The notion of a statistically adequate model – a probabilistic description that accounts for the statistical regularities in the data – provides data with 'a voice of its own', separate from the theory in question; see Spanos (2006). This error statistical notion can help the author in a multitude of ways to carry out his primary objective of 'grounding economic claims more firmly in evidence', including the following:

(a) Substantiate his bold claims in [4] cited above (p. 183).
(b) Weed out unwarranted 'empirical adequacy' claims made by many authors of the theory-based explanatory mechanisms he discusses in chapters 6–9, instead taking them at face value (see Spanos 2006).
(c) Untangle the Instrumental Variables method (chapter 7) from the clutches of the 'pre-eminence of theory' perspective by bringing out the role of the probabilistic structure of the data in selecting the relevant instruments (see Spanos 1986: 637–43).
(d) Render largely irrelevant the discussion in chapter 10 concerning causally based vs. counterfactual reasoning for any evidence-based evaluation, since none of the structural models used in that discussion are likely to be statistically adequate vis-à-vis any available observational or even experimental data; see Spanos (1995).

In conclusion, I recommend this monograph highly to anybody interested in applied economics/econometrics for its insightful and pertinent criticisms of the pre-eminence of theory perspective. However, its glaring omission to relate the notion of empirical evidence to data – via statistical inference reasoning – limits considerably the value of its proposals 'on how to ground economic claims more firmly in evidence'.

Aris Spanos
*Virginia Polytechnic
Institute and State University*

Rejoinder

I thank David Teira, Kevin Hoover, and Aris Spanos for their thoughtful and stimulating reviews. I can only respond to a fraction of the points made. I organize my replies around four topics: clinchers and vouchers; facts and values; theory and evidence; and Friedman and empiricism.

Clinchers and vouchers. The logic of clinchers and vouchers differs. A result produced by a clincher is certain, given the assumptions (and is

therefore conditional, as Hoover puts it). A voucher turns on facts about certain domains that induce regularities; it speaks for its conclusions on the basis of precedent but without guarantee.

Philosophers of science of the analytical bent tend to focus on bits of scientific practice about which they can make claims with reasonable precision, sometimes at the expense of relevance. Perhaps the chapter on instrumental variables (IV) is guilty of that sin, too. I was trying to understand the substantial assumptions that underlie that technique and developed a set that can be shown to be sufficient for the causal conclusion to be correct. In this understanding, the method is a clincher: if the assumptions are met, the causal conclusion will be right. But the assumptions are stringent indeed and if they are not satisfied, we know *nothing*; in particular, we do not know that our conclusion is approximately true; nor do we know that the conclusion is false, strictly or approximately.

I agree with Hoover that 'what is really needed is an account of vouching', which in this context means an account of the reliability of the IV method when its assumptions are not or only approximately met. All I can add is that such an account will probably be of considerably narrower scope and require much more detailed case-specific knowledge than its clincher counterpart. The set of assumptions I proposed is fairly generic, describing an abstract causal structure. Vouchers, by contrast, are usually domain specific. As an example take the Goldberg rule, which predicts whether a psychiatric patient is neurotic or psychotic: it is known that the rule classifies patients correctly in about 70% of cases. More work certainly needs to be done to extract rules of that kind from econometric practice.

Facts and values. Hoover agrees with me that the CPI 'embeds certain values in its design'. But he cautions: 'it is a weak form of value-ladenness', which 'is little threat to the positive/normative distinction'. Our disagreement here seems to be one of degree. It is clear that two researchers, if they want to agree on the correct value of the CPI, must agree on numerous normative issues. To use an example discussed in the book, consider indexing of pensions as the purpose of CPI measurement. Pensions are not indexed for nothing but rather to insulate pensioners from changes in the 'cost of living'. So even if there is agreement as to the indexation purpose, there may be differences about the appropriateness of this or that – normative – understanding of the target quantity. It is true that most economists have a ready answer: the cost of achieving a certain level of welfare, and welfare is to be understood in terms of preference satisfaction. But this is just one possible answer, probably not the best, and definitely an answer to a normative question. Next, we have to determine the household weights. The CPI weights are proportional to

