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
University of Alabama

Charles R. Nelson
University of Washington

Conference in Honor of Andrew Harvey - June 30, 2012

Copyright © 2012 by Jun Ma and Charles R. Nelson
Models that may take the form:

\[ y_i = \gamma \cdot g(\beta, \theta, x_i) + \epsilon_i; i = 1, \ldots, N. \]  

(1.1)

Include non-linear regression, ARMA model, GARCH, and Unobserved Components models.

\( \beta \) is of interest for testing, and \( \theta \), vector of other parameters; identified only if \( \gamma \neq 0 \).

Standard inference for \( \hat{\beta} \) uses estimated standard error to construct a Wald \( t \)-test statistic, \( t(\hat{\beta}) \).

Weak identification: identifying parameter \( \gamma \) is small relative to its estimation error.

NS (2007) show that under weak identification SE for \( \hat{\beta} \) is ‘too small,’ though size of test based on \( t(\hat{\beta}) \) will depend on details of model and DGP.

Asymptotic theory of course still holds, but it takes hold very slowly as sample size increases.
In this paper we show that an LM test works well in finite samples under weak identification.

Following in the spirit of Breusch and Pagan (1980), we linearize \( g(\beta, \theta, x_i) \) in \( \beta \) around the null hypothesis, \( H_0 : \beta = \beta_0 \), obtaining:

\[
y_i = \gamma \cdot g(\beta_0, \theta, x_i) + \lambda \cdot g_\beta(\beta_0, \theta, x_i) + e_i; \lambda = \gamma \cdot (\beta - \beta_0),
\]

where \( g_\beta = dg(\cdot)/d \beta \). The LM test may be computed in two steps.

However, following Harvey (1990), a “modified LM” test may 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 suggests that little is lost in doing one step and, we find this to be the case.

Alternatively this is a reduced form test, in the spirit of Anderson and Rubin (1949), or the test for a ratio of regression coefficients of Fieller (1954), both of which have exactly correct size when \( g(\cdot) \) is linear.
Archetypal case: \( g(.) \) is linear,
\[
y_i = \gamma \cdot (x_i + \beta \cdot z_i) + \varepsilon_i, \quad (2.2)
\]

Examples: Phillips curve of Staiger, Stock and Watson (1997), and stacked IV/2SLS equations.

A Wald test may be based on the asymptotic standard error for \( \hat{\beta} \) or the ‘delta method.’

NS: Size for \( t(\hat{\beta}) \) depends on the correlation between the reduced form coefficients.

Modified LM test for \( H_0 : \beta = 0 \) tests significance of second term in
\[
y_i = \gamma \cdot x_i + \lambda \cdot z_i + \varepsilon_i; \hat{\lambda} = \gamma \cdot \beta \quad (2.3)
\]

LM test based on \( t(\hat{\lambda}) \) has exact size since the RF is a classical linear regression.

Note \( t(\hat{\lambda}) \) 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 \( \gamma \).
Figure 1: Rejection Frequencies for tests of $H_0: \beta = 0$, $N = 100$, $\gamma = .10$.  

![Rejection Frequencies Graph](image-url)
Asymptotic theory does take hold as sample size becomes large, albeit very slowly.

Size for \( t(\hat{\beta}) \) approaches its nominal level only as \( \gamma / \sqrt{V_\gamma} \) approaches 10.

This provides a metric for ‘weak identification’ and a benchmark for standard tests to perform well.

