MPRA

Munich Personal RePEc Archive

The small multiple in econometrics – a redesign

Torsten L. Klein

PAS Research Unit

11 December 2014

THE SMALL MULTIPLE IN ECONOMETRICS – A REDESIGN

KLEIN, TORSTEN L.[*]

This brief note serves as a companion paper to Klein (2014). Small multiples incorporate graphical frameworks such as P *value plots with ease, and thus facilitate visualizing quantitative data that record parameter change from simulation experiments. Pitfalls in layout may be avoided when observing elementary design principles. To illustrate their workings the principles revise a small multiple that collects simulation results on the empirical size of procedures testing exogeneity in the bivariate probit model.*

JEL CODE: *C10, C15, C35, C52, Y10.*

KEYWORDS: *Monte Carlo, bivariate probit model, exogeneity testing; data visualization.*

1 SHAPING EVIDENCE: THE SMALL MULTIPLE AND GUIDELINES FOR EFFECTIVE DESIGN

Analyzing characteristics of statistical tests obtained from Monte Carlo simulation P value plots offer a comprehensive approach to examine quantitative results. Figure 1 gives an example. It compares the empirical size of seven testing procedures according to their P values' empirical distribution function (EDF) for a particular combination of data generating process (DGP) and sample length (N).[1] Changing these parameters will change the empirical size of tests. One method of framing variation stores the evidence to a small multiple.

[*] PAS Research Unit, B7,17, 68159 Mannheim, Germany, `tl.klein@gmx.de`. I am grateful for comments from Rainer Fath. All errors are mine.

[1] Davidson and MacKinnon (1998) introduce the concept of P value plots in detail. Figure 1 replicates simulation evidence from a Monte Carlo experiment by Monfardini and Radice (2008) on tests of exogeneity in the bivariate probit model. Additional information on both the graphical framework as well as the simulation can be found in Klein (2014).

The small multiple is constructed by duplicating a graphical object, keeping its form intact while varying content. Each duplicated object will then be arranged on the medium of communication. In a two-dimensional environment such as the printed page or the inanimate computer screen all duplicates obtained from figure 1 by changing simulation design are placed into a matrix of P value plots: while the data generating process varies along one dimension the length of the sample estimation is based on varies along the other.[2] A possible arrangement is displayed in Figure 2.

Although the small multiple creates a simple structure care needs to be taken for maintaining the convenience of its graphical building block. Figure 2 makes a case in point: the small multiple picks figure 1 as a nucleus, and copies every component to its slots. Easy to grasp it is not. The evolution of size seems clad in static accessory giving readers, whether author or audience, a hard time to part substance from supplement. Just like plain text an early draft of graphical displays may therefore benefit from sensible editing too. In order to find content an effective layout the principles mentioned in Tufte (1983, 92–100, 162–169) offer advice: put information first, kill redundancy in ink, whether data-related or not, and space. The Monte Carlo exercise generates meaningful information by contingent comparison of test size – for different parameter sets, to nominal significance, to competing procedures. The small multiple helps organizing such comparison; any revision must try and put this quality to best use.

There are two basic strategies when revising figures, one additive, the other reductive. The first one starts by showing bare facts only, subsequently placing feature after feature that helps explain them to the reader; the second one takes the opposite direction, removing from a given draft any element not deemed necessary for effectively communicating its central idea. In what follows figure 2 serves as a starting point to be stripped of all things irrelevant for understanding simulation results.

2 Further details on the small multiple can be found in Tufte (1983, 170–174) and Tufte (1990, 67–79). If more than two parameters vary one can iterate on the building principle. Each cell entry may then itself contain a matrix, constructed from another multiple, and so forth (Tufte, 1990, 29).

2 A STEP–BY–STEP REVISION OF THE SMALL MULTIPLE

The checkup of figure 2 starts with *P* value plots themselves. Some judgment is required as to whether every setting should be taken into account. A sensible strategy weighs providing more information to the reader against consuming more of limited space. Presenting the entire simulation output creates a matrix figure of 15 entries (3×5), each to store results for the seven procedures under study – a total of 105 objects arranging 22,575 data points. Which quantitative information is essential, and should therefore be kept, which is redundant, and could easily be omitted? Three tests (LM2, LM3 and LM4) show heavy size distortion for all combinations of DGP and sample length considered. Their EDF remains zero for small to medium nominal significance levels. Due to such lack of variation *P* value plots fail to contribute anything beyond the notion of under-rejection which is most efficiently stated in plain text. Hence, they are erased. In contrast to standard tabular displays of Monte Carlo exercises reporting empirical size at a few levels only, the small multiple plots *P* values over the entire domain of nominal significance. But for results this would be valuable only if it takes the $(0, 1)$ interval to establish pattern and regularity; since most of the time inference operates on the left tail of EDF, the revised small multiple in figure 3 considers evidence associated with nominal values smaller than 0.5.

