

Capitalism and Society

Volume 4, Issue 2

2009

Article 6

Comment on "Excessive Ambitions" (by Jon Elster)

David Hendry, *Oxford University*

Recommended Citation:

Hendry, David (2009) "Comment on "Excessive Ambitions" (by Jon Elster)," *Capitalism and Society*: Vol. 4: Iss. 2, Article 6.

DOI: 10.2202/1932-0213.1056

Comment on "Excessive Ambitions" (by Jon Elster)

David Hendry

Abstract

In "Excessive Ambitions," Jon Elster criticizes a wide range of social science aspirations to understand a complicated and evolving reality. Some of his analysis is to the point, but some is flawed, as explained in my comments. Crucially, however, his conclusions on empirical modeling are diametrically opposite to what is required—the problem has been a serious lack of ambition. And this is precisely the area where Elster is most guilty of 'criticizing others on the basis of third-party authorities.' Notwithstanding 'pitfalls and fallacies in statistical data analysis,' heterogeneous, high-dimensional objects like economies, which are subject to large, intermittent, and usually unanticipated, shifts, require ambitious approaches to characterize their behavior. I will try and explain how more ambitious empirical objectives can be achieved by automatic modeling methods which enhance human capabilities in tackling complicated data problems. En route, I re-emphasize the closely linked explanation for forecast failure recently discussed by myself, Michael Clements and Neil Ericsson in this journal. That leaves open the key issue as to why unanticipated shifts occur, and I speculate on that lacuna in existing economic theories, most of which omit any discussion of the mean levels of variables, and almost none address why such means might shift.

Author Notes: I am grateful to Jennifer L. Castle, Michael P. Clements, Søren Johansen, Katarina Juselius and Grayham E. Mizon for helpful comments on an earlier version.

I Introduction

Many aspects of the analysis by Jon Elster are well taken, and although some are flawed, my difficulty lies less with his arguments than in the implications that he draws therefrom. I will mainly focus on his near dismissal of empirical research, where like Summers (1991) in his *The Scientific Illusion in Empirical Macroeconomics*, Elster puts his finger on a real problem, but fails to realize either its causes or the solutions already available (see Juselius, 1993, for a response to Summers). **The crucial issue is that we have not been nearly ambitious enough.**

As a macro-econometrician, I do not have the expertise to comment extensively on his section II about economic theory. While there are undoubtedly issues that need to be addressed by economics professionals, some of which have previously been considered—as in Kirman (1989), and Lux (2008)—there seem to me to be counter-examples to the claim that ‘much work in economics and political science is devoid of empirical, aesthetic or mathematical interest, which means that it has no value at all’. For example, the vast body of auction theory has delivered several successes in real-world settings using rational game theory (see e.g., Klemperer, 2002, on the 3G auctions), as has matching theory for market design (see e.g., Roth, 2002, on matching prospective interns and hospitals). These were ambitious projects, and at least equal ambition will be required in the future to understand other complicated market phenomena: a lack of ambition in analyzing recent complex finance products probably contributed to policy failures in preventing the financial crisis.

Although I studied psychology before moving to economics,¹ the breadth and scope of behavioral economics—the subject for criticism in Elster’s section III—is far too large to be covered in a short critique of topics that vary from some which have been intensively studied since its birth to newcomers where ideas and understanding are evolving rapidly. One gap that surprised me in Elster’s coverage was the increasing link between psychology and economics in this arena, for example, the research leading to the use of economic games to uncover unexpected aspects in autistic behavior (see e.g., Frith and Frith, 2008, for a brief discussion of one of the many possibilities deriving from their work and others). Thus, I am again more optimistic than Elster on our ambitions: the quality and ingenuity of experiments appears to be on an upward trend even if a ‘general theory’ has not yet emerged.

Section IV is another matter altogether—we disagree diametrically. Notwithstanding ‘pitfalls and fallacies in statistical data analysis’, and the certainty that the unwary will fall into many ‘heffalump traps’, heterogeneous, high-dimensional interdependent objects like economies, which are subject to large, intermittent, and

¹Partly because the theories in the former at the time I was an undergraduate explained a little using a lot, whereas the latter had one simple theory that explained a great deal, albeit approximately.

usually unanticipated, shifts, require ambitious approaches to characterize their behavior. I have no objections to criticism *per se*—indeed criticism is the lifeblood of science so its general absence from major economics journals also troubles me greatly—and have not been averse to the occasional foray myself (as in Hendry, 1980, or Hendry and Ericsson, 1991), but I also believe that actions speak louder than words. Thus, I will try to explain how more ambitious objectives in empirical modeling can be achieved by automatic modeling methods, which enhance human capabilities in tackling the complicated data problems faced in economics. I will primarily cite publications in which I have been involved to avoid unnecessarily implicating others in my views, and apologize if a more obvious paper could have been used. I also apologize in advance for the brevity of many of the following arguments, where even important caveats, and most nuances, have been omitted: see Hendry (1995) for a more extensive treatment of the underlying theory.

