Simple Robust Testing of Hypotheses in Non-Linear Models

Helle Bunzel[*]

Nicholas M. Kiefer

Timothy J. Vogelsang

August 11, 2000

[*]Correspondence: Helle Bunzel, Department of Economics, Iowa State University, Ames, Iowa 50014, E-mail: hbunzel@iastate.edu, Phone: (515) 294-6163, Fax: (515) 294-6644; Nicholas M. Kiefer, Department of Economics, Department of Statistical Sciences and CAE, Cornell University and Department of Economics, CAF and CDME, University of Aarhus; Timothy J. Vogelsang, Department of Economics, Department of Statistical Sciences and CAE, Cornell University. The authors are grateful to two anonymous referees and an associate editor for helpful comments that helped improve the paper.

Abstract

We develop test statistics to test hypotheses in nonlinear, weighted regression models with serial correlation or conditional heteroskedasticity of unknown form. The novel aspect is that these tests are simple and do not require use of heteroskedasticity autocorrelation consistent (HAC) covariance matrix estimators. This new new class of tests utilize stochastic transformations to eliminate nuisance parameters as a substitute for consistently estimating the nuisance parameters. We derive the limiting null distributions of these new tests in a general nonlinear setting, and show that while the tests have nonstandard distributions, the distributions depend only upon the number of restrictions being tested. We perform some simulations on a simple model and we apply the new method of testing to an empirical example and illustrate that the size of the new test is less distorted than tests utilizing HAC covariance matrix estimators.

1 Introduction

It is a well known result that in models with errors that have autocorrelation or heteroskedasticity of unknown form, standard estimators remain consistent and are asymptotically normally distributed under weak regularity conditions. However, the standard results required for testing hypotheses in the usual manner no longer hold. In this paper we develop new test statistics in weighted, nonlinear regression models that are robust to serial correlation or heteroskedasticity of unknown form in the errors. Included in this class of models are instrumental variables (IV) estimation of nonlinear models and some Quasi-likelihood models. Our results extend those obtained in Kiefer,Vogelsang, and Bunzel (2000) for the linear regression model.

When the error covariance structure is known, the model can be transformed and standard testing results can be obtained using generalized least squares (GLS) methods. This is usually not possible in practice, as the serial correlation or heteroskedasticity encountered is frequently of unknown form. To obtain valid testing procedures, the most common approach in the literature to date has been to estimate the variance-covariance matrix of the parameter estimates. Often, nonparametric spectral methods are used to construct heteroskedasticity autocorrelation consistent (HAC) estimates the variance-covariance matrix. Using these HAC estimators, standard tests (i.e. t and Wald tests) are constructed based on the asymptotic normal distribution of the weighted NLS estimator.

The HAC literature has grown out of the literature on estimation of standard errors robust to heteroskedasticity of unknown form (e.g. Eicker (1967), Huber (1967) and White (1980)) and extended that literature to allow for autocorrelation. In the econometrics literature HAC estimators and their properties have recently attracted considerable attention. Among the important contributions are Andrews (1991), Andrews and Monahan (1992), Gallant (1987), Hansen (1992), Hong and Lee (1999), Newey and West (1987), Robinson (1991) and White (1984). The direct contribution of the HAC literature has been the development of asymptotically valid tests that are robust to serial correlation and/or heteroskedasticity of unknown form. This literature builds upon and extends the classic results from the spectral density estimation literature (see Priestley (1981)).

A theoretical limitation of the HAC approach is that, while the variance-covariance matrix is consistently estimated, the resulting variation in finite samples is not taken into account. Asymptotically, this clearly is not a problem; in fact, once the variance-covariance matrix has been estimated, it can be assumed to be known. In finite samples, however, this can cause substantial size distortions. In this paper, we develop an alternative approach that does not require a direct estimate of the variance-covariance matrix.

A practical limitation of the HAC approach is that in order to obtain HAC estimates, a truncation lag for a spectral density estimator must be chosen. Although asymptotic theory dictates the rate at which the truncation lag must increase as the sample size grows, no concrete guidance is provided. In fact, any choice of truncation lag can be justified for a sample of any size by cleverly choosing the function that relates the truncation lag to the sample size. Thus, a practitioners is faced with a choice that is ultimately arbitrary. It is perhaps for this reason that no standard of practice has emerged for the computation of

HAC robust standard errors. Our approach provides an elegant solution to this practical problem by avoiding the need to consistently estimate the variance-covariance matrix and thus removing the need to choose a truncation lag.

Intuitively, the approach we take is similar in spirit to Fisher's classic construction of the t test. A data-dependent transformation is applied to the NLS estimates of the parameters of interest. This transformation is chosen such that it ensures that the asymptotic distribution of the transformed estimator does not depend on nuisance parameters. The transformed estimator can then be used to construct a test for general hypotheses on the parameters of interest. The asymptotic distribution of the resulting test statistic turns out to be symmetric, but with fatter tails than the normal distribution; it is not normal, but has the form of a scale mixture of normals. Furthermore, it depends only on the number of restrictions that are being tested. We are therefore able to tabulate the critical values in the usual manner as a function of the number of restrictions and the level of the test.

To gain insight about the performance of the new tests in a simple environment, we perform simulations for the simplest special case of our framework: a location model with AR(1) serially correlated data. These simulations illustrate that our new approach can outperform traditional tests in terms of the accuracy of the asymptotic approximation. We also provide an empirical example illustrating our new tests in a non-linear model. Specifically, we examine the effect of U.S. Gross Domestic Product (GDP) growth on the growth of aggregate restaurant revenues. We use quarterly data over 26 years for the analysis, and there is reason to suspect the presence of autocorrelation and heteroskedasticity. We do not, however, have any knowledge of the specific forms of autocorrelation and heteroskedasticity we may encounter in this data set, hence making it an excellent candidate for the application of both the HAC based tests and the new tests. We also report results from finite sample simulations calibrated to the empirical example. These simulations confirm the superior finite sample size of the new tests and show that finite sample power of the new tests is competitive.

The rest of the paper is organized as follows. In section 2, we introduce the statistical model and derive basic asymptotic results. In section 3 we develop the new test statistics. In section 4 we report the results of the finite sample simulations for the simple location model. Section 5 contains the empirical example and associated finite sample simulations. Section 6 concludes, and proofs are collected in an appendix.

