*Valid Inference for a Class of Models Where Standard Inference Performs Poorly; Including Nonlinear Regression, ARMA, GARCH, and Unobserved Components***

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

Charles R. Nelson*
Department of Economics
University of Washington

First draft June 18, 2006

This draft September 15, 2008

JEL *classification*: C120, C220, C330

Keywords: ARMA, Unobserved Components, State Space, GARCH, Zero-Information-Limit-Condition.

Abstract

Standard inference works poorly in models of the form $y = \gamma \bullet g(\beta, x) + \varepsilon$, because the standard error for $\hat{\beta}$ depends on $\hat{\gamma}$. In this paper we show that this problem is usefully studied by working with the linearization of $g(.)$ and the resulting reduced form regression. Bias and dispersion in $\hat{\beta}$ depend on correlation between the 'regressors' and on γ, as does the size of the *t*-test. A reduced form test however is exact when $g(.)$ is linear and has nearly correct size in examples from non-linear regression, ARMA, GARCH, and Unobserved Components models. Further, its distribution does not depend on the identifying restriction $\gamma \neq 0$.

* Corresponding author: Box 353330, University of Washington, Seattle, WA 98195, USA. Tel: +1 206 543 5955; fax +1 206 685 7477.
E-mail address: cnelson@u.washington.edu.

** The authors gratefully acknowledge support from the Ford and Louisa Van Voorhis Endowment at the University of Washington. We thank the following for helpful comments: Michael Dueker, Walter Enders, Evan Koenig, Junsoo Lee, James Morley, Christian Murray, David Papell, Ruxandra Prodan, Bent Sørensen, Richard Startz, and Eric Zivot, but responsibility for all errors is ours.

1. Introduction

This paper is concerned with inference in the class of models that have a representation of the form

$$y_i = \gamma \bullet g(\beta, x_i) + \varepsilon_i ; i = 1,..., N . \qquad (1.1)$$

The parameter of interest is β which is identified only if $\gamma \neq 0$. Additional regressors and parameters would often be present in practice. We assume that errors ε_i are *i.i.d.* $N(0,\sigma^2)$ so that Maximum Likelihood estimates of $\hat{\gamma}$ and $\hat{\beta}$ are obtained by non-linear least squares, given data *y* and $\mathbf{x}$. In addition to non-linear regression models, this class includes the workhorse ARMA model, where data **x** are lagged observations. By extension, the GARCH model and Unobserved Components State Space models for trend and cycle decomposition fall into this class as well. What these models have in common is that standard inference based on asymptotic theory often works poorly in finite samples, essentially because the estimated standard error for $\hat{\beta}$ depends on $\hat{\gamma}$. Further, the distribution of $\hat{\beta}$ will generally be displaced away from the true value. Nelson and Startz (2007) – hereafter NS - show that the estimated standard error for $\hat{\beta}$ is generally too small. Although these two effects might seem to imply that the *t*-statistic would be over-sized, NS show that size distortion may go either way. In this paper we demonstrate that the linear approximation of $g(\beta, \mathbf{x}_i)$ is useful in understanding and predicting the direction of bias in $\hat{\beta}$ as well as test size distortion, and it provides an alternative *t*-test that works well in situations where the standard *t*-test performs poorly.

The paper is organized as follows. Section 2 studies the case where $g(\beta, x_i)$ is linear, the archetype of this class for which useful analytical results are available. We examine the sources of bias in $\hat{\beta}$ and distortion in the size of the *t*-test, comparing its size and power with that of the reduced form test. Section 3 studies how well the findings in the linear case hold in non-linear models where the reduced form is only a linear approximation, in particular nonlinear regression, ARMA, Unobserved Components model of trend and cycle, and in GARCH models. Section 4 concludes.

2. Bias and test size when *g(.)* is linear.

In the case that *g(.)* in (1.1) is linear the model takes the form:

$$y_i = \gamma \bullet (x_i + \beta \bullet z_i) + \varepsilon_i \qquad (2.1)$$

where x_i and z_i denote regressors. This model is not only an archetype of the class we are interested in, but is also of interest in practice, for example, the Phillips curve model of Staiger, Stock and Watson (1997) where *y* is the change in inflation, $g = (x_i + \beta)$, where *x* is actual unemployment and β is the natural rate. The reduced form of (2.1) is

$$y_i = \gamma \bullet x_i + \lambda \bullet z_i + \varepsilon_i \qquad (2.2)$$

where $\lambda = \gamma \bullet \beta$. The least squares estimate from (2.1) is equal to the indirect least squares estimate from the reduced form, thus $\hat{\beta} = \hat{\lambda} / \hat{\gamma}$.

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:

$$\hat{\lambda} = \alpha + \kappa \bullet \hat{\gamma} + v \tag{2.3}$$

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:

$$\hat{\beta} = -\rho + \rho \bullet \left(\frac{\gamma}{\hat{\gamma}} \right) + \frac{v}{\hat{\gamma}} \tag{2.4}$$

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 (2.4) 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.) These effects are apparent in the Monte Carlo results that follow.