**Table 1: Effect of Sample Size on Distribution of \( \hat{\beta} \) & Size for \( t(\hat{\beta}) \), Orthogonal Regressors.**

<table>
<thead>
<tr>
<th>Sample Size ( N )</th>
<th>100</th>
<th>10,000</th>
<th>1,000,000</th>
<th>10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>True ( \gamma )</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.1</td>
</tr>
<tr>
<td>Asymptotic ( \gamma / \sqrt{V_\gamma} )</td>
<td>.1</td>
<td>1</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Median ( \hat{\beta} )</td>
<td>.10</td>
<td>.01</td>
<td>-.00</td>
<td>.00</td>
</tr>
<tr>
<td>Range ( (.25, .75) )</td>
<td>(-.95, 1.17)</td>
<td>(-.64, .63)</td>
<td>(-.07, .07)</td>
<td>(-.07, .07)</td>
</tr>
<tr>
<td>Size for ( t(\hat{\beta}) )</td>
<td>.0001</td>
<td>.0006</td>
<td>.043</td>
<td>.045</td>
</tr>
</tbody>
</table>
How well does the LM work out in practice

when \( g(.) \) is not linear and the test is not expected to have correct size?
Non-linear Regression: A Production Function.

Consider Hicks-neutral Cobb-Douglas production function:

\[ y_i = \gamma \cdot x_i^\beta + \varepsilon_i; \gamma \neq 0 \]  \hspace{1cm} (3.1.1)

with linear approximation

\[ y_i = \gamma \cdot x_i^{\beta_0} + \lambda \cdot x_i^{\beta_0} \log(x_i) + e_i \]  \hspace{1cm} (3.1.2)

where \( \lambda = \gamma \cdot (\beta - \beta_0) \).

We expect \( \hat{\beta} \) and the size of \( t(\hat{\beta}) \) to be biased in directions indicted by the correlation between

\[ x_i^\beta \] and \( x_i^\beta \log(x_i) \), corresponding to \( g(\beta, x_i) \) and \( g_\beta(\beta, x_i) \).

LM test based on \( t(\hat{\lambda}) \) expected to have close to correct size.
Table 4: Small Sample Distribution of $\hat{\beta}$ and Test Size, True $\gamma = .01$, $N = 100$.

<table>
<thead>
<tr>
<th>True $\beta$</th>
<th>0</th>
<th>.1</th>
<th>.5</th>
<th>.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{g(\beta),g_\gamma(\beta)}$</td>
<td>.07</td>
<td>.29</td>
<td>.77</td>
<td>.92</td>
</tr>
<tr>
<td>Asymptotic $\gamma/\sqrt{V_{\hat{\beta}}}$</td>
<td>.10</td>
<td>.10</td>
<td>.09</td>
<td>.11</td>
</tr>
<tr>
<td>Median $\hat{\beta}$</td>
<td>-0.04</td>
<td>-0.09</td>
<td>-0.05</td>
<td>0.12</td>
</tr>
<tr>
<td>Range (.25, .75)</td>
<td>(-.53, .50)</td>
<td>(-.59, .42)</td>
<td>(-.56, .48)</td>
<td>(-.41, .71)</td>
</tr>
<tr>
<td>Size $t(\hat{\beta})$</td>
<td>0.027</td>
<td>0.037</td>
<td>0.114</td>
<td>0.179</td>
</tr>
<tr>
<td>Size $t(\hat{\lambda})$</td>
<td>0.053</td>
<td>0.054</td>
<td>0.054</td>
<td>0.054</td>
</tr>
</tbody>
</table>
Table 5: Small Sample Distribution of $\hat{\beta}$ and Test Size, $N = 100$, true $\beta = .5$

<table>
<thead>
<tr>
<th>True $\gamma$</th>
<th>.01</th>
<th>.1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptotic $\gamma/\sqrt{V_{\hat{\gamma}}}$</td>
<td>.09</td>
<td>.91</td>
<td>9.10</td>
</tr>
<tr>
<td>Median $\hat{\beta}$</td>
<td>-.05</td>
<td>.27</td>
<td>.50</td>
</tr>
<tr>
<td>Range (.25, .75)</td>
<td>(-.56, .48)</td>
<td>(-.26, .64)</td>
<td>(.46, .54)</td>
</tr>
<tr>
<td>Size $t(\hat{\beta})$</td>
<td>.114</td>
<td>.103</td>
<td>.052</td>
</tr>
<tr>
<td>Size $t(\hat{\lambda})$</td>
<td>.054</td>
<td>.054</td>
<td>.054</td>
</tr>
</tbody>
</table>

