Lear
Laboratorio di economia, antitrust, regolamentazione

Lear Competition Note

STATISTICAL SIGNIFICANCE AND THE STANDARD OF PROOF IN ANTITRUST DAMAGE QUANTIFICATION

Nov 2013

Econometric techniques can play an important role in damage quantification cases for breaches of antitrust law. However, lots of care should be devoted to applying these techniques correctly in the context of civil litigation, accounting for the specificities of the legal process.

In this short note, starting from a *prima facie* simple convention on how to interpret the estimated regression coefficients, we will discuss how a deep understanding of both the specific legal framework and the relevant technical aspects (economic principles and statistical techniques) is important in order to correctly evaluate econometric evidence in the context of civil court litigation for antitrust damage quantification. Forensic economists can help lawyers and judges bridge the gap between the legal world and the scientific practice.

Background

On 11 June 2013, the European Commission adopted a "Proposal for a Directive on antitrust damages actions for breaches of EU competition law" ("Directive")[1] and published a working document titled "Practical Guide on Quantifying Harm" ("Guide").[2]

The Guide describes a number of approaches to quantifying the harm caused by the infringement of Article 101 or 102 TFEU. These approaches can be implemented with varying degrees of sophistication: from simple comparisons between averages to regression analyses controlling for multiple factors.

On the one hand, the Commission warns that nothing in its Guide should be understood as raising or lowering the standard of proof (¶9), and that the legal framework in which Courts deal with the quantification of harm is defined by the European and national law, not by the Commission. On the other hand, the very same Guide clearly puts forward the (appropriate) idea that "regression analysis can considerably refine the damages estimation" (¶85) and "that econometric techniques can increase the degree of accuracy of a damages estimate and may thus help in meeting a higher standard of proof if required under applicable rules" (¶92).

Indeed, econometric models may be useful in: (i) determining whether a particular effect is present; (ii) measuring the magnitude of a particular effect; (iii) forecasting what a particular effect would have been, but for an intervening event, and pinpoint causality.

As the Commission itself acknowledges, "to date, little experience exists with econometric analysis in actions for antitrust damages before courts in the EU" (¶94).

Following the adoption of the Guide, however, national judges should expect to be increasingly confronted with econometric models.

Most often, the parties will present models leading to conflicting results. As the Guide correctly remarks, "it is normally not appropriate to simply take the average of the two results, nor would it be appropriate to consider that the contradictory results cancel each other out in the sense that both methods should be disregarded" (¶125).

Hence, judges (possibly with the help of experts) are required to carefully assess whether the model submitted by the claimant meets the standard of proof or, at the opposite, if the model submitted by the defendant suffices for overturning the claimants' conclusions.

The ability to understand an econometric analysis does not require advanced mathematics. Expert economists can guide judges and lawyers in identifying the important features in a model that drive the results. However, while the broad intuition can often be easily grasped by non-specialists, a great deal of experience and a well-trained clinic eye is indispensable to understand more subtle technical aspects and to evaluate the role played by the assumptions inherent in econometric modelling. Furthermore, in estimating antitrust damages, economic principles and statistical techniques must be applied within a particular legal framework.

To illustrate our point, we will start from the final stage in the damage estimation phase and discuss an apparently simple issue we have encountered in our consulting activities.

Given a multiple regression model, what is the meaning, in the context of an action for damages, of a not statistically significant coefficient for the variable of interest, e.g. the overcharge caused by an alleged cartel?

A stylized description of the issue

Broadly speaking, multiple regression is a statistical tool used to understand the relationship between two or more variables. Regression techniques allow the researcher to produce estimates of the values of the parameters of a particular model basing on available data.

For example, in a damage action following an ascertained cartel, the claimant might build a model where the final price paid for its purchases (over a time span longer than the cartel itself) is explained by some cost variables, some demand variables and a "dummy" variable identifying whether the cartel was active at the specific time. Subsequently, the analysis would focus on the dummy variable and use its value to determine (after appropriate calculations) the damage suffered as a consequence of the cartel.

The claimant would then produce as evidence in Court the output of its regression, namely a series of coefficients (each of them being a point estimate associated with a particular explanatory variable) accompanied by a number of statistical tests.

At the very least, judges can expect to be handed a table containing: the list of the model's parameters, the estimated coefficients, and a number or little star symbols (*) indicating whether each coefficient is statistically significant at a specific level α (1% ***, 5% ** or 10% *).

The choice of the specific level for α amounts to a convention, the 5% one being the most widely adopted in the academic world.

Econometric software routinely determine the level of statistical significance of the coefficients. This kind of automatic test is aimed at verifying if there is sufficient evidence against the hypothesis that the estimated coefficient is equal to zero (the so-called null hypothesis, or hypothesis of no effect).

Testing for the statistical significance of the estimated coefficients has become a consolidated practice in many scientific areas and in a sense amounts to a hallmark of

scientific practice. But is this approach always correct in the context of civil courts litigation?

To answer this question, we need to clarify the meaning of statistical significance first.