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:

$$V_{\hat{\beta}} = \frac{1}{\gamma^2} \bullet \frac{\sigma^2}{N} \bullet \frac{m_{xx} + 2\beta \bullet m_{xz} + \beta^2 \bullet m_{zz}}{m_{xx} \bullet m_{zz} - m_{xz}^2} \tag{2.5}$$

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:

$$t_{\hat{\beta}}^2 = (\hat{\beta} - \beta_0)^2 \bullet \left[\hat{\gamma}^2 \bullet \frac{N}{\hat{\sigma}^2} \bullet \frac{m_{xx} \bullet m_{zz} - m_{xz}^2}{m_{xx} + 2\hat{\beta} \bullet m_{xz} + \hat{\beta}^2 \bullet m_{zz}} \right] \tag{2.6}$$

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:

$$t_{\hat{\beta}}^2 = \frac{\hat{\lambda}^2}{\hat{\sigma}^2} \bullet N \bullet (1 - \rho^2) \bullet \frac{1}{1 + 2\hat{\beta} \bullet \rho + \hat{\beta}^2} = t_{\hat{\lambda}}^2 \bullet \frac{1}{1 + 2\hat{\beta} \bullet \rho + \hat{\beta}^2} \tag{2.7}$$

Since the reduced form is a classical linear regression, $t_{\hat{\lambda}}$ 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 any 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 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 $t_{\hat{\beta}}^2$ arbitrarily larger than $t_{\hat{\lambda}}^2$. 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.

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

To illustrate displacement of $\hat{\beta}$ away from its true value in the direction of $-\rho$, its concentration around that value, and the relative performance of $t_{\hat{\beta}}$ and $t_{\hat{\lambda}}$, we report a series of Monte Carlo experiments where true $\beta = 0$, the regressors have unit variance and are fixed in repeated samples, and errors are *i.i.d. N*(0,1). The size of the reduced form test $t_{\hat{\lambda}}$ is of course exactly its nominal size (we focus on .05) since the reduced form is a classical linear regression for this model, so we do not report its empirical size. Estimation is done in EViews™ using the non-linear regression routine, so the calculation of $t_{\hat{\beta}}$ is representative of what would be reported in applied work. The number of replications is 10,000 in all experiments in this paper, and the standard deviation of estimated size is .002 when the true size is .05.

Table 1 explores the effect of γ on inference when the regressors are orthogonal, the case where we do not expect to find displacement of $\hat{\beta}$ away from zero or concentration, but we do expect the standard *t*-test to reject less frequently than its .05

nominal level. Since it is the magnitude of γ relative to its sampling error that plays a critical role in inference, the second line reports the ratio $\gamma / \sqrt{V_{\hat{\gamma}}}$ as a metric for spurious inference. The next two lines present the median of $\hat{\beta}$ and its inter-quartile range as measures of location and dispersion respectively. The results confirm that $\hat{\beta}$ is centered around zero and becomes less disperse for larger values of γ as asymptotic theory becomes a better approximation to the actual distribution. The last row reports the empirical size of $t_{\hat{\beta}}$, and confirms that it is undersized, becoming less so with larger values of the metric $\gamma / \sqrt{V_{\hat{\gamma}}}$. The fact that the distribution of $t_{\hat{\beta}}$ depends on the unknown true value of γ means that it is non-pivotal in finite samples, its distribution depending on this nuisance parameter. As mentioned above, we do not report the empirical size of $t_{\hat{\lambda}}$ since the actual size is exactly .05.

Table 1: The Effect of γ on the Distribution of $\hat{\beta}$ and Size of $t_{\hat{\beta}}$ with Orthogonal Regressors; *N* = 100

True γ	.01	.10	.5	1.0
Asymptotic $\gamma \big/ \sqrt{V_{\hat{\gamma}}}$	.1	1	5	10
Median $\hat{\beta}$	.10	.03	-.00	-.00
Range (.25, .75)	(-.95, 1.17)	(-.65, .68)	(-.14, .13)	(-.06, .06)
Size of $t_{\hat{\beta}}$	.0001	.0002	.038	.050

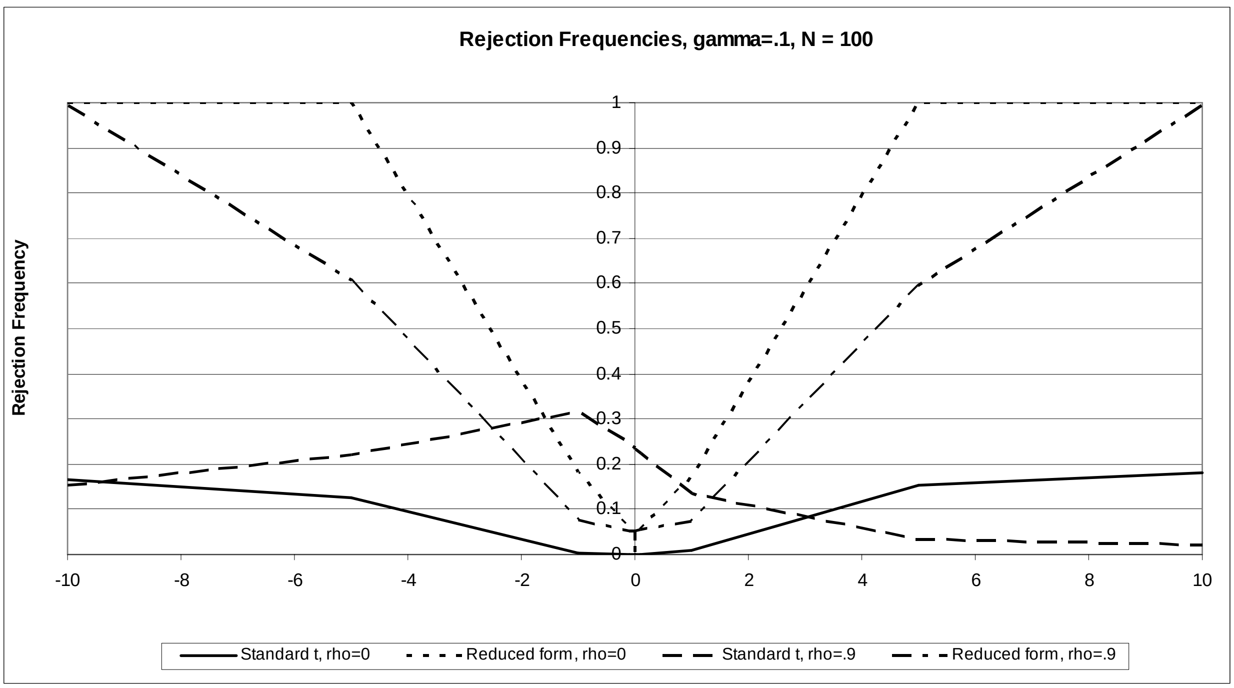
Table 2 explores the effect of correlation ρ between regressors on the distribution of $\hat{\beta}$ and on the size of $t_{\hat{\beta}}$ when the true value of γ is a 'moderate' .10. Note that as the metric $\gamma / \sqrt{V_{\hat{\gamma}}}$ is smaller for larger ρ, the distribution of $\hat{\beta}$ becomes more concentrated around a value closer to $-\rho$, and test size goes from too small to excessive, as expected.