2 The Model and Some Asymptotic Results

Consider the nonlinear regression model given by

$$y_t = f(X_t; \beta) + u_t = f_t(\beta) + u_t; \ t = 1; ...; T; \tag{1}$$

where f denotes the nonlinear function of regressors and parameters. β is a $(k \times 1)$ vector of parameters and X_t is a $(k_2 \times 1)$ vector of exogenous variables and conditional on X_t; u_t is a mean zero random process: We assume that u_t does not have a unit root, but u_t may be serially correlated and/or conditionally heteroskedastic. At times, it will be useful to stack the equations in (1) and rewrite them as

$$y = f(\beta) + u: \tag{2}$$

We will use weighted non-linear least squares to obtain an estimate of β. The estimate, $\hat{\beta}$, is defined as

$$\hat{\beta} = \arg\min_{\beta} (y - f(\beta))' W (y - f(\beta)) \tag{3}$$

where W is a symmetric, positive definite $T-$dimensional weighting matrix.

Depending on the choice of W, the following are examples of estimation techniques covered by this framework:

Example 1: Linear regression.

If $f(\beta)$ takes the form $X\beta$ and W is the diagonal matrix, then (3) simply provides the standard least squares estimator:

$$\hat{\beta} = \arg\min_{\beta} (y - X\beta)' (y - X\beta).$$

Example 2: Nonlinear Least Squares.

If we let W be the identity matrix, (3) takes the well-known form

$$\hat{\beta} = \arg\min_{\beta} (y - f(\beta))' (y - f(\beta)).$$

This is the case of standard, non-linear least squares.

Example 3: Non-linear IV estimation, Lagged Dependent Variables.

If we have a model corresponding to (2), and a $T \times l$ matrix of instruments Z, $l \geq k$, with the matrix projecting onto the space spanned by the instruments defined as $P_Z = Z(Z'Z)^{-1}Z'$, the IV estimator takes the form

$$\hat{\beta}_{IV} = \arg\min_{\beta} (y - f(\beta))' P_Z (y - f(\beta))$$

corresponding to $W = P_Z$ in (3).

A special case is models including lagged dependent variables as regressors. Here we are interested in a model

$$y = X\beta + Y_2\gamma + U; \tag{4}$$

where Y_2 is a matrix containing lagged values of y while β and γ are parameters. If we let $A = [X;\ Y_2]$ and $\delta = (\beta;\ \gamma)'$; we can rewrite (4) as

$$y = A\delta + U.$$

Using the method of instrumental variables with instruments Z; the estimate of δ is defined as

$$\hat{\delta} = \arg\min_{\delta} (y - A\delta)' P_Z (y - A\delta) \tag{5}$$

where $P_Z = Z(Z'Z)^{-1}Z'$. Comparing (5) and (3), we see that $\hat{\delta}$ is the weighted least squares estimator in a model with weighting matrix P_Z.

These examples illustrate that several well-known models and estimation techniques are special cases of the framework we use.

In developing the tests, we also work with the following transformed model

$$W^{\frac{1}{2}}y = W^{\frac{1}{2}}f(\beta) + W^{\frac{1}{2}}u$$

or, simplifying the notation

$$\tilde{y} = \tilde{f}(\beta) + \tilde{u} \tag{6}$$

where $\tilde{y}_t$; $\tilde{f}_t$ and $\tilde{u}_t$ are defined in the natural way. The following additional notation is used throughout the paper. Let the $k \times 1$ vector $\tilde{F}_t(\beta)$ denote the derivative of $\tilde{f}_t(\beta)$ with respect to β and $F(\beta)$ the derivative of $f(\beta)$ with respect to β. In addition, let $v_t \equiv \tilde{F}_t(\beta)\tilde{u}_t$ and define $\Omega = \Lambda\Lambda' = \Gamma_0 + \sum_{j=1}^{\infty}\left[\Gamma_j + \Gamma_j'\right]$ where $\Gamma_j = E\left[v_t v_{t-j}'\right]$. For later use, note that Ω is equal to 2π times the spectral density matrix of v_t evaluated at frequency zero. Define $S_t = \sum_{j=1}^{t} v_j$ and let $W_k(r)$ denote a k-vector of independent Wiener precesses, and let $[rT]$ denote the integer part of rT, where $r \in [0,1]$. We let β_0 denote the true value of the parameter β. We use $\Rightarrow$ to denote weak convergence. The following two assumptions are sufficient to obtain the main results of the paper.

Assumption 1 $\text{plim}\left[T^{-1}\sum_{t=1}^{[rT]}\tilde{F}_t(\beta_0)\tilde{F}_t'(\beta_0)\right] = rQ \ \forall r \in [0,1]$, where Q is invertible. Furthermore, $\frac{\partial}{\partial\beta'}\tilde{F}_t(\beta_0)$ exists, is uniformly continuous in β for all β in a small open ball around β_0 and $\frac{\partial}{\partial\beta'}\tilde{F}_t(\beta_0)$ is bounded in probability.

Assumption 2 $T^{-\frac{1}{2}}S_{[rT]} = T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\tilde{F}_t(\beta_0)\tilde{u}_t = T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}v_t \Rightarrow \Lambda W_k(r)$.

Assumptions 1 and 2 are more than sufficient for obtaining an asymptotic normality result for $\hat{\beta}$. Assumption 1 is fairly standard and rules out trends in the linearized regression function of the transformed model, but not necessarily in the X_t process. Assumption 1 also implies the standard assumption that $\text{plim}\left[\frac{1}{T}F'(\beta_0)WF(\beta_0)\right] = Q$. On the surface, assumption 2 appears to be stronger than what is necessary compared to the standard approach. Indeed, to obtain an asymptotic normality result, a conventional central limit theorem for $T^{-\frac{1}{2}}\sum_{t=1}^{T}v_t$ is all that would be required. However, according to the standard approach, asymptotically valid testing requires consistent estimation of Ω, i.e. an HAC estimator. Typical regularity conditions for HAC estimators to be consistent are that the $\{v_t\}$ process be fourth order stationary and satisfy well known mixing conditions (see Andrews (1991) for details). Interestingly, our assumption 2 holds under the weaker condition that the $\{v_t\}$ process is only second order $(2+\varepsilon,\ \varepsilon>0)$ stationary and mixing (see Phillips and Durlauf (1986) for details). Therefore, we are able to obtain asymptotically valid tests under weaker regularity conditions than the standard approach.