I.1 Some background

Three Mantras need to be recited daily by all empirical economic modelers. First, *change is the norm*. Second, *nothing is correct till everything is*. Third, and consequently, *never force the finally chosen empirical model to match a theory*. Let me explain.

Breaks have intermittently affected all economic variables, as I will illustrate immediately below using some historical data (also see *inter alia*, Stock and Watson, 1996, Barrell, 2001, and Clements and Hendry, 2001). This has been obvious over the last two years of dramatic asset and commodity price changes. As Clements and Hendry (2008) show, such shifts may be unanticipated relative to pre-existing information, yet nevertheless differ substantively from the so-called black swans to which Taleb (2007) refers, as they do not relate to ‘fat-tailed’ distributions, but to shifts in means. Indeed, Elster suggests such a setting: ‘I believe that the assumption of an unchanging (but unknown) state of the world is often used without sufficient justification’. Rather, as Heraclitus is purported to have argued, *change is the norm*. Section II digresses to discuss breaks and their impact on economic forecasting.

Next, relationships between variables are both directly and indirectly altered by such changes, and because most economic variables are inter-correlated, it follows that under-specified empirical models will fail whenever any omitted variable shifts. Thus, in an empirical model *nothing is correct till everything is*. The way forward is to allow at the outset for everything that might matter substantively, including multiple breaks, then eliminate the chaff (effects that transpire to be irrelevant, non-systematic or negligible) from the wheat (relevant variables) by a general-to-specific (*Gets*) search (see Campos, Ericsson and Hendry, 2004, for a survey).

Progress in automatic modeling is essential to handle all these genuine complexities, and has proceeded apace: see Hoover (1995b), Hoover and Perez (1999), Krolzig and Hendry (2001), Hendry and Krolzig (2001, 2004, 2005), Doornik (2009), and Castle, Doornik and Hendry (2009a). Section III below addresses this issue, and discusses how to adapt *Gets* to a setting where there are more variables than observations—which will transpire to be the ‘normal’ situation in our approach.

Non-constancies certainly make empirical modeling hard—but they also render theories incorrect as they cannot apply equally well both before and after a break. Moreover, economic theory has progressed and is progressing (see e.g., Kirman, 1995), so today’s best theory will be discarded in due course. Thirdly, theories deliberately abstract from all but their essential components, and rely on dozens of implicit *ceteris paribus* clauses, which will be invalidated in any empirical setting. Those are three powerful reasons to *never force the finally chosen empirical model to match a theory*. That does not preclude fitting a theory-model to establish the costs of doing so, nor using the best available theory to guide the empirical analysis. Further, a sufficiently well-specified formulation could be embedded as a ‘core’ in a more general initial specification, but what matters empirically needs to be determined empirically. Section IV below addresses that issue.

The three mantras suggest that we are hopelessly *under* ambitious in data analysis, and combined with the complexity of the reality that is being investigated in macro-economics, unless that is addressed, failed models are almost certain to result. The next section focuses on change, our first mantra, before we turn to the second mantra in section III.

II A digression on economic forecasting

First, the prevalence of breaks. Figure 1 plots changes in the UK price level, ΔP_t , from about 1865 to 2000. Would you have liked to forecast ΔP_{T+1} in figure 1 from most time points T ?

Notice how what was apparently ‘normal variation’ in one epoch can later be seen as tiny. There is no natural scale for a variable like prices by which to judge the present against the future. I concur with some of Elster’s criticisms in that all too many empirical models use levels of variables, and pay scant attention to the impossibility that their units could be ‘timeless’ in the way most theories are. But even in logs (denoted by lower case), where changes become percentage shifts, a similar phenomenon can be observed in many time series. Figure 2 shows Δp_t , aka inflation, which also exhibits marked non-stationarities over the four sub-periods, with changes in means and variances, in persistence, and in the range of values observed. Many other examples are presented in Hendry (2005, 2009).