Another task when revising concerns legibility. If procedures exhibit similar behavior under given simulation conditions, e.g. in large samples, their *P* value plots become clustered and therefore hard to distinguish. Though disorder means comparability in performance and so contains valuable information Davidson and MacKinnon (1998, 3) recommend using not the empirical distribution function, but *P* value discrepancy plots: the difference between EDF and nominal significance level.[3] Challenges for legibility arise in the opposite direction as well. The small multiple collects evidence over a sequence of fifteen parameter variations generating among procedures marked size distortion as well as almost ideal behavior. Treated one by one, each of the associated sub-figures could receive a different

[3] The use of *P* value discrepancies also facilitates incorporating a variant of the Kolmogorov-Smirnov test on whether EDF is compatible with its theoretical counterpart, the uniform distribution, cf. figure 3 and Davidson and MacKinnon (1998, 11).

range of scaling, yet mutual comparability calls for fitting the same scale on all of them. One option resorts to identical, but non-linear scaling that compresses large discrepancies and amplifies smaller ones: figure 3 employs a simple power transformation to achieve both ends.[4] The price to pay comes with two connected drawbacks. Non-linear scaling labels evenly spaced ticks irregularly, and complicates assessment of points on the EDF that fall in-between. However, the trade-off seems acceptable, because of the comparative focus in the simulation exercise.

P value plots channel all quantitative information to the empirical distribution function. Unless essential to understanding, everything else in figure 2 can – and should – be disposed of. Frames and labels and legend surrounding each sub-figure obstruct comparison between results of different parameter sets; grid lines compete with EDF for space and recognition in the respective display area. Instead, ticks may be detached from their frame and placed in the margin between sub-figures to guide the reader. Allowing better comparison of empirical size between rows ticks are set in column-specific ranges that mark for each sample length the maximum over- and under-rejection. Adequate coloring can bring relevant information to the fore while restraining supplementary elements. Finally, the small multiple may be shrunk in size without compromising accessibility of information, an adjustment which helps to realign display with the general orientation of the body text.

Not every obstacle to communicating quantitative information will be removed by applying the principles stated in the previous section. If there is a lot of data to report shortening the EDF's interval on display, employing discrepancy plots or converting scales will not always produce a design easy to read. Hence, there exists a limit as to how many objects one figure may incorporate. Also, small multiples are most effective when they tell a unidirectional story. Map is one-to-one: moving further in one direction of the parameter domain associates with moving further in the same direction of the results' target set, for all objects under study. From the small multiple in figure 3, for example, the reader may recognize at a glance that irrespective of the data generating process simulating with larger samples will reduce over-rejection for all testing procedures. On the other hand, changing the data generating process affects different tests in a differ-

---

4 Other remedies include using multiple or split scales in one figure, or cutting off dispensable observations.

ent manner.[5] Here, examining the small multiple requires more time and effort to disentangle the opposing consequences. Readability suffers.

REFERENCES

Davidson, Russell and James G. MacKinnon, 1998, "Graphical methods for investigating the size and power of hypothesis tests", *Manchester School* 66(1), 1–26.

Klein, Torsten L., 2014, "Communicating quantitative information: tables vs graphs", mimeo, PAS Research Unit.

Monfardini, Chiara and Rosalba Radice, 2008, "Testing exogeneity in the bivariate probit model: a Monte Carlo study", *Oxford Bulletin of Economics and Statistics* 70(2), 271–282.

Tufte, Edward R., 1983, *The visual display of quantitative information*, Cheshire CN: Graphics Press.

Tufte, Edward R., 1990, *Envisioning information*, Cheshire CN: Graphics Press.