Using assumptions 1 and 2, we obtain the well known asymptotic normality of $\hat{\beta}$:

$$\sqrt{T}\left(\hat{\beta}-\beta_0\right) \Rightarrow Q^{-1}\Lambda W_k(1) \sim N\left(0, Q^{-1}\Lambda\Lambda' Q^{-1}\right) = N\left(0, Q^{-1}\Omega Q^{-1}\right) = N(0, V). \quad (7)$$

The asymptotic distribution is a k-variate normal distribution with mean zero and variance covariance matrix $V = Q^{-1}\Omega Q^{-1}$. The asymptotic distribution of $\hat{\beta}$ can be used to test hypothesis about β. In the standard approach an estimate of V (and therefore Q^{-1} and Ω) is required. A natural estimate of Q is $\widehat{Q} = \frac{1}{T}F'\left(\hat{\beta}\right)WF\left(\hat{\beta}\right)$. Ω can be estimated by a HAC estimator, $\hat{\Omega}$. Letting $\hat{u}_t$ be the residuals of the transformed model, the HAC estimate would utilize $\hat{v}_t = F_t\left(\hat{\beta}\right)'\hat{u}_t$ to estimate nonparametrically the spectral density of v_t at frequency zero, and hence Ω. A typical estimator takes the form

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

with

$$\widehat{\Gamma}_j = T^{-1}\sum_{t=j+1}^{T} \hat{v}_t\hat{v}_{t-j}' \quad \text{for } j \geq 0; \quad \widehat{\Gamma}_j = T^{-1}\sum_{t=-j+1}^{T} \hat{v}_{t+j}\hat{v}_t' \quad \text{for } j<0;$$

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 $\int_{-\infty}^{\infty} k^2(x)dx < \infty$. The parameter, M, is often called the truncation lag. A typical condition for consistency of $\widehat{\Omega}$ is that $M \to \infty$ as $T \to \infty$ but $M/T \to 0$. For example, the rule $M = cT^{1/3}$ where c is any positive constant satisfies this asymptotic requirement. Because the constant, c, is arbitrary, it can always to chosen to justify any choice of M for a given sample size. Our approach removes the need to choose M at all because a consistent estimate of Ω is not required.

To test hypotheses about β in the standard approach $\hat{V} = \hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1}$ is used to transform $\sqrt{T}\left(\hat{\beta} - \beta_0\right)$ to obtain

$$\hat{V}^{-\frac{1}{2}}\sqrt{T}\left(\hat{\beta} - \beta_0\right) \Rightarrow \Omega^{-\frac{1}{2}}QQ^{-1}\Lambda W_k(1) = W_k(1) \sim N(0; I_k): \quad (8)$$

Using (8); hypotheses can be tested in the usual manner using t or Wald tests.

In the new approach we use a method that is similar; we also transform $\sqrt{T}\left(\hat{\beta} - \beta_0\right)$ in such a manner that the asymptotic distribution no longer depends on unknown parameters. The essential difference between the two approaches is that our approach does not require an explicit estimate of Ω and takes the additional sampling variation associated with not knowing the covariance matrix into account in the asymptotic approximation. HAC estimates, on the other hand, treat the variance-covariance matrix as known asymptotically and ignore the impact of finite sample variability from $\hat{\Omega}$ on the distribution of the test statistics.

We now proceed to obtain the relevant transformation. Consider the following scaled partial sum empirical process

$$T^{-\frac{1}{2}}\hat{S}_{[rT]} = T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\hat{v}_t = T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}F_t\left(\hat{\beta}\right)\hat{u}_t:$$

In the appendix, we prove the following lemma:

Lemma 1 $T^{-\frac{1}{2}}\hat{S}_{[rT]} = T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}F_t(\beta_0)\hat{u}_t + l_{[rT]}$ where $l_{[rT]}$ is a residual term with the property that $l_{[rT]} = o_p(1)$:

Using this lemma it follows that

$$
\begin{aligned}
T^{-\frac{1}{2}}\hat{S}_{[rT]} &= T^{-\frac{1}{2}}\sum_{t=1}^{[rT]} F_t(\beta_0)\hat{u}_t + I_{[rT]} \\
&= T^{-\frac{1}{2}}\sum_{t=1}^{[rT]} F_t(\beta_0)\left[f_t(\beta_0) + u_t - f_t\left(\hat{\beta}\right)\right] + I_{[rT]} \\
&= T^{-\frac{1}{2}}\sum_{t=1}^{[rT]} F_t(\beta_0)u_t + T^{-\frac{1}{2}}\sum_{t=1}^{[rT]} F_t(\beta_0)\left[f_t(\beta_0) - f_t\left(\hat{\beta}\right)\right] + I_{[rT]}.
\end{aligned}
$$

Using a Taylor expansion of $f_t(\hat{\beta})$ around β_0, we see $f_t(\hat{\beta}) = f_t(\beta_0) + F_t'\left(\bar{\beta}\right)(\hat{\beta} - \beta_0)$, where $\bar{\beta} \in \left[\hat{\beta}, \beta_0\right]$. This expansion allows us to write

$$
\begin{aligned}
T^{-\frac{1}{2}}\hat{S}_{[rT]} &= T^{-\frac{1}{2}}S_{[rT]} - T^{-\frac{1}{2}}\sum_{t=1}^{[rT]} F_t(\beta_0)F_t'\left(\bar{\beta}\right)\left(\hat{\beta} - \beta_0\right) + I_{[rT]} \qquad (9)\\
&= T^{-\frac{1}{2}}S_{[rT]} - r\left[\frac{1}{rT}\sum_{t=1}^{[rT]} F_t(\beta_0)F_t'\left(\bar{\beta}\right)\right]\sqrt{T}\left(\hat{\beta} - \beta_0\right) + o_p(1).
\end{aligned}
$$

Applying assumptions 1 and 2 and equation (7) to (9) gives the asymptotic result

$$
T^{-\frac{1}{2}}\hat{S}_{[rT]} \Rightarrow \Lambda W_k(r) - rQ\left[Q^{-1}\Lambda W_k(1)\right] = \Lambda\left(W_k(r) - rW_k(1)\right). \qquad (10)
$$

Note that $(W_k(r) - rW_k(1))$ is a k-dimensional Brownian Bridge.

Now consider $\hat{C} = T^{-2}\sum_{t=1}^{T}\hat{S}_t\hat{S}_t'$. From (10) and the continuous mapping theorem, it follows that

$$
\hat{C} = T^{-1}\sum_{t=1}^{T}\left(T^{-\frac{1}{2}}\hat{S}_t\right)\left(T^{-\frac{1}{2}}\hat{S}_t\right)' \Rightarrow \Lambda\int_0^1 (W_k(r) - rW_k(1))(W_k(r) - rW_k(1))'\,dr\,\Lambda'.
$$

Define $P_k \equiv \int_0^1 (W_k(r) - rW_k(1))(W_k(r) - rW_k(1))'\,dr$. The asymptotic distribution of $\hat{C}$ can then be written as $\Lambda P_k\Lambda'$. We have now obtained a matrix whose asymptotic distribution is a quadratic form in Λ. In what follows, this enables us to apply a transformation to

$\sqrt{T}\left(\hat{\beta} - \beta_0\right)$ eliminating Λ from the asymptotic distribution. To that end, note that because P_k is constructed as the integral of the outer product of a k-dimensional Brownian Bridge, it is positive definite. This permits us to use the Cholesky decomposition to write $P_k = Z_k Z_k'$.