Table 2: The Effect of Correlation ρ Between Regressors on Distribution of $\hat{\beta}$ and on Size of $t_{\hat{\beta}}$; $\gamma = .1$, *N* = 100.

Correlation ρ	0	.50	.90	.99
Asymptotic $\gamma / \sqrt{V_{\hat{\gamma}}}$	1	.87	.44	.14
Median $\hat{\beta}$	.03	-.21	-.71	-.96
Range (.25, .75)	(-.65, .68)	(-.84, .54)	(-1.27, -.14)	(-1.17, -.75)
Size of $t_{\hat{\beta}}$	.0002	0.019	.235	.565

Figure 1 explores the response of rejection frequencies for $t_{\hat{\beta}}$ and $t_{\hat{\lambda}}$ to departures of the true value of β from the null value of zero when the true value of γ is again .10, when the independent variables are, alternatively, orthogonal and strongly correlated. Since test 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 for $t_{\hat{\beta}}$ increases very slowly as a function of the true β. In contrast, the power of the correctly sized test $t_{\hat{\lambda}}$ rises steeply as the true β departs from zero. The corresponding comparison when the independent variables are strongly correlated reveals that for $t_{\hat{\beta}}$ rejections 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 reduced form test $t_{\hat{\lambda}}$ does increase as expected as the null departs from the true value. Thus we conclude that the standard test is not only poorly sized but contains little if any information about the null hypothesis.

Figure 1: Rejection Frequencies for tests of $H_0 : \beta = 0$, *N* = 100, γ = .10.



Asymptotic theory does take hold as sample size becomes large, albeit very slowly, as is evident in Table 3 below. We note that the size of $t_{\hat{\beta}}$ just approaches to its correct level only as the quantity $\gamma/\sqrt{V_{\hat{\gamma}}}$ increases to 10, requiring a sample size as large as 10,000 for $\gamma = .1$, and 1,000,000 for $\gamma = .01$!

Table 3: 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	.01	.01	.01	.1
Asymptotic $\gamma/\sqrt{V_{\hat{\gamma}}}$	.1	1	10	10
Median $\hat{\beta}$	.10	.01	-.00	.00
Range (.25, .75)	(-.95, 1.17)	(-.64, .63)	(-.07, .07)	(-.07, .07)
Size of $t_{\hat{\beta}}$	.0001	.0006	.043	.045

To sum up, in the archetypal case where *g(.)* is linear it is clear how the displacement of $\hat{\beta}$ away from the true value and distortion in the size of the standard *t*-test depend on correlation between the regressors, while the reduced form test has exact size. The remainder of the paper is concerned with models where *g(.)* is not linear and the reduced form test is based on a linear approximation.

3. A reduced form test for non-linear *g*(.) and relative performance in four models.

3.0 The Reduced Form Test in a Linear Approximation.

More generally *g*(.) will not be linear as in section 2 but a reduced form test can be based on the linear approximation of $g(\beta, x_i)$ around β_0:

$$y_i = \gamma \bullet [g(\beta_0, x_i) + (\beta - \beta_0) \bullet g_\beta(\beta_0, x_i)] + e_i \qquad (3.0.1)$$

where $g_\beta = dg(.)/d\beta$ and e_i includes a remainder. The corresponding reduced form is:

$$y_i = \gamma \bullet g(\beta_0, x_i) + \lambda \bullet g_\beta(\beta_0, x_i) + e_i; \text{ where } \lambda = \gamma \cdot (\beta - \beta_0). \qquad (3.0.2)$$

The least squares estimate conditional on β_0 in (3.0.2) is equivalent to the indirect least squares estimate $\hat{\beta} = \beta_0 + (\hat{\lambda}/\hat{\gamma})$ and the implication of the null hypothesis $\beta = \beta_0$ is $\lambda = 0$. Intuitively, the first term captures the contribution of β to the model if the null is correct, the second term being required only if it is wrong. Since the reduced form test cannot be 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 2 we showed that bias in $\hat{\beta}$ as well as the size of the standard *t*-test 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 *t*-test., but not the reduced form test which relies on evaluation the $g(.)$ and $g_{\beta}(.)$ under the null hypothesis.