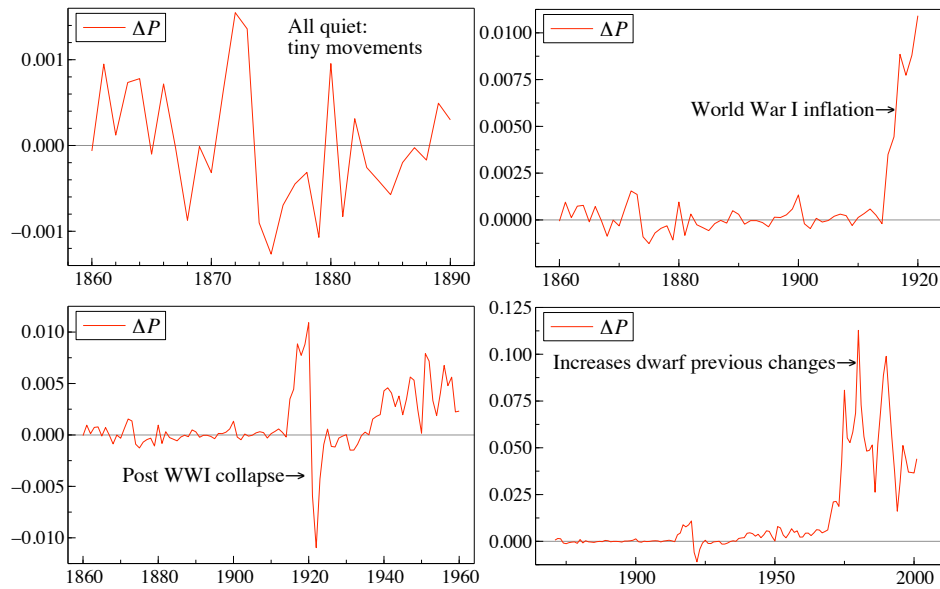


Figure 1: ΔP over four historical epochs

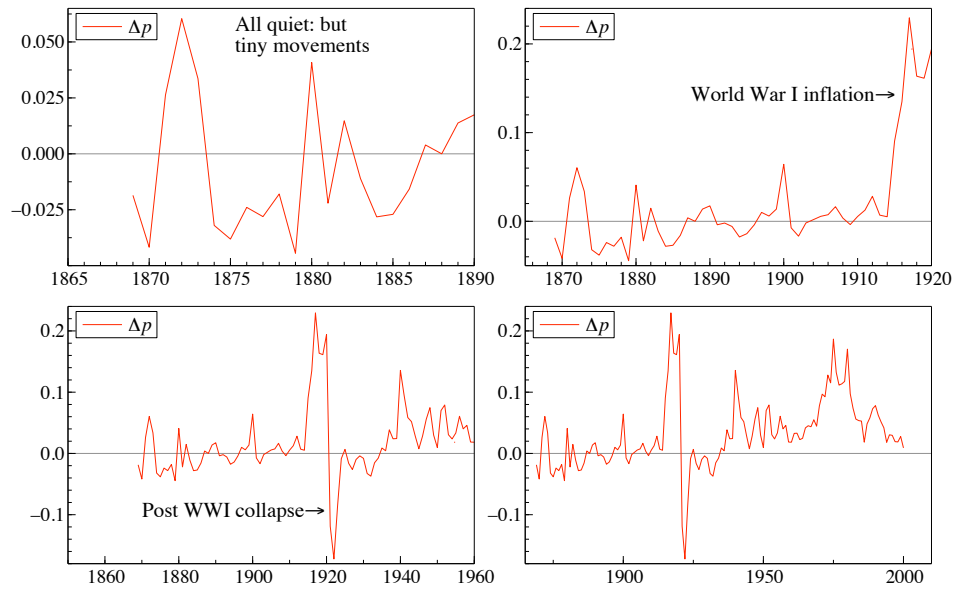


Figure 2: Δp over four historical epochs

II.1 Forecast failure

Forecast failure is a significant deterioration in accuracy relative to the anticipated outcome. Macro-economic forecast failure occurs all too frequently, and we have

witnessed many such events in the last decade alone. However, to summarize Clements and Hendry (2008) (as an accessible source for present readers): forecast failure depends on forecast-period events, which were usually unanticipated when the forecast was made; its occurrence need not invalidate the underlying theory or model, nor be predictable from any feasible in-sample statistical tests; and it need not be either avoided or induced, by the ‘goodness’ or ‘badness’ of the forecasting model. As an analogy, consider a rocket fired to the moon, which is predicted to land on 4th July, but en route is hit by a meteor and knocked off course. Clearly, the original forecast is systematically and badly wrong—indeed, if the rocket never gets to the moon, an infinite error is made. That outcome is not due to a poor forecasting model, and definitely does not refute the Newtonian gravitational theory that underpinned the initial forecast. Both implications would be seriously mis-leading non sequiturs, albeit that I see many investigators drawing equivalent conclusions about econometric models when they fail to forecast accurately. What is learned from the failure is that the forecast scenario was incomplete, here a combination of insufficient engineering to protect from a meteor impact, and an absence of knowledge about its existence. Neither may be correctable, so the risk of future disasters remains; but the next rocket is likely to use the same Newtonian theory and Kalman-filter forecasting algorithms. Strange how that logical error is not made by physical scientists, but often is in other areas....

