The Superiority of the LM Test in a Class of Models Where the Wald Test Performs Poorly; Including Nonlinear Regression, ARMA, GARCH, and Unobserved Components

Jun Ma*
Department of Economics, Finance and Legal Studies
University of Alabama

Charles R. Nelson
Department of Economics
University of Washington

September 18, 2012

JEL *classification*: C120, C220, C330.

Keywords: LM test, Non-linear regression, ARMA, Unobserved Components, State Space, GARCH, Zero-Information-Limit-Condition.

Abstract

In models having a representation $y = \gamma \bullet g(\beta, x) + \varepsilon$ the standard Wald test for $\hat{\beta}$ has systematically wrong size in finite samples when the identifying parameter γ is small relative to its estimation error; Nelson and Startz (2007). Here we study the LM test based on linearization of $g(.)$ which may be interpreted as an approximation to the exact test of Fieller (1954) for a ratio of regression coefficients, or as a reduced form test which is exact when g(.) is linear. We show that this test has nearly correct size in non-linear regression, ARMA, GARCH, and Unobserved Components models when the Wald test performs poorly.

* Corresponding author: jma@cba.ua.edu.

** The authors wish to thank Manfred Deistler, Michael Dueker, Walter Enders, Walter Fisher, Andrew Harvey, Evan Koenig, Junsoo Lee, James Morley, Christian Murray, David Papell, Hashem Pesaran, Benedikt Pötscher, Leopold Soegner, Richard Startz, James Stock, Martin Wagner, Arnold Zellner and Eric Zivot, for helpful comments, but responsibility for all errors is ours. Support from the Ford and Louisa Van Voorhis Endowment at the University of Washington is gratefully acknowledged.

1. Introduction

This paper is concerned with inference in models that can take the form:

$$y_i = \gamma \bullet g(\beta, \boldsymbol{\theta}, \mathbf{x}_i) + \varepsilon_i ; i = 1,..., N . \tag{1.1}$$

Parameters β, of interest for testing, and $\boldsymbol{\theta}$, a vector of other parameters, are identified only if $\gamma \neq 0$. This class is broader than may appear at first, including, in addition to non-linear regression models, the workhorse ARMA model, the GARCH model, and Unobserved Components models. When errors ε_i are *i.i.d.* $N(0, \sigma^2)$, Maximum Likelihood estimates are obtained by non-linear least squares. Standard inference for $\hat{\beta}$ uses the estimated asymptotic standard error to construct a Wald *t*-test statistic, denoted $t(\hat{\beta})$, or confidence interval. What models in this class have in common is that standard inference works poorly in finite samples when the identifying parameter γ is small relative to its estimation error. Nelson and Startz (2007) – hereafter NS - show that the estimated standard error for $\hat{\beta}$ is generally smaller than implied by asymptotic theory, essentially because it depends on $\hat{\gamma}$. Further, the distribution of $\hat{\beta}$ will generally be displaced away from the true value. Although these two effects might seem to imply that the size of the test based on $t(\hat{\beta})$would be excessive, NS show that it may be too large or too small since the numerator and denominator are correlated in finite samples, and the direction depends on specifics of the model and data. Further, while asymptotic theory is valid when the identification condition is met, the limiting distribution of $t(\hat{\beta})$ takes hold slowly as sample size increases in the range relevant for applied work in economics.

In this paper we consider an LM-type test based on linearization of $g(\beta, \mathbf{\theta}, \mathbf{x}_i)$, in the spirit of Breusch and Pagan (1980) and Harvey (1990). The test may, alternatively, be interpreted as Fieller's (1954) test for a ratio of regression coefficients in the linearized model, or as a reduced form test, in the spirit of Anderson and Rubin (1949), both of which have correct size (and are equivalent) if $g(.)$ is linear. While the LM test is asymptotically equivalent to Wald and LR, its relation to tests that have correct size in the linear case suggests that the LM test will be close to exact when the linear approximation to $g(\beta, \mathbf{\theta}, \mathbf{x}_i)$ is close. Section 2 presents the "modified LM' test and compares its size and power in small samples with the standard Wald test in the archetypal case of linearity. Section 3 studies the small sample performance of the LM test in several non-linear models that are of considerable interest in practice, namely non-linear regression, ARMA, the Unobserved Components model of trend and cycle, and GARCH models. Section 4 concludes.

2. The modified LM test as an approximation to a test that is exact in finite samples.

We begin with the model in (1.1) and linearize $g(\beta, \mathbf{\theta}, \mathbf{x}_i)$ in β around the null hypothesis, $H_0 : \beta = \beta_0$, obtaining:

$$y_i = \gamma \bullet g(\beta_0, \mathbf{\theta}, x_i) + \lambda \bullet g_\beta(\beta_0, \mathbf{\theta}, x_i) + e_i ; \lambda = \gamma \bullet (\beta - \beta_0), \tag{2.1}$$

where $g_\beta = dg(.)/d\beta$ and e_i includes a remainder. The LM test may be computed in two-steps:

(1) Fit the restricted model $y_i = \gamma \bullet g(\beta_0, \mathbf{\theta}, \mathbf{x}_i) + e_i$, obtaining the restricted estimates $\tilde{\gamma}$ and $\tilde{\mathbf{\theta}}$ and restricted residuals: $\tilde{e}_i = y_i - \tilde{\gamma} \bullet g(\beta_0, \tilde{\mathbf{\theta}}, x_i)$.

(2) Regress $\tilde{e}_i$ on $g(\beta_0, \tilde{\mathbf{\theta}}, x_i)$ and $g_\beta(\beta_0, \tilde{\mathbf{\theta}}, x_i)$ to see if the second term adds significant explanatory power.

However, following Harvey (1990), a "modified LM" test can be done in one step by running regression (2.1), noting the significance of the second term in a Wald *t*-test, using statistic $t(\hat{\lambda})$. He notes that in the case of linear regression the resulting statistic will be essentially the F-statistic for additional variables, suggesting that little is lost in doing one step in the non-linear case. Indeed, we find this to be the case.

Note too that since the null hypothesis $\beta = \beta_0$ implies $\lambda = 0$, the modified LM test may also be thought of as a reduced form (RF) test. The RF test will have exact size in finite samples when $g(\beta, \boldsymbol{\theta}, \mathbf{x}_i)$ is linear, since in that case (2.1) is a classical linear regression. Alternatively, Fieller (1954) noted that an exact test may be obtained for a ratio of regression coefficients from the reduced form, in this case by making use of the relation $\lambda = \gamma \bullet (\beta_0 + \Delta\beta)$ where $\Delta\beta$ is the amount by which β differs from its value under the null hypothesis. Thus the three approaches all use the reduced form regression obtained from the first order expansion of $g(\beta, \mathbf{x}_i)$ and the resulting Wald *t*-statistic, $t(\hat{\lambda})$.

Recall that standard inference for $\hat{\beta}$ uses the estimated asymptotic standard error, obtained from the last iteration of non-linear least squares or by the 'delta method,' to construct a Wald *t*-statistic, $t(\hat{\beta})$. Because the modified LM is also based on a Wald *t*-statistic, $t(\hat{\lambda})$ in the RF, we try to reduce the potential confusion by referring to them as the 'standard test' and 'LM test' respectively. To demonstrate the problems with the standard test in this setting, and thus the need for an alternative test, it is useful to consider the archetypal model,

$$y_i = \gamma \bullet (x_i + \beta \bullet z_i) + \varepsilon_i \quad . \tag{2.2}$$

An example in practice is the Phillips curve model of Staiger, Stock and Watson (1997), where y is the change in inflation, $g = (x_i + \beta)$, x is actual unemployment and β is the natural rate of unemployment, or NAIRU. The standard test for β may be based on the asymptotic standard error for $\hat{\beta}$ or equivalently the 'delta method.' Alternatively, to test the hypothesis $\beta = 0$ using the modified LM approach we expand $g(.)$, obtaining the reduced form

$$y_i = \gamma \bullet x_i + \lambda \bullet z_i + \varepsilon_i \tag{2.3}$$

and test the hypothesis $\lambda = \gamma \bullet \beta = 0$.

NS showed that the size of the Wald test using $t(\hat{\beta})$ will depend on the correlation between the reduced form coefficients as determined by correlation between the regressors. When the regressors are orthogonal the size is too small, in fact for any given sample size $t(\hat{\beta}) < t(\hat{\lambda})$. In contrast, when correlation between the regressors is strong, this inequality reverses; despite the fact that the asymptotic standard error for $\hat{\beta}$ is always downward biased; see Appendix A for further details. The modified LM test based on $t(\hat{\lambda})$ has exact size since the RF is a classical linear regression.

The identification condition $\gamma \neq 0$ is a maintained hypothesis underlying the asymptotic standard error and *t*-statistic for $\hat{\beta}$. If it does not hold then the information matrix for the model is singular. However, $t(\hat{\lambda})$ still has an exact *t*-distribution when identification fails because the reduced form (2.3) is a properly specified classical regression regardless of the value of γ. Of course the data do not contain information about β in that case so the test has no power.

To illustrate the relative performance of the two tests for this archetypal model, we report Monte Carlo experiments where the true $\beta = 0$, the regressors have unit variance and fixed in repeated samples, and errors are *i.i.d.* $N(0,1)$. Estimation is done in EViews™ using the non-linear regression routine, so the evaluation of $t(\hat{\beta})$ is representative of what would be reported in applied research. The number of replications is 10,000 in all experiments in this paper, and we note that the standard deviation of estimated size is .002 when the true size is .05.

