SINNING IN THE BASEMENT: WHAT ARE THE RULES? THE TEN COMMANDMENTS OF APPLIED ECONOMETRICS

Peter E. Kennedy

Simon Fraser University

Abstract. Unpleasant realities of real-world data force applied econometricians to violate the prescriptions of econometric theory as taught by our textbooks. Leamer (1978) vividly describes this behavior as wanton sinning in the basement, with sinners' metamorphizing into high priests as they ascend to the third floor to teach econometric theory. But this sinning is not completely wanton — applied econometricians do (or should) follow some unwritten rules of behavior, in effect bounding the sinning and promoting a brand of honor among sinners. This paper exposits these rules, and culls from them an unauthorized list of the Ten Commandments of applied econometrics.

Keywords. Applied econometrics; Methodology; Data mining.

1. Introduction

It is no secret that there is a world of difference between applied and theoretical econometrics. This difference is dramatically illustrated by a frequently quoted passage from Leamer (1978, p.vi):

> As it happens, the econometric modeling was done in the basement of the building and the econometric theory courses were taught on the top floor (the third). I was perplexed by the fact that the same language was used in both places. Even more amazing was the transmogrification of particular individuals who wantonly sinned in the basement and metamorphosed into the highest of high priests as they ascended to the third floor.

In defense of applied econometricians, this paper maintains that the sinning committed in the profession's basement is not entirely wanton — a set of unwritten rules governs these transgressions, creating a code of honor among econometric sinners. Indeed, 'unwritten' may be too strong a word here: although not all these rules can be found in the econometrics literature, the missing rules appear, albeit piecemeal, in various branches of the applied statistics literature. The purpose of this paper is to bring these rules together, and from them to create an unauthorized list of the ten commandments of applied econometrics. No claim to novelty is made, but a heightened awareness of these rules should improve applied econometric

0950-0804/02/04 0569–21 JOURNAL OF ECONOMIC SURVEYS Vol. 16, No. 4

work, prompt instructors to incorporate them into applied econometrics courses, and convince textbook authors to add them to new editions.[1]

These rules are fundamental to doing competent applied econometric work, so fundamental that to avoid misunderstanding it is necessary to begin by offering some perspective on where they fit in the hierarchy of econometric material. In Leamer's metaphor, econometric theory is taught on the third floor. This consists of things like proving the Gauss-Markov theorem, deriving the Cramer-Rao lower bound, and teaching the details of asymptotics. Theorem-proving and algebraic derivations define this floor.

From the point of view of applied econometricians there are several concerns with this third-floor activity. Leamer (1988) complains that there is too much asymptotics, and that reliance on asymptotics is very dangerous in applied econometrics — practitioners work with finite sample sizes, for which asymptotic results can mislead. Worswick (1972, p. 79) complains that

> Econometricians are not, it seems to me, engaged in forging tools to arrange and measure actual facts so much as making a marvelous array of pretend-tools which would perform wonders if ever a set of facts should turn up in the right form.

The implication of such complaints for applied econometricians is best expressed by Pagan's (1999, p. 374) story of the Zen master and his student, concluding that

> It's not what you know about something that is important, but rather how you use it.

On the second floor we move into the world of applied econometrics, in which students are taught, in hands-on fashion, how to undertake a wide variety of advanced econometric techniques, such as testing for unit roots and cointegration, correcting for sample selection bias, performing Hausman tests, and estimating using Tobit, Poisson, and ordered probit models. To follow Pagan's Zen master story, this is about using the third-floor results, but at an advanced level — the focus is on techniques, not the basics that define basement activity. To many, such as Chatfield (1991, p. 247), this appears odd:

> It is unfortunate that most textbooks concentrate on the easy estimation stage, when trouble is more likely to occur at the earlier specification stage.

Intriligator, Bodkin and Hsiao (1996, p.xiv) agree:

> At least 80% of the material in most of the existing textbooks in econometrics focuses purely on econometric techniques. By contrast, practicing econometricians typically spend 20% or less of their time and effort on econometric techniques per se; the remainder is spent on other aspects of the study, particularly on the construction of a relevant econometric model and the development of appropriate data before estimation and the interpretation of results after estimation.

Moving down to the first floor, we find econometrics courses of a more elementary nature, covering theoretical and applied topics such as the use and interpretation of dummy variables, the logic of F and chi-square tests, and testing and correcting for nonspherical errors. On this floor, as on the higher floors, a common complaint (Wild, 1994; Chatfield, 1995a) is that teaching is technique-oriented rather than problem-oriented. In Bailar's (1988, p. 7) words, instructors 'were not teaching statistical reasoning. They were teaching mechanical manipulations'. An examination of typical econometrics textbooks does little to dispel this view — assignments in these texts mainly set students up to apply a technique, rather than creating for them a real-world scenario in which, undirected, they must address an empirical issue.

Judging by our econometrics textbooks, the first floor is the bottom floor of the econometrics building — there is no floor below this, only a separate building housing introductory statistics courses. But what about the basement of the econometrics building? According to Leamer, this is where real-world econometric modeling is done. Does what we teach on the upper floors adequately prepare students for working in the basement? One theme of this paper is that our students are not well prepared for real-world econometric work. We do not alert them to the true fundamentals of doing applied work (the 'rules for sinning'), prerequisites for successful application of the concepts and techniques they learn on other floors.

It is not hard to find rumblings of discontent among applied econometricians regarding the way in which the subject is taught. Magnus (1999, p. 60) identifies a major concern:

> My worry as an econometric theorist is not that there is tension between us (the theorists) and them (the applied economists). On the contrary, such tension can be healthy and inspiring. My worry is rather the lack of tension. There are two camps, a gap between them, and little communication.

Magnus and Morgan (1999, p. 379) conclude that 'There is clearly a big problem of middle-level instruction'. And Wooldridge (2000, p. iii), notes 'There is a widening gap between the way in which introductory econometrics is taught and the way empirical researchers think about, apply, and interpret econometric methods'. These complaints are directed more at first-and second-floor activity than at basement activity. In contrast, this paper focuses on the basement, proposing rules of applied econometrics behavior that will annoy econometricians — they will view these rules as either unteachable[2] or so elementary that both undergraduate and graduate students surely will have learned them at an earlier stage in their academic career. Readers can judge the former issue for themselves; my opinion is that regardless of teachability, we have a moral obligation to inform students of these rules, and, through suitable assignments, socialize them to incorporate them into the standard operating procedures they follow when doing empirical work. On the latter issue, my many years of experience teaching, practicing, supervising, refereeing and editing applied econometrics (but not my reading of econometrics textbooks!) have led me to believe that these rules are far more important than instructors believe, and that students at all levels do not accord them the respect they deserve. Magnus and Morgan (1999a, p. 1) claim that 'the establishment of a credible applied econometrics is possibly the most important task of econometrics today'. Identifying the 'rules of sinning' is surely a prerequisite to moving the profession in this direction.[3]

2. The Rules of Sinning

Rule #1: Use common sense and economic theory.
The reason for this rule is that common sense is not all that common. Indeed, sometimes it appears that not much thought has gone into empirical work, let alone good thought, as Preece (1987, p. 397) emphasizes:

> The procedures of good statistical practice are founded on experience and commonsense; it is good practice to stop and think before running a regression.