5 Take CM and LM1: for the former a move from DGP1 to DGP2 results in lower size distortion than one from DGP1 to DGP3, for the latter the reverse is true. The non-uniform behavior seems more likely to occur if, as is the case here, more than one simulation parameter changes while moving from one matrix entry to the next, cf. Klein (2014, table 4).
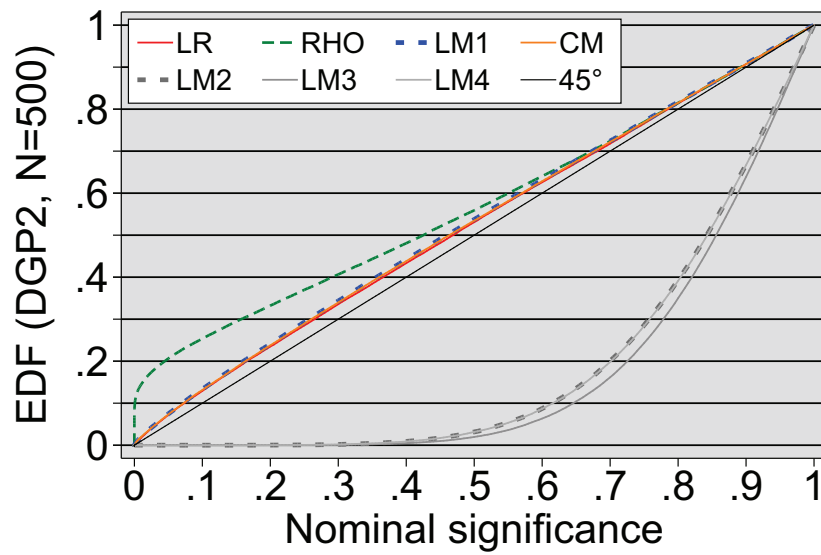
Figure 1: *P* value plots – comparing empirical size of testing proce-
dures.

Source: Klein (2014)

Notes: The figure depicts for a specific combination of data generating process
(DGP) and sample length (*N*) the empirical distribution function (EDF) of proce-
dures testing exogeneity in the bivariate probit model analysed by Monfardini and
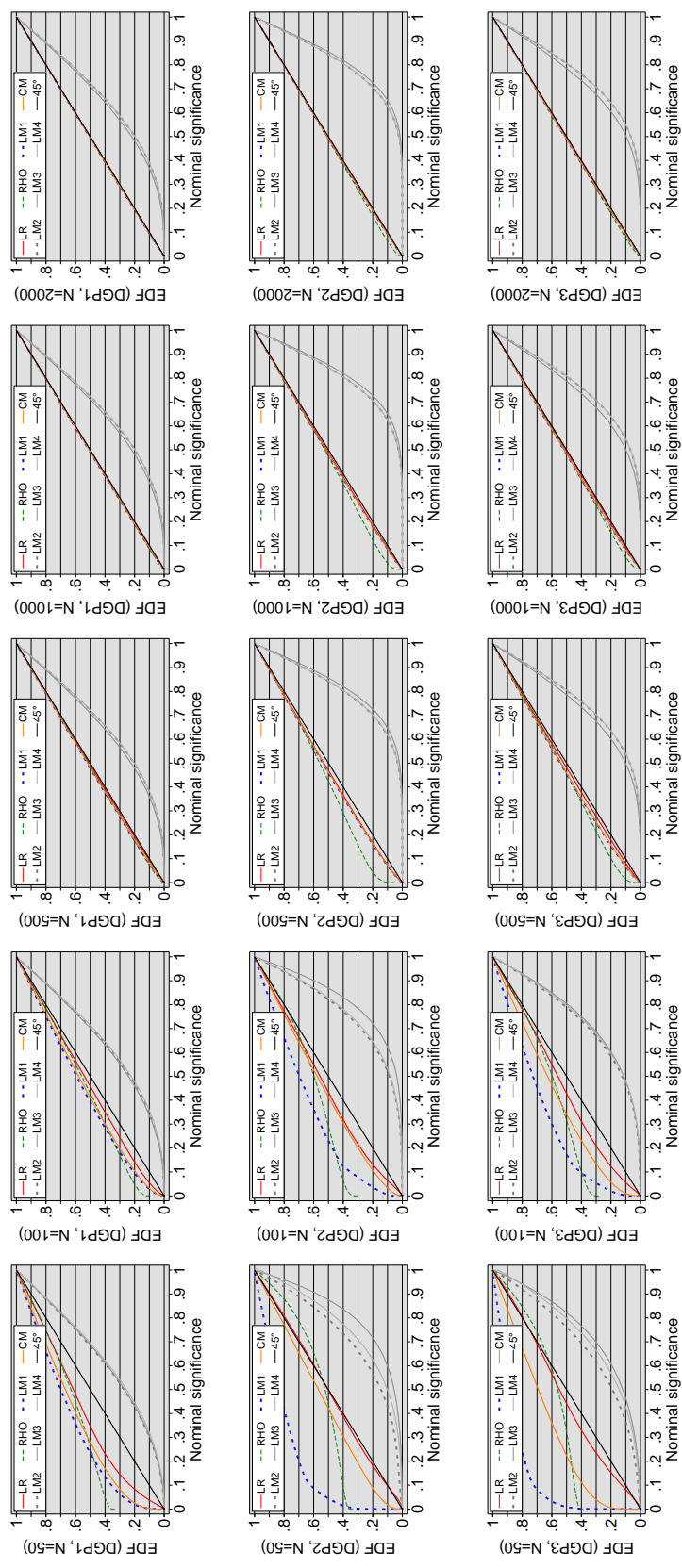Radice (2008, section IV).