What is the meaning of "statistical significance"?

Let's stick to our example. The hypothetical econometric model considered produces an estimate of the overcharge paid by buyers due to the cartel. The number obtained is an approximation of the unknown true, from which it will differ. The "distance" between the estimate and the true value depends on several factors (e.g. sample size, selected variables, measurement errors). For these reasons, there could be cases in which the ascertained cartel had no effect on prices (i.e. the true overcharge is equal to zero) but the estimated overcharge is greater than zero.

Testing for statistical significance amounts to applying a conventional rule to help controlling for the risk of incorrectly interpreting the evidence at hand (i.e. concluding that the overcharge was greater than zero when in fact it was zero).[3]

If, in the context of the chosen model and given the data at hand, it is sufficiently unlikely to come up with an estimated overcharge as high or higher than the one at hand when the true overcharge is instead zero, then the estimated value is said to be statistically significant.

The "conventional" part of the test consists in choosing how unlikely the rejection of a true zero overcharge (null hypothesis) needs to be in order for the estimates to be considered statistically significant (i.e. statistically different from zero).

The judicial context

To translate the idea of testing for statistical significance in the context of a cartel litigation, the null hypothesis is normally framed so that its rejection is associated with the defendant being liable. Obtaining a statically significant coefficient for the overcharge is considered to be a strong evidence against the hypothesis that the cartel caused no overcharge

But how strong this evidence needs to be?
From a judicial perspective, civil court cases are assessed under the preponderance of evidence rule. Neither proof beyond reasonable doubt nor clear and convincing evidence is required.

Testing for statistical significance at an α, of 10%, 5% or 1% amounts to a convention, even though a consolidated one. It could be argued that, under some circumstances, relaxing the test might be appropriate.

Moreover it is important to stress that, in order to carry the significance test, the null is assumed to be true. Hence, a failure to reject the null hypothesis (at *any* chosen level of significance) does not prove that what is just an initial assumption is indeed true. In our example, a not statistically significant coefficient at the chosen significance level implies that the particular model and data do not provide strong enough evidence to reject the hypothesis of no overcharge. It does not prove the zero overcharge thesis.

Last but not least, statistical significance is determined, in part, by the number of observations in the dataset. When the sample size is small, it is also possible to obtain results that are practically significant (because an effect is indeed present) but fail to achieve statistical significance (not enough confidence on the measurement).

Thus, what role should statistical significance play in the context of actions for damages?

Possible approaches

There have been some attempts to reconcile the ideas of testing for statistical significance and the different standard of proof (possibly identifying specific thresholds, an idea that fortunately did not gain much traction).[4]

The Commission stresses that "even very sophisticated regression equations rely on a range of assumptions and will only be able to deliver estimates" (¶85). "One way to deal with the uncertainty of the estimate is to

indicate the results not as a point estimate but as an interval", a confidence interval.[5] Another way is to "refer to the notion of 'statistical significance', which is a standard way of testing whether the results obtained in a regression analysis are due to a coincidence or whether they reflect in fact a genuine correlation" (¶87).

Rubinfeld has long been a proponent of an "instrumentalist conception of the statistical burden of proof" affirming that "if significance levels are to be used, it is inappropriate to set a fixed statistical standard irrespective of the substantive nature of the litigation." [6]

The same author in the Reference Manual on Scientific Reference (2010) argues that "multiple regression results can be interpreted in purely statistical terms, through the use of significance tests, or they can be interpreted in a more practical, nonstatistical manner. Although an evaluation of the practical significance[7] of regression results is almost always relevant in the courtroom, tests of statistical significance are appropriate only in particular circumstances".[8]

Rubinfeld's statements can be interpreted as a warning against the mechanical application of the statistical significance test. This is indeed very appropriate, as in general it is possible for a significance test to reject the null hypothesis even though the likelihood of the null hypothesis being correct is greater than the likelihood that any single alternative hypothesis is true.[9]

Forensic economics

In our opinion focusing on the choice of a particular threshold for statistical significance testing is incorrect. Damage estimation requires the application of economic principles within a particular legal framework. However, it is inappropriate to stretch the scientific principles to accommodate the legal process. Scientific evidence, even in the context of litigation, must be evaluated within its own realm, according to scientific criteria. The scientific approach, however, does not amount to a set of receipts to be mechanically applied, nor to a sort of cookbook. On the contrary, science is more like a process, a method of enquire.

Within this process, statistical significance provides a relevant piece of information that needs to be assessed together with the available evidence, and cannot be considered an end in itself.

So the important point to bear in mind is that any econometric exercise has to be validated. This includes assessing the reliability of the underlying data, the appropriateness of the model itself (both the functional form and the chosen variables), and robustness to variations in both assumptions and data considered.

A similar assessment requires a solid understanding of the underlying economic principles, a great deal of technical expertise, sensibility in dealing with real world data (which is always messier and provides less clear-cut evidence than one could expect) and, finally, the ability to speak the language of lawyers and courts in order to pinpoint the most relevant information and to apply the scientific best practices within the relevant legal framework.