Such thinking does not need to involve complicated econometrics, as is evident in this commentary from Trosset (1998, p. 23):

> I was struck by how often I provided a service without doing anything that an academic researcher would recognize as statistics. Time and again I was thanked (and paid) for asking questions and suggesting perspectives that seemed to me to be little more than common sense. This highly developed common sense is an easily overlooked, but extraordinarily valuable commodity.

In the econometric context, this thinking should cause researchers to match per capita variables with per capita variables, use real exchange rates to explain real imports/exports, employ nominal interest rates to explain real money demand, select appropriate functional forms for dependent variables constrained to lie between zero and one, resist trying to explain a trendless variable with a trended variable, avoid cardinalizing ordered qualitative explanatory variables, beware of the regression fallacy (Friedman, 1992), and never infer causation from correlation.

Hotelling *et al*. (1948, p. 103) have captured the essence of this rule nicely:

> Unfortunately, too many people like to do their statistical work as they say their prayers — merely substitute in a formula found in a highly respected book.

Rule #2: Avoid type III errors.
A type III error, introduced in Kimball (1957), occurs when a researcher produces the right answer to the wrong question.[4] A corollary of this rule, as noted by Chatfield (1995, p. 9), is that an approximate answer to the right question is worth a great deal more than a precise answer to the wrong question. Tukey (1962, p. 13–14) states this in a more accusatory fashion:

> Far better an approximate answer to the *right* question, which is often vague, than an *exact* answer to the wrong question, which can always be made precise.

The phenomenon at issue here is that the relevant objective/hypothesis/specification may be completely different from what is initially suggested. Econometricians experience this regularly when colleagues or students stop by for advice, prefacing their request with words to the effect that they do not want to take up much of the econometrician's time so they will explain just the technical detail with which they want help. Acquiescing to this is usually a mistake, as articulated by Joiner (1982, p. 333)

> We found repeatedly that simple questions about seemingly minor details often bring to light misunderstandings of important issues.

Magnus (1999, pp. 56–7) offers an example, and opines that 'asking the right question is a non-trivial and difficult aspect of research'.

The main lesson here is a blunt one: Ask questions, especially seemingly foolish questions, to ensure that you have a full understanding of the context of the 'technical detail' being discussed; often it turns out that your colleague has not formulated her/his research question appropriately.

Rule #3: Know the context.

This rule is a natural extension of the previous rule. It is crucial that one become intimately familiar with the phenomenon being investigated — its history, institutions, operating constraints, measurement peculiarities, cultural customs, and so on, going beyond a thorough literature review. As Zellner (1992) emphasizes, 'GET THE FACTS'. Again, questions must be asked: Exactly how were the data gathered? Did government agencies impute the data using unknown formulas? What were the rules governing the auction? How were the interviewees selected? What instructions were given to the participants? What accounting conventions were followed? How were the variables defined? What is the precise wording of the questionnaire? How closely do measured variables match their theoretical counterparts?

Data are numbers with a context (Moore, 1990); know the context! Tweedie *et al*. (1998) and Pfannkuch and Wild (2000) provide examples of how a careful examination of the data-generating procedure has led to substantive insights. Burdekin and Burkett (1998) and Wilcox (1992) are examples of how not knowing the context can lead to error.

Belsley and Welch (1988, p. 447) have summarized this neatly:

> Don't try to model without understanding the nonstatistical aspects of the real-life system you are trying to subject to statistical analysis. Statistical analysis done in ignorance of the subject matter is just that — ignorant statistical analysis.

Rule #4: Inspect the data.

Even if a researcher knows the context, s/he needs to become intimately familiar with the specific data with which s/he is working. Economists are particularly prone to the complaint that researchers do not know their data very well, a phenomenon made worse by the computer revolution, allowing researchers to obtain and work with data electronically by pushing buttons. Typical of these complaints are Reuter (1982, p. 137)

> Economists are unique among social scientists in that they are trained only to analyse, not collect, data ... One consequence is a lack of skepticism about the quality of data ...

and Maier (1999, p. 5)

> It is a recurring complaint in the social sciences that researchers, from the student in training to the advanced scholar, do not know enough about the data they use.

Inspecting the data involves summary statistics, graphs, and data cleaning, to both check and 'get a feel for' the data. Summary statistics can be very simple, such as calculating means, standard errors, maximums, minimums, and correlation matrices, or more complicated, such as computing condition indices (Belsley, 1991) and influential observation diagnostics (Belsley, Kuh and Welch, 1980).

The advantage of graphing is that graphics broadcast whereas statistics narrowcast, or, as Tukey (1977, p.vi) notes: 'The greatest value of a picture is when it forces us to notice what we never expected to see'. With all due respect to Tukey, however, his exploratory data analysis (EDA), cannot be recommended — it is evident that many statisticians (Ehrenberg, 1979), and especially econometricians, simply will not use the range of techniques he advocates. But the spirit or 'attitude' of EDA, as described by Cobb (1987, p. 329) is crucial:

> I find it useful to distinguish exploratory techniques such as stem-and-leaf diagrams and box plots, from exploratory attitudes: Does an author pay attention to such things as residuals, outliers, and the possible value of transforming? The former (techniques) are comparatively superficial, but the latter (residuals, outliers, transforming) lie close to the heart of data analysis.

What is recommended here is that the attitude of EDA be adopted — researchers should supplement[5] their summary statistics with simple graphs: histograms, residual plots, scatterplots of residualized data, and graphs against time.[6]

Data cleaning looks for inconsistencies in the data — are any observations impossible, unrealistic, or suspicious? According to Rao (1997, p. 152), 'Every number is guilty unless proved innocent'. Day and Liebowitz (1998) is a classic example of such detective work. The questions here are mostly simple, but could become more complicated in a particular context. Do you know how missing data were coded? Are dummies all coded zero or one? Are any observations born in two different months? Do all observations obey logical constraints they must satisfy? Hamermesh (2000, p. 366) laments that 'Physicians bury their medical mistakes in the ground; we bury mistakes in our data under a welter of econometric technique', and urges (p. 364) econometricians to put 'data cleanliness ahead of econometric godliness'.