We now turn our attention to the matrix $\hat{B} = \hat{Q}^{-1}\hat{C}\hat{Q}^{-1}$. Note that $\hat{B}$ is constructed analogously to $\hat{V}$ except that $\hat{C}$ is used in place of $\hat{\Omega}$. Now define $\hat{M} = \hat{Q}^{-1}\hat{C}^{\frac{1}{2}}$, where $\hat{C}^{\frac{1}{2}}$ is the Cholesky decomposition of $\hat{C}$. Therefore, $\hat{M}\hat{M}' = \hat{B}$. Because $\hat{C} \Rightarrow \Lambda P_k \Lambda'$, it follows that $\hat{M} \Rightarrow Q^{-1}\Lambda Z_k$.

We are now ready to introduce our transformation of $\sqrt{T}\left(\hat{\beta} - \beta_0\right)$ using $\hat{M}^{-1}$. As $T \to \infty$, it evidently follows that

$$\hat{M}^{-1}\sqrt{T}\left(\hat{\beta} - \beta_0\right) \Rightarrow [Q\Lambda Z_k]^{-1} Q^{-1}\Lambda W_k(1) = Z_k^{-1} W_k(1). \tag{11}$$

The limiting distribution given by (11) does not depend on the nuisance parameters Q and Ω. It is trivial to show that $W_k(1)$ and P_k^{-1} are independent, so conditional on Z_k, $\hat{M}^{-1}\sqrt{T}\left(\hat{\beta} - \beta_0\right)$ is distributed as $N\left(0, P_k^{-1}\right)$. If we denote the density function of P_k by $p(P_k)$, the unconditional distribution of $\hat{M}^{-1}\sqrt{T}\left(\hat{\beta} - \beta_0\right)$ is $\int_0^1 N\left(0, P_k^{-1}\right) p(P_k)\, dP_k$. This is a mixture of normals, which is symmetric, but has thicker tails than the normal distribution. It is important to note that $\hat{M}$ is easy to compute from data and does not require a choice of truncation lag.

The above derivation is similar to Fisher's classic development of the t-statistic. Fisher utilized a data dependent transformation to avoid unknown variance parameters and obtained a distribution with fatter tails than the normal distribution. Although our approach is similar, note that we do not obtain the finite sample distribution of our test statistic, and that $Z_k^{-1} W_k(1)$ is not a multivariate t distribution.

3 Inference

3.1 Simple Hypotheses for Individual Parameters

In order to construct a test statistic for hypotheses about the individual β's, we let the square root of the diagonal elements of $T^{-1}\hat{B}$ assume the role. Suppose the null hypothesis is $H_0 : \beta_i = \beta_{0i}$. Define the t statistic $t^* = \left(\hat{\beta}_i - \beta_i\right) / \sqrt{T^{-1}\hat{B}_{ii}}$ where $\hat{B}_{ii}$ is the i^{th} diagonal element of the matrix $\hat{B}$. Because the t^* statistic is invariant to the ordering of the individual β's, its asymptotic distribution is given by the first element in the vector $Z_k^{-1} W_k(1)$. Making use of the fact that Z_k^{-1} is lower triangular, it is straightforward to show that

$$t^* \Rightarrow \frac{W_1(1)}{\sqrt{\int_0^1 \left(W_1(r) - rW_1(1)\right)^2 dr}} \equiv \frac{W_1(1)}{\sqrt{P_1}}. \tag{12}$$

The density of the limiting random variable in (12) has been derived by Abadir and Paruolo (1997). For convenience, we reproduce the critical values in Table I.

3.2 General Nonlinear Hypotheses

Suppose we are interested in testing general non-linear hypotheses. We examine hypotheses of the form

$$H_0 : r(\beta_0) = 0; \qquad H_1 : r(\beta_0) \neq 0;$$

where $r(\cdot) : \mathbb{R}^k \to \mathbb{R}^q$ imposes q restrictions on the parameter vector β. We restrict our attention to hypotheses where $r(\cdot)$ is twice continuously differentiable with bounded second derivatives near β_0; and $R(\cdot) \equiv \frac{\partial}{\partial \beta'} r(\cdot)$ has full rank q in a neighborhood around β_0; implying that there are genuinely q restrictions. For later use, define $\hat{R} = R\left(\hat{\beta}\right)$ and $R_0 = R(\beta_0)$.

The relevant test statistic is analogous to the standard Wald (or F) statistic for non-linear models. We simply substitute $\hat{B}$ for the standard variance-covariance matrix to obtain

$$F^* = T\, r\left(\hat{\beta}\right)' \left[\hat{R}\hat{B}\hat{R}'\right]^{-1} r\left(\hat{\beta}\right)/q.$$

The asymptotic distribution of F^* follows from application of the delta method to $r\left(\hat{\beta}\right)$ and is given by the following theorem which is proved in the appendix.

Theorem 2 Suppose that Assumptions 1 and 2 hold. Then under the null hypothesis $H_0 : r(\beta_0) = 0$, $F^* \Rightarrow W_q(1)' P_q^{-1} W_q(1)/q$ as $T \to \infty$.

With F^* we have constructed a test for general nonlinear hypotheses of the parameters in a broad class of non-linear models. The asymptotic distribution depends only on the number of restrictions (as would a standard Wald test). The density of $W_q(1)' P_q^{-1} W_q(1)/q$ has not been derived to our knowledge. However, critical values are easily simulated, and we tabulate critical values in Table II for $q = 1, 2, 3, \ldots, 30$.

4 Finite Sample Evidence in a Simple Location Model

In this section, we consider the finite sample performance of the t^* statistic in a simple special case of our more general framework. Consider a simple location model with AR(1) serially correlated data:

$$y_t = \beta + u_t, \qquad t = 1, \ldots, T, \tag{13}$$

$$u_t = \rho u_{t-1} + \varepsilon_t,$$

$$\varepsilon_t \sim \text{iid } N(0,1).$$

Inference regarding β even in this simple model can often be related to interesting empirical applications. For example, Zambrano and Vogelsang (2000) used this simple model to test the Law of Demand which is one of the core ideas in economics. Another example is testing whether the radon level in the basement of a house exceeds the limit recommended by the Environmental Protection Agency. When radon levels are measured in a house, typically hourly readings are taken over some short time period like 48 hours. This generates time series data that is highly serially correlated. The relevant question is whether the "average" radon level exceeds the cutoff. In other words, the question is whether an estimate of β in model (13) is consistent with a true value of β above the cutoff.