In addition, econometric evidence must be assessed within the context of economic evidence as a whole, including documentary and witness evidence. Once again, forensic economists can spot contradictions or help lawyers and judges reconcile quantitative and qualitative evidence.

As a general principle, we believe courts must steer away from adopting pre-cooked receipts or rule of thumbs in evaluating scientific evidence.

Any data analysis exercise must be carefully tailored on the specific problem under scrutiny. Moreover, it is paramount to evaluate the robustness of the conclusion by examining different models and by testing the impact on the results of small variations of the controversial assumptions.

Ultimately, an important point stemming both from the EC Guide and the consolidated practice in the United States is that, in evaluating competing econometric evidence, it is not appropriate to dismiss models basing on the claim that some shortcomings have been adopted, for instance that some relevant variables have been omitted; one needs to show that the omitted variables have an actual impact. Indeed, the Guide clearly states that "pointing out that a model relies on seemingly simplifying assumptions should therefore on its own not be sufficient to dismiss it; rather, one should consider how some of the simplifying assumptions are likely to affect its results" (¶104). Hence, our warning about the complexities of the economic analysis shall not be subverted to affirm that simpler techniques (e.g. simple averages or, even worse, rules of thumb) are better. Quite the opposite: econometric models, if properly handled, are superior insofar they allow to provide damages estimates that already control for the influence of other factors and to measure the precision of these estimates (for example allowing the construction of lower and upper bounds on the estimated damages).

Conclusions

This note discussed the role of testing for statistical significance in the context of antitrust damage quantification cases and sought to show the importance of steering away from "cookbook recipes" and "automatic procedures".

We believe that scientific evidence, even in the context of litigation, must be valuated within its own realm, according to scientific criteria. Furthermore, we strongly believe in the importance of assessing econometric evidence within the context of all the available economic evidence.

In general, correctly applying econometric analysis within the relevant legal framework requires a deep knowledge of both technical and legal issues. Forensic economists can play an important role in helping lawyers and judges bridge the gap between the legal world and the scientific practice.

If you would like to discuss those issues further, or if you would like more information about what our forensic economists can do for you, please contact us.

Tel:+39 06 68 300 530

Email: lear@learlab.com

Notes

[1] Proposal for a Directive of the European Parliament and of the Council on certain rules governing actions for damages under national law for infringements of the competition law provisions of the Member States and of the European Union. COM(2013) 404, 11.6.2013

[2] Commission Staff Working Document – Practical Guide on Quantifying Harm in Actions for damages based on breaches of Article 101 or 102 of the Treaty on the Functioning of the European Union SWD(2013) 205, 11.6.2013

[3] Rejecting the null hypothesis when this is in fact true is called a type I error, whereas not rejecting the null hypothesis when this is in fact false is called a type II error. The researcher can control the probability a type I error by choosing the level of significance α. The probability of rejecting the null hypothesis when it is false (power of the test), instead, gives an indication of how effective a test is in detecting deviations from a null hypothesis.

[4] For a discussion of the similarities between the structure of decision-making under uncertainty used in hypothesis testing in empirical science and the formal structure of decision-making under uncertainty used in legal trials refer to Saks, M. J. and Neufeld, S. L., 2011. Convergent evolution in law and science: the structure of decision-making under uncertainty. Law, Probability and Risk, 10(2), pp.133–148.

For a reply suggesting that "the role played by the required level of statistical significance in the scientific method should not be seen as analogous to that played by the standard of proof in the legal process" refer to Miller, C., 2012. A comment on Saks and Neufeld: "Convergent evolution in law and science: the structure of decision making under uncertainty." Law, Probability and Risk, 11(1), pp.101–104."

[5] A confidence interval is another concept related to null hypothesis significance testing. See Verbeek, M., 2004. *A Guide to Modern Econometrics* (2nd edition). Wiley, p.25. "A confidence interval can be defined as the interval of all values for β_k^0 for which the null hypothesis that $\beta_k = \beta_k^0$is not rejected by the t-tests. […] In repeated sampling, 95% of these intervals will contain the true value $\beta_k$which is a fixed but unknown number." The confidence level does not express the chance that repeated estimates would fall into the confidence interval!

[6] Rubinfeld, D. L., 1985. *Econometrics in the Courtroom*. Colum. L. Rev., 85, p.1048.

[7] "Practical significance" means that the magnitude of the effect being studied is not *de minimis*—it is sufficiently important substantively for the court to be concerned. The issue we are discussing here, although related, is slightly different from the usual dispute between "practical" and "statistical" significance. In the context of antitrust damage quantification *any* effect is relevant. The latest EC Directive reaffirms the *acquis communautaire* of the right on the injured party to full compensation of the harm suffered.

[8] Federal Judicial Center and National Research Council, 2011, *Reference Manual on Scientific Evidence* (3rd edition), National Academies Press, Washington.

[9] Kaye, D. H., 1983. *Statistical Significance and the Burden of Persuasion*. Law & Contemp. Probs., 46, p.13.