Chatfield (1985, p. 230) summarizes the main message of this rule:

> Students should be taught that instead of asking 'what technique shall I use here?' they should ask 'How can I summarize and understand the main features of this set of data?'

Rule #5: Keep it sensibly simple.

This KISS rule, based on Zellner's (1992, 2001) 'keep it sophisticatedly simple' rule, should not, as Zellner warns, be confused with the commercial 'keep it simple, stupid' rule, because some simple models are stupid, containing logical errors or being at variance with facts. Zellner notes that, as in science, progress in economics results from beginning with simple models, seeing how they work in applications, and then modifying them if necessary. As examples he cites functional form specifications of Nobel Laureates — Tinbergen's social welfare functions, Arrow's and Solow's work on the CES production function, Friedman's, Becker's, Tobin's and Modigliani's consumer models, and Lucas's rational expectations model.

Beginning an analysis with simple models can be defended on many grounds. Doing so is consistent with the history of scientific inference/progress; simple models are characterized by stronger priors (Zellner, 2001); model construction costs are lower; the dirty, nonexperimental data of economics require simple models that do not place unrealistic demands on the data, or, as noted by Griliches (1985, p. 199), estimation of sophisticated econometric models is 'much more likely to be sensitive to errors and inconsistencies in the data'; empirical evidence in the forecasting literature shows that simple models outforecast more complicated models (Makridakis and Hibon, 2000); sources of model failure are easier to detect; learning how and why a simple model performs poorly is important information for the model development process; 'simple analyses are easier to explain than complex ones — and are often less likely to lead to serious blunders or oversights' (Joiner, 1982, p. 339); and subjective insights, essential ingredients of discovery, are facilitated, as noted by Wainer (1997, p. 3) 'revelation accompanies simplicity' and Keuzenkamp and McAleer, (1995, p. 16) 'in empirical investigations, simple models can inspire knowledge, while complex models rarely do'.

Keuzenkamp and McAleer (1995) present a thorough and persuasive defense of simplicity, addressing a variety of thorny issues such as how simplicity should be defined. Zellner (2001) also addresses the issue of defining simplicity.

The role of simple specifications in econometrics is most dramatically illustrated by explaining how econometricians develop their specifications. Most econometricians will say that they believe in the 'top-down' or 'general-to-specific' approach — begin with a large/complex model and through judicious testing pare it down to a final specification. The problem with this approach, in Magnus's (1999, pp. 61–2) words,

> ... is that it does not work. If you try to estimate such a large model, which has everything in it that you can think of, you get nonsensical results.

> Everyone who has done empirical econometric work knows this. The second problem is that you cannot really discover anything new and interesting in this way, because the interesting bit is the construction of the top model and we are not told how this is done. Therefore no applied economist proceeds in this way. Instead they follow the bottom-up approach. In the bottom-up approach one starts with a simple model and builds up from there. This is, in fact, how scientists in other disciplines work.

Magnus goes on to lament how this phenomenon has given rise to an unfortunate tendency for results to be presented as though they had come from a top-down approach.

The main advantage of the general-to-specific approach is that if the general model incorporates the true model generating the data, then testing is unbiased. But no such true model can ever be found, so this advantage is questionable. Furthermore, as Magnus notes above, it is not realistic to think that we can estimate a general model incorporating all conceivable explanatory variables and functional forms. Keuzenkamp and McAleer (1995, pp. 15–18) provide a cogent critique of the general-to-specific methodology.

With so many practitioners claiming to be using the top-down approach, but appearing to be using a bottom-up approach, suspicions of serious sinning abound. This sinning can be roughly described as follows. Practitioners begin with simple models which are expanded whenever they fail. Failures are identified through misspecification tests such as evaluation of out-of-sample forecasts. Expansions are on the one hand modest in that they introduce one extra layer of complexity (a new variable, for example), but on the other hand quite general in that they cover a range of possible roles for the new element (generous lags, for example), as degrees of freedom allow. Testing down is undertaken to create a new simple model which is subjected to misspecification tests, and this process of discovery is repeated. In this way simplicity is combined with the general-to-specific methodology, producing a compromise process which, judged by its wide application, is viewed as acceptable sinning. (The related sin of data mining is discussed in Rule 7 below.)

The conflict between simplicity and complexity arises in another context. Replacing Zellner's 'sophisticatedly' with 'sensibly' is to guard against the former being interpreted as license to employ the latest, most sophisticated econometric techniques, something that frequently happens because such techniques are novel and available, not because they are appropriate. Only when faced with obvious problems such as simultaneity or selection bias should more advanced techniques be employed, and then, as emphasized by Hamermesh (2000, p. 378), only after a benefit-cost calculation has been applied, as he illustrates from his own work. Wilkinson *et al.* (1999, p. 598) underline this view:

> Do not choose an analytic method to impress your readers or to deflect criticism. If the assumptions and strength of a simpler method are reasonable for your data and research problem, use it. Occam's razor applies to methods as well as to theories.

Rule #6: Use the interocular trauma test.

Output from modern empirical work typically fills many pages, as researchers try a variety of functional forms and sets of explanatory variables. This rule cautions researchers to look long and hard at this plethora of results: look at the results until the answer hits you between the eyes! Part of this rule is to check that the results make sense. Are the signs of coefficients as expected? Are important variables statistically significant? Are coefficient magnitudes reasonable? Are the implications of the results consistent with theory? Are there any anomalies? Are any obvious restrictions evident? Apply the 'laugh' test, what Hamermesh (2000, p. 374) calls the 'sniff' test: 'ask oneself whether, if the findings were carefully explained to a thoughtful layperson, that listener could avoid laughing'.

But another part of this rule is more subtle, and subjective. By looking long and hard at reams of computer output, a researcher should eventually, through both conscious and subconscious means, recognize the message they are conveying (which could be a negative message) and become comfortable with it. This subjective procedure should be viewed as separate from and complementary to formal statistical testing procedures used to investigate what is going on. Indeed, the results of such testing procedures form part of the mass of statistical output one is trying to interpret.[7]

Rule #7: Understand the costs and benefits of data mining.

Mukherjee *et al*. (1998, p. 30) claim that historically

> ... any attempt to allow data to play a role in model specification ... amounted to data mining, which was the greatest sin any researcher could commit.

On the other hand, Hoover (1995, p. 243) maintains that

> ... data mining is misunderstood, and once it is properly understood, it is seen to be no sin at all.