3.1. Non-linear Regression: A Production Function.

Consider the Hicks-neutral Cobb-Douglas production function:

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

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

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

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 $g_{\beta}(\beta, x_i)$. The alternative 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 a sample of x_i from the log-normal distribution and pair 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 4 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 $x_i^{\beta} \log(x_i)$. If the true value of β were .9, and the regressors were evaluated at .9, then the correlation is .92 and $\hat{\beta}$ would be biased downward since correlation and bias vary inversely. In the iteration of Gauss-Newton, these regressors are re-evaluated at the successive provisional estimates until convergence is reached. As we see below, the point estimates are strongly biased downward as expected. Note also that the standard *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 in the expected direction, it is not as dramatic as in section 2 even as true β becomes as large. This attenuation is attributable to the fact that the correlation is attenuated when regressors are evaluated at the downward biased point estimates. Finally, we also report the size of the reduced form test of $H_0 : \lambda = 0$, in (3.1.2) which is close to correct in all cases.

Table 4: Small Sample Distribution of $\hat{\beta}$ and Test Size, True γ = .01, *N* = 100.

True β	0	.1	.5	.9
$\rho_{g(\beta),g_\beta(\beta)}$	.07	.29	.77	.92
Asymptotic $\gamma/\sqrt{V_{\hat{\gamma}}}$	.10	.10	.09	.11
Median $\hat{\beta}$	-0.04	-0.09	-0.05	0.12
Range (.25, .75)	(-.53, .50)	(-.59, .42)	(-.56, .48)	(-.41, .71)
Size of $t_{\hat{\beta}}$	0.027	0.037	0.114	0.179
Size of $t_{\hat{\lambda}}$	0.053	0.054	0.054	0.054

We report in Table 5 the corresponding results as γ increases from .01 to 1 for a fixed value of true β at .5. As the key metric $\gamma/\sqrt{V_{\hat{\gamma}}}$ approaches 10, the asymptotic distribution gradually takes hold and the actual size of the standard test is correct. We again find that a value of 10 for this metric seems to be a rough rule of thumb for correct size of the standard test. The reduced-form test, however, maintains about the correct size across the range of parameter values.

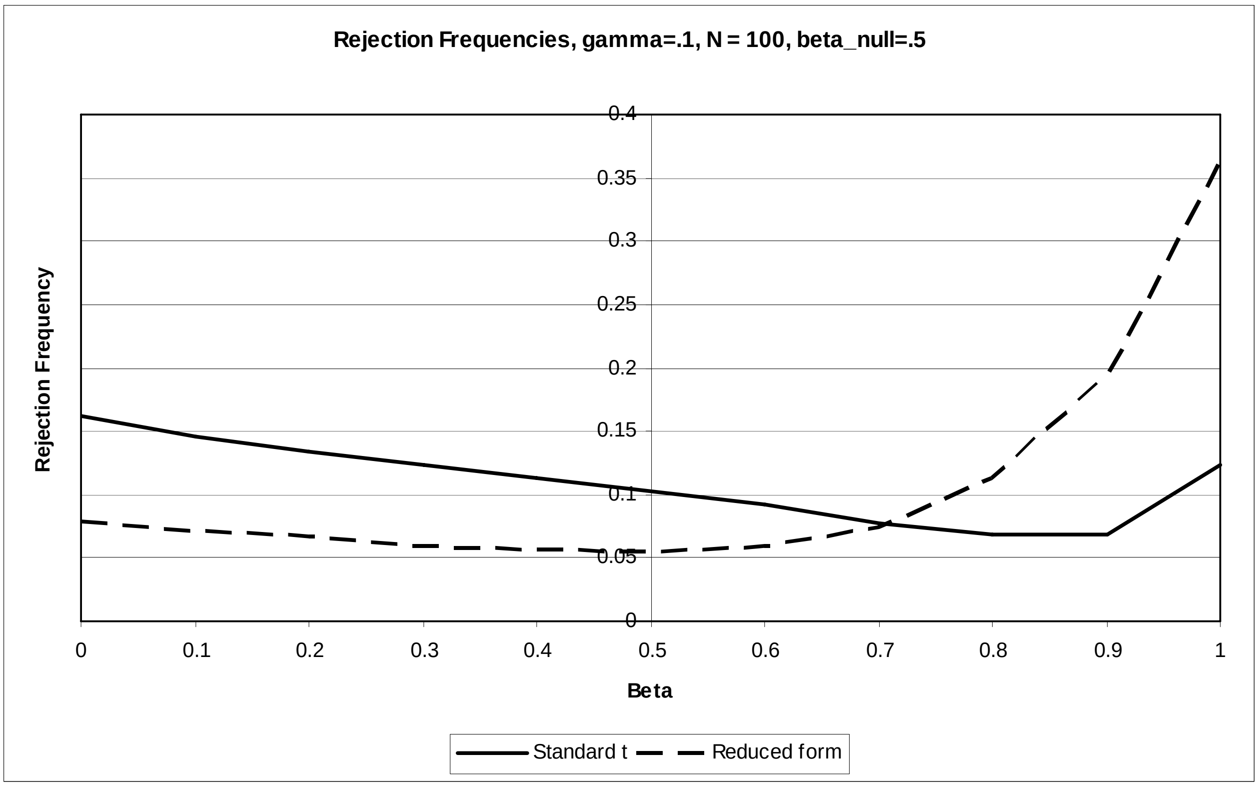
Table 5: Small Sample Distribution of $\hat{\beta}$ and Test Size , *N* = 100, true β = .5