Figure 2.1 explores the response of rejection frequencies for tests based on $t(\hat{\beta})$ and $t(\hat{\lambda})$ to departures of the true value of β from the null hypothesis value of zero. The true value of γ =.10, and the independent variables are, alternatively, orthogonal and strongly correlated (correlation 0.9). Since size is not correct for $t(\hat{\beta})$ this response can at best suggest whether the test conveys some information about the null hypothesis. For the case of orthogonal regressors, what we see is that the frequency of rejection using $t(\hat{\beta})$ increases very slowly as a function of the true β. In contrast, the power of the correctly sized test based on $t(\hat{\lambda})$ rises steeply as the true β departs from zero. The corresponding comparison when the independent variables are strongly correlated reveals that rejections by the standard test become *less* frequent as the true value of β departs farther from the null of zero rather than more frequent. In contrast, the power of the correctly sized LM test does increase as expected as the null departs from the true value. Thus we conclude that the Wald test is not only poorly sized but contains little if any information about the null hypothesis. This echoes the findings of Staiger, Stock, and Watson (1996) who report that in Monte Carlo experiments in the more restrictive NAIRU model 'the coverage rate of the delta method confidence interval is poorly controlled over empirically relevant portions of the parameter space.' Based on this finding, the confidence intervals for the NAIRU presented in their well-known 1997 paper use Fieller's method, and are wider than intervals using delta method standard errors.

[Insert Figure 2.1 here]

Asymptotic theory does, necessarily, take hold as sample size becomes large, albeit very slowly, as is evident in Table 2.1. The size of the test based on $t(\hat{\beta})$ approaches its nominal level only as the quantity $\gamma / \sqrt{V_{\hat{\gamma}}}$ approaches 10, requiring a sample size as large as 10,000 for $\gamma = .1$, and 1,000,000 for $\gamma = .01$! This provides a metric for 'weak identification' and a benchmark for where we expect asymptotic tests to perform satisfactorily. Distribution and testing theory for weakly identified models is an active area of research in econometrics and the focus of a rapidly growing literature. Among others, D. W. K. Andrews and Cheng (2010) study the asymptotic properties of a range of estimators in a class of models in which parameters are unidentified or weakly identified in some parts of the parameter space. I. Andrews and Mikusheva (2011) investigate robust approaches to testing in the context of a DSGE model.

[Insert Table 2.1 here]

Since the LM test is not expected to have exact size when $g(.)$ is not linear we use simulation to evaluate its performance relative to the standard *t*-test in four models of practical interest in Section 3. Based on NS we expect that the size of the standard test will depend on the correlation between $g(.)$ and $g_{\beta}(.)$ which are fixed in the linear case. In the non-linear case an estimation routine like Gauss-Newton iterates on β_0 to obtain least squares estimates, the final standard errors and resulting *t*-statistic being based on evaluation of $g(.)$ and $g_{\beta}(.)$ at $\beta = \hat{\beta}$. Thus the correlation between the 'regressors' is not fixed in the general case but rather depends on the provisional value of β at each iteration. As we see below, this co-determination affects the distribution of the point estimate and the size of the standard test, but not the reduced form test which relies on evaluation of the $g(.)$ and $g_{\beta}(.)$ under the null hypothesis.

3. Small sample performance of the modified LM test in four models.

3.1. Non-linear Regression: A Production Function.

Consider the Hicks-neutral Cobb-Douglas production function:

$$y_i = \gamma \bullet x_i^{\beta} + \varepsilon_i ; \gamma \neq 0 \qquad (3.1.1)$$

where y_i and x_i are per capita output and capital input respectively, γ is Total Factor Productivity, and β the share of capital input. The linear reduced form approximation is

$$y_t = \gamma \bullet x_i^{\beta_0} + \lambda \bullet x_i^{\beta_0} \log(x_i) + e_i \qquad (3.1.2)$$

where $\lambda = \gamma \bullet (\beta - \beta_0)$. Based on the analysis of the linear model we expect the point estimate $\hat{\beta}$ and the size of the standard *t*-test to be biased in directions indicted by the correlation between x_i^{β} and $x_i^{\beta} \log(x_i)$, corresponding to $g(\beta, x_i)$ and $g_{\beta}(\beta, x_i)$. Alternatively, the LM test will be based on the reduced form coefficient λ which we expect to have close to correct size. To see if these implications hold, we drew one sample of x_i from the log-normal distribution and paired it with 10,000 paths of standard Normal ε_i, each of sample size 100. Estimation is done in EViews™ using the nonlinear regression routine.

Table 3.1.1 reports estimation results for values of β in the economically relevant range, zero to .9, with γ =.01. The second line is the un-centered correlation between the 'regressors' in the linear reduced form, x_i^{β} and $x_i^{\beta} \log(x_i)$ which is of interest because analysis of the linear case suggests that bias in $\hat{\beta}$ should vary inversely with it; see Appendix A. Thus when the true value of β is .9, and the regressors are evaluated at .9 the correlation between regressors is .92 and $\hat{\beta}$ should be strongly biased downward. As we see below, $\hat{\beta}$ is biased downward as expected, but the relation of bias to true β is attenuated because Gauss-Newton evaluates the regressors at the estimated rather than true value of β. Note also that the standard Wald *t*-test rejects the null too infrequently when the true β is zero but rejects too often when true β is large. While the relation of size distortion to correlation is again in the expected direction based on the linear case, it is attenuated because regressors are evaluated at the downwardly biased point estimates. Finally, we also report the size of the LM test which is close to correct in all cases.

[Insert Table 3.1.1 here]

We report in Table 3.1.2 the corresponding results as γ increases from .01 to 1, fixing the true value of β at .5. As $\gamma/\sqrt{V_{\hat{\gamma}}}$ increases identification becomes better, the asymptotic distribution of $t(\hat{\beta})$ gradually takes hold, and its size approaches the nominal level of .05. We find that a value of 10 for the metric $\gamma/\sqrt{V_{\hat{\gamma}}}$ seems to provide a rough rule of thumb for good identification and correct size. In contrast, the LM test, maintains approximately the correct size across the range of parameter values.

[Insert Table 3.1.2 here]

A special case of interest is $\gamma = 0$ corresponding to failure of the identification condition for β, when the asymptotic theory underling the Wald standard error and *t*-statistic for $\hat{\beta}$ is not valid. However, the reduced form test does not depend on that assumption, and we find that its empirical size is 0.054, close to its nominal size.

Figure 3.1.1 compares rejection frequencies for the two tests as the true β departs from the fixed null $H_0 : \beta = .5$. The standard test using $t(\hat{\beta})$ starts with a higher level of rejection frequency when the true β is .5, reflecting its size distortion, and rejections decline as true β increases from .5. In contrast, the LM test based on $t(\hat{\lambda})$ starts with correct size when the null is true and power increases *monotonically* as true β deviates from the null. Neither test is sensitive to departure from the null in the direction of zero; we surmise that the non-linearity of the model accounts for this asymmetry.

[Insert Figure 3.1.1 here]

3.2. The ARMA (1,1) Model.

Ansley and Newbold (1980) in their study of ARMA model estimation noted that in the case of near parameter redundancy standard Wald confidence intervals are too narrow. NS also study the failure of Wald inference in that situation. Fortunately, ARMA models fall into the class we are concerned with here. We begin with the workhorse ARMA(1,1) and inference for the moving average coefficient. The results are then extended to the autoregressive coefficient and higher order models. Consider then:

$$\begin{aligned} y_t &= \phi \bullet y_{t-1} + \varepsilon_t - \theta \bullet \varepsilon_{t-1}; t = 1,...,T \\ \varepsilon_t &\sim i.i.d.N(0,\sigma_\varepsilon^2), |\phi|<1, |\theta|<1 \end{aligned} \tag{3.2.1}$$

Given invertibility of the moving average, we may express it in the form:

$$y_t = \gamma \bullet g(\theta, \bar{y}_{t-1}) + \varepsilon_t \tag{3.2.2}$$

where, $\gamma = (\phi - \theta)$, $g(\theta, \bar{y}_{t-1}) = \sum_{i=1}^{\infty} \theta^{i-1} y_{t-i}$ and $\bar{y}_{t-1} = (y_{t-1}, y_{t-2}, ...)$. NS show that when γ is small relative to the sample variation the estimated standard error for either $\hat{\phi}$ or $\hat{\theta}$ is too small and the standard *t*-test rejects the null too often.

Linearizing the *g(.)* around the null to obtain the LM test for θ:

$$y_t = \gamma \bullet g(\theta_0, \bar{y}_{t-1}) + \lambda \bullet g_\theta(\theta_0, \bar{y}_{t-1}) + e_t, \tag{3.2.3}$$

where $g_\theta(\theta, \bar{y}_{t-1}) = \frac{\partial g(\theta, \bar{y}_{t-1})}{\partial \theta} = \sum_{i=2}^{\infty} (i-1) \bullet \theta^{i-2} y_{t-i}$, $\lambda = \gamma \bullet (\theta - \theta_0)$ and e_i incorporates a remainder term. If the null $\theta = \theta_0$ is correct, the second term in regression (3.2.3) should not be significant. In practice, to evaluate the regressors [$g(\theta_0, \bar{y}_{t-1})$, $g_\theta(\theta_0, \bar{y}_{t-1})$], we set y_t at its unconditional mean for all $t \le 0$. Once again we are interested in comparing the small sample performance of the usual Wald test based on asymptotic theory, $t(\hat{\theta})$, with the LM best based on reduced form equation (3.2.3), using $t_{\hat{\lambda}}$; Monte Carlo simulations were done in EViews™.

Table 3.2.1 explores the effect of γ for true $\theta = 0$ with a sample size $T = 1000$. When γ is small relative to sample variation, as indicated by a small value of the metric $\gamma / \sqrt{V_{\hat{\gamma}}}$ the Wald *t*-test rejects the null too often. As γ gets larger and the key metric $\gamma / \sqrt{V_{\hat{\gamma}}}$ approaches 10, asymptotic theory gradually takes hold and the size of conventional *t*-test based on $t(\hat{\theta})$ becomes closer to its nominal level, 0.05. The fact that the sampling distribution of $t(\hat{\theta})$ depends on the nuisance parameter γ implies that the test is not pivotal. Note that the LM test based on $t(\hat{\lambda})$ in this case is equivalent to testing the second lag in an AR(2) regression, which is approximately the Box-Ljung *Q*-test with one lag for the residuals from an AR(1) regression. The estimated size of the LM test is correct within sampling error.

One may wonder how the LM test performs when true γ is zero, corresponding to the case of identification failure. The empirical size of the LM test in this case is 0.0509, close to correct.