NB: Case $\gamma = 0$ corresponds to failure of identification condition, asymptotic theory underling the standard error and $t$-statistic not valid. However, the LM test does not depend on that assumption. Empirical size is 0.054.
Figure 2: Rejection Frequencies for the test $H_0 : \beta = .5$, $N = 100$, True $\gamma = .1$. 

![Graph showing rejection frequencies for different beta values.](image)
The ARMA \((1,1)\) Model.

\[
y_t = \phi \cdot y_{t-1} + \varepsilon_t - \theta \cdot \varepsilon_{t-1}; t = 1, \ldots, T
\]

\[
\varepsilon_t \sim i.i.d. N(0, \sigma^2 \varepsilon), |\phi| < 1, |\theta| < 1
\]  

(3.2.1)

\[
y_t = \gamma \cdot g(\theta, \bar{y}_{t-1}) + \varepsilon_t
\]  

(3.2.2)

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

Ansley and Newbold (1980) and NS show that when \(\gamma\) is small, the standard error for \(\hat{\theta}\) is too small and the standard test based on \(t(\hat{\theta})\) rejects too often.
Linearize the $g(.)$ around the null $\theta = \theta_0$:

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

where $g_\theta(\theta, \bar{y}_{t-1}) = \frac{\partial g(\theta, \bar{y}_{t-1})}{\partial \theta} = \sum_{i=2}^{\infty} (i - 1) \cdot \theta^{i-2} y_{t-i}$, $\lambda = \gamma \cdot (\theta - \theta_0)$.

If the null $\theta = \theta_0$ is correct, the second term should not be significant.

Equivalent to testing second lag in an AR(2) regression, and is approximately Box-Ljung $Q$-test with one lag from an AR(1) regression.

The estimated size of the reduced form test is correct within sampling error.

Case $\gamma = 0$ fails identification but LM test still well defined and size close to correct.
Table 6: Effect of $\gamma$ on Inference for ARMA (1,1), True $\theta = 0$, $T = 1,000$. 

<table>
<thead>
<tr>
<th>True $\gamma (= \phi)$</th>
<th>.01</th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptotic $\gamma / \sqrt{\hat{\gamma}}$</td>
<td>.32</td>
<td>3.16</td>
<td>6.32</td>
<td>9.49</td>
</tr>
<tr>
<td>Median $\hat{\theta}$</td>
<td>-.02</td>
<td>-.01</td>
<td>-.00</td>
<td>-.00</td>
</tr>
<tr>
<td>Range (.25, .75)</td>
<td>(-.65, .64)</td>
<td>(-.26, .24)</td>
<td>(-.11, .11)</td>
<td>(-.07, .07)</td>
</tr>
<tr>
<td>Size $t(\hat{\theta})$</td>
<td>0.46</td>
<td>0.22</td>
<td>0.11</td>
<td>0.07</td>
</tr>
<tr>
<td>Size $t(\hat{\lambda})$</td>
<td>0.051</td>
<td>0.052</td>
<td>0.053</td>
<td>0.052</td>
</tr>
</tbody>
</table>
Figure 3: Histogram of $\hat{\phi}$ in the Monte Carlo. True $\gamma = .01$, $\theta = 0$, $T = 1,000$. 
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 $\gamma = .01$, $\theta = 0$, $T = 1,000$. 
Table 7: Sample Size and Inference in the ARMA (1, 1), True \( \theta = 0 \).