True γ	.01	.1	1
Asymptotic $\gamma/\sqrt{V_{\hat{\gamma}}}$	.09	.91	9.10
Median $\hat{\beta}$	-.05	.27	.50
Range (.25, .75)	(-.56, .48)	(-.26, .64)	(.46, .54)
Size of $t_{\hat{\beta}}$	.114	.103	.052
Size of $t_{\hat{\lambda}}$	.054	.054	.054

The case that $\gamma = 0$ corresponds to failure of the identification condition for β, so the asymptotic theory underling the 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 both to its nominal size and what we observed over the range of γ above.

Figure 2 below presents a comparison about how well the reduced-form test and the standard *t*-test can detect the departure of true β walking away from the null $H_0 : \beta = .5$ to both ends. The standard $t_{\hat{\beta}}$-test starts with a higher level of rejection frequency when the true β is .5, reflecting its size distortion and starts to decline as true β becomes higher than .5, only beginning climbing up as true β reaches as high as .9. The reduced form test using $t_{\hat{\lambda}}$, however, starts with a correct size and has a higher level of rejection frequency *monotonically* as true β deviates further away from the null. As true β heads toward the left of the null, rejections by the standard *t*-test rise but not significantly more rapidly than the reduced form test. Neither test is very sensitive to departure from the null in the direction of zero. We surmise that the non-linearity of the model accounts for this asymmetry.

Figure 2: Rejection Frequencies for the test $H_0 : \beta = .5$, *N* = 100, True γ = .1.



3.2. The ARMA (1,1) Model.

ARMA models also belong to the class we are interested in. 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 term, we may express it in the form:

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

where, $\gamma = (\phi - \theta)$, $g(\theta, \vec{y}_{t-1}) = \sum_{i=1}^{\infty} \theta^{i-1} y_{t-i}$ and $\vec{y}_{t-1} = (y_{t-1}, y_{t-2}, ...)$. 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.

In light of the discussions in Section 2, we linearize the nonlinear *g(.)* around the null to achieve the reduced-form test for θ:

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

where $g_\theta(\theta, \vec{y}_{t-1}) = \frac{\partial g(\theta, \vec{y}_{t-1})}{\partial \theta} = \sum_{i=2}^{\infty} (i-1) \bullet \theta^{i-2} y_{t-i}$, $\lambda = \gamma \bullet (\theta - \theta_0)$ and e_i incorporates a remainder term. If the null $\theta = \theta_0$ is correct, the second term in reduced form regression (3.2.3) should not be significant. In practice, to evaluate the regressors [$g(\theta_0, \vec{y}_{t-1})$, $g_\theta(\theta_0, \vec{y}_{t-1})$], we set y_t at its unconditional mean for all $t \le 0$. To compare the reduced form test to the conventional *t*-test, we estimated the model in EViews™.

Table 6 explores the effect of γ for true $\theta = 0$ with a sample size $T = 1000$. Clearly, when γ is small relative to sample variation as indicated by a small value of the metric $\gamma/\sqrt{V_{\hat{\gamma}}}$ the conventional *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 gets closer to the nominal level 0.05. The fact that the sampling distribution of the conventional *t*-test statistic depends on the nuisance parameter γ implies again that the test is not pivotal. Note that the reduced-form test $t_{\hat{\lambda}}$ 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 reduced form test is correct within sampling error.

One may wonder how the reduced-form test performs when true γ is zero, corresponding to the failure of identification. As we pointed out, the reduced-form test is still well defined in this case and the estimated size of it in the Monte Carlo experiment is 0.0509, close to correct.

Table 6: Effect of γ on Inference for ARMA (1,1), True θ = 0, *T* = 1,000.

True $\gamma(=\phi)$	.01	.1	.2	.3
Asymptotic $\gamma/\sqrt{V_{\hat{\gamma}}}$	.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 of $t_{\hat{\theta}}$	0.4585	0.2237	0.1051	0.0734
Size of $t_{\hat{\lambda}}$	0.0506	0.0518	0.0526	0.0522

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 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. Figure 4 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 when $\hat{\theta}$ falls far from zero, not simply the too-small standard error as surmised by NS. The relative success of the reduced form test comes from the fact that it evaluates the test statistic under the null hypothesis $\theta_0 = 0$ instead of at $\hat{\theta}$.

Figure 3: Histogram of $\hat{\theta}$ in the Monte Carlo. True γ = .01, θ = 0, T = 1,000.