Who is right here — is data mining a sin, or not? Both are right — some variants of 'data mining' can be classified as the greatest of the basement sins, but other variants of 'data mining' can be viewed as important ingredients in data analysis. Unfortunately, these two variants usually are not mutually exclusive and so frequently conflict in the sense that to gain the benefits of the latter, one runs the risk of incurring the costs of the former.

Hoover and Perez (2000, p. 196) offer a general definition of data mining as referring to 'a broad class of activities that have in common a search over different ways to process or package data statistically or econometrically with the purpose of making the final presentation meet certain design criteria'. Two markedly different views of data mining lie within the scope of this general definition. One view of 'data mining' is that it refers to experimenting with (or 'fishing through') the data to produce a specification. Kramer and Runde (1997) present an instructive example. The problem with this, and why it is viewed as a sin, is that such a procedure is almost guaranteed to produce a specification tailored to the peculiarities of that particular data set, and consequently will be misleading in terms of what it says about the underlying process generating the data. Furthermore, traditional testing procedures used to 'sanctify' the specification are no longer legitimate, because these data, since they have been used to generate the specification, cannot be judged impartial if used to test that specification.

This objection to data mining is the basis for complaints about Hendry's (1980, p. 403) edict that 'The three golden rules of econometrics are test, test, and test'. Applied indiscriminately, these rules, and their associated 'general to specific' specification search methodology, can lead to problems such as pretest bias and distortion of type I error rates, jokes about having more test statistics than observations, and reference to Ronald Coase's oft-cited comment that 'If you torture the data long enough, Nature will confess'. A more sympathetic view recognizes (as Hendry has always maintained), that model specification should not blindly follow testing procedures — that it needs to be a well-thought-out combination of theory and data, and that testing procedures used in such specification searches should be designed to minimize the costs of data mining. Examples of such procedures are setting aside data for out-of-sample prediction tests, adjusting significance levels, and avoiding questionable criteria such as maximizing R^2. Hoover and Perez (1999), and associated commentary, provide a good summary of recent innovations on this front, and of related criticisms. Hendry and Mizon (1990) is an excellent discussion of the issues surrounding the testing ingredient of model specification.

An alternative view of 'data mining' is that it refers to experimenting with (or 'fishing through') the data to discover empirical regularities that can inform economic theory. This approach to data mining, likened by some to Exploratory Data Analysis (Tukey, 1977), has been welcomed into the mainstream of statistical analysis by the recent launching of the journal *Data Mining and Knowledge Recovery*. Hand *et al.* (2000) describe data mining as the process of seeking interesting or valuable information in large data sets. Its greatest virtue is that it can uncover empirical regularities that point to errors/omissions in theoretical specifications, an example of which is described by Kennedy (1998, p. 87). The spirit of this approach is captured by Thaler's (2000, p. 139) remark that 'Some economists seem to feel that data-driven theory is, somehow, unscientific. Of course, just the opposite is true'. Pena (2001, p. 9) quotes George Box as emphasizing 'the necessity to continually change the model as one's understanding develops'. The art of the applied econometrician is to allow for data-driven theory while avoiding the considerable dangers inherent in data mining.

In summary, this second type of 'data mining' identifies regularities in or characteristics of the data that should be accounted for and understood in the context of the underlying theory. This may suggest the need to rethink the theory behind one's model, resulting in a new specification founded on a more broad-based understanding. This is to be distinguished from a new specification created by mechanically remolding the old specification to fit the data; this would risk incurring the costs described earlier when discussing the first variant of 'data mining'.

The issue here is how should the model specification be chosen? As usual, Leamer (1996, p. 189) has an amusing view:

> As you wander through the thicket of models, you may come to question the meaning of the Econometric Scripture that presumes the model is given to you at birth by a wise and beneficent Holy Spirit.

In practice, model specifications come from both theory and data, and given the absence of Leamer's Holy Spirit, properly so. An important ingredient in specification searches, often forgotten, is captured by Smith's (1998) question 'Why are you doing this?' and spelled out by Magnus (1999, p. 61)

> The best we can hope for is that a model is valid locally. This implies that the model should depend on the central question which the researcher wishes to answer ... Everything else — your model, the data that you need, your estimation method — depends on it. Now, this may seem obvious, but it is not obvious to most econometricians.

If this is not obvious to econometricians, it quickly becomes obvious the moment they undertake applied work, as noted by Feldstein (1982, p. 829):

> The applied econometrician, like the theorist, soon discovers from experience that a useful model is not one that is 'true' or 'realistic' but one that is parsimonious, plausible, and informative.

The process by which a specification is developed, blending economic theory, common sense, and a judicious mixture of both bottom-up and top-down, clearly incorporates elements of 'data mining', a terminology with strong emotive content. In preliminary versions of this paper econometricians complained about this rule, originally titled 'Data Mine with Care', more than any other. The intent of this rule is to admit that data mining is here to stay, and in light of that, ensure that practitioners are fully aware of its potential costs. This view is not new to the economics literature. Backhouse and Morgan (2000, p. 176) summarize a symposium on data mining, published in the June 2000 issue of the *Journal of Economic Methodology*, by noting that some of the papers in this symposium advocate an increase in data mining but 'hedge this with strong warnings about the need for such data mining to be undertaken in a suitable manner'.

Rule #8: Be prepared to compromise.

This rule reflects Trosset's (1998, p. 24) observation that

> In virtually every application, there is a gap (often a vast gulf) between the details of the application and standard statistical theory ... gaps between application and theory require that compromise be made.

Leamer (1997, p. 552) lends it special emphasis when listing his choices for the three most important aspects of real data analyses: 'compromise, compromise, compromise'. Its most eloquent expression in the econometrics literature, however, is due to Valavanis (1959, p. 83):

> Econometric theory is like an exquisitely balanced French recipe, spelling out precisely with how many turns to mix the sauce, how many carats of spice to add, and for how many milliseconds to bake the mixture at exactly 474 degrees of temperature. But when the statistical cook turns to raw materials, he finds that hearts of cactus fruit are unavailable, so he substitutes chunks of cantaloupe; where the recipe calls for vermicelli he used shredded wheat; and he substitutes green garment dye for curry, ping-pong balls for turtle's eggs, and for Chalifougnac vintage 1883, a can of turpentine.

The issue here is that on the third floor students are taught standard solutions to standard problems, but, as noted by Joiner (1982, p. 335) 'In practice, there are no standard problems, only standard solutions', and echoed by McCabe (1982, p. 373) 'Usually when I identify a situation as having a clear problem with a precise optimal solution, I have made a mistake'.