To make matters concrete suppose the null hypothesis of interest is $H_0 : \beta \leq 0$ and the alternative hypothesis is $H_0 : \beta > 0$. Suppose we estimate β by least squares to obtain $\hat{\beta} = T^{-1}\sum_{t=1}^{T} y_t$. We compare the performance of t^* to several HAC based t statistics. Note that all of the t statistics can be written as $\hat{\beta}/se(\hat{\beta})$. The first statistic, labelled, t_{HAC}, uses a HAC standard error calculated using the quadratic spectral kernel as recommended by Andrews (1991). The truncation lag is chosen using the data dependent method recommended by Andrews (1991) based on an AR(1) plug-in formula (see Andrews (1991) for details). The second statistic, t_{PARM} uses a standard error calculated using a parametric estimate of Ω given by $\hat{\Omega} = s_\varepsilon^2/(1-\hat{\rho})^2$ where $\hat{\rho} = \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} / \sum_{t=2}^{T} \hat{u}_{t-1}^2$, $\hat{u}_t = y_t - \hat{\beta}$ and $s_\varepsilon^2 = (T-1)^{-1}\sum_{t=2}^{T} \hat{\varepsilon}_t^2$ where $\hat{\varepsilon}_t = \hat{u}_t - \hat{\rho}\hat{u}_{t-1}$. Note that in practice, the form of serial correlation is usually unknown so that t_{PARM} is not often feasible. It is included here as a benchmark. Some authors in the HAC literature, most notably Andrews and Monahan (1992), have recommended using prewhitened spectral density estimators. Therefore, we also implement t^* and t_{HAC} using AR(1) prewhitening. The idea is to take $\hat{v}_t$ and then calculate $\hat{\Omega}$ (or $\hat{C}$ for the case of t^*) and then "recolor" these estimates by multiplying by $(1-\hat{\rho})^{-2}$. The prewhitened statistics are denoted by t_{HAC-PW} and t^*_{PW}.

We generated data according to (13) under the null hypothesis of $\beta = 0$. We computed empirical null rejection probabilities of the t tests using 5% (right tail) asymptotic critical values. We report results for $\rho = -0.5, -0.3, 0.0, 0.3, 0.5, 0.7, 0.9, 0.95, 0.99$ and sample sizes of $T = 25, 50, 100, 200$. We used 10,000 replications in all cases. The results are given in Table III. In nearly all cases, the t^* statistic has empirical rejection probabilities closest to the nominal level of 0.05. When prewhitening is used and the sample size is 200, as long as $\rho \leq 0.7$, t^* has empirical rejection probabilities almost exactly equal to 0.05 unlike the other t tests. When ρ is close to one, all of the statistics are size distorted, but the distortions for t^* are always much less.

Why is the asymptotic approximation so much better for t^* compared to the more standard t statistics? Consider the standard tests. Asymptotic standard normality of these tests follows by appealing to consistency of the estimated asymptotic variance. In a sense, merely appealing to consistency of the estimated asymptotic variance leads to a poor asymptotic approximation: If one is interested in testing hypotheses about β, then an asymptotic normality result for $\sqrt{T}(\hat{\beta} - \beta)$ is required. This is a first order asymptotic approximation for the sampling behavior of $\hat{b}$. But, this approximation cannot be used itself because there is an unknown asymptotic variance. Simply replacing the unknown asymptotic variance with an estimator, i.e. studentizing, and appealing to consistency of that estimator does not fully preserve the first order asymptotic approximation (consistency is a less than first order approximation). In other words, the sampling variability of the variance estimator is ignored in the asymptotics (the unknown variance is treated as known). Therefore, the first order asymptotic approximation is not "complete" for the standard tests. On the other hand, the asymptotic approximation for t^* captures the sampling variability of the standard error to first order through the random variable, $\sqrt{P_1}$, that appears in the denominator. Thus, the first order asymptotic approximation for t^* can be viewed as "complete" and is hence more accurate in finite samples. Note that we are not claiming that our asymptotic approximation for t^* is more accurate than first order (i.e. second order). There may still be room for further improvements in accuracy that could perhaps be obtained via bootstrapping or other resampling methods.

5 Empirical Nonlinear Regression: Example and Further Finite Sample Evidence

In this section we illustrate the new tests using an empirical example that includes a nonlinear regression model. We wish to examine the effects of growth in U.S. Gross Domestic Product (GDP) on the growth of aggregate restaurant revenues. Let ΔRR denote the first difference of the natural logarithm of real, seasonally adjusted aggregate restaurant revenues for the United States, and ΔGDP denote the first difference of the natural logarithm of real, seasonally adjusted GDP. Initially, we consider the basic regression model

$$\Delta RR_t = \beta_1 + \beta_2 \Delta GDP_t + u_t; \tag{14}$$

where β_1 is an intercept term, while β_2 is the parameter measuring the long run effect of GDP growth on restaurant revenues. It is well known that macroeconomic aggregates are serially correlated over time. To eliminate some of the autocorrelation in the error structure, we consider an AR(1) transformation of the model. The idea is to "soak" up some of the autocorrelation using the AR(1) transformation and then use HAC robust t tests or t^* to deal with any remaining autocorrelation in the model. Therefore, consider the model:

$$\begin{aligned} \Delta RR_t - \rho \Delta RR_{t-1} &= \beta_1 (1-\rho) + \beta_2 (\Delta GDP_t - \rho \Delta GDP_{t-1}) + u_t; \quad t \geq 2 \\ \left[1-\rho^2\right]^{\frac{1}{2}} \Delta RR_1 &= \left[1-\rho^2\right]^{\frac{1}{2}} \beta_1 + \left[1-\rho^2\right]^{\frac{1}{2}} \beta_2 \Delta GDP_1 + u_1 \end{aligned} \tag{15}$$

where ΔRR_{t-1} and ΔGDP_{t-1} are lagged values of first differences of real log-restaurant revenue and real log-GDP respectively. Model (15) is a nonlinear regression that can be estimated by NLS using a grid search over values of ρ.

We use quarterly data from 1971 through 1996 (a total of 104 observations). The restaurant revenues are total for all sectors and the source is Current Business Reports, published by the Bureau of the Census. The real GDP series was obtained from the Survey of Current Business published by the Bureau of Economic Analysis, US Department of Commerce and is seasonally adjusted. We converted nominal restaurant revenue to real revenue by dividing by the GDP deflator price index also obtained from the Survey of Current Business. We are interested in β_2, the long-run effect of GDP growth on the growth of restaurant revenues. We computed 95% confidence intervals for β_2 using t^* and t_{HAC}. We implemented t_{HAC} as recommended by Andrews (1991) using the quadratic spectral kernel with a data dependent truncation lag based on the $VAR(1)$ plug-in method (see Andrews (1991) for details). Table IV summarizes the results. We see that the different methods of calculating confidence intervals result in different intervals, and that t^* provides a tighter confidence interval than t_{HAC}.