[Insert Table 3.2.1 here]

We note that the median and inter-quartile range of $\hat{\theta}$ suggest that the sampling distribution of $\hat{\theta}$ is centered on zero. However, the histogram of $\hat{\theta}$ in Figure 3.2.1 for the case γ = .01 shows that the estimates tend to be concentrated close to boundaries of the parameter space, reflecting the well-known 'pile-up' effect in ARMA models; see Hannan (1982) and Hauser, Pötscher, and Reschenhofer (1999). Figure 3.2.2 plots the un-centered correlation ρ between the 'regressors' $g(\theta, \vec{y}_{t-1})$ and $g_\theta(\theta, \vec{y}_{t-1})$ as a function of provisional estimate θ. At $\theta = 0$ the correlation is zero but becomes larger in absolute value as θ moves away from zero in either direction and toward the boundaries where $\hat{\theta}$ occurs with greatest frequency. The excessive size of the test based on $t_{\hat{\theta}}$ reflects this strong correlation between the reduced-form regressors when $\hat{\theta}$ falls far from zero and moves toward boundaries, not simply the too-small standard error as surmised by NS. The relative success of the LM test comes from the fact that it evaluates the test statistic under the null hypothesis $\theta_0 = 0$ instead of at $\hat{\theta}$.

[Insert Figure 3.2.1 here]

[Insert Figure 3.2.2 here]

Table 3.2.2 explores the effect of increasing sample size when true γ = .01. Asymptotic theory does take hold, but the conventional *t*-test approaches correct size very slowly (requiring a sample size as large as 10,000 for γ = .1!). In contrast, the reduced form test consistently has correct size within sampling error.

[Insert Table 3.2.2 here]

Often it is the AR root ϕ that is of a great economic interest since it measures persistence. For instance, in an influential work Bansal and Yaron (2000, 2004) show that if consumption growth g_t follows an ARMA (1,1) process a large value of ϕ implying a very persistent consumption growth expectation, namely the long-run risk, may explain the infamous equity premium puzzle. Ma (2012) finds that the $\hat{\gamma}$ in the estimated ARMA(1,1) is small relative to its sampling variance and explores the implications of possible test size distortion in the conventional test as well as valid inference following the strategy suggested in this paper. In another influential paper, Binsbergen and Koijen (2010) propose the AR(1) process to model the conditional expected returns as a latent variable resulting in an ARMA(1,1) process for the realized returns. In their estimation results, the persistence parameter is estimated to be very high with a rather small standard error. However, their estimated variance of the shock to the conditional expected returns process is very small, resulting in very close AR and MA roots, which raises a concern of the validity of conventional inferences. Ma and Wohar (2012) extend the LM test in current work to reexamine the issue of Binsbergen and Koijen and find that the corrected confidence interval for the persistence parameter of the expected returns is much wider.

In the Appendix B.1, we offer details about how to obtain the reduced-form and LM test for AR coefficient ϕ. For the case $\gamma = 0.1, \phi = 0$ and T = 100 the rejection frequency of the LM test is 0.046 in contrast to 0.423 for the standard *t*-test.

The reduced-form test can also be generalized to address any higher order ARMA model, and the extensions are illustrated in the Appendix B.2. For the ARMA(2,2) model with parameter values $\phi_1 = 0.01, \phi_2 = 0.01, \theta_1 = 0, \theta_2 = 0$ and T = 100 we find that the standard *t*-test for $\hat{\theta}_1$ and $\hat{\theta}_2$ has empirical sizes of 0.571 and 0.698. In contrast the reduced-form test gives rejection frequencies of 0.049 and 0.049 respectively.

3.3. The Unobserved Component Model for Decomposing Trend and Cycle

The Unobserved Component model (hereafter UC) of Harvey (1985) and Clark (1987) is widely used to decompose the log of real GDP into trend and cycle. Thus:

$$y_t = \tau_t + c_t , \tag{3.3.1}$$

where trend is assumed to be a random walk with drift:

$$\tau_t = \tau_{t-1} + \mu + \eta_t, \eta_t \sim i.i.d.\, N(0, \sigma_\eta^2) , \tag{3.3.2}$$

and cycle has a stationary AR representation:

$$\phi(L)c_t = \varepsilon_t, \varepsilon_t \sim i.i.d.\, N(0, \sigma_\varepsilon^2) . \tag{3.3.3}$$

The UC model is estimated by maximizing the likelihood computed using the Kalman filter under the assumption that trend and cycle shocks are uncorrelated. In practice the largest AR root is estimated to be close to unity, implying that the cycle is very persistent, and the trend variance is estimated to be very small, implying that the trend is very smooth. The question we wish to investigate here is whether standard inference about cycle persistence may be spurious and whether the approach in this paper can provide a correctly sized test.

To simplify we focus on the case that the cycle is AR(1). Following Morley, Nelson and Zivot (2003), we note that the univariate representation of this particular UC model is ARMA(1,1) with parameters implied by the equality:

$$(1-\phi L)\Delta y_t = \mu(1-\phi) + (1-\phi L)\eta_t + \varepsilon_t - \varepsilon_{t-1} = \mu(1-\phi) + u_t - \theta u_{t-1} \qquad (3.3.4)$$

Where $u_t \sim i.i.d.\, N(0,\sigma_u^2)$. Thus the AR coefficient of the ARMA(1,1) is simply ϕ, while the MA parameter θ is identified (under the restriction $\sigma_{\eta,\varepsilon} = 0$) by matching the zero and first-order autocovariances of the equivalent MA parts:

$$\psi_0 = (1+\phi^2)\sigma_\eta^2 + 2\sigma_\varepsilon^2 + 2(1+\phi)\sigma_{\eta\varepsilon}^2 = (1+\theta^2)\sigma_u^2 \qquad (3.3.5)$$

$$\psi_1 = -\phi\sigma_\eta^2 - \sigma_\varepsilon^2 - (1+\phi)\sigma_{\eta\varepsilon}^2 = -\theta\sigma_u^2 \qquad (3.3.6)$$

We may then solve for a unique θ by imposing invertibility, obtaining:

$$\theta = \frac{(1+\phi^2)+2(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2})+2(1+\phi)\rho_{\eta\varepsilon}(\frac{\sigma_\varepsilon}{\sigma_\eta}) - \sqrt{[(1+\phi)^2+4(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2})+4(1+\phi)\rho_{\eta\varepsilon}(\frac{\sigma_\varepsilon}{\sigma_\eta})]\bullet[(1-\phi)^2]}}{2[\phi+\frac{\sigma_\varepsilon^2}{\sigma_\eta^2}+(1+\phi)\rho_{\eta\varepsilon}(\frac{\sigma_\varepsilon}{\sigma_\eta})]} \qquad (3.3.7)$$

It is straightforward to show that θ becomes arbitrarily close to ϕ as ${\sigma_\varepsilon}/{\sigma_\eta}$ approaches zero. By an analogy to the ARMA (1,1) model, the estimated standard error for $\hat{\phi}$ may be too small when ${\sigma_\varepsilon}/{\sigma_\eta}$ is small relative to sampling variation, and a standard *t*-test may be incorrectly sized.

To visualize spurious inference in this case, we implement a Monte Carlo experiment. Data is generated from the UC model given by (3.3.1) – (3.3.3) with true parameter values $\mu = 0.8, \phi = 0, \sigma_\eta^2 = 0.95, \sigma_\varepsilon^2 = 0.05$, corresponding roughly to quarterly U.S. GDP if almost all the variation were due to trend while the cycle is small with no persistence at all. Estimation is done in MATLAB 6.1 and the routine is available on request. Sample size T is 200, approximately what is encountered in practice for postwar data. To avoid local maxima, various starting values are used.

The standard *t*-test for $\hat{\phi}$ indeed rejects the null much too often; size is 0.481. This is partly because the standard error for $\hat{\phi}$ is underestimated; the median is 0.2852 compared with its true value 1.4815. Furthermore, $\hat{\phi}$ is upward biased as illustrated in Figure 3.3.1, its median being 0.58. Many $\hat{\phi}'s$ occur close to the positive boundary. This is consistent with Nelson's (1988) finding that a UC model with persistent cycle variation fits better than the true model even when all variation is due to stochastic trend, the case where $\sigma_\varepsilon^2 = 0$.

At the same time, the cycle innovation variance estimate $\hat{\sigma}_\varepsilon^2$ is upward biased, having a median of .20, while the trend innovation variance estimate $\hat{\sigma}_\eta^2$ is instead downward biased, with a median of .73. What is the underlying driving force for the upward bias of $\hat{\phi}$ and $\hat{\sigma}_\varepsilon^2$ and the downward bias of $\hat{\sigma}_\eta^2$? The scatter plot in Figure 3.3.2 shows that there is a positive co-movement between $\hat{\phi}$ and $\hat{\sigma}_\varepsilon^2$, thus persistence in the estimated cycle tends to occur in samples that also show large variance in the cycle. This is driven by the necessity that the model must account for the small amount of serial correlation in the data generating process for Δy_t. Setting the auto-covariance at lag one equal to the true value for the sake of illustration, one obtains the restriction

$-\frac{1-\phi}{1+\phi}\bullet\sigma_{\varepsilon}^{2}=-.05$. One solution is the combination of true values, $\phi=0;\sigma_{\varepsilon}^{2}=.05$, but another is $\phi=.9;\sigma_{\varepsilon}^{2}=.95$. Thus $\hat{\sigma}_{\varepsilon}^{2}$ will be far greater than its true value when $\hat{\phi}$ is close to its positive boundary, implying a dominating persistent cycle that tends to mimic the true underlying stochastic trend. Finally, Figures 3.3.1 and 3.3.2 show that large negative values of $\hat{\phi}$ are possible but infrequent because positive variances place restrictions on the parameter space.

