607
LM tests, diagnostic testing

Jon Faust

http://e105.org/e607

November 9, 2015

- **Miscellaneous testing**
  - There are two main reasons to run tests
    - 1: To shed light on question of interest
    - 2: To shed light on auxiliary issues such as the validity of auxiliary assumptions.
      Auxiliary issues are not of primary interest, but may be relevant to interpreting the statistics for questions of interest
  - These tests are often called diagnostic tests
    Diagnostic here refers to the purpose the test is run. From a generic standpoint, these tests are the same as any statistical test.
- **Auxiliary tests, examples**
  - If you assume that residuals in a regression are Gaussian, or not serially correlated, or homoskedastic, you might test to see if these assumptions appear to be roughly sastisfied.
  - You do the test not because the question is of inherent interest but because the validity of later inferences may depend on whether these assumptions are roughly satisfied.
- **Thus,**
  - Diagnostic tests are tests of features of the data that may affect the small or large sample properties of your inferences of fundamental interest.
  - Everything we have said about tests in general applies here.
- **Quick review: Wald test**

- Standard case,

$$\sqrt{T}(\hat{\beta} - \beta) \overset{a}{\sim} N(0, V)$$

  for some V which we can consistently estimate by, say, $\hat{V}$

- Suppose null is $H_0 : R\beta = r$
- Under null,

$$\sqrt{T}(R\hat{\beta} - r) \overset{a}{\sim} N(0, RVR')$$

- Wald statistic is,

$$W = T(R\hat{\beta} - r)(R\hat{V}R')^{-1}(R\hat{\beta} - r)$$

▶ **Wald test**

- Under null,

$$W \overset{a}{\sim} \chi^2_{(p)}$$

  p is number of (independent) restrictions.

- We reject the null if, $W > c$
- Size of the test is

$$\sup pr(W > c)$$

  where sup is over all models in which null is true.

- Usually don't know size, so we approximate it by *nominal size* chosen in some standard ways: conventional asymptotics and bootstraps.

▶ **LM tests**

- Diagnotic tests often formed as LM tests
- Why does this seem sensible?
- In LM tests, we only estimate under the restricted model.
- In the case of diagnostic tests of maintained assumptions, this means we only estimate under the maintained assumption.
- We may know how to estimate the model well udner the maintained assumptions, but it may be more complicated to do so dropping those assumptions.
- Thus, LM tests may be easier than Wald or LR tests in this context.

▶ **Aside:: Well, actually**

- In simple linear regression contexts, everything is pretty easy and this advantage of the LM test is really moot.

- But more generally, this advantage can be a big deal.

▶ **Thus in diagnostic testing using LM tests, we estimate the model under the maintained assumption and test for symptoms that our maintained assumptions don't hold very well.**

▶ **Terminology and LM tests in ML estimation**

- In ML estimation, we try to set the score to zero.
- If there are overidentifying restrictions, we won't be able to set the score exactly to zero

  We will get, e.g.,

$$\partial \mathcal{L}(\hat{\theta}^{MLE})/\partial\theta \equiv q(\hat{\theta}^{MLE}) = \lambda \ldots$$

- The LM test in ML context checks how close the score is to zero

  If it is too far from zero, we take this as evidence that the restrictions are costly. That is, the assumption that the restrictions hold in reality is implausible.

- Thus, LM tests are often called score tests or efficient score tests.

▶ **LM tests in LS probs.**

- Suppose the restricted model is

$$Y = X\beta + \varepsilon$$

  where β is $K \times 1$, and the more general, unrestricted model is

$$Y = X\beta + Z\gamma + \varepsilon$$

  where γ is $p \times 1$.

- We could formulate the estimation of the model under the restriction that $\gamma = 0$ as a Lagragean problem minimizing the sum of squares under the general model and imposting the restriction.
- You should be able to see that the $\hat{\beta}$ that will emerge is simply the $\hat{\beta}$ from estimating the restricted model: $\hat{\beta} = (X'X)^{-1}X'Y$.
- But to understand the LM test, let's do it the cumbersome way with the Lagrangean.

▶ **The Lagrangean**

- The general model

$$\begin{aligned} Y &= X\beta + Z\gamma + \varepsilon \\ Y &= W\theta + \varepsilon \end{aligned}$$

  where $W = [X : Z]$ and $\theta = (\beta', \gamma')'$.

We'll assume X includes a constant for simplicity.

▶ **Lagrangean**

- OLS on the restricted model is equivalent to restricted least squares:

$$\begin{aligned}\hat{\theta}^r &= \underset{\theta}{\text{argmin}}\, SSR(\theta) \\ &= \underset{\theta}{\text{argmin}} (Y - W\theta)'(Y - W\theta)\end{aligned}$$

s.t. $R\theta = r$

$R\theta = r$ imposes $\gamma = 0$

▶ **Lagrangean**

- The Lagrangean:

$$\hat{\theta}^r = argmin_\theta SSR(\theta) + (R\theta - r)'\lambda$$

where λ is a $(p \times 1)$ vector of Lagrange multipliers (γ is $p \times 1$).