<table>
<thead>
<tr>
<th>Sample size</th>
<th>100</th>
<th>1000</th>
<th>10,000</th>
<th>10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>True ( \gamma(= \phi) )</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.1</td>
</tr>
<tr>
<td>Asymptotic ( \gamma / \sqrt{V_\hat{\gamma}} )</td>
<td>0.1</td>
<td>0.32</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Median ( \hat{\theta} )</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.00</td>
</tr>
<tr>
<td>Range (.25, .75)</td>
<td>(-.69, .67)</td>
<td>(-.65, .64)</td>
<td>(-.58, .55)</td>
<td>(-.07, .07)</td>
</tr>
<tr>
<td>Size ( t(\hat{\theta}) )</td>
<td>.483</td>
<td>.458</td>
<td>.399</td>
<td>.066</td>
</tr>
<tr>
<td>Size ( t(\hat{\lambda}) )</td>
<td>.051</td>
<td>.051</td>
<td>.049</td>
<td>.048</td>
</tr>
</tbody>
</table>
Often AR root $\phi$ of greater economic interest since it measures persistence.

LM test is easily extended to AR parameter and results are qualitatively the same as MA.


Ma (2007) finds that estimated ARMA(1,1) implies small $\gamma$ relative to sampling error and explores possible test size distortion.

The reduced-form test can also be generalized to ARMA model of arbitrary order.

For ARMA(2,2) with $\phi_1 = 0.01, \phi_2 = 0.01, \theta_1 = 0, \theta_2 = 0$ and $T = 100$

standard $t$-test for $\hat{\theta}_1$ and $\hat{\theta}_2$ have sizes 0.571 and 0.698, vs. 0.049 and 0.049.
The Unobserved Component Model for Decomposing Trend and Cycle

Unobserved Component model of Harvey (1985) and Clark (1987):

\[ y_t = \tau_t + c_t, \quad (3.3.1) \]

Trend assumed to be random walk with drift:

\[ \tau_t = \tau_{t-1} + \mu + \eta_t, \eta_t \sim i.i.d. N(0, \sigma^2_\eta), \quad (3.3.2) \]

and cycle a stationary AR:

\[ \phi(L)c_t = \varepsilon_t, \varepsilon_t \sim i.i.d. N(0, \sigma^2_\varepsilon). \quad (3.3.3) \]

In practice, largest AR root is estimated close to unity, implying the cycle is very persistent, and the trend variance is estimated to be very small, implying that the trend is very smooth.
We focus on the case that cycle is AR(1).

Following Morley, Nelson and Zivot (2003), univariate representation is ARMA(1,1):

\[(1 - \phi L)\Delta y_i = \mu(1 - \phi) + (1 - \phi L)\eta_i + \varepsilon_i - \varepsilon_{i-1} = \mu(1 - \phi) + u_i - \theta u_{i-1} \tag{3.3.4}\]

MA parameter \(\theta\) is identified (under the restriction \(\sigma_{n,\varepsilon} = 0\)) by matching autocovariances:

\[\psi_0 = (1 + \phi^2)\sigma_\eta^2 + 2\sigma_\varepsilon^2 + 2(1 + \phi)\sigma_{n,\varepsilon}^2 = (1 + \theta^2)\sigma_u^2 \tag{3.3.5}\]

\[\psi_1 = -\phi\sigma_\eta^2 - \sigma_\varepsilon^2 - (1 + \phi)\sigma_{n,\varepsilon}^2 = -\theta\sigma_u^2 \tag{3.3.6}\]

Solve for unique \(\theta\) by imposing invertibility:

\[
\theta = \frac{(1 + \phi^2) + 2\left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2}\right) + 2(1 + \phi)\rho_{\eta,\varepsilon}\left(\frac{\sigma_\varepsilon}{\sigma_\eta}\right) - \sqrt{\left[(1 + \phi^2) + 4\left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2}\right) + 4(1 + \phi)\rho_{\eta,\varepsilon}\left(\frac{\sigma_\varepsilon}{\sigma_\eta}\right)\right]\cdot[(1 - \phi)^2]}}{2[\phi + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2} + (1 + \phi)\rho_{\eta,\varepsilon}\left(\frac{\sigma_\varepsilon}{\sigma_\eta}\right)]} \tag{3.3.7}\]
It is straightforward that $\theta$ becomes arbitrarily close to $\phi$ as $\frac{\sigma_\varepsilon}{\sigma_\eta}$ approaches zero.