Using this empirical example to calibrate realistic data generating processes, we compared finite sample size and power of t^* and t_{HAC} using simulations. To this end, we fit the residuals from (14) to several different ARMA models and found that an AR(4) model provides a good fit. We also fit ΔGDP to several ARMA processes and found that an AR(1) model provides a good fit. To perform the simulations, we generated data according to two models. Model A used the actual real GDP data to generate pseudo data for restaurant revenues: $\Delta RR_t = \beta_2 \Delta GDP_t + u_t$. In model B, we simulated pseudo data for ΔGDP_t using the model $\Delta GDP_t = -0.21 \Delta GDP_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, 0.007888)$. In both models, the error term was generated according to $u_t = -0.343u_{t-1} - 0.330u_{t-2} - 0.269u_{t-3} + 0.595u_{t-4} + \xi_t$, $\xi_t \sim iidN(0, 0.0367746)$. We take the null hypothesis to be $H_0 : \beta_2 = 0$. We generated data using $\beta_2 = 0.0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2$. We calculated empirical rejection probabilities using 5% asymptotic critical values. Results for $\beta_2 = 0$ correspond to size while results for $\beta_2 > 0$ correspond to power. 2,000 replications were used in all cases.

The results are summarized in Table V. We see that t^* dominates t_{HAC} with respect to size. There is significantly less distortion in the sense that the finite sample size of t^* is closer to the nominal level of 5% than for the other statistics. The fact that t_{HAC} is so undersized explains why the empirical confidence intervals are wider. No single test dominates in terms of power and in many cases, t^* has the highest power.

6 Conclusion

In this paper, we have developed test statistics to test possibly non-linear hypotheses in nonlinear, weighted regression models with errors that have serial correlation and/or heteroskedasticity of unknown form. These tests are simple and do not require use of heteroskedasticity autocorrelation consistent (HAC) estimators. Thus, our approach avoids the pitfalls associated with the choice of truncation lag required when using HAC estimators. We derived the limiting null distributions of these new tests in a general nonlinear setting, and showed that while the tests have nonstandard distributions, the distributions depend only upon the number of restrictions, and critical values were easily obtained using simulations. Monte Carlo simulations show that in finite samples the first order asymptotic approximation for the new tests is more accurate than the first order asymptotic approximation for standard tests. The simulations also show that the new tests have competitive power.

7 Appendix

Proof of Lemma 1: First we will make use of the following Taylor expansion of $F_t\left(\hat{\beta}\right)$ around β_0:

$$F_t\left(\hat{\beta}\right) = F_t(\beta_0) + \frac{\partial}{\partial \beta'} F_t(\delta)'\left(\hat{\beta} - \beta_0\right); \quad \text{where} \quad \delta \in \left[\hat{\beta}, \beta_0\right]$$

Plugging this into the expression for $\hat{S}_{[rT]}$; we obtain

$$\begin{aligned}
T^{-\frac{1}{2}}\hat{S}_{[rT]} &= T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} F_t\left(\hat{\beta}\right) \hat{u}_t \\
&= T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} F_t(\beta_0)\hat{u}_t + T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} \frac{\partial}{\partial \beta'} F_t(\delta)'\left(\hat{\beta} - \beta_0\right)\hat{u}_t \\
&= T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} F_t(\beta_0)\hat{u}_t + I_{[rT]};
\end{aligned}$$

where

$$I_{[rT]} = T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} \frac{\partial}{\partial \beta'} F_t(\delta)'\left(\hat{\beta} - \beta_0\right)\hat{u}_t. \tag{16}$$

What remains to be shown is that $I_{[rT]}$ is $o_p(1)$. We rewrite $\hat{u}_t$ in the following manner:

$$\hat{u}_t = f_t(\beta_0) + u_t - f_t\left(\hat{\beta}\right)$$

and plugging this into (16) gives

$$\begin{aligned}
I_{[rT]} &= T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} \frac{\partial}{\partial \beta'} F_t(\delta)'\hat{u}_t\left(\hat{\beta} - \beta_0\right) \\
&= T^{-1} \sum_{t=1}^{[rT]} \frac{\partial}{\partial \beta'} F_t(\delta)' u_t \left\{\sqrt{T}\left(\hat{\beta} - \beta_0\right)\right\} \\
&\quad + T^{-1} \sum_{t=1}^{[rT]} \frac{\partial}{\partial \beta'} F_t(\delta)'\left(f_t(\beta_0) - f_t\left(\hat{\beta}\right)\right)\left\{\sqrt{T}\left(\hat{\beta} - \beta_0\right)\right\}.
\end{aligned}$$

Now, by assumption 1, $T^{-1}\sum_{t=1}^{[rT]}\frac{\partial}{\partial\beta'}F_t(\cdot)\varepsilon_t = o_p(1)$ and we know that $\sqrt{T}\left(\hat\beta-\beta_0\right) = O_p(1)$. This implies that

$$I_{[rT]} = T^{-1}\sum_{t=1}^{[rT]}\frac{\partial}{\partial\beta'}F_t(\cdot)\left(f_t(\beta_0)-f_t\left(\hat\beta\right)\right)\left\{\sqrt{T}\left(\hat\beta-\beta_0\right)\right\}+o_p(1). \qquad (17)$$

To determine the behavior of the last remaining term, we expand $f_t\left(\hat\beta\right)$ around β_0:

$$f_t\left(\hat\beta\right) = f_t(\beta_0)+F_t\left(\ddot\beta\right)'\left(\hat\beta-\beta_0\right); \quad \text{where} \quad \ddot\beta\in\left[\hat\beta,\beta_0\right],$$

$$f_t(\beta_0)-f_t\left(\hat\beta\right) = -F_t\left(\ddot\beta\right)'\left(\hat\beta-\beta_0\right)$$

This expansion together with (17) provides us with the expression

$$\begin{aligned} I_{[rT]} &= -T^{-1}\sum_{t=1}^{[rT]}\frac{\partial}{\partial\beta'}F_t(\cdot)F_t\left(\ddot\beta\right)'\left(\hat\beta-\beta_0\right)\left\{\sqrt{T}\left(\hat\beta-\beta_0\right)\right\}+o_p(1) \\ &= -T^{-\frac{1}{2}}\left(T^{-1}\sum_{t=1}^{[rT]}\frac{\partial}{\partial\beta'}F_t(\cdot)F_t\left(\ddot\beta\right)'\sqrt{T}\left(\hat\beta-\beta_0\right)\right)\sqrt{T}\left(\hat\beta-\beta_0\right)+o_p(1)\end{aligned}$$