▶ **Lagrangean**

- The F.O.C. says to choose $\hat{\theta}$ to satisfy:

$$\begin{aligned}\partial SSR(\hat{\theta})/\partial\theta &= R'\lambda \\ W'\hat{\varepsilon} &= R'\lambda\end{aligned}$$

$\hat{\varepsilon}$ are the residuals when $\hat{\theta}$ is the parameter.

▶ **Aside:: Intuition**

- In unrestricted OLS, we choose the parameters to force the sample covariance between the $\hat{\varepsilon}$s and the regressors to be zero.
- Under restrictions, we make this covariance as small as possible.

▶ **Back to the main argument**

- We have that

$$W'\hat{\varepsilon} = R'\lambda$$

- To test whether $\lambda = 0$ in population, we can test whether $W'\hat{\varepsilon} \approx 0$.
- That is, we are testing whether the regressors (all of them, not just Zs) are orthogonal to $\hat{\varepsilon}$.
- But how do we test whether some single variable is orthogonal to a vector of variables W?

The natural way is to run a regression of the variable on the others and perform a Wald test whether all the coefficients are jointly zero.

- ▶ **The LM test**
  - LM test of the restrictions can be seen as a Wald test on an auxiliary regression.

    Specifically, a test that all coefficients on stochastic variables are zero in an auxiliary regression of $\hat{\varepsilon}$ on W.
  - The standard χ^2 variant of the Wald test works fine or equivalently we can use the regression F test that all the coefficients are zero.
- ▶ **Aside:: Notice 1**
  - The auxiliary regression has all the regressors, Xs and Zs.

    Any association of the Zs with the εs may only be apparent when the Xs are included.
- ▶ **Aside:: Notice 2:**
  - The asymp. dist. of Wald test in this auxiliary regression has d.f. equal to p, the number of γs, not the number of stochastic regressors in the auxiliary regression $(p + K - 1)$.
  - Hint: We know at the outset that the LHS variable, $\hat{\varepsilon}$ is orthogonal to the Xs.

    We really have only p new degrees of freedom consumed.
  - Deeper: the χ^2 result comes from the quadratic form of the Wald statistic, say, $h'Ah$ where h is mean zero and normal and A is idempotent. The rank of A determines the d.f. If you do the linear algebra in this case, you will see where the hint gives rise to a quadratic form where the rank is p
- ▶ **A useful observation**
  - The Wald statistic on this auxiliary regression can be written as a simple function of the R^2 of the auxiliary regression:

    $$TR^2/(1 - R^2)$$

    You should be able to verify this.
  - Under the null, the denominator has a plim of 1 (the R^2 goes to zero).
  - Thus,

    $$TR^2$$

    has the same asymptotic distribution as $TR^2/(1 - R^2)$.
  - Thus, one often sees reference to LM tests of this type being TR^2 tests.

▶ **Aside:: Useful observation, cont.**

- Alternatively, we could use the standard regression F test to check if all coefs (but the const) are zero.
- This test in this context can be written:

$$LMF = \frac{T}{p}\frac{R^2}{1-R^2}$$

- We can take this to be $F_{(p,T-(p+k))}$ where k is the number of Xs.

  This is often called the F-form of the Wald statistic.

- Remember (e.g., from the problem sets), the relation between F and χ^2 distributions when the denominator degrees of freedom in the F go to infinity and the numerator degrees of freedom are fixed.
- The F form is generally preferred on grounds of accuracy of implied nominal size.

  The F form is conservative relative to the χ^2 form.

▶ **General result**

▶ **General result**

- Whenever you want a diagnostic for a problem that can be cast as adding regressors to the restricted model the TR^2 logic can be applied.
- Now some concrete cases.

▶ **Example 1: serial correlation**

- Reminder:The restricted and gen. models:

$$\begin{aligned} Y &= X\beta + \varepsilon \\ Y &= X\beta + \gamma Z + \varepsilon \end{aligned}$$

- For autocorr, the Zs are lags 1 through p of the $\hat{\varepsilon}$s.

  Thus, for the aux. regn. run $\hat{\varepsilon}$ on the original Xs and lags 1 through p of the $\hat{\varepsilon}$s.

