ECOM40006/ECOM90013 Econometrics 3
Department of Economics
University of Melbourne

GMM-Based Tests of Over-Identifying Restrictions

Semester 1, 2025

Version: May 12, 2025

We will skip over this area relatively lightly because you already know the basics of it. These basics are the same as those for maximum likelihood. Here are the main principles that you should understand.

1. Our estimators are formed my optimizing some estimation criterion. In maximum likelihood we maximize the likelihood. In least squares, be it ordinary or generalized least squares, in instrumental variables estimation we explicitly minimize a sum of squares residuals. In GMM we we implicitly minimize a quadratic form, that often reduces to what looks like a (weighted) sum of squared residuals. By the way, in a Normal world the likelihood maximization process typically reduces to minimizing a sum of squared residuals, at least for the regression coefficients.

2. The optimization process, in a regular problem produces some first-order conditions that we need to solve in order to obtain our estimator. These first-order conditions have essentially the same structure as do the sample moment conditions of a GMM estimator. These *estimating equations* can typically be solve analytically of they are linear in form. Where they are non-linear they will need to be solved numerically. As an aside, many techniques for solving non-linear equations essentially reduce to using a first-order Taylor approximation to linearize the equations and then proceed iteratively from there.

3. When it comes time to test hypotheses there are 3 primary approaches.

   (a) We can compare restricted and unrestricted models on the basis of the optimized value of the estimation criterion. In the context of likelihood theory this leads to the likelihood ratio test. Everywhere else it reduces to a likelihood ratio-like test.

   (b) We can compare the values of the estimates themselves. In a likelihood framework this yields Wald tests. Everywhere else you can think of them as Wald-like tests.

   (c) Finally, we can check that unrestricted estimating equation (the score in the likelihood world) still equals zero when evaluated at the restricted parameter estimates. Clearly, this idea leads to score tests or Lagrange multiplier tests in the likelihood-world. We can think of them as LM-like tests everywhere else.

The interesting thing about our test procedures is that they produce essentially the same results when testing linear restrictions in linear models. For example, if testing that a single regression coefficient is equal to some number then all three of them can reduce to a something like a t-test (The LM test uses a slightly different variance estimate to the other two, but they are identical in this situation.) If we restrict ourselves to asymptotic arguments then, when the null hypothesis is true, then we compare that statistic to critical values from a standard Normal distribution. If we choose to square that statistic then we compare it to a chi-square distribution.[1] If you have more restrictions to test then the t-type tests are no longer appropriate and all procedures lead to something that is asymptotically chi-squared under the null.

Turning attention to the '-like' tests, we see that the estimators that these tests are based upon have similar asymptotic distributions to our maximum-likelihood estimators. Consequently, we can follow similar arguments to generate similar tests based upon these new estimators. For instance LM-like tests can typically be generated by any number of auxiliary regressions, often as nR^2 statistics,[2] although this is contingent upon using an opg form of the information matrix (or equivalent for a non-mle estimator).[3] Wald-like tests reduced to a t statistic (estimator divided by estimated standard error) in the case of a single restriction or quadratic forms in asymptotically Normal random variables otherwise. Finally, LR-like tests can be based on the implicit estimation criterion being optimize by, say the GMM estimator. For example, we can evaluate

$$S_n^{GMM}(\beta) = (y - X\beta)'W(W'\Omega W)^{-1}W'(y - X\beta). \tag{1}$$

at

$$\tilde{\beta}_n^{GMM} = (X'W(W'\Omega W)^{-1}W'X)^{-1}X'W(W'\Omega W)^{-1}W'y. \tag{2}$$

and, say, the HAC estimator for Ω.[4] As the HAC estimator is consistent, but like $\hat{\sigma}^2$ will converge at a faster rate than $\tilde{\beta}_n^{GMM}$, or rather its feasible counterpart $\tilde{\beta}_n^{FGMM}$,

$$S_n^{FGMM}(\tilde{\beta}_n^{FGMM}, \widehat{\Omega}_{HAC}) = (y - X\tilde{\beta}_n^{FGMM})'W(W'\widehat{\Omega}_{HAC}W)^{-1}W'(y - X\tilde{\beta}_n^{FGMM}) \tag{3}$$

will have the same asymptotic distribution as you would have if you know Ω, which will be χ^2 with some degrees of freedom.