Applied econometricians are continually faced with awkward compromises, and must be willing to make ad hoc modifications to standard solutions. Should a proxy be used? Can sample attrition be ignored? Should these unit root tests be believed? Is aggregation legitimate here?

Rule #9: Do not confuse statistical significance with meaningful magnitude. Very large sample sizes, such as those that have become common in cross-sectional data, thanks to the computer revolution, can give rise to estimated coefficients with very small standard errors. A consequence of this is that coefficients of trivial magnitude may test significantly different from zero. In the psychometric literature this problem has given rise to books entitled 'What if there were no significance tests?' (Harlow *et al*., 1997) and journal policies not to publish papers that do not report effect size (the magnitude of a treatment's impact, usually measured in terms of standard deviations of the phenomenon in question). The flavor of this controversy can be summed up by Rosnow and Rosenthal's (1989, p. 1277) comment that 'Surely, God loves the .06 nearly as much as the .05', and by Loftus's (1993, p. 250) opinion that 'hypothesis testing is overstated, overused and practically useless as a means of illuminating what the data in some experiment are trying to tell us'. Nester (1996) has a collection of similar quotes berating significance testing.

In econometrics, Leamer (1978) has given formal representation to this phenomenon, by relating significance level to sample size, and McCloskey (1998, chap.8) summarizes her several papers on the subject, chastising the profession for its tendency to pay undue homage to significance testing. Tukey (1969) views this use of significance testing as 'sanctification' of a theory, with a resulting unfortunate tendency for researchers to stop looking for further insights. McCloskey and Ziliak (1996, p. 112) cogently sum up this view as follows:

> No economist has achieved scientific success as a result of a statistically significant coefficient. Massed observations, clever common sense, elegant

> theorems, new policies, sagacious economic reasoning, historical perspective, relevant accounting: these have all led to scientific success. Statistical significance has not.

Leamer (1996, p. 176) believes that an important reason for this state of affairs is that

> A distressing amount of data analysis proceeds without benefit of any clear questions. When the questions are not explicitly on the table it is easy to fantasize that models are either true or false when in fact they are sometimes useful for answering the questions and sometimes misleading.

He maintains that the interesting questions are not sharp hypotheses in which the null takes the form of a specific parameter value (because all such hypotheses are surely false), but rather that the questions economists deal with, or should be dealing with, are neighborhoods of specific parameter values. Recognizing this can alleviate many of the criticisms of significance testing.

A promising attitude towards this problem is articulated by Milton Friedman, as quoted by Hamermesh (2000, p. 376):

> I have long had relatively little faith in judging statistical results by formal tests of significance. I believe that it is much more important to base conclusions on a wide range of evidence coming from different sources over a long period of time.

Sanctification via significance testing should be replaced by searches for additional evidence, both corroborating evidence, and, especially, disconfirming evidence. If your theory is correct, are there testable implications? Can you explain a range of interconnected findings? Can you find a bundle of evidence consistent with your hypothesis but inconsistent with alternative hypotheses? Abelson (1995, p. 186) offers some examples. A related concept is encompassing: Can your theory encompass its rivals in the sense that it can explain other models' results? See Hendry (1988). Zellner (1992) recommends making a special effort to obtain unusual facts.

Rule #10: Report a sensitivity analysis.
Everyone knows, as expressed by Weick (1992, p. 100), that

> Presentations of research findings are usually notoriously misleading accounts of how the research itself was conducted.

Because of this it is very difficult for readers of research papers to judge the extent to which data mining may have unduly influenced the results. Indeed, results tainted by subjective specification decisions undertaken during the heat of econometric battle should be considered the rule, as John Maynard Keynes (1940, p. 155) recognized long ago:

> It will be remembered that the seventy translators of the Septuagint were shut up in seventy separate rooms with the Hebrew text and brought out with

> them, when they emerged, seventy identical translations. Would the same miracle be vouchsafed if seventy multiple correlators were shut up with the same statistical material?

From these observations stem two important prescriptions. First, researchers should explain fully their specification search so that readers can judge for themselves how the results may have been affected. Mayer (1993, chap. 10) makes an amusing suggestion that captures the flavor of this debate — researchers should publish all the regressions that were run, not just the good ones! This is basically an 'honesty is the best policy' approach, advocated by Leamer (1978, p. vi):

> Sinners are not expected to avoid sins; they need only confess their errors openly.

Second, a sensitivity analysis should be reported, indicating to what extent the substantive results of the research are affected by adopting different specifications about which reasonable people might disagree. For example, are the results sensitive to the sample period, the functional form, the set of explanatory variables, or measurement of or proxies for the variables? Levine and Renelt (1992) is a notorious example. Abelson (1995) stresses that anticipation of criticism is fundamental to good research and data analysis.

3. The Ten Commandments of Applied Econometrics

With apologies to Thomas (1976) and to Driscoll (1977), I have taken the liberty of culling from these rules Ten Commandments for applied econometricians.

1. **Thou shalt use common sense and economic theory.**
   Corollary: Thou shalt not do thy econometrics as thou sayest thy prayers.
2. **Thou shalt ask the right questions.**
   Corollary: Thou shalt place relevance before mathematical elegance.
3. **Thou shalt know the context**.
   Corollary: Thou shalt not perform ignorant statistical analyses.
4. **Thou shalt inspect the data**.
   Corollary: Thou shalt place data cleanliness ahead of econometric godliness.
5. **Thou shalt not worship complexity**.
   Corollary: Thou shalt not apply asymptotic approximations in vain.
   Corollary: Thou shalt not talk Greek without knowing the English translation.
6. **Thou shalt look long and hard at thy results**.
   Corollary: Thou shalt apply the laugh test.
7. **Thou shalt beware the costs of data mining**.
   Corollary: Thou shalt not worship R^2.
   Corollary: Thou shalt not hunt statistical significance with a shotgun.
   Corollary: Thou shalt not worship the 0.05% significance level.

8. **Thou shalt be willing to compromise**.
   Corollary: Thou shalt not worship textbook prescriptions.
9. **Thou shalt not confuse significance with substance**.
   Corollary: Thou shalt not ignore power.
   Corollary: Thou shalt not test sharp hypotheses.
   Corollary: Thou shalt seek additional evidence.
10. **Thou shalt confess in the presence of sensitivity**.
    Corollary: Thou shalt anticipate criticism.