Monte Carlo experiment: $\mu = 0.8, \phi = 0, \sigma_\eta^2 = 0.95, \sigma_\varepsilon^2 = 0.05$, $T=200$.

Almost all the variation due to trend while cycle is small with no persistence at all.

The standard $t$-test for $\phi$ indeed rejects the null much too often; size is 0.481.

Standard error for $\hat{\phi}$ is underestimated; the median is 0.29 compared true value 1.48.

Further, $\hat{\phi}$ is upward biased, its median being 0.58, often close to the positive boundary; Fig 5.

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.
Also, the cycle innovation variance estimate is upward biased, while the trend innovation variance estimate is instead downward biased.

Persistence in estimated cycle tends to occur in samples that also show large variance in the cycle.

Why?

Model must account for the small amount of serial correlation in data generating process.

Roughly, \[ \frac{1 - \phi}{1 + \phi} \cdot \sigma_\varepsilon^2 = -0.05. \]

One solution is the combination of true values, \( \phi = 0; \sigma_\varepsilon^2 = 0.05 \),

but another is \( \phi = 0.9; \sigma_\varepsilon^2 = 0.95 \).

Large negative values of \( \hat{\phi} \) are possible but infrequent since variance non-negative.
LM test: 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 $\tilde{u}$, in the reduced-form ARMA(1,1).

Size of reduced-form test for $\phi$ is 0.054.

When all variation is due to stochastic trend, i.e., $\sigma_\epsilon^2 = 0$, identification for $\phi$ fails.

However, LM test works well with size 0.058.

LM test can be generalized to UC model with higher AR order.
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

$\mu = 0.8, \phi = 0, \sigma_\eta^2 = 0.95, \sigma_\varepsilon^2 = 0.05$
The GARCH(1,1) Model.

The GARCH(1,1) model of Bollerslev (1986):

\[ \varepsilon_t = \sqrt{h_t} \cdot \xi_t, \xi_t \sim i.i.d. N(0,1) \]  \hspace{1cm} (3.4.1)

\[ h_t = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot h_{t-1} \]  \hspace{1cm} (3.4.2)

Analogy to the ARMA (1,1) model:

\[ \varepsilon_t^2 = \omega + (\alpha + \beta) \cdot \varepsilon_{t-1}^2 + w_t - \beta \cdot w_{t-1} \]  \hspace{1cm} (3.4.3)

Ma, Nelson and Startz (2007) show that when \( \alpha \) is small relative to its sampling error, 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.
State Space form of GARCH:

State equation describes the dynamic evolution of unobserved volatility:

\[ h_t = \omega + (\alpha + \beta) \cdot h_{t-1} + \alpha \cdot w_{t-1} \]  

(3.4.4)

\[ w_{t-1} = \varepsilon_{t-1}^2 - h_{t-1} . \]

The measurement equation is simply:

\[ \varepsilon_t^2 = h_t + w_t \]  

(3.4.5)

State equation shock (\( \alpha \cdot w_{t-1} \)) is one period lag of measurement equation shock (\( w_t \)).

The essence of SS model is to filter out noise (\( w_t \)) to extract signal (\( h_t \)).

Parameter \( \alpha \) reflects size of the signal shock relative to that of noise.