Finally, let us think briefly about what to do in models with over-identifying restrictions. Our treatment is based on that of Davidson and MacKinnon (2004, Section 8.6). Suppose that you have an over-identified model, which comes about when you have more estimating equations than you have parameters to estimate. In terms of the notation from earlier we are thinking about the case where $p > k_2$. It transpires that you can test some, but not all of these restrictions.

By way of analogy, if we go back to a very simple model of estimating a sample mean and sample variance, in order to calculate a variance we had first to estimate a sample mean. After that process we were left with only $n - 1$ degrees of freedom because

[1] Both of theses null distributions are only asymptotic approximations and you would be well within your right to compare the statistics to the t and F distributions that you originally learned about. They are slightly more conservative approximations in the sense that you need slightly larger values for the statistic in order to reject the null hypothesis.

[2] Important early contributions include Newey (1985) and Tauchen (1985).

[3] Remember that the information matrix is only there because it is the variance of the score. With other estimators simply use the equivalent matrix to weight the quadratic form.

[4] Note that you will need to use the HAC estimator in (2) as well in order to make it a feasible estimator.

we chewed up one of our pieces of sample information in estimating the mean. So s_n^2 is proportional to a chi-squared random variable with $n-1$ degrees of freedom and a t statistic for testing hypotheses about the mean has a t distribution with only $n-1$ degrees of freedom. You can't test the validity of your formula for the mean because everything that you are doing is conditional upon it. It is just a given.

In the case of GMM with $p > k$ moment conditions but only k parameters to estimate, you need to use k of those equations to estimate the parameters and are left with $p-k$ over-identifying restrictions. You can test these over-identifying restrictions but not the k notionally used in the construction of the estimator. Just to be clear, we don't ever pull out a set of k moment conditions, leaving $p-k$ behind, whose validity we can then test. However, conceptually, we can test the impact of the excess restrictions or moment conditions that we do use in estimation. We do this by comparing $S_n^{FGMM}(\tilde{\beta}_n^{FGMM}, \widehat{\Omega}_{HAC})$ with critical values from a χ^2_{p-k} distribution and reject the null hypothesis of the validity of the over-identifying restrictions for large values of the statistic. In the context of GMM, this test has names like Hansen's over-identification statistic or Hansen's J statistic.[5] Davidson and MacKinnon (2004, p.367) refer to it as the Hansen-Sargan test, as it is essentially the same as Sargen's test of over-identifying restrictions in the GIVE context that was introduced many years earlier (Sargan, 1958).[6]

Bibliography

Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods*. Oxford University Press, New York. 3

Newey, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica* 53, 1047–1073. 2

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica* 26(3), 393–415. 3

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30, 415–443. 2

[5] Lars Peter Hansen (1952–), has spent most of his academic career at the University of Chicago. Best known for his work on GMM, he is also a distinguished macroeconomist. He was a recipient of the 2013 Nobel Prize in Economics. (The other recipients that year were Eugene Fama and Robert Shiller.) He was a PhD student of Christopher Albert Sims (1942–) at the University of Minnesota. Sims won the Nobel prize in 2011 along with Thomas John Sargent (1943–).

[6] John Denis Sargan (1924–1996) was a dominant figure in British Econometrics from the late 1950s until the time of his retirement in 1984, which coincided with the period where Britain and, in particular, the London School of Economics, where Sargan worked, was probably the leading Department for econometrics anywhere in the world. Apart from a string of significant papers he also supervised a string of extremely successful PhD students. First and foremost of these is Peter Phillips, but also David Hendry and Manuel Arellano count among their number, to name but a few. Gone but not forgotten, Sargan can be seen in action at http://www.ncer.edu.au/resources/historical-archive.php.