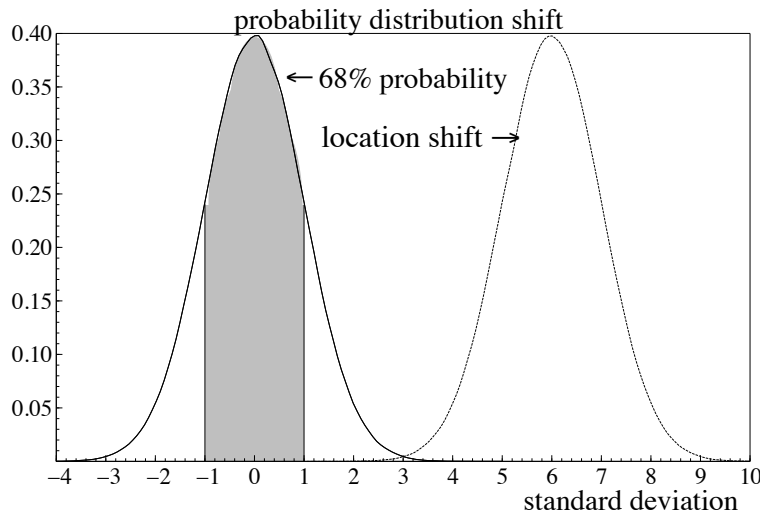


Figure 3: A mean shift in a distribution

Rather, unmodeled shifts in the preceding unconditional mean—called location shifts—are the primary cause of forecast failure: see Clements and Hendry (2008) and Ericsson (2008). Figure 3 illustrates. The main forecasting devices fall in the class of equilibrium-correction models (EqCMs), which is vast and contains all

regression equations, autoregressions and dynamic systems, vector autoregressions (VARs), dynamic stochastic general equilibrium models (DSGEs), autoregressive conditional heteroskedasticity (ARCH) and generalized ARCH (GARCH), as well as some other volatility models. All models in the equilibrium-correction class fail systematically following unmodeled changes in the mean of the data $\{x_t\}$, denoted by its expectation $E[x_t] = \mu$, from μ to μ^* , because the model forces convergence back to μ , irrespective of the new value of μ^* . This is a pervasive and pernicious problem for all forecasters, and obviously influences how economic data should be modeled and forecasted.

II.2 Folklore versus mathematical theory

Many readers of Clements and Hendry (1999) have been startled that our book is almost solid mathematics, filled with proofs and derivations seeking to establish a general theory applicable to a real world of change, forecasting with imperfect models, estimated from inaccurate data. We now understand vastly more about forecasting in non-stationary processes than we did 20 years ago, as Ericsson (2008) summarizes. Previously, much of forecasting ‘wisdom’ was folklore, some of it misleading, most unreliable. It took ambition and a decade of research to construct a new theory with numerous surprising findings, including overturning most previous results, which transpired to be relevant only in stationary worlds. Yet almost every research grant application was rejected on the grounds that ‘it could not be done’—but it was. We are now trying to develop methods that might help forecast breaks (see e.g., Castle, Fawcett and Hendry, 2009c), and are facing similar reactions. This is possibly *pluralistic ignorance*, as Elster suggests, but I suspect the more mundane outcome of others feeling (like him) that we are over-reaching: yet a lack of ambition would have stopped neolithic people leaving home.

III Empirical modeling

It is impossible to disagree more with any conclusion than that drawn by Jon Elster in relation to data analysis and empirical modeling. Not only is there a manifest lack of expertise in his analysis, it is based on the beliefs, not proofs, of others while he is seemingly unaware of key substantive developments. There is even a suggestion of biased selection as to whom to quote, namely ‘scholars who combine a high reputation among their peers with deep skepticism of the ways in which statistical analysis is routinely applied’. What about distinguished scholars who exhibit less skepticism?

Data analysis involves a multitude of decisions about what to model, how to do so, conditional on which variables, at what significance levels, and using what methods, applied to which data.... Every decision that is made entails selection. Thus, selection is inevitable and unavoidable, albeit often undocumented, and indeed sometimes not even seen as selection. There are many ways to make selection decisions, most of them unstructured and with unknown consequences or costs. What some economists call ‘data mining’ is an unstructured search across a range of alternative models to find the one that is most ‘liked’. It is perfectly possible, even quite common, to learn nothing of value from a data analysis—on that Elster and I agree, as do many others. However, the ‘goodness’ of a model is intrinsic to it, and does not depend on how it was found, so random search is culpable primarily as an inefficient way of proceeding. The question is: why do clever individuals waste time doing such analyses?

III.1 Deconstructing Elster’s list

I think such a worry about waste is what lies behind his statement: ‘As I understand data analysis, it has an almost infinite number of potential temptations, pitfalls and fallacies. Let me cite a few: data snooping (shopping around for independent variables until one gets a good fit), curvefitting (shopping around for a functional form that yields a good fit), arbitrariness in the measurement of independent or dependent variables, sample heterogeneity, the exclusion or inclusion of ‘outliers’, selection biases, the choice of the proper level of significance, the choice between one-tailed and two-tailed tests, the use of lagged variables, the problem of distinguishing correlation from causation, and that of identifying the direction of causation.’ In essence, this is a critique of over-ambitious interpretations and inferences from incoherent approaches—the solution to which is a more ambitious empirical methodology.