Because $T^{-1}\sum_{t=1}^{[rT]}\frac{\partial}{\partial\beta'}F_t(\cdot)F_t\left(\ddot\beta\right)'$ and $\sqrt{T}\left(\hat\beta-\beta_0\right)$ both are $O_p(1)$, $I_{[rT]} = o_p(1)$ and the proof is complete. ■

Proof of Theorem 1: We wish to find the asymptotic distribution of

$$F^* = Tr\left(\hat\beta\right)'\left[\hat R\hat B\hat R'\right]^{-1}r\left(\hat\beta\right)/q.$$

First note that using the delta method it follows that $\sqrt{T}r\left(\hat\beta\right) \Rightarrow R_0Q^{-1}\Lambda W_k(1)$; implying that

$$\begin{aligned} F^* &= \sqrt{T}r\left(\hat\beta\right)'\left[\hat R\hat B\hat R'\right]^{-1}\sqrt{T}r\left(\hat\beta\right)/q \\ &\Rightarrow \left[R_0Q^{-1}\Lambda W_k(1)\right]'\left[R_0Q^{-1}\Lambda P_k\Lambda'Q^{-1}R_0'\right]^{-1}\left[R_0Q^{-1}\Lambda W_k(1)\right]/q.\end{aligned}$$

Because $R_0 Q^{-1} \Lambda$ has rank q and $W_k(1)$ is a vector of independent Wiener processes that are Gaussian, we can rewrite $R_0 Q^{-1} \Lambda W_k(1)$ as $\Lambda^* W_q(1)$, where $W_q(1)$ is a $q-$dimensional vector of independent Wiener processes, and Λ^* is the $q \times q$ matrix square root of $R_0 Q^{-1} \Lambda \Lambda' Q^{-1} R_0'$. Note that this square root exist because $R_0 Q^{-1} \Lambda \Lambda' Q^{-1} R_0'$ is a full rank $q \times q$ matrix. Using the same arguments, note that

$$
\begin{aligned}
& R_0 Q^{-1} \Lambda P_k \Lambda' Q^{-1} R_0' \\
= \; & R_0 Q^{-1} \Lambda \int_0^1 (W_k(r) - r W_k(1)) (W_k(r) - r W_k(1))' dr \Lambda' Q^{-1} R_0' \\
= \; & \int_0^1 \left(R_0 Q^{-1} \Lambda W_k(r) - r R_0 Q^{-1} \Lambda W_k(1) \right) \left(R_0 Q^{-1} \Lambda W_k(r) - r R_0 Q^{-1} \Lambda W_k(1) \right)' dr \\
= \; & \int_0^1 (\Lambda^* W_q(1) - r \Lambda^* W_q(1)) (\Lambda^* W_q(1) - r \Lambda^* W_q(1))' dr \\
= \; & \Lambda^* \int_0^1 (W_q(1) - r W_q(1)) (W_q(1) - r W_q(1))' dr \, (\Lambda^*)' \\
= \; & \Lambda^* P_q (\Lambda^*)'.
\end{aligned}
$$

Therefore it directly follows that

$$
\begin{aligned}
F^* \; \Rightarrow \; & [\Lambda^* W_q(1)]' \left[\Lambda^* P_q (\Lambda^*)' \right]^{-1} [\Lambda^* W_q(1)] / q \\
= \; & W_q(1)' P_q^{-1} W_q(1) / q.
\end{aligned}
$$

■

References

[1] Abadir, K.M. and P. Paruolo (1997), "Two Mixed Normal Densities From Cointegration Analysis," Econometrica, 65, 671-680.

[2] Andrews, D.W.K. (1991), “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” Econometrica, 59, 817-858.

[3] Andrews, D.W.K. and J.C. Monahan (1992), “An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator,” Econometrica, 60, 953-966.

[4] Davidson, R. and J.G. MacKinnon (1993), Estimation and Inference in Econometrics, New York: Oxford University Press.

[5] Eicker, F. (1967), ”Limit Theorems for Regressions with Unequal and Dependent Errors,” Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 59-82, Berkeley, University of California Press.

[6] Gallant, A.R. (1987), Nonlinear Statistical Models, New York: Wiley.

[7] Hansen, B.E. (1992), “Consistent Covariance Matrix Estimation for Dependent Heterogenous Processes,” Econometrica, 60, 967-972.

[8] Hong, Y. and J. Lee, “Wavelet-based Estimation of Heteroscedasticity and Autocorrelation Consistent Covariance Matrices”, Cornell Working paper, October 1999.

[9] Huber, P.J. (1967), ”The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions,” Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 221-233, Berkeley, University of California Press.

[10] Kiefer, N.M., Vogelsang, T.J, and H. Bunzel (2000), “Simple Robust Testing of Regression Hypotheses”, Econometrica, 68, 695-714.

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

[12] Phillips, P.C.B. and S.N. Durlauf (1986), "Multiple Time Series Regression with Integrated Processes," Review of Economic Studies, 53, 473-495.

[13] Priestley, M.B. (1981), Spectral Analysis and Time Series, Vol. 1, Academic Press, NewYork.

[14] Robinson, P.M (1991), "Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models", Econometrica, 59, 1329-1363.

[15] White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and Direct Test for Heterskedasticity," Econometrica, 48, 817-838.

[16] White, H. (1984), Asymptotic Theory for Econometricians, Orlando: Academic Press.

[17] Zambrano, E. and T.J. Vogelsang (2000), "A Simple Test of the Law of Demand for the United States," Econometrica, 68, 1013-1022.

Table I: Asymptotic Critical values of t^*

1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95%	97.5%	99.0%
-8.613	-6.747	-5.323	-3.875	0.000	3.875	5.323	6.747	8.613

Source: Line 1 of Table I from Abadir and Paruolo (1997, p. 677).

Table II: Asymptotic Critical values of F^*.

