

A Comment on *Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology*

David F. Hendry*

Economics Department, Oxford University

June 29, 2005

Abstract

Florax et al. (2003) undertake a simulation study designed to assess the properties of various selection strategies applied to a spatial modelling problem. Unfortunately, a serious flaw in their experimental design vitiates their main conclusions. A proposal for how to rectify the problem suggests that the opposite conclusion might well result, particularly for more interesting settings than the simple one they considered.

Keywords: Model selection; Specification strategy; Hendry methodology; Monte Carlo methods

JEL classification: C12; C13; C15

1 Introduction

Florax, Folmer and Rey (2003) seek to examine the properties of specification searches in spatial econometrics using a simulation study. To quote:

Our main conclusion is that the classical forward stepwise approach outperforms the Hendry strategy in terms of finding the true data generating process as well as in the observed accuracy of the estimators for spatial and non-spatial parameters. It also dominates concurrent forward stepwise strategies recently suggested in the literature.

Unfortunately, a serious flaw in their experimental design vitiates this 'conclusion', and indeed the analysis thereof suggests that the opposite conclusion may well hold. Precisely, they seek to compare the rejection frequencies under the alternative of tests with different null rejection frequencies. Section 2 demonstrates the invalidity of their approach in a simple setting. Section 3 shows that the solution is both well known and trivial to implement. Section 4 concludes and notes some difficulties in generalizing their proposed approach beyond the simple setting they simulated, even if it remained the best of the methods after an appropriate comparison.

*email: david.hendry@economics.ox.ac.uk

2 Invalid comparisons

Consider a conventional test on a scalar hypothesis of interest, denoted H_0 , based on a statistic τ . Under the null, the test is normally distributed with mean zero and variance σ_τ^2 , denoted:

$$\tau \sim N[0, \sigma_\tau^2], \quad (1)$$

so that for any significance level α and associated 2-sided critical value c_α :

$$\Pr[|\tau| > c_\alpha \mid H_0] = \alpha.$$

As is well known, given (1), $c_{0.05} \simeq 2\sigma_\tau$ and $c_{0.01} \simeq 2.58\sigma_\tau$.

A second test, denoted κ , is also used and is believed to be distributed as Student's t in large samples. To check its finite sample distribution, a simulation study is undertaken, and it is found that:

$$\Pr[|\kappa| > c_\alpha \mid H_0] = 0.$$

If, for example, $\sigma_\tau = 10$ so $c_{0.05} \simeq 20$, such an outcome is almost certain. Nevertheless, there exist 2-sided critical values $d_\alpha < c_\alpha$ such that:

$$\Pr[|\kappa| > d_\alpha \mid H_0] = \alpha,$$

at which critical values, the two tests have the same null rejection frequency for every choice of α : here $d_{0.05} \simeq 2$ and $d_{0.01} \simeq 2.58$. Indeed, to place the tests on a comparable basis, one could either base each on their appropriate critical values (derived analytically or calculated as null rejection frequencies by simulation), rescale κ , or more simply, normalize τ by letting:

$$\tau^* = \frac{\tau}{\sigma_\tau} \sim N[0, 1], \quad (2)$$

so $c_{0.05} \simeq 2$ and $c_{0.01} \simeq 2.58$ as well.

Under the alternative that $E[\tau] = \psi \neq 0$, the power $p_{\tau,\alpha}(\psi/\sigma_\tau)$ of the τ test is easily derived from (1) for any given c_α , and in this setting is monotonically non-decreasing in ψ for any σ_τ , monotonically non-increasing in σ_τ for any ψ , and monotonically non-decreasing in α for any ψ/σ_τ :

$$\Pr[|\tau| > c_\alpha \mid \psi \neq 0] = p_{\tau,\alpha}(\psi/\sigma_\tau) > \alpha.$$

Numerically, when $\psi = 20$ and $\sigma_\tau = 10$ with $c_{0.05} \simeq 20$ (neglecting the tiny probability that one draws a value $\tau < -20$):

$$\Pr[\tau > 20 \mid \psi = 20] = 0.5.$$

Consider what happens in the simulation when investigating the rejection frequency of κ under the alternative, but using $c_{0.05}$ as the critical value:

$$\Pr\left[\kappa > 20 \mid \frac{\psi}{\sigma_\tau} = 2\right] \simeq 0.$$

Apparently τ is vastly more powerful than κ even though it is clear from (2) that they are essentially the same statistic, with:

$$\Pr \left[|\kappa| > d_{0.05} \mid \frac{\psi}{\sigma_\tau} = 2 \right] \simeq 0.5.$$

So-called ‘power’ comparisons are meaningless without standardizing null rejection frequencies. Unfortunately, that is precisely what Florax *et al.* (2003) do in their paper. Consider their table 1: the procedure they denote by ‘Hendry’ has a zero null rejection frequency at the critical value they input, whereas the other methods have the correct 5% which Florax *et al.* (2003) seem to want to use. Thus, their ‘power’ comparisons are as meaningless as those just described, and no conclusions as to the choice of an appropriate method can follow.