household expenditure. This gives wealthier households a greater weight. An alternative is 'democratic weighting' where every household receives the same weight. This is a normative choice. Next we have to determine how to deal with change in the economy. Old goods leave the market, new goods come in. We cannot always assume that consumers prefer the new goods to the old (because their market share increases), but if we cannot do that then we need to evaluate the value of quality changes, which again is a normative task. These changes occur at the level of the individual good, at the level of the mix of available goods, at the level of the distribution channel, at the level of the environment within which it is consumed and so on. Next we have to determine how many indices to compute. Shall we invest more taxpayers' money and compute a larger number of more accurate indices or is one or a small number enough? It thus appears that normativity is all over the place in measuring the CPI.

Second, does this threaten the positive/normative distinction we have grown up with? I do not see how it could not. The backbone of that distinction is that disputes about facts can be settled on the basis of a neutral, external arbiter called 'evidence'. Of course, evidence plays a role in settling disputes about economic variables. But this case shows that the evidence itself is often infused with values, and disagreements about values can in turn lead to disagreements about facts.

Theory and evidence. I indeed fail to discuss theories of evidence in any detail in the book and am therefore grateful to Spanos to give me the opportunity to clarify my position. But when Spanos writes that *Error in Economics* suffers from a 'crucial omission' because 'the theory of evidence implicitly adopted in this book lacks any meaningful anchoring in data through statistical modeling and inference' he either mistakes the book's project or what his preferred theory of evidence – the Error Statistical approach – can deliver.

Error in Economics does not aim to provide a general theory of evidence or confirmation. The characterizations of prima facie, valid and sound evidence of the first chapter may have the form of a general theory but they are meant to provide a conceptual framework for thinking about the issues treated in the book. The substance of the 'theory' adopted is to be found, if anywhere, in the context of the detailed case studies. What the case studies aim to do is to give what one might call 'theories of the mid-range': rules that, in the context of certain kinds of investigations, help to make inferences more reliable. They are meant to be more general than individual cases but much more specific than the highly general epistemological principles such as 'update beliefs using Bayes' rule' or 'subject hypotheses to severe tests'.

Perhaps, though, Spanos wants to make a different point: a general theory of evidence is necessary to underwrite mid-range methodological

claims. Thus, in a joint article Mayo and Spanos write (Mayo and Spanos 2006: 325):

> We read, for instance, in a recent article in *Statistical Science*: 'professional agreement on statistical philosophy is not on the immediate horizon, but this should not stop us from agreeing on methodology' (Berger [2003], p. 2). However, the latter question, we think, turns on the former.

Now, when I was writing *Error in Economics* I did attempt to use such philosophical resources for my purposes. But the attempt failed. Bayesianism, for instance, I found too abstract to yield any detailed prescriptions of the kind I was looking for. The error-correction approach, by contrast, is too narrow. Despite Mayo's and Spanos's repeated insistence that its 'idea may be generalized to inferring the presence of 'an error' or flaw, very generally conceived' (2006: 329), it is best suited to statistical hypothesis-testing. While Spanos seems to agree with the thrust of my arguments regarding the role theory currently plays in economics, the three concrete points he makes in (b)–(d) demonstrate that adopting a narrow statistical perspective brushes over many of the substantial issues the book was concerned with:

(b) Lack of empirical adequacy is indeed a problem of theory-based explanatory mechanisms, and it was taken up in chapter 6. I could have made a similar argument using the error-statistical perspective (or Popperianism or verificationism or other a priori philosophies) but instead adopted a contextualist approach. One advantage of that approach (and there are others) is that brings to the fore many important issues that have nothing to do with empirical adequacy, for instance issues regarding the appropriateness of mechanistic claims – no matter how well substantiated empirically – to help address certain purposes economists pursue such as predicting variables of interest, inferring causal relations among aggregate variables or designing policy interventions.

(c) 'No causes in/no causes out' is a principle that is now generally accepted among scholars working on causal inference. Chapter 7 shows that this principle applies to instrumental variables. I only give sufficient but no necessary conditions and therefore it is well possible and even likely that correct inferences can be drawn using other sets of assumptions. I would be very surprised, however, if there is a set that does not include assumptions about the *causal* structure of the system studied.