4. Conclusions

Knowing the 'rules of sinning' presented above is not enough, as the anecdote presented earlier in footnote 6 illustrates. Inspecting the data requires knowing how to inspect, what to look for and how to interpret what is found, not to mention remembering to look; the interocular trauma test seems trivial, but is hard to perform; knowing that it is necessary to compromise does not mean that a researcher knows how to compromise. Magnus and Morgan (1999, p. 378) sum this up nicely:

> Learning to be an applied econometrician apparently involves learning the appropriate rules of scientific procedure and when to ignore them, growing the expertise which enables choices, decisions and judgements to be made almost unconsciously, and learning the tacit skills to undertake the work.

The bottom line here is that much of the skill of sinning in acceptable fashion is judgemental, subjective, and difficult, as articulated by Welch (1986, p. 405):

> Even with a vast arsenal of diagnostics, it is very hard to write down rules that can be used to guide a data analysis. So much is really subjective and subtle ... A great deal of what we teach in applied statistics is not written down, let alone in a form suitable for formal encoding. It is just simply 'lore'.

It seems that such 'lore' can only be learned by experience and by watching the masters[8] — in short, by sweating in the basement. But even this may not be enough, as Pagan (1987, p. 20) warns us:

> Few would deny that in the hands of the masters the methodologies perform impressively, but in the hands of their disciples it is all much less convincing.

For those of us who are disciples, learning the 'rules of sinning' is a crucial prerequisite to doing quality applied econometrics. Judging by econometrics textbooks, and by what students from around the world have told me about what happens in their 'applied' econometrics classes, the profession is placing inadequate emphasis on the fundamentals as practiced in the basement. How has this state of affairs come about? Four possible reasons follow.

First, in the words of Hendry and Richard (1983, p. 112), doing applied work is difficult:

> The data generation process is complicated, data are scarce and of uncertain relevance, experimentation is uncontrolled, and available theories are highly abstract and rarely uncontroversial.

Verbeek (2000, p. 1) has a blunter way of expressing this: 'Econometrics is much easier without data'. These observations explain comments such as the following by Wickens (1997, p. 523):

> After a mere 30 years as an applied econometrician (a lowly subspecies rarely to be found in the halls of fame of *Econometrica*) I am still wrestling with the problem of how to combine theory with evidence.

The Wickens quote points to a second reason for lack of emphasis on the basics of applied work — doing quality applied work brings little prestige in the profession, a fact lamented by Hendry and Leamer: 'As a profession we don't value empirical work very highly' (Hendry *et al*., 1990, p. 180).

A third reason is apparent from an earlier quote from Magnus (1999), lamenting the lack of tension, and thus lack of communication, between econometric theorists and applied economists. It is natural, and common, for economics departments to assign econometric theorists to teach applied econometrics courses. If there is little communication between econometric theorists and applied econometricians, it is not surprising that the 'rules of sinning' are not taught.

Fourth, as noted by J. Bailar (1994, p. 18), 'We teach what we enjoy teaching and what we know how to teach, not what the world needs'. This is probably the crux of the problem: econometrics instructors believe their students don't need to be taught the 'rules of sinning' described above, and in any event they would be uncomfortable teaching them because these rules don't have the intellectual rigor and clean answers so prized by econometric theorists. A more controversial view is expressed by B. Bailar (1988, p. 7):

> Part of the problem is that many of the people who teach, whether in undergraduate or graduate courses, have little or no practical experience in what they teach: Their horizons are largely limited to what appears in journals and textbooks.

The teaching task is unquestionably difficult. Tukey (1969, p. 90) expresses this difficulty eloquently:

> Divert them from depending upon the 'authority' of standard textbook solutions, but without being able to substitute a second religion for the first. Stimulate intelligent problem formulation, without being able to say quite how this is done. Demand high standards of statistical reasoning, but without specifying a simple model of statistics which might serve as a criterion for quality of reasoning.

There is some hope on the horizon, however. As advances in computer technology lower the cost of doing and teaching applied work, as more journals provide access to the data used in their published articles, and as more authors feel an obligation to include CD ROMs full of data with their textbooks, teaching of applied econometrics may move toward adopting the 'rules of sinning' as important topics. But it will not be enough simply to list these rules, as has been done here, or provide empirical examples with real-world data. Instructors will have to design assignments that force students to fend for themselves in real-world contexts, with vague, general instructions, rather than specific step-by-step directions telling them what to do. These assignments are not easy to design — they must be focussed on specific issues that should become apparent to students as they investigate the data, and they must have model answers showing the steps that students should have taken and the results that they should have uncovered. Such learning by doing is how students will acquire the 'mental habits' advocated by Wild (1994), reflecting the insight of Magnus and Morgan (1999, p. 377) that 'we learn by being socialized, not instructed'.

Acknowledgements

For comments on earlier drafts of this paper I am indebted to Bill Becker, David Belsley, Chris Chatfield, David Colander, David Hendry, Ed Leamer, Lonnie Magee, Jan Magnus, Michael McAleer, Les Oxley, Ron Smith, Michael Veall, Jeff Wooldridge, and Arnold Zellner, some of whom objected strongly to some of its contents.

Notes

1. Econometricians seeing early drafts of this paper have wondered if it is meant to be serious. It is not a serious paper in the sense that it contributes to the scholarly debate on econometric methodology, a realm in which sinning is not allowed. It is a serious paper in the sense that it offers commentary and advice relevant to the real world of applied econometrics, where sinning is ubiquitous.
2. Remarkably, some believe that applied statistics should not be taught because it is unteachable, better learned on the job after graduation, and has too high an opportunity cost; see Anderson and Loynes (1987) for references and rebuttals.
3. Smith (1998) presents some similar rules, but in a context of advising researchers to ask themselves questions such as 'Why am I doing this?' and 'How much would you bet on the predictions of this model?' He believes (personal correspondence) 'that the most prevalent sin in applied econometrics is mechanical application of rules and procedures in inappropriate settings' and so is against proposing rules, however well-intentioned, for fear 'that teachers will quickly convert them into specific injunctions, which students will then apply inappropriately'. My intention is that these rules be generic guidelines for behavior, to be applied in context.
4. This is not to be confused with psychologists' type III error, introduced in Kaiser (1960), concerned with concluding significance in the wrong direction. Political scientists have yet another definition (unwritten) of type III error — confusing type I and type II errors! A type IV error is providing the wrong answer to the wrong question.
5. Graphing is advocated as a supplement to summary statistics, not as a substitute.

McAleer (1999) argues the advantages of statistical diagnostics over graphs, noting, among other things, that eyes can play tricks. On the other hand, Hartwig and Dearing (1979, p. 16) note that 'It should be remembered that numeric summaries are just that; they summarize characteristics of distributions. Thus, the analysis should begin with the data, not with summaries of the data'.

