econometrics

MDPI

Short Note

On Using the *t*-Ratio as a Diagnostic

Jan R. Magnus

Department of Econometrics and Operations Research, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands; jan@janmagnus.nl

Received: 8 February 2019; Accepted: 28 May 2019; Published: 29 May 2019

Abstract: The *t*-ratio has not one but two uses in econometrics, which should be carefully distinguished. It is used as a test and also as a diagnostic. I emphasize that the commonly-used estimators are in fact pretest estimators, and argue in favor of an improved (continuous) version of pretesting, called model averaging.

Keywords: *t*-statistic; pretest estimator; model averaging

1. Introduction

The concept of 'statistical significance' appears in almost all scientific papers in order to form or strengthen conclusions, and the *p*-value or the *t*-ratio are commonly used to quantify this concept. Unfortunately, there is a lot of confusion about statistical significance. Almost 25 years have passed since Hugo Keuzenkamp and I wrote on this issue (Keuzenkamp and Magnus 1995), but the confusion persists and does not seem to diminish over time.

Most importantly, despite many warnings in textbooks, there is confusion about the difference between significance and importance. Statistical significance does not imply importance. This and other misuses of the *p*-value were recently well summarized by (Wasserstein and Lazar 2016).

In this note, which draws heavily on my recent undergraduate textbook (Magnus 2017), I concentrate on another (mis)use of the *t*-ratio (or of the *p*-value)—one which is not mentioned in (Wasserstein and Lazar 2016), but also needs attention and warning. This concerns the role of the *t*-ratio as a diagnostic. My aim is to explain that the *t*-ratio has not one but two uses in econometrics, which should be carefully distinguished; to emphasize (again) the difference between significance and importance; to show that the estimators that are used in practice are pretest (or post-selection) estimators (Leeb and Pötscher 2005); and to argue in favor of an improved (continuous) version of pretesting, called model averaging.

2. Two Uses of the *t*-Ratio

The *t*-ratio can be viewed in two ways. We could, for example, be interested in testing the hypothesis that $\beta_j = 0$ in the linear model $y = X\beta + u$. In that case the *t*-ratio t_j can be fruitfully employed, because under certain assumptions (such as normality) t_j follows Student's *t*-distribution under the null hypothesis and if we fix the significance level of the test (say at 5%) then we can reject or not reject the hypothesis.

The *t*-ratio, however, is also commonly employed in a different manner. Suppose we are primarily interested in the value of another β-coefficient, say β_i. Then, t_j is often used as a diagnostic rather than as a test statistic in order to decide whether we wish to keep the jth regressor x_j in the model or not. In this situation the 5% level is also typically used, but why? The two situations are quite different because in the first case we are interested in β_j while in the second case we are interested in β_i. In the first case we ask: Is it true that $\beta_j = 0$? In the second case: Does inclusion of the jth regressor improve the estimator of β_i? These are two different questions and they require different approaches.

3. Significance and Importance

Suppose you are an econometrician working on a problem and some famous expert comes by, looks over your shoulder, and tells you that she knows the data-generation process (DGP). Of course, you yourself do not know the DGP. You use models but you do not know the truth; this expert does. Not only does the expert know the DGP but she is also willing to tell you, that is, she tells you the specification, not the actual parameter values. So now, you actually have the true model. What next? Is this the model that you are going to estimate?

The answer, surprisingly perhaps, is no. The truth, in general, is complex and contains many parameters, nonlinearities, and so on. All of these need to be estimated and this will produce large standard errors. There will be no bias if our model happens to coincide with the truth, but there will be large standard errors. A smaller model will have biased estimates but also smaller standard errors. Now, if we have a parameter in the true model whose value is small (so that the associated regressor is unimportant), then setting this parameter to zero will cause a small bias, because the size of the bias depends on the size of the deleted parameter. Setting this unimportant parameter to zero also means that we don't have to estimate it. The variance of the parameters of interest will therefore decrease, and this decrease does not depend on the size of the deleted parameter. Thus, deleting a small unimportant parameter from the model is generally a good idea, because we will incur a small bias but may gain much precision.

This is true even if the estimated parameter happens to be highly 'significant', that is, has a large t-ratio. Significance indicates that we have managed to estimate the parameter rather precisely, possibly because we have many observations. It does not mean that the parameter is important.

Note the proviso 'if our model happens to coincide with the truth' in the second paragraph. When we omit relevant variables we get biased estimators (which is bad), but a smaller variance (which is good). This, however, is only true when we compare the restricted model with an unrestricted model which coincides with the DGP. If, which is much more likely, we compare two models one of which is small (the restricted model) and the other is somewhat larger (the unrestricted model), but both are smaller than the DGP, then the estimator from the unrestricted model is also biased and, in fact, this bias may be larger than the bias from the restricted model; see (De Luca et al. 2018).