In light of the connection between UC model and ARMA model we implement the LM test in the following steps: first impose the null $\phi=\phi_0$ and estimate all other parameters in the UC model; secondly, impute from (3.3.7) the restricted estimate $\tilde{\theta}$ and $\tilde{u}_t$ in the reduced-form ARMA(1,1) model; lastly, compute the reduced form *t*-test statistic by following the strategy in Appendix B.1. Using the same set of simulated data as above for true parameter values $\mu=0.8,\phi=0,\sigma_{\eta}^{2}=0.95,\sigma_{\varepsilon}^{2}=0.05$, the rejection frequency for the LM test for ϕ is 0.054.

One may also be interested in the case when all variation is due to stochastic trend, i.e., $\sigma_{\varepsilon}^{2}=0$. For this case, the identification for ϕ fails and the standard *t*-test is not well-defined. However, the reduced-form LM test works well and gives estimated size 0.0581 in the Monte Carlo with true parameter values $\mu=0.8,\phi=0,\sigma_{\eta}^{2}=1,\sigma_{\varepsilon}^{2}=0$.

The reduced-form test can also be generalized to address a UC model with higher AR orders in the cycle by following the strategy discussed in Appendix B.2.

[Insert Figure 3.3.1 here]

[Insert Figure 3.3.2 here]

3.4. The GARCH(1,1) Model.

The GARCH model developed by Bollerslev (1986) is perhaps one of the most popular approaches in capturing the time-varying volatility for time series data. The archetypal GARCH (1,1) may be written:

$$\varepsilon_t = \sqrt{h_t} \bullet \xi_t, \xi_t \sim i.i.d.N(0,1) \tag{3.4.1}$$

$$h_t = \omega + \alpha \bullet \varepsilon_{t-1}^2 + \beta \bullet h_{t-1} \tag{3.4.2}$$

To see why GARCH is among the models we are concerned with, write out its ARMA representation and make an analogy to the ARMA (1,1) model:

$$\varepsilon_t^2 = \omega + (\alpha + \beta) \bullet \varepsilon_{t-1}^2 + w_t - \beta \bullet w_{t-1} \tag{3.4.3}$$

The innovation $w_t = \varepsilon_t^2 - h_t = h_t(\xi_t^2 - 1)$ is a Martingale Difference Sequence (MDS) with time-varying variance. Since $\alpha + \beta$ and β correspond to the AR and MA roots respectively, α controls the information about β. Ma, Nelson and Startz (2007) show that when α is small relative to its sampling variation, the standard error for $\hat{\beta}$ is underestimated and the standard *t*-test rejects the null too often, implying a significant GARCH effect even when there is none.

It is also useful to write out the state-space representation of the GARCH model to see the root cause of the weak identification and its likely implications for statistical inference. Rewrite (3.4.3) to give the state equation that describes the dynamic evolution of the unobserved volatility:

$$h_t = \omega + (\alpha + \beta) \bullet h_{t-1} + \alpha \bullet w_{t-1} \tag{3.4.4}$$

Where $w_{t-1} = \varepsilon_{t-1}^2 - h_{t-1}$. The measurement equation is simply:

$$\varepsilon_t^2 = h_t + w_t \tag{3.4.5}$$

This is a particular state-space model since the shock in the state equation ($\alpha \bullet w_{t-1}$) is one period lag of the shock in the measurement equation (w_t). The essence of the state-space model is to filter out the noise (w_t) in order to extract the unobservable signal sequence (h_t), and the parameter α in the GARCH model measures the size of the signal shock relative to that of the noise shock. When the signal is small relative to noise, the uncertainty of the signal dynamics ought to be fairly large and a correct test must produce results consistent with this intuition.

The LM test can be easily extended to test the null $\beta = \beta_0$. Defining $g(\beta, \vec{\varepsilon}_{t-1}^{\,2}) = \sum_{i=1}^{\infty} \beta^{i-1} \varepsilon_{t-i}^2$ and $\vec{\varepsilon}_{t-1}^{\,2} = (\varepsilon_{t-1}^2, \varepsilon_{t-2}^2, ...)$ to rewrite (3.4.2) one obtains:

$$h_t = \frac{\omega}{1-\beta} + \alpha \bullet g(\beta, \vec{\varepsilon}_{t-1}^{\,2}) \qquad (3.4.6)$$

Taking a linear expansion of nonlinear *g*(.) around the null, defining $c = \frac{\omega}{1-\beta}$,

$\lambda = \alpha \bullet (\beta - \beta_0)$ and $g_\beta(\beta, \vec{\varepsilon}_{t-1}^{\,2}) = \sum_{i=2}^{\infty} (i-1) \bullet \beta^{i-2} \varepsilon_{t-i}^2$, we have:

$$h_t = c + \alpha \bullet g(\beta_0, \vec{\varepsilon}_{t-1}^{\,2}) + \lambda \bullet g_\beta(\beta_0, \vec{\varepsilon}_{t-1}^{\,2}) \qquad (3.4.7)$$

The LM test then is the *t*-statistic for the null $\lambda = 0$ in (3.4.7).

Table 3.4.1 presents the comparison of the LM test based on $t(\hat{\lambda})$ and the standard *t*-test based on $t(\hat{\beta})$ for a range of values of α where $\beta = 0$ and sample size $T = 1000$ (the Matlab 6.1 code is available on request). When the key metric $\gamma / \sqrt{V_{\hat{\gamma}}}$ is small, the standard *t*-test rejects the null too often. The LM test however has consistently better size.

In the case $\alpha = 0$ identification fails and the standard *t*-test does not have the usual asymptotic distribution. The LM test, however, is still valid and has estimated size of 0.076 for true $\beta = 0$ and $T = 1{,}000$.

The sum $\alpha + \beta$ is of particular economic interest since it measures the persistence of volatility. Bansal and Yaron (2000, 2004) show that a large value of $\alpha + \beta$ for consumption volatility, interpreted as long run risk, may help to resolve the equity premium puzzle. Appendix C gives details about how to obtain a valid test for $\hat{\alpha} + \hat{\beta}$ and evaluates its performance; see Ma (2012) for further discussion.

We have applied the LM test to monthly S&P 500 index returns from the DRI Database for the period January 1947 to September 1984 studied in Bollerslev (1987). The GARCH estimates with Bollerslev and Wooldridge's (1992) robust standard errors, after accounting for the "Working (1960) effect," are:

$$\hat{\omega} = 0.16 \bullet 10^{-3}(0.14 \cdot 10^{-3}),\ \hat{\alpha} = 0.077(0.048),\ \hat{\beta} = 0.773(0.169)$$

The standard *t*-test implies a significant and large GARCH effect as indicated by the 95% confidence interval for β: [0.44, 1). However, the small value of $\hat{\alpha}$ (the upper bound for α at a 95% significance level 0.173) relative to the sample size $T = 453$, raises concern about the possibility of spurious inference for β. To obtain a confidence interval for β based on the LM test, we numerically invert $t(\hat{\lambda})$ over a grid of values for β_0, compute the corresponding $t(\hat{\lambda})$, and plot the latter against the former in Figure 3.4.1. The resulting 95% confidence interval for β based on the reduced-form test is [-0.95, 0.87], which covers almost the entire parameter space. That this can happen in practice should not surprise us, in light of the theorem of Dufour (1997) that the probability that a valid confidence interval covers the entire parameter space must be greater than zero if identification is weak enough.

[Insert Figure 3.4.1 here]

4. Summary and Conclusions

This paper presents a modified LM test as an alternative to the standard Wald *t*-test in a class of models where the latter is associated with spurious inference when identification is weak. There are models that have a representation of the form $y = \gamma \bullet g(\beta, x) + \varepsilon$, where β is the parameter of interest and the amount of information about that parameter available from the data depends on the unknown identifying parameter γ. This class includes not only the obvious non-linear regression model but also the workhorse ARMA model and, by extension, GARCH and Unobserved Components models. NS showed that inference is problematic because the standard error for $\hat{\beta}$ depends on $\hat{\gamma}$. While the estimated standard error is downward biased in finite samples in the class of models that satisfy the Zero-Information-Limit-Condition, the *t*-statistic can be either too large or too small depending on the data generating process. In this paper we show that small sample inference in this class is usefully studied by working with the approximation:

$$g(\beta, x) \approx g(\beta_0, x) + (\beta - \beta_0) \bullet g_\beta(\beta_0, x)$$

and the corresponding reduced form regression:

$$y_i = \gamma \bullet g(\beta_0, x_i) + \lambda \bullet g_\beta(\beta_0, x_i) + e_i; \text{ where } \lambda = \gamma \bullet (\beta - \beta_0).$$

A *t*-test that exploits the fact that the null hypothesis $\beta = \beta_0$ implies $\lambda = 0$ and thus is exact when $g(.)$ is linear. We show that is has nearly correct size when the reduced form model is only an approximation. This test may be interpreted as a modified LM test in the spirit of Harvey (1990). The paper illustrates the superior performance of the LM test in the presence of weak identification with examples from non-linear regression, ARMA, GARCH, and Unobserved Components models.

Table 2.1. The Effect of Sample Size N on the Distribution of $\hat{\beta}$ and Size of $t(\hat{\beta})$ with Orthogonal Regressors.

Sample Size N	100	10,000	1,000,000	10,000
True γ	0.01	0.01	0.01	0.1
Asymptotic $\gamma/\sqrt{V_{\hat{\gamma}}}$	0.1	1	10	10
Median $\hat{\beta}$	0.1	0.01	-0.00	0.00
Range (.25, .75)	(-0.95, 1.17)	(-0.64, 0.63)	(-0.07, 0.07)	(-0.07, 0.07)
Size for $t(\hat{\beta})$	0.0001	0.0006	0.043	0.045

Table 3.1.1. Small Sample Distribution of $\hat{\beta}$ and Test Sizes, True γ = .01, *N* = 100.