3 A solution

The solution is trivial: in the simulations conducted under the null hypothesis, record the ‘empirical’ critical values at which 5% rejections actually occurred for a large simulation sample, and use these in the second stage ‘power’ comparisons. This is hardly a novel suggestion – see e.g. Aneuryn-Evans and Savin (1982), Hendry (1984) and Engle, Hendry and Trumbull (1985). For a chi-squared type test, this will be the upper tail value \varkappa such that no more than 5% of the drawings exceeded \varkappa . Then the simulations under the alternative can be conducted conditional on \varkappa , which is why a large simulation sample size is needed at the first stage to ensure accurate determination of \varkappa . All practical applications of the test will need such critical values in any case, and response surface methods exist to determine useful programmable functions: see e.g. Hendry (1984) and Ericsson and MacKinnon (2002).

4 Conclusion

Not to standardize tests for their null rejection frequency is a fatal flaw in Florax *et al.* (2003), vitiating any ‘power’ comparisons. This is so well known it is surprising such a mistake could be published. Moreover, a minimal knowledge of Monte Carlo methods is needed to develop a solution. To proceed to compare rejection frequencies under the alternative when one of the tests has a zero null rejection frequency precludes drawing any conclusions. Indeed, based on the simple analogy with which this note began, the opposite conclusion to theirs could well hold after appropriate standardization of the null rejection frequencies of the tests involved, although I have not conducted such experiments.¹

More generally, the study in Florax *et al.* (2003) does not take account of the many recent advances in general-to-simple selection procedures—in any case, they do not reference them. Accessible examples include Hoover and Perez (1999, 2004), Hendry and Krolzig (1999, 2001, 2004a, 2004b, 2005) and Krolzig and Hendry (2001). In the case Florax *et al.* (2003) examine by simulation, the number of tests is known in advance, and no selection is required across regressors. This is not representative of the selection problems confronting spatial modelling in general. Some methods, such as general-to-simple,

¹The first author unfortunately failed to respond to several requests for clarification of their simulations which could have made such a comparison feasible.

will generalize, whereas others, such as simple to general, will not work well in the realistic setting where other features of a model need to be selected, and the number of possible tests is not known in advance: see e.g., Anderson (1962, 1971).

Acknowledgments

Financial support from the ESRC under a Professorial Research Fellowship, RES051270035, is gratefully acknowledged.

References

- Anderson, T. W. (1962). The choice of the degree of a polynomial regression as a multiple-decision problem. *Annals of Mathematical Statistics*, **33**, 255–265.
- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. New York: John Wiley & Sons.
- Aneuryn-Evans, G. B., and Savin, N. E. (1982). Conflict among the criteria revisited: The W, LR and LM tests. *Econometrica*, **50**, 737–748.
- Engle, R. F., Hendry, D. F., and Trumbull, D. (1985). Small sample properties of ARCH estimators and tests. *Canadian Journal of Economics*, **43**, 66–93.
- Ericsson, N. R., and MacKinnon, J. G. (2002). Distributions of error correction tests for cointegration. *Econometrics Journal*, **5**, 285–318.
- Florax, R. J. G. M., Folmer, H., and Rey, S. J. (2003). Specification searches in spatial econometrics: The relevance of Hendry’s methodology. *Regional Science and Urban Economics*, **33**, 557–579.
- Hendry, D. F. (1984). Monte Carlo experimentation in econometrics. In Griliches, Z., and Intriligator, M. D. (eds.), *Handbook of Econometrics*, Vol. 2, Ch. 16, pp. 937–976. Amsterdam: North-Holland.
- Hendry, D. F., and Krolzig, H.-M. (1999). Improving on ‘Data mining reconsidered’ by K.D. Hoover and S.J. Perez. *Econometrics Journal*, **2**, 202–219.
- Hendry, D. F., and Krolzig, H.-M. (2001). *Automatic Econometric Model Selection*. London: Timberlake Consultants Press.
- Hendry, D. F., and Krolzig, H. M. (2004a). Automatic model selection: A new instrument for social science. *Electoral Studies*, **23**, 525–544.
- Hendry, D. F., and Krolzig, H. M. (2004b). We ran one regression. *Oxford Bulletin of Economics and Statistics*, **66**, 799–810.
- Hendry, D. F., and Krolzig, H.-M. (2005). The properties of automatic Gets modelling. *Economic Journal*, **115**, C32–C61.
- Hoover, K. D., and Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, **2**, 167–191.

- Hoover, K. D., and Perez, S. J. (2004). Truth and robustness in cross-country growth regressions. *Oxford Bulletin of Economics and Statistics*, **66**, 765–798.
- Krolzig, H.-M., and Hendry, D. F. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, **25**, 831–866.