We should therefore omit from the model all aspects that have little impact, so that we end up with a small model—one, which captures the essence of our problem.

4. Pretesting

Let us consider the situation where t_j is used as a diagnostic in more detail. In fact, we have three estimators of β_i: the estimator from the unrestricted model, $\hat{\beta}_{iu}$; the estimator from the restricted model (where $\beta_j = 0$), $\hat{\beta}_{ir}$; and the estimator after a preliminary test,

$$b_i = w\hat{\beta}_{iu} + (1-w)\hat{\beta}_{ir}, \qquad w = \begin{cases} 1 & \text{if } |t_j| > c, \\ 0 & \text{if } |t_j| \leq c, \end{cases}$$

for some $c > 0$, such as $c = 1.96$ or $c = 1$. The estimator b_i is called the pretest estimator.

The estimators $\hat{\beta}_{ir}$ and $\hat{\beta}_{iu}$ are linear and (under standard assumptions) normally distributed, but b_i is nonlinear, because its distribution depends on a random restriction. The pretest estimator is therefore much more complicated than the other two estimators. But it is the pretest estimator that is commonly used in applied econometrics, because in applied econometrics we typically use t- and F-statistics as diagnostics to select the most suitable model. That in itself is not ideal, but what is worse is that we typically ignore the model selection aspect when reporting properties of our estimators.

The pretest estimator is kinked and therefore inadmissible. Its poor features are well-studied; see for example (Magnus 1999). Surely we should be able to come up with an estimator which performs better than the pretest estimator. This is where model averaging comes in.

5. Model Averaging

In (Magnus 2017) I tell the following story.

> A King has twelve advisors. He wishes to forecast next year's inflation and calls each of the advisors in for his or her opinion. He knows his advisors and obviously has more faith in some than in others. All twelve deliver their forecast, and the King is left with twelve numbers. How to choose from these twelve numbers? The King could argue: which advisor do I trust most, who do I believe is most competent? Then I take his or her advice. The King could also argue: all advisors have something useful to say, although not in the same degree. Some are more clever and better informed than others and their forecast should get a higher weight. Which way of thinking is better?

Intuitively most people, and I also, prefer the second method (model averaging), where all pieces of advice are taken into account. In standard econometrics, however, it is the first method (pretesting) which dominates.

There are theoretical and practical problems with the pretest estimator. One practical problem is the property that—if 1.96 is our cut-off point—for $t_j = 1.95$ we would choose one estimator and for $t_j = 1.97$ another, while in fact there is little difference between 1.95 and 1.97. This is not satisfactory.

These and other considerations lead us to reconsider the estimator

$$b_1 = w\hat{\beta}_{iu} + (1 - w)\hat{\beta}_{ir}$$

by allowing w to be a smoothly increasing function of $|t_j|$. This is model averaging in its simplest form, and we see that it is just the continuous counterpart to pretesting. In model averaging we give weight to all models of interest, but not in the same degree, while in pretesting we select one model after a preliminary test, precisely as the King in the story above.

In practice, econometricians use not one but many models. One of these is the largest and one is the smallest. Neither is probably the most suitable for the question at hand. If we use diagnostic tests to search for the best-fitting model, then we need to take into account not only the uncertainty of the estimates in the selected model, but also the fact that we have used the data to select a model. In other words, model selection and estimation should be seen as a combined effort, not as two separate efforts. This is what model averaging does. It incorporates the uncertainty arising from estimation and model selection jointly. Failure to do so may lead to misleadingly precise estimates.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflicts of interest.

References

De Luca, Giuseppe, Jan R. Magnus, and Franco Peracchi. 2018. Balanced variable addition in linear models. *Journal of Economic Surveys* 32: 1183–200.

Keuzenkamp, Hugo A., and Jan R. Magnus. 1995. On tests and significance in econometrics. *Journal of Econometrics* 67: 5–24.

Leeb, Hannes, and Benedikt M. Pötscher. 2005. Model selection and inference: Facts and fiction. *Econometric Theory* 21: 21–59.

Magnus, Jan R. 1999. The traditional pretest estimator. *Theory of Probability and Its Applications* 44: 293–308.

Magnus, Jan R. 2017. *Introduction to the Theory of Econometrics*. Amsterdam: VU University Press.

Wasserstein, Ronald L., and Nicole A. Lazar. 2016. The ASA's statement on *p*-values: Context, process, and purpose. *The American Statistician* 70: 129–33.