True β	0	0.1	0.5	0.9
$\rho_{g(\beta),g_{\beta}(\beta)}$	0.07	0.29	0.77	0.92
Asymptotic $\gamma \big/ \sqrt{V_{\hat{\gamma}}}$	0.10	0.10	0.09	0.11
Median $\hat{\beta}$	-0.04	-0.09	-0.05	0.12
Range (.25, .75)	(-0.53, 0.50)	(-0.59, 0.42)	(-0.56, 0.48)	(-0.41, 0.71)
Size for $t(\hat{\beta})$	0.027	0.037	0.114	0.179
Size for $t(\hat{\lambda})$	0.053	0.054	0.054	0.054

Table 3.1.2. Small Sample Distribution of $\hat{\beta}$ and Test Sizes, N = 100, true β = .5

True γ	0.01	0.1	1
Asymptotic $\gamma / \sqrt{V_{\hat{\gamma}}}$	0.09	0.91	9.10
Median $\hat{\beta}$	-0.05	0.27	0.50
Range (.25, .75)	(-0.56, 0.48)	(-0.26, 0.64)	(0.46, 0.54)
Size for $t(\hat{\beta})$	0.114	0.103	0.052
Size for $t(\hat{\lambda})$	0.054	0.054	0.054

Table 3.2.1. Effect of γ on Inference for ARMA (1,1), True θ = 0, T = 1,000.

True $\gamma(=\phi)$	0.01	0.1	0.2	0.3
Asymptotic $\gamma/\sqrt{V_{\hat{\gamma}}}$	0.32	3.16	6.32	9.49
Median $\hat{\theta}$	-.02	-.01	-.00	-.00
Range (.25, .75)	(-.65, .64)	(-.26, .24)	(-.11, .11)	(-.07, .07)
Size for $t(\hat{\theta})$	0.4585	0.2237	0.1051	0.0734
Size for $t(\hat{\lambda})$	0.0506	0.0518	0.0526	0.0522

Table 3.2.2. Sample Size and Inference in the ARMA (1, 1), True θ = 0.

Sample size	100	1000	10,000	10,000
True $\gamma(=\phi)$	0.01	0.01	0.01	0.1
Asymptotic $\gamma\Big/\sqrt{V_{\hat{\gamma}}}$	0.1	0.32	1	10
Median $\hat{\theta}$	-0.04	-0.02	-0.02	-0.00
Range (.25, .75)	(-0.69, 0.67)	(-0.65, 0.64)	(-0.58, 0.55)	(-0.07, 0.07)
Size for $t(\hat{\theta})$	0.483	0.458	0.399	0.066
Size for $t(\hat{\lambda})$	0.051	0.051	0.049	0.048

Table 3.4.1. LM test and standard *t*-tests for GARCH(1,1): True β = 0, *T* = 1,000.

True $\gamma(=\alpha)$	0.01	0.05	0.1	0.2
Asymptotic $\gamma / \sqrt{V_{\hat{\gamma}}}$	0.32	1.59	3.19	6.60
Median $\hat{\beta}$	0.33	0.08	-0.00	-0.01
Range (.25, .75)	(-0.30, 0.74)	(-0.31, 0.49)	(-0.22, 0.22)	(-0.11, 0.09)
Size of $t(\hat{\beta})$	0.470	0.344	0.198	0.106
Size for $t(\hat{\lambda})$	0.078	0.074	0.076	0.096

Figure 2.1: Rejection Frequencies for tests of $H_0 : \beta = 0$, N = 100, γ = .10.

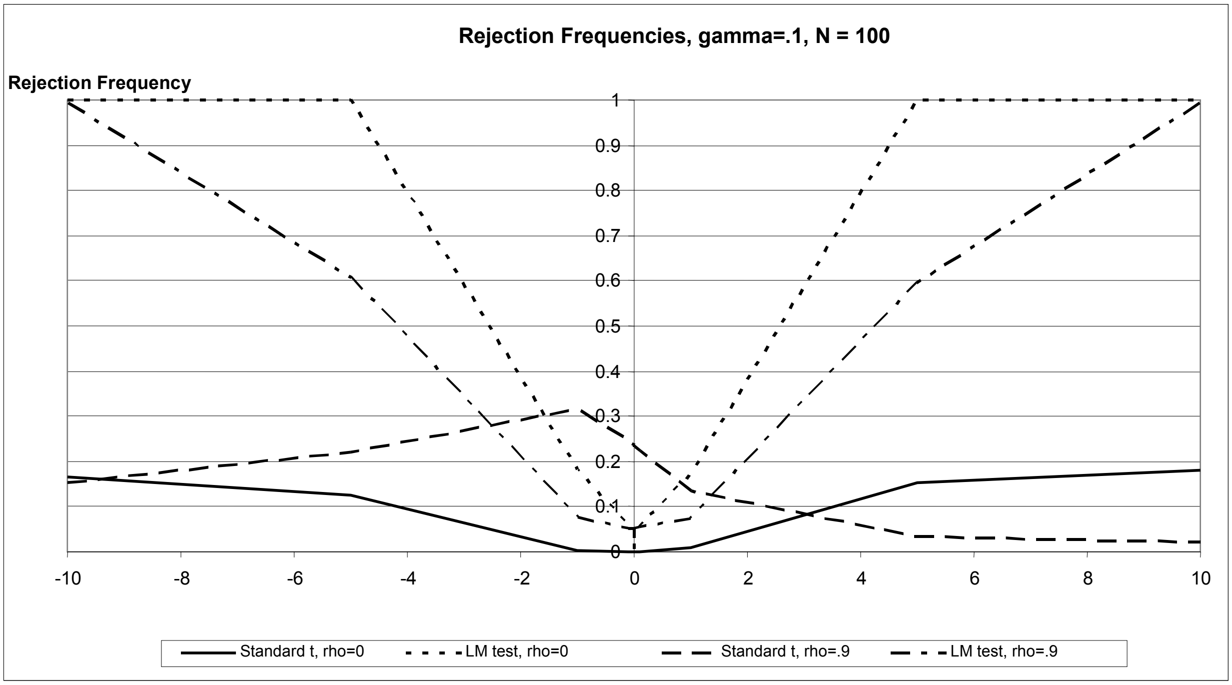


Figure 3.1.1: Rejection Frequencies for the test $H_0 : \beta = .5$, N = 100, True γ = .1.

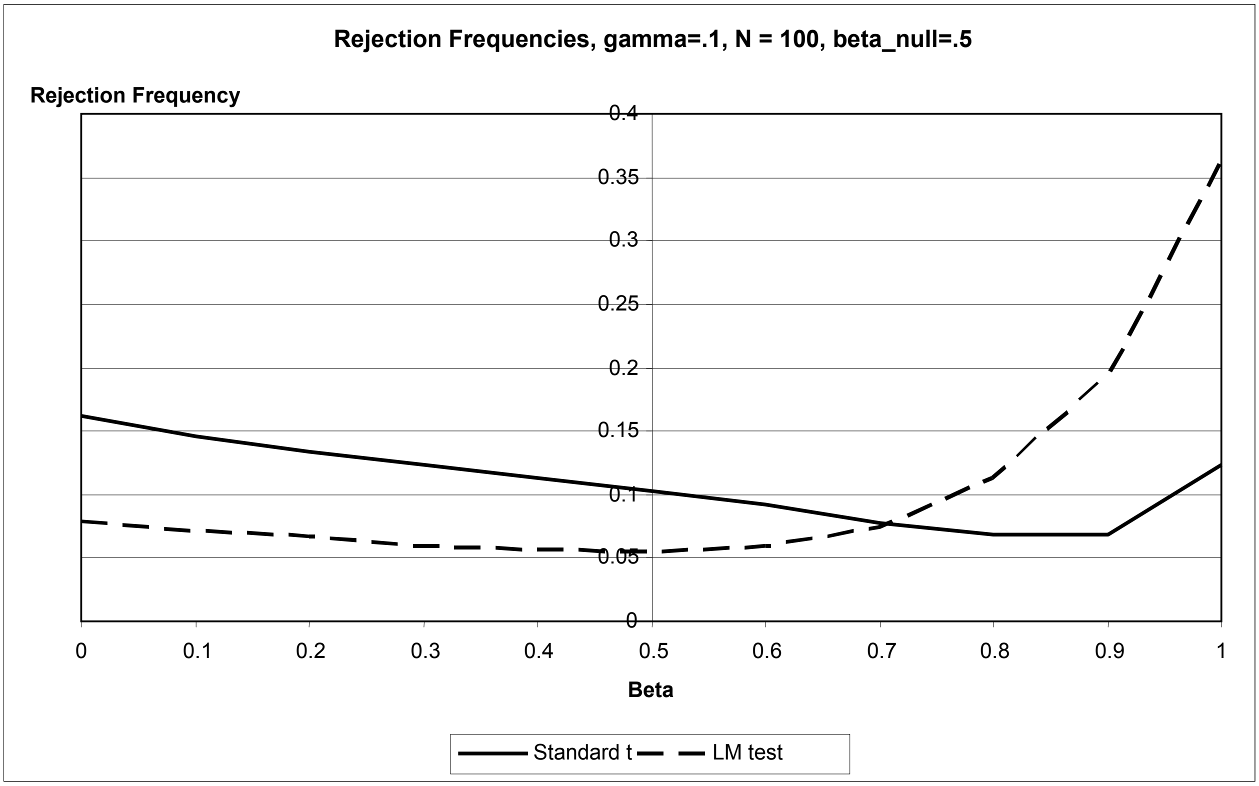


Figure 3.2.1: Histogram of $\hat{\theta}$ in the Monte Carlo. True γ = .01, θ = 0, T = 1,000.