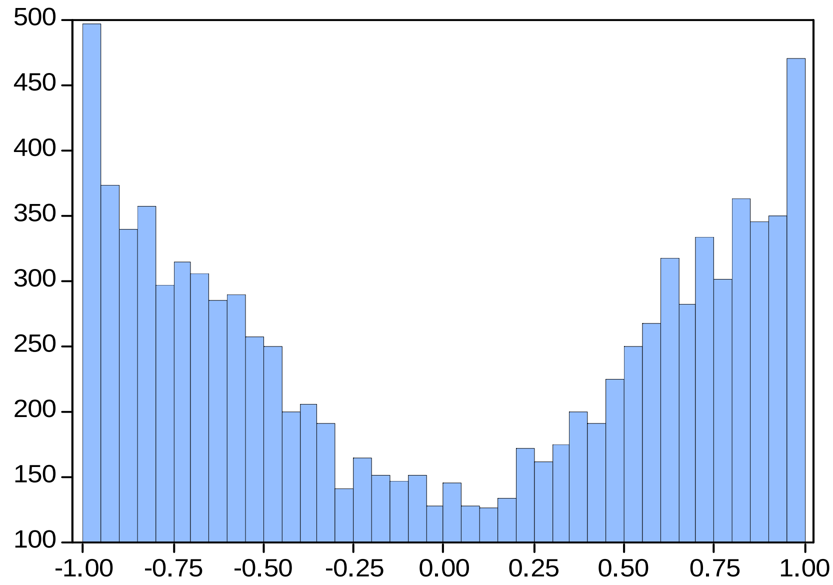


Figure 4: Computed un-centered correlation between $g(\theta, \vec{y}_{t-1})$ and $g_\theta(\theta, \vec{y}_{t-1})$ based on one sample draw. True γ = .01, θ = 0, T = 1,000.

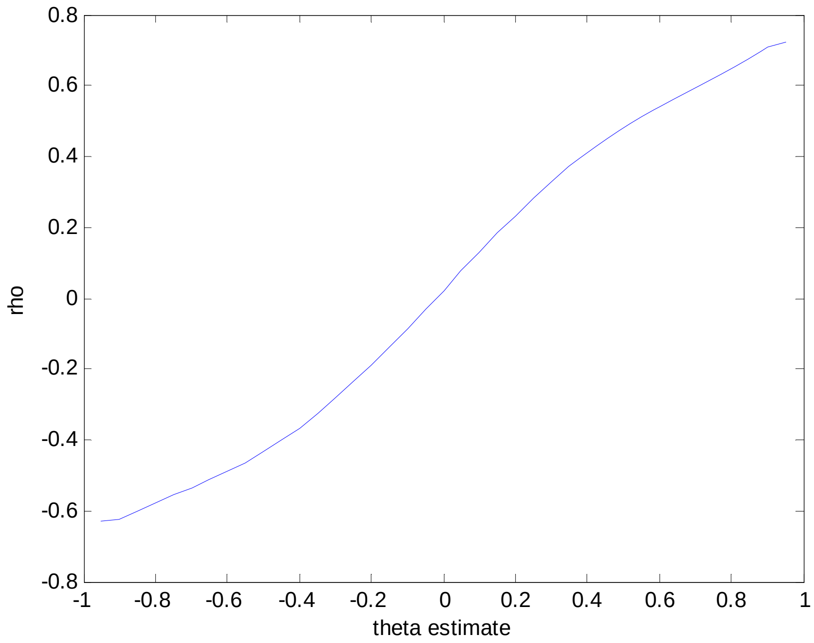


Table 7 explores the effect of increasing sample size when true γ = .01. Asymptotic theory does take hold, but the conventional *t*-test approaches correct size very slowly (requiring a sample size as large as 10,000 for γ = .1!). In contrast, the reduced form test consistently has correct size within sampling error.

Table 7: Sample Size and Inference in the ARMA (1, 1), True θ = 0.

Sample size	100	1000	10,000	10,000
True $\gamma(=\phi)$	.01	.01	.01	.1
Asymptotic $\gamma/\sqrt{V_{\hat{\gamma}}}$	0.1	0.32	1	10
Median $\hat{\theta}$	-.04	-.02	-.02	-.00
Range (.25, .75)	(-.69, .67)	(-.65, .64)	(-.58, .55)	(-.07, .07)
Size of $t_{\hat{\theta}}$	.483	.458	.399	.066
Size of $t_{\hat{\lambda}}$	.051	.051	.049	.048

Often it is the AR root ϕ that is of a greater economic interest since it measures persistence. For example, if consumption growth g_t follows an ARMA (1,1) process, a large value of ϕ implies that any shock to the economy has a long-lasting impact to the economic agent's conditional expectation of future consumption growth. Recently, Bansal and Yaron (2000, 2004) show that such a high level of persistence, interpreted as long-run risk, may explain the equity premium puzzle of Mehra and Prescott (1985). Ma (2007) finds that the estimated ARMA(1,1) implies a small estimated γ 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.

The reduced-form test for ϕ turns out to require an extra step and we offer details in the Appendix A.1. For the case $\gamma = 0.1, \phi = 0$ and T = 100 the rejection frequency of the reduced form test is 0.046 in contrast to 0.423 of the standard *t*-test.

The reduced-form test can also be generalized to address an ARMA model of arbitrary order, of which we provide a general treatment in the Appendix A.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 *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 5, 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 6 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 our 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 5 and 6 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 suggest implementing the reduced-form 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 reduce-form test statistic by following the strategy in Appendix A.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 of reduced-form 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 test works well and gives estimated size 0.0581 in the Monte Carlo with true parameter values $\mu = 0.8, \phi = 0, \sigma_{\eta}^{2} = 1, \sigma_{\varepsilon}^{2} = 0$.

This 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 A.2.