Elster’s list comprises the four main issues of model specification, selection, inference, and evaluation. The first concerns the set of potential variables, their lag lengths (in time series or panels), their functional forms and associated transformations, and breaks or outliers. The second concerns how to select a representative, or group of representatives (as in model averaging), what criteria to use in selection, and how to determine when those criteria are satisfied, including significance levels, and how to obtain (near) unbiased estimates. The third refers to the basis on which inferences are to be made, such as independent sampling (or Martingale difference errors), homogeneity, and distributional form (often Normal, so Student t-tests are used in finite samples), beautifully explained in Spanos (1986) among others. The fourth relates to testing the final selection (or selections), and involves checking its congruence, such that the model matches the evidence in all evaluated directions,

encompassing of rival models (see e.g., Mizon and Richard, 1986, as well as the recent special issue on that topic edited by Hendry, Marcellino and Mizon, 2008, related to the views in Elster's section VI), and testing for 'causation', or more precisely, invariance of the model to outside changes so that predicted responses occur as anticipated, which is the requirement of super exogeneity (see *inter alia*, Engle, Hendry and Richard, 1983, Hendry, 2004, and Hendry and Santos, 2009). If you expected that list to involve forecasting, please re-read section II now. And if you think that random sampling is sufficient to justify a data analysis, write the number 1 on a thousand pieces of paper, place them in a hat, and draw a random sample (with replacement if you wish). Hidden dependence is a serious problem to add to Elster's list—but like all the others is testable and refutable when absent in reality (or in a sequentially-factorized model): see e.g., Hendry (2009) for a more detailed analysis. Statistics is a difficult subject, but not impossible to learn: try Hendry and Nielsen (2007).

Selecting an empirical model inevitably involves uncertainty over the choice of variables, their lag responses and functional forms, the occurrences and timings of multiple location shifts, and the adequacy of the available data. To successfully determine what matters and how it enters, all potentially important determinants need to be included, since omitting key variables adversely affects the goodness of fit, biases the included factors' effects, and in a world of inter-correlated variables with non-stationarities induced by breaks and stochastic trends, leads to non-constant estimated models. Small unsystematic factors can usually be omitted without great loss—but large models are needed to help ensure such occurs. That is the basis of a *Gets* approach, and adopting it provides a structured analysis within which decisions are controlled with *ex ante* known costs when the initial model formulation is sufficiently general to provide a congruent approximation to the processes generating the data.

III.2 Structured model selection

Consider modeling aggregate expenditure on food over time—Hendry (2009) offers a more extensive discussion. Many correct decisions are needed for successful modeling, since food expenditure potentially depends on a large number of relevant variables including incomes, own and other prices, food quality and composition, interest rates, taxes, demography, etc. Their effects could vary with changes in outside factors such as legislation, policy regimes, financial innovation, health scares or increased nutritional knowledge, etc. Dependence could be linear or non-linear; and short-run, long-run and seasonal responses may all differ—indeed everything about these relationships could evolve over time. The level of aggregation also matters: national or regional, by income levels, categories of transactions, etc. Data

may be poorly measured, not timely, etc., but data contamination is also tackled in our approach.

Since those processes are high-dimensional, complicated, and subject to breaks, the initial model has got to be large if it is to nest the underlying processes. By itself, largeness in models is unproblematic in our approach, as can be established in computer simulations, or Monte Carlo experiments. In those, the process generating the data is known to the investigator, but not used by the selection algorithm. Many random trials can be conducted and the distributions of the results can then be analyzed to ascertain the match with the supposed theory. Castle *et al.* (2009a) report Monte Carlo simulation results (based on 1000 random replications) for selecting a regression when there are $N = 1000$ candidate variables, of which $n = 10$ are actually relevant, when $T = 2000$ (and the regressors are orthogonal). Nevertheless, the scale *per se* did not preclude the simulation, and the results matched the theory. First, the frequency of retaining irrelevant variables (called the *gauge*) equaled the nominal significance level, α . Secondly, retention rates for relevant variables (the *potency*) also matched the anticipated test powers from theory. And those despite the existence of $2^{1000} \simeq 10^{301}$ possible models in every replication, incomputable in the lifetime of the universe even at a million models per nano-second.

However, in our approach to be explained shortly, there are bound to be more variables N in total than the number of observations T , so in practice all cannot be entered from the outset. This conundrum is addressed by Castle *et al.* (2009a), based on the technical analysis in Johansen and Nielsen (2009), building on Hendry, Johansen and Santos (2008), who derive the distributions of reported parameter estimates in the canonical case of allowing an impulse indicator for every observation, so there are T additional zero-one dummies (called ‘impulse saturation’), as well as all candidate stochastic regressors, lags and functional forms. Impulse saturation is specifically designed to address the problem of location shifts jointly with what is more usually designated ‘model selection’, precisely because non-stationarities entail that any mis-specifications have deleterious effects in a world of breaks. Without a computer program to undertake the calculations, humans cannot investigate such situations—they far exceed most ambitions—yet are almost surely essential if models are to characterize the realities we face. When there are k basic variables, there are $k(k+1)(k+5)/6$ non-linear functions just up to a cubic, each potentially requiring perhaps r lags plus T indicators: for $k = 10$ linear variables after creating lags, and $T = 100$ say, there are $N = 385$ possible regressors. Thus, both expanding and contracting searches are used in *Autometrics* as described in Doornik (2009). While no optimal approach is known as yet, *Autometrics* works well in the situations we have investigated to date, including such small-sample settings as $N = 40$ when $T = 20$.