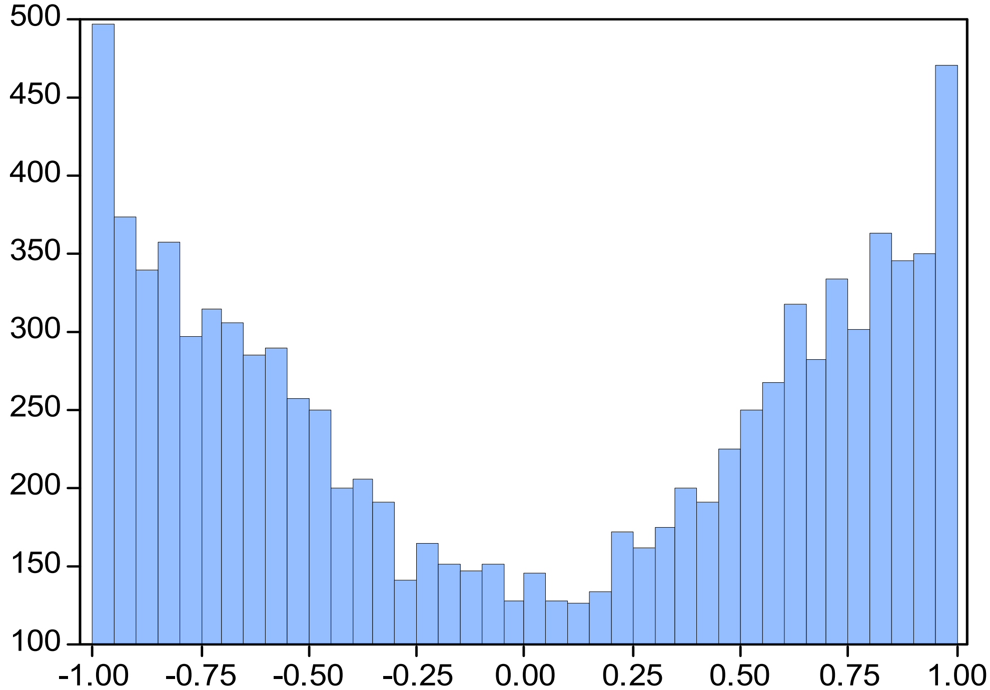


Figure 3.2.2: Computed un-centered correlation between $g(\theta, \vec{y}_{t-1})$ and $g_\theta(\theta, \vec{y}_{t-1})$ based on one sample draw. True γ = .01, θ = 0, T = 1,000.

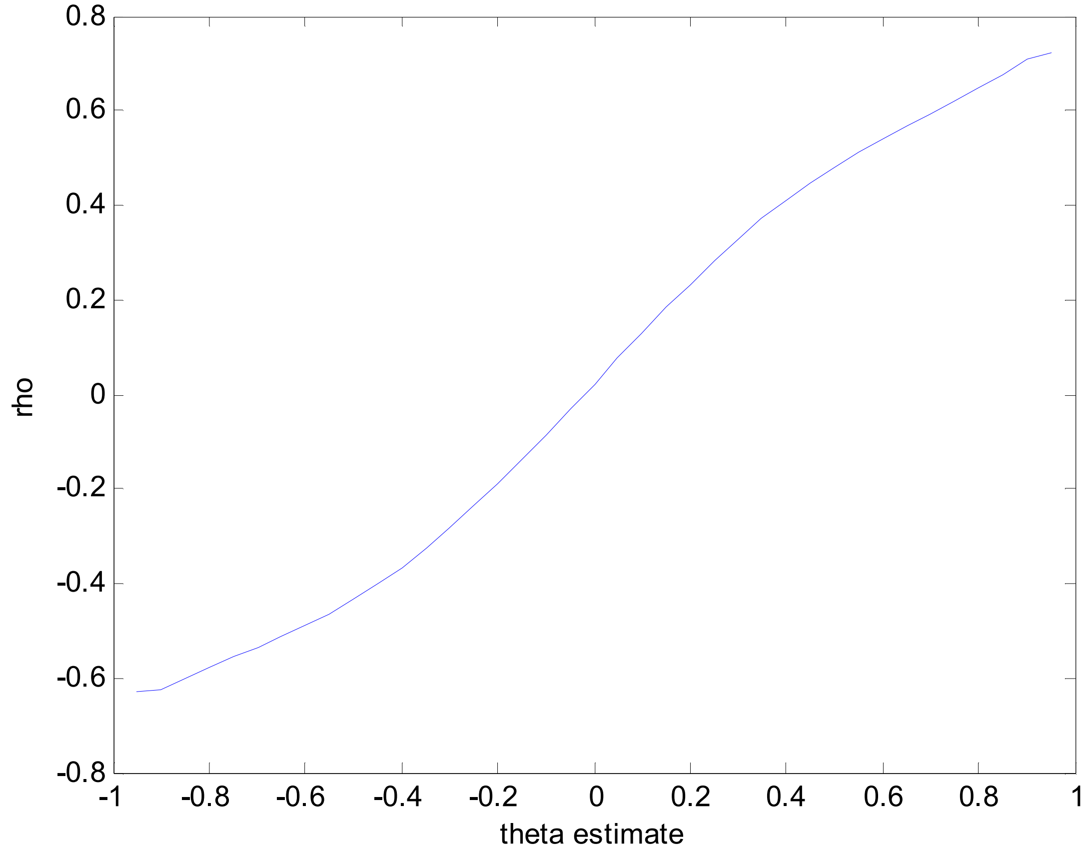


Figure 3.3.1: Plot of $\hat{\phi}$ in the Monte Carlo Experiment with true parameter

$$\mu = 0.8, \phi = 0, \sigma_{\eta}^{2} = 0.95, \sigma_{\varepsilon}^{2} = 0.05$$

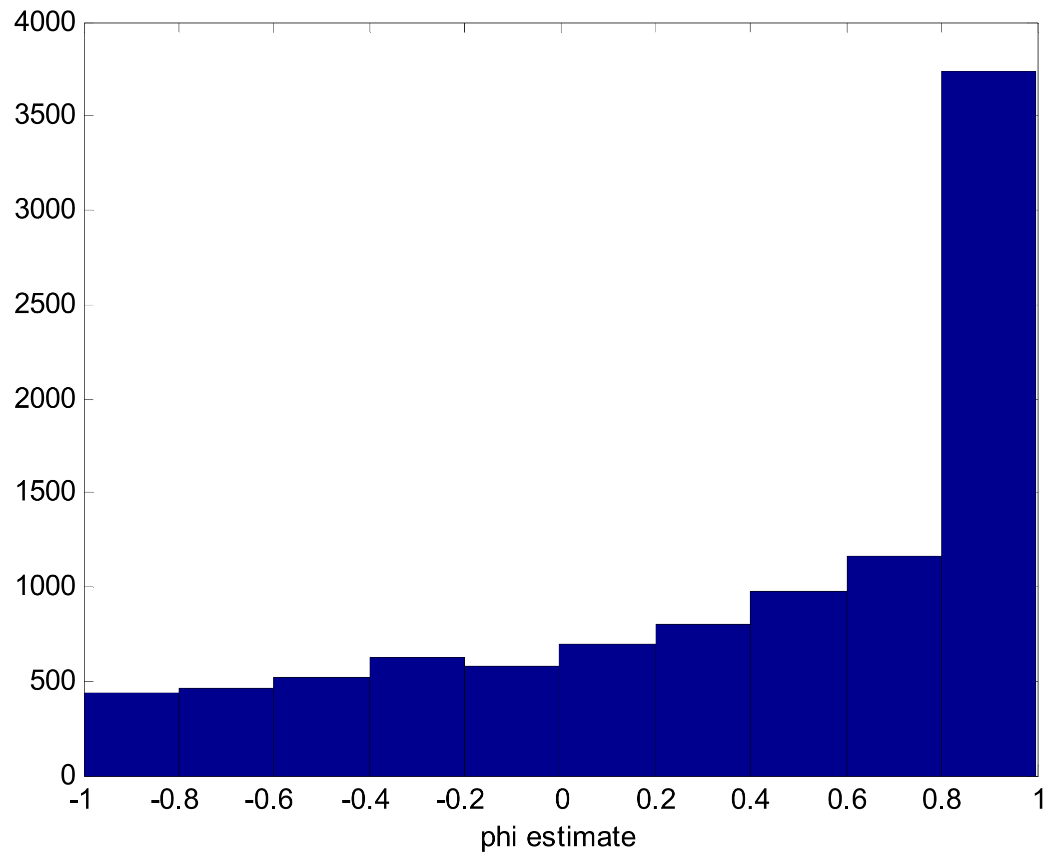


Figure 3.3.2: Scatter Plot of $\hat{\phi}$ and $\hat{\sigma}_{\varepsilon}^{2}$ in the Monte Carlo Experiment with true parameter $\mu = 0.8, \phi = 0, \sigma_{\eta}^{2} = 0.95, \sigma_{\varepsilon}^{2} = 0.05$

