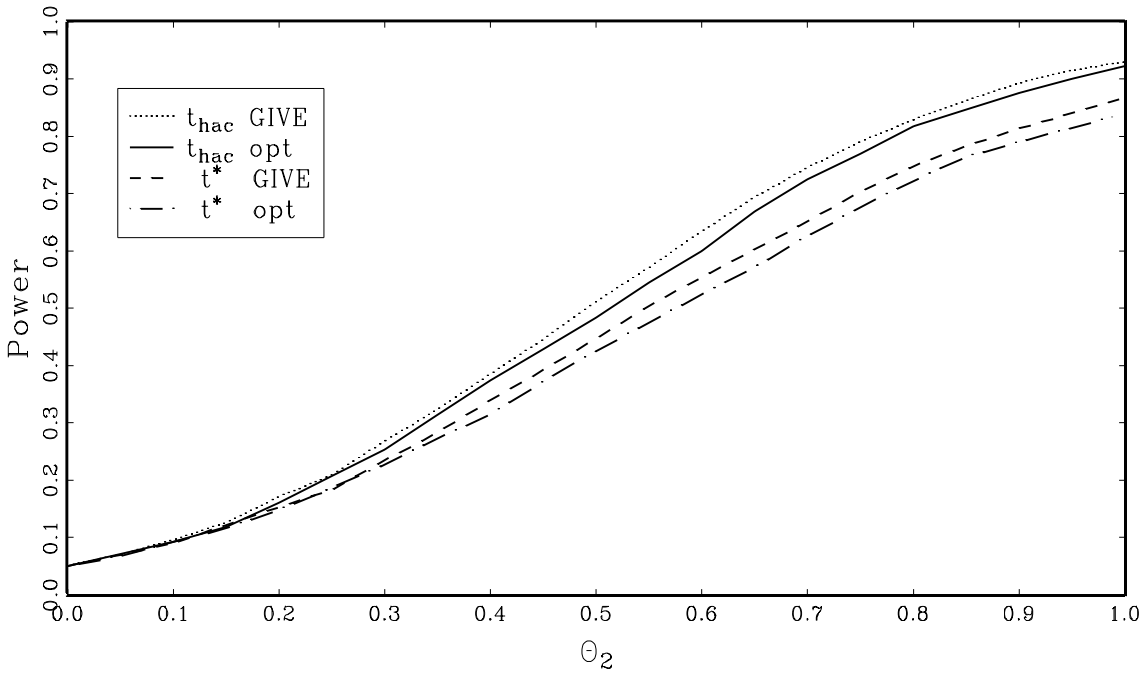Model (11), T=100, AR(1) Errors, $\rho_1=0.7$.

Figure 2: Finite Sample Power (size adjusted, 5% level)
Model (11), T=100, AR(2) Errors, $\rho_1=1.3$, $\rho_2=-0.6$.