III.3 From model selection to model discovery

Impulse saturation is a form of robust estimation, so tackles potential data contamination, and equally handles ‘fat-tailed’ distributions by removing outliers, as well as location shifts and innovation outliers. Importantly, it is almost costless to add T impulses to a selection exercise, since when the significance level for retaining an impulse is set at $\alpha = 1/T$, then on average under the null, **the cost is just one observation**. I can hear the disbelief, yet that outcome is intuitively obvious after a moment’s thought: under the null, there are no outliers, breaks, or contaminants, so the probability that an estimated impulse’s t-value exceeds (in absolute value) the corresponding critical value c_α by chance must be α , which equals $1/T$, and when that one impulse is retained, its observation is ‘dummied out’. Thus, despite adding T extra variables to the selection process, a trivial cost is sustained: see the proof in Johansen and Nielsen (2009), for a wide range of models including regressions and autoregressions (possibly with unit roots).

Now remember what you learned in elementary statistics: adding any variable costs a ‘degree of freedom’, and so the same cost seems to apply to irrelevant regressor variables. When we undertook our $N = 1000$ -variable Monte Carlo study, we set $\alpha = 1/1000 = 0.1\%$ so $c_\alpha \simeq 3.8$, and hence one irrelevant variable was retained by chance out of the 990 candidates. Our procedure for ensuring that the reported parameter estimates on retained variables are unbiased (see Hendry and Krolzig, 2005) then drives the estimate on any adventitiously retained irrelevant variable to near zero, as there is a low probability that its t-value will be much larger than c_α (in absolute value). I suspect 1000 variables being searched across far exceeds any number Elster may have dreamt of (the quote of his views above suggests nightmare may be a better description), yet almost all irrelevant variables are correctly eliminated, and the reported estimates are nearly unbiased, including the ‘goodness-of-fit’ as measured by the residual standard deviation.

The following four-fold distinction tries to summarize the process of model discovery for a relationship between an observed variable y which is postulated to depend via a parameter β on a set of candidate variables \mathbf{x} , when a sample of T observations is available, possibly contaminated, or with breaks:

Classical econometrics: obtain the ‘best’ estimate of β , given the correct \mathbf{x} and T ;

Model selection: find β and the relevant variables as a subset of a given correct \mathbf{x} and T ;

Robust statistics: find a ‘robust’ estimate of β and T , given the correct set of relevant variables \mathbf{x} ;

Model discovery: find β , $\mathbf{f}(\mathbf{x}_1)$, \mathbf{d} and T jointly, where $\mathbf{f}(\mathbf{x}_1)$ is the appropriate function of a subset of the basic variables \mathbf{x} , \mathbf{d} are indicators for breaks, outliers and data contamination, and v is the unexplained component, when the finally chosen

model is the congruent representation $y = \beta'f(x_1) + \gamma'd + v$ (including establishing the validity of conditioning on any contemporaneous variables, perhaps used as instruments, and sequential factorization).

III.4 On the ‘magical’ 5% significance level

The significance level, α , for any empirical data analysis needs careful consideration of the trade-off between excess retention of irrelevant effects (so depends on N), and low retention of relevant (so depends on T , and on the unknown n through the non-centralities of their test statistics): there is certainly nothing ‘magical’ about 5%. The most important consideration is that the model must provide a congruent representation of the evidence so that the actual inferences will have the properties assumed by whatever choice is made for α : failure here could mean that the nominal choice of α (including 5%) is wildly wrong (even being 80% and upwards), and that is the main source of spurious and nonsense regressions. The choice of α should also depend on the purpose of the analysis, whether it be modeling for understanding, for testing theories, for policy, or for forecasting.

Taking these in reverse order, forecasting is different—as section II has shown—and loose significance levels make sense in that context: for example, $\alpha = 16\%$ is close to the implicit value in the Akaike criterion (often denoted AIC: see Akaike, 1973), and even $\alpha = 25\%$ can be justified, when the process is wide-sense stationary. Next, for policy, the situation is less clear, and although one might argue that α should be derived from the trade-offs entailed by the objective function of the policy maker, that rarely reflects all the complexities of modeling, specifically multiple breaks and the essential requirement of checking that the policy variables are super exogenous for the parameters in the policy model. Thus, we tend to propose more stringent α values of 1% or less, emphasizing again that bias correction will drive the estimates of the adventitiously-retained irrelevant variables towards zero so reducing any potentially false policies, as well as providing near unbiased estimates of the parameters of relevant variables. Also, you should by now have deduced a fourth mantra: *never judge a policy model by its forecast accuracy* (see e.g., Hendry and Mizon, 2000, 2005). For testing theories, α is *au choix*: but congruence is not—again many major mistakes derive from testing theories in non-congruent models. Finally, I have already discussed the considerations for setting α when modeling.