(d) Chapter 10 discusses work on counterfactuals for policy by scholars such as James Heckman, Stephen LeRoy, and Judea Pearl. We argued that many working on counterfactuals in econometrics have the relationship between causal claims and counterfactuals wrong and outlined a causal-model theory of counterfactuals to replace that Hume-inspired thinking. It is certainly the case that if the approach we put forward were to

be implemented, issues of statistical adequacy of causal models will eventually have to be dealt with. But these were not topic of that chapter, and unless I am presented with an impossibility result I see no reason to reject out-of-hand attempts to build models that have the right properties to underwrite counterfactual claims and are statistically adequate at the same time.

Friedman and empiricism. I am in complete agreement with Teira that the kinds of value commitments we make when measuring variables of interest reduce the potential for the community of scientists to find agreements on economic matters. I now see that the optimistic stance taken in the book – though cautious indeed – should have been even more limited or perhaps absent altogether. What I disagree with is the implicit suggestion that my methodological proposal brings values into economic inquiries. The values are there whether we want it or not. What I tried to do is to make that point explicit and to suggest some ways of how to deal with it.

Hoover writes, and Teira seems to concur, that Milton Friedman was an empiricist just as I think we ought to be. Teira then challenges me to defend my optimism with respect to (the potential for) progress in economics in the light of the fact that Friedman's agenda seems to have failed – partly for reasons I spend a lot of time elaborating, viz. that the very data base all empirical programmes have to take as their starting point can and has been disputed. Succinctly, we may put the challenge thus: how can a research agenda be (as) purely data-driven (as possible) if these data are themselves highly theory and value laden?

I happen to disagree with the claim that Friedman was an empiricist just as I think we ought to be, partly because I think research should be purpose- rather than data-driven. But this is probably a side-issue. To the extent that I find research should be more empiricist I think it is possible to reduce theoretical commitments in the areas I am most concerned with. To take an example in the area of causal inference, Angrist's use of the random-sequence number as an instrument to determine whether veteran status causes earnings losses has been criticized on the basis of the claim that the instrument causes losses via an independent route, viz. firms' training investment decisions (see chapter 7). If we look at the 'evidence base' of that claim, we see that it is a 'behavioural model': a mathematical model of the kind criticized in chapter 6 in which it is optimal for firms to take employee's random-sequence numbers into account. There is no empirical investigation showing that companies indeed act as the model predicts. It is *that* practice that I criticize – hardly controversially – wherever I found it in the case studies that I looked at. Since it is fairly common, however, my judgment that a more evidence-based economics is possible and desirable should not be so surprising.

If Friedman's research agenda has failed that may well have to do with other issues. There is no doubt that Friedman had a very strong political agenda and people may have resisted his more empirical work because of that alone. And even if in some sense data driven, Friedman's empirical methodology had its own problems as I briefly mention in chapter 6 and as discussed in detail by eminent econometricians such as Hendry, Ericsson and others.

Julian Reiss
Erasmus University Rotterdam

REFERENCES

Mayo, D. G. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Mayo, D. G. and A. Spanos 2004. Methodology in practice: statistical misspecification testing. *Philosophy of Science* 71: 1007–25.

Mayo, D. G. and A. Spanos 2006. Severe testing as a basic concept in a Neyman–Pearson philosophy of induction. *British Journal for the Philosophy of Science* 57: 323–57.

Ricardo, D. 1817. *Principles of Political Economy and Taxation*. Vol. 1 of *The Collected Works of Davie Ricardo*, ed. P. Sraffa and M. Dobb. Cambridge: Cambridge University Press.

Spanos, A. 1986. *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

Spanos, A. 1995. On theory testing in econometrics: modeling with nonexperimental data. *Journal of Econometrics* 67: 189–226.

Spanos, A. 2006. Econometrics in retrospect and prospect. In *New Palgrave Handbook of Econometrics*, Vol. 1, ed. T. C. Mills and K. Patterson. Basingstoke: Macmillan.

Student 1931. The Lancashire milk experiment. *Biometrika* 23: 398–406.

Teira, D. Forthcoming. Why Friedman's methodology did not generate consensus among economists? *Journal of the History of Economic Thought*.