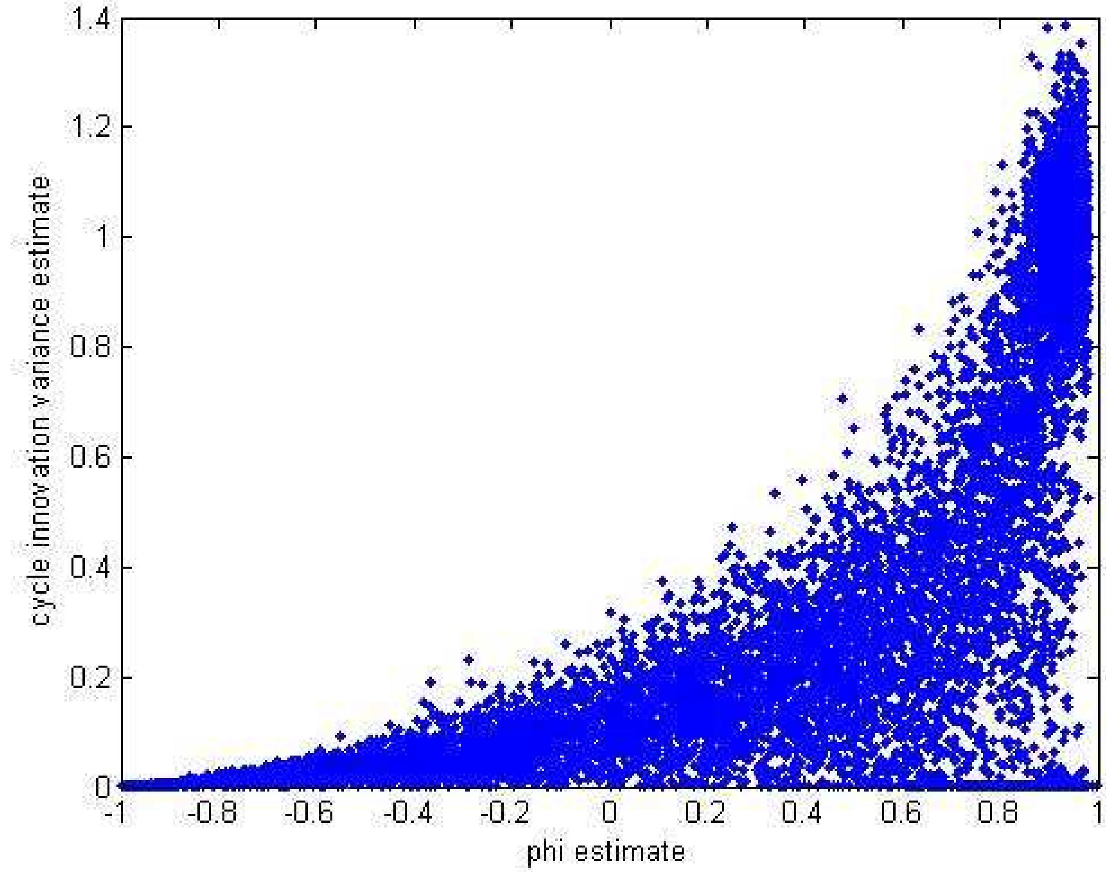
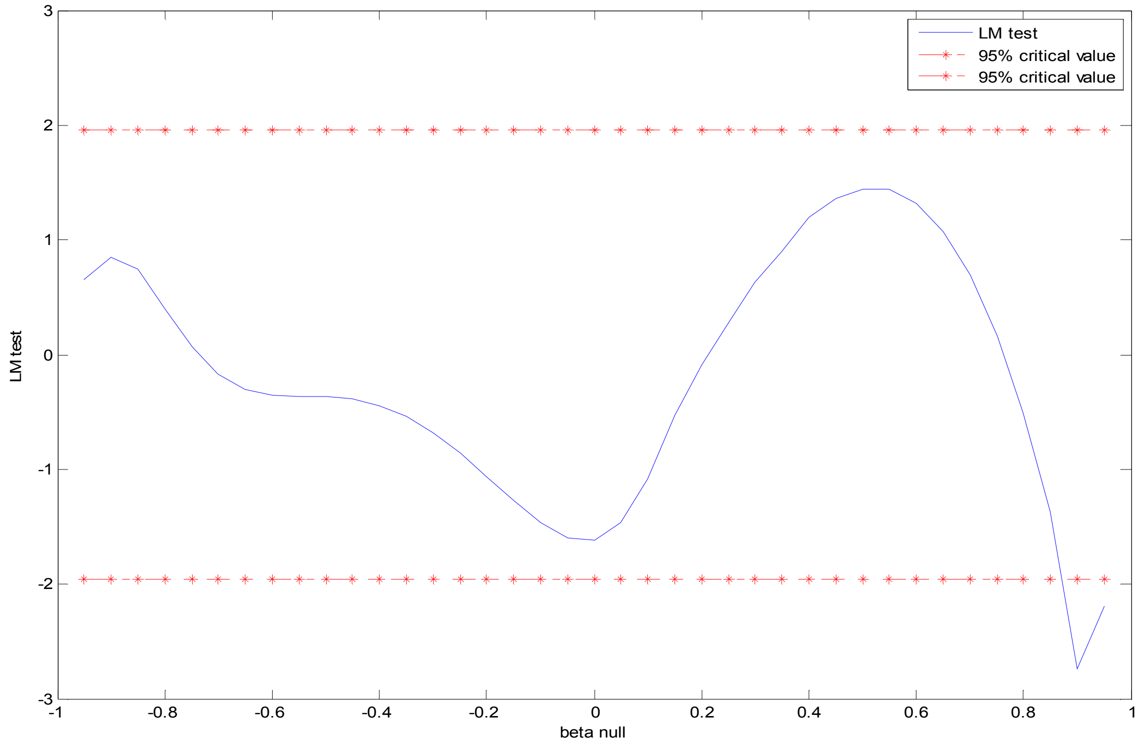


Figure 3.4.1: The 95% Confidence Interval for $\hat{\beta}$ based on the LM test for the monthly S&P 500 stock return data



Appendix A

Consider the regression model (2.1). The parameter of interest β may be estimated as the ratio of two parameters estimates in the reduced form (2.2) $\hat{\beta} = \hat{\lambda} / \hat{\gamma}$. Although the moments of the ratio of normal random variables do not in general exist, see Fieller (1932) and Hinckley (1969), we can nevertheless draw some conclusions about the sampling distribution of $\hat{\beta}$. Noting that $\hat{\lambda}$ and $\hat{\gamma}$ are jointly Normal across samples, the conditional mean of the former given the latter implies:

$$\hat{\lambda} = \alpha + \kappa \bullet \hat{\gamma} + v \tag{A.1}$$

where α and κ are parameters and v is Normal and uncorrelated with $\hat{\gamma}$ by construction. To simplify exposition we focus on the case $\beta = 0$ and standardized regressors with sample correlation ρ. It is straightforward to show that $\alpha = \rho \bullet \gamma$, $\kappa = -\rho$, and the variance of v is σ^2/N. Making these substitutions and dividing by $\hat{\gamma}$ one obtains:

$$\hat{\beta} = -\rho + \rho \bullet \left(\frac{\gamma}{\hat{\gamma}} \right) + \frac{v}{\hat{\gamma}} \tag{A.2}$$

Consider now how the distribution of $\hat{\beta}$ is affected by γ, which controls the amount of information in the data about β, and by correlation between regressors ρ, for given sample size. A larger value of γ means that the ratio $\gamma/\hat{\gamma}$ tends to be closer to unity, since the standard deviation of $\hat{\gamma}$, given by $\sqrt{\sigma^2 \bullet N^{-1}/(1-\rho^2)}$, is not a function of γ. The second term in (A.2) will tend toward ρ, canceling out the first term, and the third term will be relatively small, so the sampling distribution of $\hat{\beta}$ will be located more tightly around its true value, zero. However, a smaller value of γ means that $\gamma/\hat{\gamma}$ will typically be small, thus locating the sampling distribution of $\hat{\beta}$ around $-\rho$ but with greater dispersion since the third term will typically be large. Shifting now to the effect of ρ, stronger correlation will increase sampling variation in $\hat{\gamma}$, so the second and third terms will tend to be small, concentrating the distribution of $\hat{\beta}$ around $-\rho$. (In this paper we refer to these shifts of central tendency away from the true value as 'bias' for the sake of brevity.)

Turning now to hypothesis testing, the asymptotic variance of $\hat{\beta}$ derived either from the information matrix for (2.1) under maximum likelihood, or using the 'delta method' for indirect least squares, is given by:

$$V_{\hat{\beta}} = \frac{1}{\gamma^2} \bullet \frac{\sigma^2}{N} \bullet \frac{m_{xx} + 2\beta \bullet m_{xz} + \beta^2 \bullet m_{zz}}{m_{xx} \bullet m_{zz} - m_{xz}^2} \tag{A.3}$$

where '*m*' denotes the raw sample second moment of the subscripted variables. In practice the parameters are unknown and are replaced in standard software packages by the point estimates. Thus the reported *t*-statistic for $\hat{\beta}$ is:

$$t(\hat{\beta})^2 = (\hat{\beta} - \beta_0)^2 \bullet \left[\hat{\gamma}^2 \bullet \frac{N}{\hat{\sigma}^2} \bullet \frac{m_{xx} \bullet m_{zz} - m_{xz}^2}{m_{xx} + 2\hat{\beta} \bullet m_{xz} + \hat{\beta}^2 \bullet m_{zz}} \right] \tag{A.4}$$

where the null hypothesis is $\beta = \beta_0$. We confine our attention to the case $\beta_0 = 0$, noting that a non-zero value of β_0 simply corresponds to a transformed model. In the standardized regressors case the *t*-statistic for $\hat{\beta}$ is given by:

$$t(\hat{\beta})^2 = \frac{\hat{\lambda}^2}{\hat{\sigma}^2} \bullet N \bullet (1-\rho^2) \bullet \frac{1}{1 + 2\hat{\beta} \bullet \rho + \hat{\beta}^2} = t_\lambda^2 \bullet \frac{1}{1 + 2\hat{\beta} \bullet \rho + \hat{\beta}^2} \tag{A.5}$$

Since the reduced form is a classical linear regression, a test based on $t(\hat{\lambda})$ has correct size and so provides an alternative test of the null hypothesis $\beta = 0$ with correct size. Indeed this is the exact test of Fieller (1954) for a ratio of regression coefficients. As noted by NS, if the two explanatory variables are orthogonal, then in any given sample $t(\hat{\beta})^2 < t(\hat{\lambda})^2$ since the last term in (A.5) must be less than one. In contrast, the effect of strong correlation between *x* and *z*, working through the concentration of $\hat{\beta}$ around the value $-\rho$, is to drive $(1 + 2\hat{\beta} \bullet \rho + \hat{\beta}^2)$ close to zero, making $t(\hat{\beta})$ arbitrarily larger than $t(\hat{\lambda})$. Thus, whether test size is too large or too small depends on the correlation between the regressors, strong correlation of either sign producing an over-sized *t*-test.

Appendix B.1

To obtain the reduced-form test for ϕ, we may re-write the ARMA(1,1):

$$y_t = \gamma \bullet g(\phi, \vec{\varepsilon}_{t-1}) + \varepsilon_t \tag{B.1.1}$$

Where $g(\phi, \vec{\varepsilon}_{t-1}) = \sum_{i=1}^{\infty} \phi^{i-1} \varepsilon_{t-i}$ and $\vec{\varepsilon}_{t-1} = (\varepsilon_{t-1}, \varepsilon_{t-2}, ...)$. Take a linear approximation of *g(.)* around the null, and the reduced-form test is a *t*-test for $\lambda = 0$ in the following regression:

$$y_t = \gamma \bullet g(\phi_0, \vec{\varepsilon}_{t-1}) + \lambda \bullet g_\phi(\phi_0, \vec{\varepsilon}_{t-1}) + e_t \tag{B.1.2}$$

Where $g_\phi(\phi, \vec{\varepsilon}_{t-1}) = \frac{\partial g(\phi, \vec{\varepsilon}_{t-1})}{\partial \phi} = \sum_{i=2}^{\infty} (i-1) \bullet \phi^{i-2} \varepsilon_{t-i}$, and $\lambda = \gamma \bullet (\phi - \phi_0)$.