- The TR^2 is Godfrey's test for serial correlation of the εs.
- This test will be consistent against both MA and AR-type correlation in the underlying model, see Greene or Hamilton

▶ **Example 2: General nonlinearity**

- Suppose general model is,

$$y_t = f(X'_{t.}\beta) + \varepsilon_t$$

- If f is smooth, we can expand around $X'_t\beta = 0$:

$$y_t = \gamma_0 + \gamma_1 X'_{t.}\beta + \sum_j \gamma_{j=2}(X'_{t.}\beta)^j + \varepsilon_t$$

Where the γ_js, $j > 0$, are the relevant partial derivatives

- (Note: if X has a constant, we can drop the γ_0, folding it into the γ_1 term.)
- Under the null that the model is linear, we collapse to,

$$y_t = \gamma_1 X'_{t.}\beta + \varepsilon_t$$

- The auxiliary regressors in the general model are:

$$Z_j = (X'_{t.}\hat{\beta})^j \quad j = 1, \ldots, \bar{j}$$

- Note that under the null, the predicted value of y from the regression is,

$$\hat{y}_t = X'_{t.}\hat{\beta}$$

- Thus, as auxiliary regressors we can use,

$$Z_j = \hat{y}^j$$

- The TR^2-style LM test in this case is called the RESET test
- In large samples, will be appropriate against any smooth form of nonlinearity.

▶ **Note:**

- In Godfrey's test we had to choose the number of lags to include.
- In the reset test, we have to choose the number of powers of $\hat{y}$ to include
- We will return to this choice below.

▶ **Example 3: Heterosked. related to Xs**

- The only heterosked. that makes the conventional OLS variance-covariance matrix inconsistent is when the variance of ε is related to the Xs.
- White proposed and LM-style test based on auxiliary regression:

$$\hat{\varepsilon}^2 = X\xi_x + Z\xi_z + \varepsilon$$

where the columns of Zs are the squares and cross products of the Xs.

- This directly tests whether the squared residuals, our proxy for the true εs, are associated with levels and squares of regressors.

▶ **Note about d.f. in White's test**

- Unlike the other LM tests we have described, in this case, the test is $\chi^2_{(k)}$ where k is number of all regressors in the auxiliary regression except the const.
- Why? We are running the regression on the $\hat{\varepsilon}$s-squared. While the $\hat{\varepsilon}$s are orthogonal to the xs, the $\hat{\varepsilon}$s-squared need not be.

▶ **A practical comment on LM diagnostics**

▶ **Practical diagnostic testing**

- This isn't rocket science: run the auxiliary regression with the original regressors and some stuff that might come in under some more general model.
- As noted above, these tests involve a variable selection problem for the auxiliary regression

  What extra stuff do you put in the auxiliary regression?

▶ **Who cares?**

- In the limit, adding regressors must lower power of the test.

  so long as size is controlled

- As number of additional regressors goes to number of observations, R^2 in the auxiliary regression goes to 1 whether or not the null is true.

  Thus, either your size becomes distorted from the desired level or power goes down (converging to size)

▶ **But of course,**

- Leaving out potentially important auxiliary variables also lowers power.

  Why? If you leave out variables that are in fact important, you obviously may not discover the problem.

- Thus, we are in one of those cases where the best advice is: 'Choose wisely.'
- Aside: Essentially we face the same tradeoff here as faced when picking moment conditions in GMM

  leaving out important ones is bad, but so is using too many

▶ **How to respond to diagnostic results**

▶ **What do you do about a failed diagnostic**

- Answer 1: Robustify.
- Be sure your methods are robust to the problems you detect

  e.g. use a HAC estimate of the variance-covariance matrix of coefficients if you find H or AC of the residuals.

- Answer 2: re-specify your model in some way to 'take account of' the problem you discovered.
- Once again we face the question of whether we robustify or move to a more complicated estimator that may fix the problem.

▶ **Do we really care about diagnostic tests**

- One view says, don't bother to test, always robustify.

  Robustify, don't test

- Thus, many macroeconometricians tend robustify and report few auxiliary tests.
- We might call this the "see no evil—but robustify just in case" approach.
- Before discussing this position, let me sketch polar opposite.

▶ **The Hendry or LSE method**

- The Hendry or LSE approach is the polar opposite
- Dennis Sargan at the LSE was key to developing this view, and David Hendry of the LSE and subsequently of Oxford pushed it much further.
- Key idea: before addressing question of interest, construct a parsimonious model that passes a large battery of diagnostic tests

▶ **Thus,**

- Don't proceed to question of interest until you have what, according to your tests, is a well-behaved statistical model.
- After you have this model, assess the question of interest in much the same way as in other approaches.

▶ **In practice,**

- In practice many folks end up somewhere on the continuum between the two polar cases.
- For a discussion of my view on the Hendry method, see,

Faust and Whitman, General-to-Specific Procedures for Fitting a Data-Admissible, Theory-Inspired, Congruent, Parsimonious, Encompassing, Weakly-Exogenous, Identified, Structural Model to the DGP: A Translation and Critique. go

http://www.sciencedirect.com/science/article/B6V8D-3SX8BK6-7/2/062f330eb42b74795a4aaba2c5032c97