Figure 2: Small multiple of *P* value plots – tracing empirical size through changes of simulation parameters.
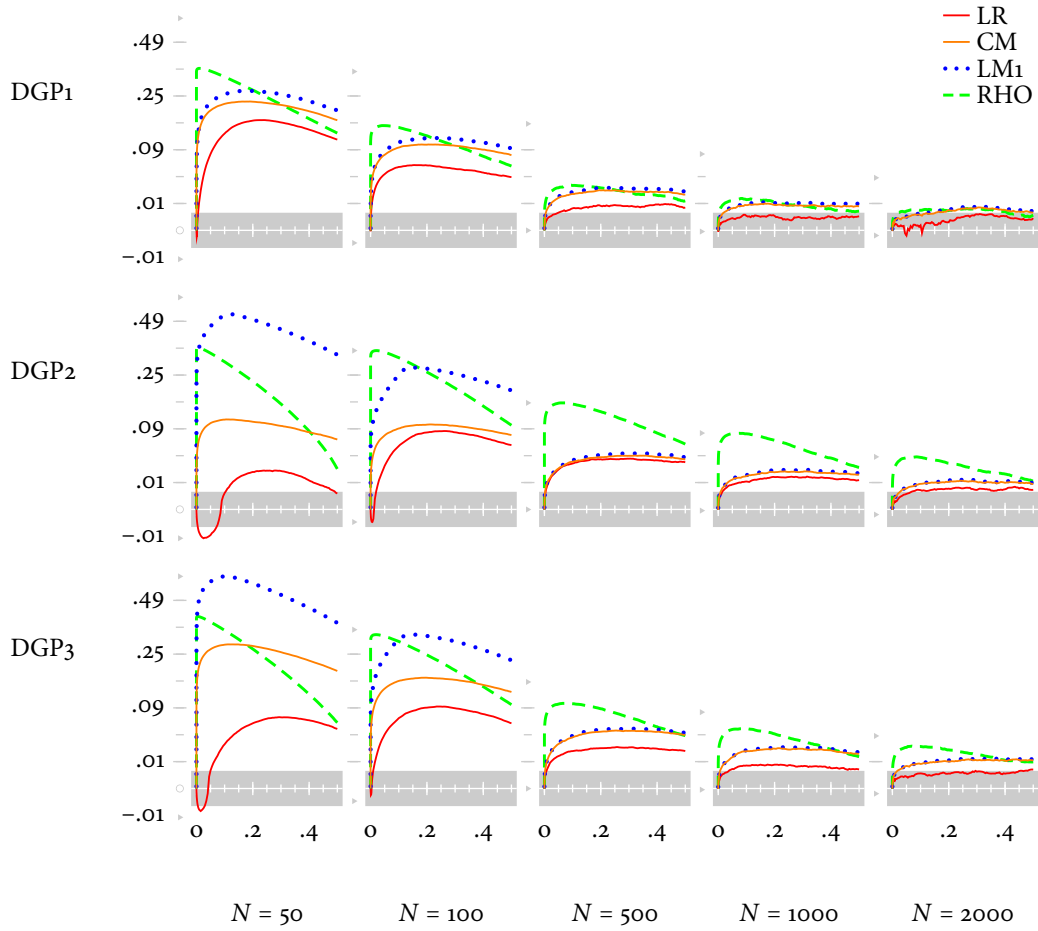
Notes: Cf. the notes to figure 1.

Figure 3: The small multiple – a redesign of figure 2 using *P* value discrepancies.

Notes: *P* value discrepancies are calculated by taking the difference between the *P* values' empirical distribution function $\hat{F}(x_i)$ and nodes x_i inserted for evaluation. Each ordinate is scaled non-linearly according to the power transformation g: $y \mapsto \text{sgn}(y) \cdot |y|^{0.5}$. The shaded area enveloping abscissae indicates the non-rejection region of the Kolmogorov-Smirnov test at 0.05 critical values.