When signal is small relative to noise, uncertainty about signal is large and identification is weak.
LM test: defining \( g(\beta, \varepsilon_{t-1}^2) = \sum_{i=1}^{\infty} \beta_{t-i}^2 \varepsilon_{t-i}^2 \) and \( \varepsilon_{t-1}^2 = (\varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots) \) one obtains:

\[
h_t = \frac{\omega}{1 - \beta} + \alpha \cdot g(\beta, \varepsilon_{t-1}^2) \quad (3.4.4)
\]

Linear expansion of nonlinear \( g(.) \) around the null,

defining \( c = \frac{\omega}{1 - \beta}, \quad \lambda = \alpha \cdot (\beta - \beta_0) \) and \( g_\beta(\beta, \varepsilon_{t-1}^2) = \sum_{i=2}^{\infty} (i-1) \beta_{t-i}^2 \varepsilon_{t-i}^2 \):

\[
h_t = c + \alpha \cdot g(\beta_0, \varepsilon_{t-1}^2) + \lambda \cdot g_\beta(\beta_0, \varepsilon_{t-1}^2) \quad (3.4.5)
\]

The LM test then uses \( t(\hat{\lambda}) \) to test the null \( \lambda = 0 \) in (3.4.5).
Table 8: Reduced form and standard $t$-tests for GARCH(1,1): True $\beta = 0$, $T = 1,000$.

<table>
<thead>
<tr>
<th>True $\gamma (= \alpha)$</th>
<th>.01</th>
<th>.05</th>
<th>.1</th>
<th>.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptotic $\gamma/\sqrt{V_\gamma}$</td>
<td>0.32</td>
<td>1.59</td>
<td>3.19</td>
<td>6.60</td>
</tr>
<tr>
<td>Median $\hat{\beta}$</td>
<td>0.33</td>
<td>0.08</td>
<td>-0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>Range (.25, .75)</td>
<td>(-0.30, 0.74)</td>
<td>(-0.31, 0.49)</td>
<td>(-0.22, 0.22)</td>
<td>(-0.11, 0.09)</td>
</tr>
<tr>
<td>Size $t(\hat{\beta})$</td>
<td>0.470</td>
<td>0.344</td>
<td>0.198</td>
<td>0.106</td>
</tr>
<tr>
<td>Size $t(\hat{\lambda})$</td>
<td>0.078</td>
<td>0.074</td>
<td>0.076</td>
<td>0.096</td>
</tr>
</tbody>
</table>
For the case $\alpha = 0$ identification fails and standard $t$-test does not have usual asymptotic distribution. The reduced-form test, however, is still valid and has size of 0.076 for true $\beta = 0$ and $T = 1,000$.

The sum $\alpha + \beta$ is of potentially greater economic interest.

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.

Similar results follow in this case; see Ma (2007) for further discussion.

GARCH estimates with Bollerslev and Wooldridge’s (1992) robust standard errors:

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

Implies a significant and large GARCH effect,

95% confidence interval for \( \hat{\beta} \): [0.44, 1).

Monte Carlo suggests small value of \( \hat{\alpha} \) relative to sample size. Does it matter?

To obtain confidence interval we numerically invert the LM test statistic,

create a grid of \( \beta_0 \)'s, compute the corresponding \( t(\hat{\lambda}) \), and plot the latter against the former.

Resulting 95% confidence interval is [-0.95, 0.87], which covers almost entire parameter space.

Consistent with ‘impossibility theorems’ of Dufour (1997) for weakly identified models.
Figure 7: The 95% Confidence Interval for $\hat{\beta}$ based on the LM test for Monthly S&P 500 stock return data
Summary and Conclusions

While the three classical tests, Wald LR and LM, are equivalent asymptotically, they may produce *dramatically* different results under weak identification.

Size matters!

Standard Wald test size depends on model details, the DGP and nuisance parameters.

Size of modified LM test in spirit of Harvey (1990) is, in contrast, robustly close to correct,

This is because it is exact when the linear approximation to $g(.)$ used in LM is exact.

Its computational convenience and robustness have been illustrated here in four models that are widely employed in applied econometrics.

THANK YOU ANDREW!