% / q	1	2	3	4	5	6	7	8	9	10
90.0	28.88	35.68	42.39	48.79	55.02	61.18	67.37	73.10	78.52	83.84
95.0	46.39	51.41	58.17	65.33	71.69	78.70	84.63	90.89	96.38	101.8
97.5	65.94	69.76	76.07	83.35	89.65	96.53	102.7	109.8	114.2	120.0
99.0	101.2	96.82	100.7	108.4	114.2	121.2	126.9	134.4	139.6	144.9

% / q	11	12	13	14	15	16	17	18	19	20
90.0	89.39	94.47	100.1	105.3	110.3	115.5	121.2	126.6	131.5	136.5
95.0	107.7	113.6	119.9	125.3	131.5	136.6	141.4	147.1	152.9	158.0
97.5	127.2	132.9	138.8	145.2	151.0	155.9	161.1	167.6	174.0	179.8
99.0	152.6	157.8	163.8	169.7	174.7	181.6	188.8	194.8	203.2	208.5

% / q	21	22	23	24	25	26	27	28	29	30
90.0	141.9	146.6	152.1	157.0	161.8	167.2	171.6	177.0	181.6	187.0
95.0	163.6	169.3	174.7	180.3	184.9	190.7	196.0	201.5	206.4	211.4
97.5	186.0	191.2	197.0	202.3	207.5	213.3	218.9	224.4	229.1	236.0
99.0	214.0	219.3	224.6	230.1	236.3	242.4	246.9	252.9	259.8	266.3

Notes: q is the number of restrictions being tested. The critical values were simulated with 100,000 replications using normalized partial sums of 1000 iid $N(0;1)$ random deviates to approximate the Wiener processes in the limiting distribution. In addition to this table James MacKinnon provides on his webpage,

http://qed.econ.queensu.ca/pub/faculty/mackinnon,

a program that calculates critical values and the p-values for both the t^* the F^* statistics using response surface methods.

Table III: Empirical Null Rejection Probabilities in Simple Location Model
$y_t = \beta + u_t$; $H_0 : \beta \leq 0$; $H_1 : \beta > 0$; 5% Nominal Level

T = 25					
ρ	t_{HAC}	t^*	t_{HAC-PW}	t^*_{PW}	t_{PARM}
-0.5	0.072	0.028	0.078	0.045	0.071
-0.3	0.065	0.035	0.075	0.045	0.070
0.00	0.070	0.047	0.077	0.045	0.077
0.30	0.098	0.058	0.085	0.047	0.087
0.50	0.124	0.073	0.095	0.052	0.098
0.70	0.175	0.104	0.120	0.067	0.127
0.90	0.275	0.180	0.192	0.110	0.199
0.95	0.317	0.207	0.222	0.124	0.229
0.99	0.351	0.226	0.242	0.121	0.251

T = 100					
ρ	t_{HAC}	t^*	t_{HAC-PW}	t^*_{PW}	t_{PARM}
-0.5	0.057	0.043	0.058	0.047	0.056
-0.3	0.054	0.046	0.057	0.047	0.056
0.00	0.057	0.050	0.058	0.048	0.058
0.30	0.073	0.054	0.060	0.048	0.060
0.50	0.086	0.058	0.062	0.048	0.063
0.70	0.113	0.066	0.070	0.049	0.071
0.90	0.197	0.102	0.106	0.065	0.108
0.95	0.255	0.142	0.146	0.085	0.147
0.99	0.351	0.220	0.222	0.123	0.223

T = 50					
ρ	t_{HAC}	t^*	t_{HAC-PW}	t^*_{PW}	t_{PARM}
-0.5	0.061	0.037	0.062	0.045	0.059
-0.3	0.056	0.042	0.061	0.046	0.060
0.00	0.060	0.048	0.061	0.047	0.061
0.30	0.079	0.054	0.065	0.048	0.066
0.50	0.099	0.062	0.071	0.048	0.074
0.70	0.133	0.079	0.087	0.054	0.089
0.90	0.235	0.136	0.139	0.081	0.144
0.95	0.292	0.187	0.190	0.107	0.195
0.99	0.354	0.233	0.238	0.129	0.242

T = 200					
ρ	t_{HAC}	t^*	t_{HAC-PW}	t^*_{PW}	t_{PARM}
-0.5	0.056	0.048	0.057	0.050	0.056
-0.3	0.053	0.049	0.057	0.051	0.057
0.00	0.057	0.051	0.057	0.050	0.057
0.30	0.072	0.054	0.058	0.050	0.058
0.50	0.081	0.056	0.060	0.050	0.061
0.70	0.099	0.060	0.064	0.051	0.065
0.90	0.167	0.079	0.087	0.057	0.088
0.95	0.219	0.107	0.111	0.068	0.112
0.99	0.344	0.199	0.195	0.114	0.197

Notes: The error terms, u_t, were generated according to $u_t = \rho u_{t-1} + \varepsilon_t$, $\xi_t \sim iidN(0,1)$. The model is generated as $y_t = \beta + u_t$ with $\beta = 0$. β is estimated by OLS.

Table IV: Empirical 95% Confidence Intervals

	$\hat{\phi}_2$	$\hat{\rho}$	t^*	t_{HAC}
Model (14)	0.681	NA	[0.305, 1.056]	[0.059, 1.302]
Model (15)	0.694	−0.293	[0.385, 1.003]	[0.184, 1.203]

Table V: Empirical Rejection Probabilities, Model (14)
$T = 103$, 5% Nominal Level.

		Power is Size	Adjusted	Power not Size	Adjusted
DGP	β_0	t^*	t_{HAC}	t^*	t_{HAC}
A	0.0	0.028	0.006	0.028	0.006
	0.2	0.158	0.177	0.101	0.039
	0.4	0.429	0.504	0.339	0.187
	0.6	0.744	0.813	0.652	0.473
	0.8	0.925	0.961	0.871	0.750
	1.0	0.978	0.993	0.957	0.911
	1.2	0.994	1.000	0.988	0.979
	1.4	1.000	1.000	0.996	0.995
	1.6	1.000	1.000	1.000	1.000
	1.8	1.000	1.000	1.000	1.000
	2.0	1.000	1.000	1.000	1.000
B	0.0	0.046	0.031	0.046	0.031
	0.2	0.108	0.106	0.098	0.070
	0.4	0.252	0.289	0.233	0.223
	0.6	0.449	0.542	0.423	0.459
	0.8	0.628	0.739	0.607	0.678
	1.0	0.768	0.872	0.753	0.833
	1.2	0.867	0.941	0.854	0.916
	1.4	0.924	0.976	0.913	0.963
	1.6	0.954	0.990	0.947	0.985
	1.8	0.972	0.997	0.967	0.995
	2.0	0.980	0.998	0.977	0.997

Notes: For both models, pseudo data for ΔRR_t was generated using (14) with the error terms given by $u_t = -0.343u_{t-1} - 0.330u_{t-2} - 0.269u_{t-3} + 0.595u_{t-4} + \xi_t;\ \xi_t \sim N(0; 0.0367746)$. Model A uses the original ΔGDP_t data as the regressor, while Model B generates the GDP series according to the following process: $\Delta GDP_t = -0.21\Delta\Delta GDP_t + \zeta_t;\ \zeta_t \sim N(0; 0.007888)$. The slanted script provide empirical null rejection probabilites while the other table entries are finite sample power.