To make this test feasible, first obtain a consistent estimate for ε through estimation under the restriction of null so as to evaluate the regressors. We generate data with true parameter values $\gamma = 0.1, \phi = 0, \sigma_\varepsilon = 1$ and sample size T = 100. Estimation is done in EViews[TM]. The rejection frequency of the proposed test is 0.046, at the nominal level 0.05, in contrast to 0.423, that of the standard *t*-test.

Appendix B.2

Consider an ARMA(*p*,*q*) model:

$$[1 - \phi_p(L)] y_t = [1 - \theta_q(L)] \varepsilon_t ; t = 1, ..., T, \varepsilon_t \sim i.i.d. N(0, \sigma_\varepsilon^2) \tag{B.2.1}$$

Where $\phi_p(L) = \sum_{i=1}^{p} \phi_i L^i$, $\theta_q(L) = \sum_{i=1}^{q} \theta_i L^i$, and the roots for $1 - \phi(z) = 0$ and $1 - \theta(z) = 0$ are all outside unit circle. A general representation similar to (3.2.2) may be obtained:

$$y_t = \gamma_1 \bullet [(1 - \theta_m(L))^{-1} \bullet y_{t-1}] + \cdots + \gamma_m \bullet [(1 - \theta_m(L))^{-1} \bullet y_{t-m}] + \varepsilon_t \tag{B.2.2}$$

Where $\gamma_k = \phi_k - \theta_k, 1 \le k \le m$, $m = \max(p,q)$, and $\phi_k = 0$ for $p < k \le m$ or $\theta_k = 0$ for $q < k \le m$. To test the null $\theta_k = \theta_{k,0}, 1 \le k \le q$, simply linearize the last term associated with y_{t-m} to obtain the following regression with *q* augmented terms:

$$\begin{aligned} y_t &= \gamma_1 \bullet [(1-\theta_{m,0}(L))^{-1} \bullet y_{t-1}] + \cdots + \gamma_m \bullet [(1-\theta_{m,0}(L))^{-1} \bullet y_{t-m}] \\ &\quad + \lambda_1 \bullet [(1-\theta_{m,0}(L))^{-2} \bullet y_{t-(m+1)}] + \cdots + \lambda_q \bullet [(1-\theta_{m,0}(L))^{-2} \bullet y_{t-(m+q)}] + e_t \end{aligned} \tag{B.2.3}$$

Where $\lambda_k = \gamma_k \bullet (\theta_k - \theta_{k,0}), 1 \le k \le q$. If the null is correct the first *m* terms on the right hand side of (A.2.3) are enough to capture the serial correlation. Note to compute the regressors for nonzero $\theta_{k,0}$'s, the coefficients $\varphi_{l,j}$'s in $\sum_{j=0}^{\infty} \varphi_{l,j} L^j = (1-\theta_{m,0}(L))^{-l}, l = 1,2$ may be obtained as the (1,1) element of matrix $(F_l)^j$, where F_l is the $(l \times m)$ by $(l \times m)$ transition matrix $(1-\theta_{m,0}(L))^l, l = 1,2$ in the state-space representation of the ARMA model.

We experiment this idea on the ARMA(2,2) model. With true parameter values $\phi_1 = 0.01, \phi_2 = 0.01, \theta_1 = 0, \theta_2 = 0, \sigma_\varepsilon = 1$ and sample size *T* = 100 we find that the standard *t*-test for $\hat{\theta}_1$ and $\hat{\theta}_2$ has empirical sizes of 0.571 and 0.698 at a nominal level 0.05. In contrast the reduced-form test for $\lambda_1 = 0$ and $\lambda_2 = 0$ based on regression (B.2.3) gives rejection frequencies of 0.049 and 0.049 respectively. Notice here since the null is $\theta_1 = 0$ and $\theta_2 = 0$, our proposed test is equivalent to testing the third and fourth lag in an AR(4) regression.

Appendix C

To obtain a reduced-form test for $\hat{\alpha} + \hat{\beta}$, we may re-write the variance equation:

$$h_t = \frac{\omega}{1-\rho} + \alpha \bullet g(\rho, \vec{w}_{t-1}) \qquad \text{(C.1)}$$

Where $\rho = \alpha + \beta$, $\vec{w}_{t-1} = (w_{t-1}, w_{t-2}, ...)$. Take a linear expansion of *g(.)* around the null:

$$h_t = \frac{\omega}{1-\rho} + \alpha \bullet g(\rho_0, \vec{w}_{t-1}) + \lambda^* \bullet g_\rho(\rho_0, \vec{w}_{t-1}) \qquad \text{(C.2)}$$

Where $\lambda^* = \alpha \bullet (\rho - \rho_0)$ and the modified LM test is the *t*-stat for λ^*. To make the test feasible one needs to have a consistent estimate for w_t which is readily obtained through estimation under the restriction of null.

Using simulated data with true $\beta = 0, \gamma = 0.01$ and T = 1,000, we find that the reduced-form test for $\hat{\rho}$ has an empirical size 0.072, close to the nominal level 0.05 while the estimated size of standard *t*-test is 0.469, suffering greatly from the size distortion of similar magnitudes to that of $\hat{\beta}$.

References

Anderson, T. and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46-63.

Andrews, D. W. K., and X. Cheng (2010): "Estimation and Inference with Weak, Semi-strong, and Strong Identification," working paper, Yale University.

Andrews, I., and A. Mikusheva (2011): "Maximum Likelihood Inference in Weakly Identified Models," working paper, Massachusetts Institute of Technology.

Ansley, C. F. and P Newbold (1980): "Finite Sample Properties of Estimates for Auto-regressive Moving Average Processes, *Journal of Econometrics*, 13, 159-84.

Bansal, R., and A. Yaron (2000): "Risks for the long run: A potential resolution of asset pricing puzzles," NBER Working Paper 8059.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: a Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, LIX, 1481-1509.

Binsbergen, J.H. van, and R.S.J. Koijen (2010): "Predictive Regressions: A Present-Value Approach," *The Journal of Finance*, LXV, 1439-1471.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

Bollerslev, T. (1987): "A Conditional Heteroskedastic Time Series M for Speculative Prices and Rates of Return," *The Review of Economics and Statistics*, 69, 542-547.

Bollerslev, T., and J. Wooldridge (1992): "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances," *Econometric Reviews*, 11, 143-172.

Breusch T. S., and A. R. Pagan (1980): "The Lagrange Multiplier Test and its Application to Model Specification in Econometrics," *Review of Economic Studies*, XLVII, 239-253.

Clark, P.K. (1987): "The Cyclical Component of U.S. Economic Activity," *The Quarterly Journal of Economics*, 102, 797-814.

Dufour, J. M. (1997): "Some Impossibility Theorems in Econometrics with Application to Structural and Dynamic Models," *Econometrica*, 65, 1365-88.

Fieller, E. C. (1932): “The Distribution of the Index in a Normal Bivariate Population,” *Biometrika*, 24, 428-440.

Fieller, E. C. (1954): “Some Problems in Interval Estimation,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 16, 175-85.

Hannan, E. J. (1982): “Testing for Autocorrelation and Akaike’s Criterion,” in Gani J. M. and E. J. Hannan (eds.) *Essays in Statistical Science*, Applied Probability Trust, Sheffield, 403-412.

Harvey, A. C. (1985): “Trends and Cycles in Macroeconomic Time Series,” *Journal of Business and Economic Statistics*, 3, 216-27.

Harvey, A. C. (1990): *The Econometric Analysis of Time Series*, MIT Press Edition, The MIT Press, Cambridge Massachusetts.

Hauser, M. A., B. M. Pötscher, and E. Reschenhofer (1999): “Measuring Persistence in Aggregate Output: ARMA Models, Fractionally Integrated ARMA Models and Nonparametric Procedures,” *Empirical Economics*, 24, 243-269.

Hinkley, D. V. (1969): “On the Ratio of Two Correlated Normal Random Variables,” *Biometrika*, 56 (3), 635-639.

Ma, J. (2012): “Long-Run Risks and Its Implications for the Equity Premium Puzzle: New Evidence from a Multivariate Framework,” *Journal of Money, Credit, and Banking*, forthcoming.

Ma, J., C. R. Nelson., and R. Startz (2007): “Spurious in the GARCH(1,1) Model When It Is Weakly Identified,” *Studies in Nonlinear Dynamics & Econometrics*, Vol.11, No. 1, Article 1.

Ma, J. and M.E. Wohar (2012): “Expected Returns and Expected Dividend Growth: Time to Rethink an Established Empirical Literature,” working paper, University of Alabama.

Morley, James, C.R. Nelson and E. Zivot (2003): “Why Are Beveridge-Nelson and Unobserved Component Decomposition of GDP So Different?” *Review of Economics and Statistics*, Vol. 85, No.2, 235-243.

Nelson, C. R. (1988): “Spurious Trend and Cycle in the State Space Decomposition of a Time Series with a Unit Root," *Journal of Economic Dynamics & Control*, 12, 475-488.

Nelson, C. R., and R. Startz (2007): “The Zero-Information-Limit Condition and Spurious Inference in Weakly Identified Models,” *Journal of Econometrics*, Vol. 138, 47-62.

Staiger, D., J. H. Stock., and M. W. Watson (1996): “How Precise Are Estimates of the Natural Rate of Unemployment?,” National Bureau of Economic Research, Working Paper 5477, March 1996.

Staiger, D., J. H. Stock., and M. W. Watson (1997): “The NAIRU, Unemployment and Monetary Policy,” *The Journal of Economic Perspectives*, 11, 33-49.

Working, H. (1960): “Note on the correlation of first differences of a random chain,” *Econometrica*, 28, 916-918.

End.