Thus, data analysis can be done properly, in a structured approach with low probabilities of retaining ‘garbage’—despite selecting from a huge set of candidate variables—while still protecting against outliers and breaks, and ending with near unbiased estimates, all at low cost from searching. Thus, structured selection can greatly reduce uncertainty about model specification. The main cost is under the alternative from setting $\alpha \leq 1/T$, so lower power results to retain relevant variables.

But relative to starting with an under-specified model where a key variable or break has been omitted, so the result is nearly useless, we believe that is a small price to pay for a controlled method that avoids almost all the mud slung at data analysis in the quote by Elster above. It need not be a craft, requiring an apprenticeship of hundreds of previous applications. Yes, one needs to understand the statistical or econometric theory to sensibly undertake data analysis, and yes, subject-matter knowledge provides an essential background—but empirical modeling can be taught, so we need not be stuck with the claim that ‘the acquisition of “wisdom” is a task that is so time-consuming and demanding that it excludes the acquisition of substantive knowledge in any broad field of empirical inquiry’. Automatic modeling tools can be designed to formulate a general specification from a basic set of variables of interest to an investigator, select therefrom and then evaluate the final choice: the ‘wisdom of the ages’ can be consolidated and made available to all.

Let me re-iterate the remarks at the end of the previous section. Many economists and statisticians have reacted that model selection must have high costs; that ‘good’ models must be parsimonious; that large numbers of variables must lead to spurious findings; that the selection results must be biased; that more variables than observations cannot be handled; that identification must be known in advance, and so on—folklore has a strong hold on thinking. Moreover, our research grant applications for this program have been even more systematically rejected on the grounds that it cannot be done. It has: it works. As yet, the proofs and derivations are still only for relatively special cases—but with greater mathematical skills and efforts, and more researchers involved, I am sure general results can be established. In both these fundamental areas of forecasting and modeling, the most powerful quantitative tools have paid the highest dividends. Do not accept Elster’s invitation to return to the stone age: **be ambitious!**

IV Economic theory-based empirical models

A current penchant in economics is for corroborationism: formulate a theory, seek data with the same names as the theory, impose the theory on the data, and check that some of the theory predictions are not refuted. The methodology espoused by Friedman (1953)—that assumptions can be unrealistic provided predictions are verified—has done almost irreparable damage to empirical research by inducing a non-scientific approach. The example I have used all too often is: assume $1 = 2$; then clearly $2 = 1$; adding both sides yields $3 = 3$, which is true, proving that $1 = 2$. Science seeks refutation, not corroboration, and the former is easy here: if you accept that $1 \neq 2$, I can take a lot of money from you, very quickly. It was already known at the time of Plato that the conclusion of an argument being correct

did not confirm the validity of the argument—yet it is almost impossible to publish a paper in a so-called top economics journal unless you follow the logic just outlined. This has nothing to do with ambition, excessive or otherwise, but does match Elster's views of the present state of economics. Repeatedly finding confirmation when you seek refutation adds substantial credibility, but all 'empirical knowledge' is fallible, scientific or otherwise. Progress is the watchword, and a glance round your home reveals how dramatic that has been, even though many of the theories that helped us advance have been replaced over time—even in economics.²

In a less extreme form, successful sequential corroboration of aspects of a theory can entail its rejection when the evidence is combined—see Ericsson and Hendry (1999) for some examples. And here an under ambition has been pernicious—those who advocated real business cycle models, such as Kydland and Prescott (1990, 1991), have used minimal empirical criteria because formal tests guaranteed rejection: see e.g., Hoover (1995a). A more recent example is Ireland (2004), with powerful critiques in Johansen (2006), Juselius and Franchi (2007) and Hoover, Johansen and Juselius (2008), who together demonstrate the misleading conclusions that can result when models are based on assumptions that are at variance with the data evidence. A quote from Johansen (2006) seems apposite: 'The failure to adequately confront the models with the data means that the development of economic theories cannot be given guidance from the data. The consequence is that the profession is likely to stick to theories which are no longer useful or even have become harmful. Even worse, the profession will fail to recognize new features in the data, which go against mainstream beliefs, which could signal a change in the underlying mechanism.' And that just before the financial crisis erupted.