Figure 5: Plot of $\hat{\phi}$ in the Monte Carlo Experiment with true parameter

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

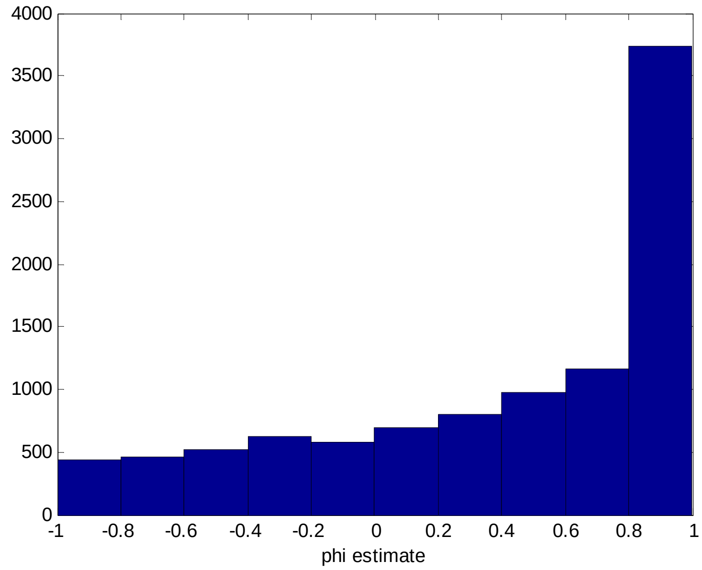
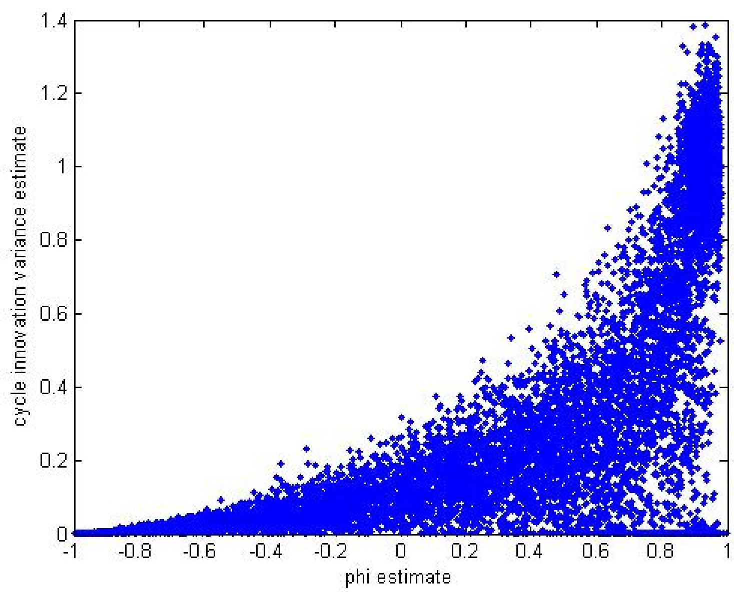


Figure 6: Scatter Plot of $\hat{\phi}$ and $\hat{\sigma}_{\varepsilon}^{2}$ in the Monte Carlo Experiment with true parameter $\mu = 0.8, \phi = 0, \sigma_{\eta}^{2} = 0.95, \sigma_{\varepsilon}^{2} = 0.05$



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:

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

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

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:

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

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, $\alpha + \beta$ and β correspond to the AR and MA roots respectively, and 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.

The reduced-form 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:

$$h_t = \frac{\omega}{1-\beta} + \alpha \bullet g(\beta, \vec{\varepsilon}_{t-1}^{\,2}) \quad (3.4.4)$$

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:

$$h_t = c + \alpha \bullet g(\beta_0, \vec{\varepsilon}_{t-1}^{\,2}) + \lambda \bullet g_\beta(\beta_0, \vec{\varepsilon}_{t-1}^{\,2}) \tag{3.4.5}$$

The reduced-form test is the *t*-stat of the null $\lambda = 0$ in (3.4.5).

Table 8 presents the comparison of the reduced-form test and the standard *t*-test as α increases when the truth is $\beta = 0$ and sample size $T = 1000$ (the Matlab 6.1 code is available on request). When the key metric $\gamma \big/ \sqrt{V_{\hat{\gamma}}}$ is small, the standard *t*-test rejects the null too often. The reduced-form test however has consistently better size.

Table 8: Reduced form and standard *t*-tests for GARCH(1,1): True β = 0, *T* = 1,000.

True $\gamma (= \alpha)$	.01	.05	.1	.2
Asymptotic $\gamma \big/ \sqrt{V_{\hat{\gamma}}}$	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 of $t_{\hat{\lambda}}$	0.078	0.074	0.076	0.096