6. An anecdote can lend some perspective to this, and to the reason why these rules need emphasis. Students in my graduate applied econometrics course have all had a third-floor theory course. Their first assignment is to perform some Chow testing on a variant of the infamous Anscombe (1973) data, for which inspecting the data reveals how misleading traditional test statistics can be. Despite my extensive first lecture on inspecting the data, invariably these students manipulate these data electronically, without looking at them!
7. On a personal note, this rule has had the biggest impact on my own work. I am continually amazed at how after several days of worrying about a mass of results, a sense of what is going on with the data becomes apparent, often in revelation-like fashion.
8. Despite Welch's claim that it is very difficult to write down rules for formal encoding, Hendry (2000, chap. 20) believes that the PCGET's software, automating the 'general to specific' model selection procedure, is very successful.

References

Abelson, R. P. (1995) *Statistics as Principled Argument*. Hillsdale, NJ: Lawrence Erlbaum.
Anderson, C. W. and Loynes, R. M. (1987) *The Teaching of Practical Statistics*. New York: Wiley.
Anscombe, F. J. (1973) Graphs in statistical analysis. *American Statistician* 27, 1, 17–21.
Backhouse, R. E. and Morgan, M. S. (2000) Introduction: is data mining a methodological problem? *Journal of Economic Methodology*, 7, 171–81.
Bailar, B. A. (1988) Statistical practice and research: the essential interactions. *Journal of the American Statistical Association*, 83, 1–8.
Bailar, J. C. (1994) A larger perspective. In J. R. Kettenring *et al*., (eds), *Modern Interdisciplinary University Statistics Education* (pp. 18–20) Washington, DC: National Academy Press.
Belsley, D. A. (1991) *Conditioning Diagnostics*. New York: Wiley.
——, Kuh, E. and Welch, R. E. (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley.
—— and Welch, R. E. (1988) Modelling energy consumption — using and abusing regression diagnostics. *Journal of Business and Economic Statistics*, 6, 442–7.
Burdekin, R. C. K. and Burkett, P. (1998) Economic history and econometrics: a cautionary note from the hyperinflation front. *Applied Economics Letters*, 5, 251–4.
Chatfield, C. (1985) The initial examination of data. *Journal of the Royal Statistical Society*, A148, 214–31.
—— (1991) Avoiding statistical pitfalls. *Statistical Science*, 6, 240–68.
—— (1995) *Problem Solving: A Statistician's Guide* (2nd edn). London: Chapman and Hall.
—— (1995a) Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society*, A158, 419–44.
Cobb, G. W. (1987) Introductory textbooks: a framework for evaluation. *Journal of the American Statistical Association*, 82, 321–39.
Day, T. E. and Liebowitz, S. J. (1998) Mortgage lending to minorities: where's the bias?' *Economic Inquiry*, 36, 3–28.

Driscoll, M. (1977) The ten commandments of statistical inference. *American Mathematical Monthly*, 84, 628.
Ehrenberg, A. S. C. (1979) Book review. *Applied Statistics*, 28, 79–83.
Feldstein, M. S. (1982) Inflation, tax rules and investment: some econometric evidence. *Econometrica*, 50, 825–62.
Friedman, M. (1992) Do old fallacies ever die? *Journal of Economic Literature*, 30, 2129–32.
Griliches, Z. (1985) Data and econometricians — the uneasy alliance. *American Economic Review, Papers and Proceedings*, 75, 196–200.
Hamermesh, D. S. (2000) The craft of labormetrics. *Industrial and Labor Relations Review*, 53, 363–80.
Hand, D. J., Blunt, G., Kelly, M. G. and Adams, N. M. (2000) Data mining for fun and profit. *Statistical Science*, 15, 111–31.
Harlow, L. L., Mulaik, S. A. and Steiger, J. H. (1997) *What If There Were No Significance Tests?* Mahwah, NJ: Lawrence Erlbaum.
Hartwig, F. and Dearing, B. E. (1979) *Exploratory Data Analysis*. Newbury Park, CA: Sage Publications.
Hendry, D. F. (1980) Econometrics — alchemy or science? *Economica*, 47, 387–406.
—— (1988) Encompassing. *National Institute Economic Review*, 125, 88–92.
—— (2000) *Econometrics: Alchemy or Science?* (new edition). Oxford, UK: Oxford University Press.
——, Leamer, E. E. and Poirier, D. J. (1990) The ET dialogue: a conversation on econometric methodology. *Econometric Theory*, 6, 171–261.
——, and Mizon, G. E. (1990) Procrustean econometrics: or stretching and squeezing data. In C. W. J. Granger (ed.), *Modelling Economic Series: Readings in Econometric Methodology* (pp. 121–36). Oxford: Oxford University Press.
——, and Richard, J. F. (1983) The econometric analysis of economic time series. *International Statistical Review*, 51, 111–48.
Hoover, K. D. (1995) In defense of data mining: some preliminary thoughts. In K. D. Hoover and S. M. Sheffrin (eds), *Monetarism and the Methodology of Econonomics: Essays in Honor of Thomas Mayer* (pp. 242–57). Aldershot: Edward Elgar.
—— and Perez, S. J. (1999) Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *Econometrics Journal* 2, 167–91.
—— (2000) Three attitudes towards data mining. *Journal of Economic Methodology*, 7, 195–210.
Hotelling, H., Bartky, W., Deming, W. E., Friedman, M. and Hoel, P. (1948) The teaching of statistics. *Annals of Mathematical Statistics*, 19, 95–115.
Intriligator, M. D., Bodkin, R. G. and Hsiao, C. (1996) *Econometric Models, Techniques, and Applications* (2nd edn). Upper Saddle River, NJ: Prentice Hall.
Joiner, B. (1982) Practicing statistics, or what they forgot to say in the classroom. In J. S. Rustagi and D. A. Wolfe (eds), *Teaching of Statistics and Statistical Consulting* (pp. 327–42). New York: Academic Press.
Kaiser, H. F. (1960) Directional statistical decisions. *Psychological Review*, 67, 160–67.
Kennedy, P. E. (1998) *A Guide to Econometrics* (4th edn). Cambridge, MA: MIT Press.
Keuzenkamp, H. A. and McAleer, M. (1995) Simplicity, scientific inference and econometreic modelling. *Economic Journal*, 105, 1–21.
Keynes, J. M. (1940) On method of statistical research: comment. *Economic Journal*, 50, 154–6.
Kimball, A. W. (1957) Errors of the third kind in statistical consulting, *Journal of the American Statistical Association*, 52, 133–42.
Kramer, W. and Runde, R. (1997) Stocks and the weather: an exercise in data mining or yet another capital market anomaly. *Empirical Economics*, 22, 637–41.
Leamer, E. E. (1978) *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: John Wiley.