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

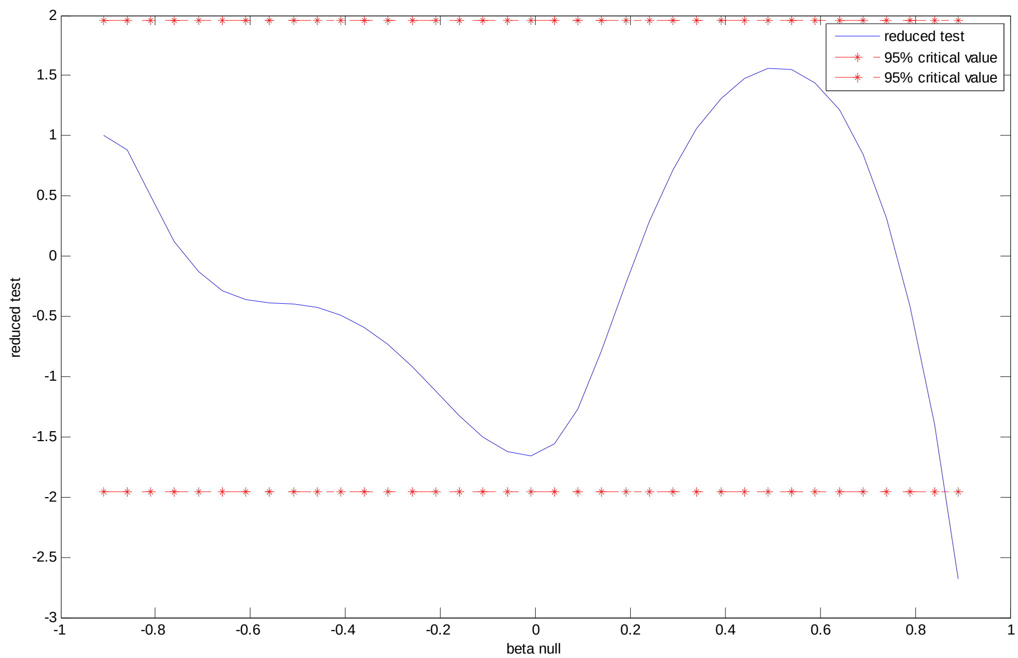
The sum $\alpha + \beta$ is of potentially greater economic interest since it measures the persistence of volatility in (3.4.3). Bansal and Yaron (2000, 2004) show that a large value of $\alpha + \beta$, interpreted as long run risk in uncertainty dynamics, may help to resolve the equity premium puzzle. Appendix B gives details about how to obtain a valid test for $\hat{\alpha} + \hat{\beta}$ and evaluates its performance; see Ma (2007) for further discussion.

In the following example we show how to apply the reduced-form test to a real dataset and obtain a confidence interval for $\hat{\beta}$ to see if it gives a different result from the standard *t*-test. The monthly S&P 500 index return data is from the Eviews 5.1 DRI Database for the sample period January 1947 to September 1984 corresponding to Bollerslev (1987). The GARCH estimates along with the Bollerslev and Wooldridge's (1992) robust standard errors (in parenthesis), after accounting for the "Working" effect (see Working (1960)), are reported below:

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

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, in light of our Monte Carlo results above, raises concern about the possibility of spurious inference for β. To numerically invert the reduced-form test statistic, we create a grid of β_0's, compute the corresponding $t_{\hat{\lambda}}$'s and plot the latter against the former in Figure 7. 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.

Figure 7: The 95% Confidence Interval for $\hat{\beta}$ based on the reduced-form test for the monthly S&P 500 stock return data



4. Summary and Conclusions

This paper considers models of the form $y = \gamma \bullet g(\beta, x) + \varepsilon$, where β is the parameter of interest. This class includes not only the obvious non-linear regression model but also the workhorse ARMA model of time series and by extension GARCH and State Space models such as those used for decomposition into trend and cycle components. Inference is problematic because the standard error for $\hat{\beta}$ depends on $\hat{\gamma}$. Nelson and Startz (2007) showed that although that standard error is downward biased in a broad class of models including this one 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 model

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

The distribution of $\hat{\beta}$ is biased in a direction determined by the correlation between the 'regressors' in the reduced form, and the distribution becomes concentrated when that correlation is strong. The distribution of the standard *t*-statistic for $\hat{\beta}$ based on asymptotic theory is also dependent on that correlation, as is the size of the *t*-test. Both of these distributions are also dependent on the true value of γ, so the conventional *t*-test is not pivotal in finite samples. A reduced form test that exploits the fact that under the null hypothesis $\beta = \beta_0$ then $\lambda = 0$ is exact when $g(.)$ is linear and we show that is has nearly correct size when the reduced form model is only an approximation. Further, its distribution does not depend on the identifying restriction $\gamma \neq 0$. The paper illustrates this with examples from non-linear regression, ARMA, GARCH, and Unobserved Components models.

References

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.

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.

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

Durfour, J. M. (1997): "Some Impossibility Theorems in Econometrics with Application to Structural and Dynamic Models," *Econometric*, 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.

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

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

Ma, J. (2007): “The Long-Run Risk in Consumption and Equity Premium Puzzle: New Evidence Based on Improved Inference,” working paper, Department of Economics, Finance and Legal Studies, University of Alabama.

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.

Mehra, R., and E. C. Prescott (1985): “The Equity Premium: A puzzle,” *Journal of Monetary Economics*, 15, 145-161.

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 (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.

Appendix A.1

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

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

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:

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

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™. The rejection frequency of the proposed test is 0.0461, at the nominal level 0.05, in contrast to 0.4233, that of the standard *t*-test.

Appendix A.2

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

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

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:

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

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:

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

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 0.5712 and 0.6981 at a nominal level 0.05. In contrast the reduced-form test for $\lambda_1 = 0$ and $\lambda_2 = 0$ based on regression (A.2.3) gives rejection frequencies of 0.0491 and 0.0487 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 B

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

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

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

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

Where $\lambda^* = \alpha \bullet (\rho - \rho_0)$ and the reduced form test is the *t*-stat for λ^*. To make the reduced form 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 0.0723, close to the nominal level 0.05 while the estimated size of standard *t*-test is 0.4690, suffering greatly from the size distortion of similar magnitudes to that of $\hat{\beta}$.