—— (1988) Things that bother me. *Economic Record*, 64, 331–5.
—— (1996) Questions, theory and data. In S. G. Medema and W. Samuels (eds), *Foundations of Research in Economics: How Do Economists Do Economics?* (pp. 175–90). Aldershot, UK: Edward Elgar.
—— (1997) Revisiting Tobin's 1950 study of food expenditure. *Journal of Applied Econometrics*, 12, 533–53.
Levine, R. and Renelt, D. (1992) A sensitivity analysis of cross-country growth regressions. *American Economic Review*, 82, 942–63.
Loftus, G. R. (1993) A picture is worth a thousand p values: on the irrelevance of hypothesis testing in the microcomputer age. *Behavior Research Methods, Instruments and Computers*, 25, 250–6.
Magnus, J. R. (1999) The success of econometrics. *De Economist*, 147, 55–71.
—— and Morgan, M. (1999) Lessons from the tacit knowledge experiment. In J. Magnus and M. Morgan (eds), *Methodology and Tacit Knowledge: Two Experiments in Econometrics* (pp. 375–81). Chichester, UK: Wiley.
—— and Morgan, M. (1999a) Introduction: applied econometrics and design of two experiments. In J. Magnus and M. Morgan (eds), *Methodology and Tacit Knowledge: Two Experiments in Econometrics* (pp. 1–10). Chichester, UK: Wiley.
Maier, M. (1999) *The Data Game* (3rd edn). Aramonk, NY: M. E. Sharpe.
Makridakis, S. and Hibon, M. (2000) The M3 competition: results, conclusions and implications. *International Journal of Forecasting*, 16, 451–76.
Mayer, T. (1993) *Truth versus Precision in Economics*. Aldershot, UK: Edward Elgar.
McAleer, M. (1997) Comments. *Journal of Applied Econometrics*, 12, 553–7.
McCabe, G. P. (1982) Statistical consulting: reaction and comments. In J. S. Rustagi and D. A. Wolfe (eds), *Teaching of Statistics and Statistical Consulting* (pp. 371–4). New York: Academic Press.
McCloskey, D. N. (1998) *The Rhetoric of Economics*. Madison, WI: University of Wisconsin Press.
McCloskey, D. N. and Ziliak, S. T. (1996) The standard error of regression. *Journal of Economic Literature* 34, 97–114.
Moore, D. A. (1990) Uncertainty. In L. A. Steen (ed.), *On the Shoulders of Giants: A New Approach to Quantitative Literacy* (pp. 95–137). Washington, DC: National Academy of Sciences.
Mukherjee, C., White, H. and Wuyts, M. (1998) *Econometrics and Data Analysis for Developing Countries*. London: Routledge.
Nester, M. R. (1996) An applied statistician's creed. *Applied Statistics*, 45, 401–10.
Pagan, A. R. (1987) Three econometric methodologies: a critical appraisal. *Journal of Economic Surveys*, 1, 3–24.
—— (1999) The Tilburg experiments: impressions of a drop-out' In J. Magnus and M. Morgan (eds), *Methodology and Tacit Knowledge: Two Experiments in Econometrics* (pp. 369–74). Chichester, UK: Wiley.
Pena, D. (2001) George Box: an interview. *International Journal of Forecasting*, 17, 1–9.
Pfannkuch, M. and Wild, C. J. (2000) Statistical thinking and statistical practice: themes gleaned from professional statisticians. *Statistical Science*, 15, 132–52.
Preece, D. A. (1987) Good statistical practice. *The Statistician*, 36, 397–408.
Rao, C. R. (1997) *Statistics and Truth: Putting Chance to Work* (2nd edn). Singapore: World Scientific.
Reuter, P. (1982) The irregular economy and the quality of macroeconomic statistics. In V. Tanzi (ed.), *The Underground Economy in the United States and Abroad* (pp. 125–43). Vito Lexington: Lexington Books.
Rosnow, R. and Rosenthal, R. (1989) Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 44, 1276–84.
Smith, R. P. (1998) Quantitative methods in peace research. *Journal of Peace Research*, 35, 419–27.

Thaler, R. H. (2000) From homo economicus to homo sapiens. *Journal of Economic Perspectives*, 14, 133–41.
Thomas, D. H. (1976) *Figuring Anthropology: First Principles of Probability and Statistics*. New York: Holt, Rinehart and Winston.
Trosset, M. W. (1998) Comment. *Statistical Science*, 13, 23–4.
Tukey, J. W. (1962) The future of data analysis. *Annals of Mathematical Statistics*, 33, 1–67.
—— (1969) Analyzing data: sanctification or detective work? *American Psychologist*, 24, 83–91.
—— (1977) *Exploratory Data Analysis*. Reading, MA: Addison Wesley.
Tweedie, R. *et al*. (1998) Consulting: real problems, real interactions, real outcomes. *Statistical Science*, 13, 1–29.
Valavanis, S. (1959) *Econometrics*. New York: McGraw-Hill.
Verbeek, M. (2000) *A Guide to Modern Econometrics*. New York: Wiley.
Wainer, H. (1997) *Visual Revelations*. New York: Copernicus.
Weick, K. E. (1992) Jolts as a synopsis of organizational studies. In P. J. Frost and R. E. Stablein (eds), *Doing Exemplary Research* (pp. 99–104). Newbury Park, CA: Sage Publications.
Welch, R. E. (1986) Comment. *Statistical Science*, 1, 403–5.
Wickens, M. (1997) Comments. *Journal of Applied Econometrics*, 12, 523–7.
Wilcox, D. W. (1992) The construction of US consumption data: some facts and their implications for empirical work. *American Economic Review*, 82, 922–41.
Wild, C. J. (1994) Embracing the 'wider view' of statistics. *American Statistician*, 48, 163–71.
Wilkinson, L. and the Task Force on Statistical Inference. (1999) Statistical methods in psychology journals. *American Psychologist*, 54, 594–604.
Wooldridge, J. M. (2000) *Introductory Econometrics*. Cincinnati: South-Western.
Worswick, G. D. N. (1972) Is progress in science possible? *Economic Journal*, 82, 73–86.
Zellner, A. (1992) Statistics, science and public policy. *Journal of the American Statistical Association*, 87, 1–6.
—— (2001) Keep it sophisticatedly simple. In H. Keuzenkamp, M. McAleer, and A. Zellner (eds), *Occam's Razor, Simplicity and Scientific Inference*. Cambridge, UK: Cambridge University Press.