IV.1 On location shifts and economic theory

We remarked above that location shifts are as much a problem for economic theory as empirical modeling. More crucially, they also reveal a fundamentally incorrect assumption underpinning all inter-temporal optimization approaches, namely that economic agents can form 'rational expectations' of future events. Once location shifts can occur, current conditional expectations cannot be shown to be even unbiased predictors, let alone proving that estimated versions thereof are minimum mean-squared error predictors (see e.g., Castle, Doornik, Hendry and Nymoen, 2009b). The belief in that result has derived from proofs which simply assumed

²I am unsure how to discuss Elster's claim that the Alfred Nobel Memorial Prize for Economic Science has never been awarded 'for confirmed empirical predictions'. Both Milton Friedman and Franco Modigliani have citations that mention demonstrating the empirical relevance of the permanent income hypothesis, and life-cycle hypothesis respectively. More importantly perhaps, the citations note that they produced findings that overturned previous views.

that distributions never shifted—universal wide-sense stationarity. Once tomorrow's mean can differ from today's, it makes no sense to calculate its expected value by integrating over today's distribution. Thus, unless one has a crystal ball that somehow reveals all future distributions today, other methods of peering into the future are required. Replicating in a different context the related diagram from Clements and Hendry (2008), figure 3 above illustrated this. There, integrating over today's distribution delivers an expected value of zero, which is not a good prediction of tomorrow's mean which lies six standard deviations away. Either agents must be able to forecast such shifts—and a few always claim to do so, though most humans do not appear to be successful very often—or they must adapt after shifts occur. Without omniscience as to the entire characteristics of every future post-shift distribution, some forecasting device other than a pre-existing conditional expectation is needed: see e.g., Hendry (2006) and Frydman and Goldberg (2007). Consequently, the calculations behind dynamic stochastic general equilibrium theory models cease to be valid: in the real world, no rational agent would assume a constant indefinite future, and hence would not try to solve the inter-temporal optimization problem in the conventionally postulated way. Here I sense accord with another of Elster's worries, already emphasized in the *Dahlem Report*.

One of the great lacunas in existing macroeconomic theories is that most of them omit any discussion of the mean levels of variables, and almost none address why such means might suddenly shift. If location shifts are the primary cause of ubiquitous forecast failures, and the magnitudes of systematic failures depend on how substantially long-run means shift, then theories of both are urgently needed. I suspect such theories will demand high-level mathematical skills and an ability to develop tools which let us manipulate evolving multivariate distributions—and great ambition to even begin such a project.

V Conclusion

Jon Elster has highlighted a number of substantive problems with theoretical and empirical approaches in many social sciences. My disagreement is about the potential solutions. A formal discussant of Yule (1926) complained about the mathematical difficulty of his paper—yet school mathematics now suffices to understand his brilliant analysis. Keynes (1939) then used that work to attack early empirical econometrics. As I noted in Hendry (1980), even by 1980 his critique was not rebutted—but by Hendry (2000) it was: we are progressing. A more recent, but equally pertinent example of the crucial need for advanced technical analysis is the late Clive Granger's discovery of cointegration: many applied economists 'knew' that linear combinations of non-stationary time series could be stationary, but until

Granger (1981) developed the mathematical theory, no-one knew that cointegration provided an endogenous explanation for the prevalence of unit roots in economic time series.

The way ahead requires more powerful tools, better analyzed methods, clearer teaching of the new approaches, and their ambitious application to new areas, in political science, climate change, and epidemiology among others. By all means criticize bad practice and poor theory; and yes, there are many situations where outcomes need not be unique, and so are not predictable: but do not suggest that the venture is not worthwhile.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N., and Csaki, F.(eds.), *Second International Symposium on Information Theory*, pp. 267–281. Budapest: Akademia Kiado.
- Barrell, R. (2001). Forecasting the world economy. In Hendry, D. F., and Ericsson, N. R.(eds.), *Understanding Economic Forecasts*, pp. 149–169. Cambridge, Mass.: MIT Press.
- Campos, J., Ericsson, N. R., and Hendry, D. F. (2004). Editors' introduction. In Campos, J., Ericsson, N. R., and Hendry, D. F.(eds.), *Readings on General-to-Specific Modeling*, pp. 1–81. Cheltenham: Edward Elgar.
- Castle, J. L., Doornik, J. A., and Hendry, D. F. (2009a). Model selection when there are multiple breaks. Working paper, Economics Department, University of Oxford.
- Castle, J. L., Doornik, J. A., Hendry, D. F., and Nymoen, R. (2009b). Testing the invariance of expectations models of inflation. Working paper, Economics Department, University of Oxford.
- Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2009c). Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, forthcoming.
- Castle, J. L., and Shephard, N.(eds.)(2009). *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.
- Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.
- and — (2001). An historical perspective on forecast errors. *National Institute Economic Review*, **177**, 100–112.

- and — (2008). Economic forecasting in a changing world. *Capitalism and Society*, **3**, 1–18.
- Doornik, J. A. (2009). Autometrics. in Castle, and Shephard (2009), pp. 88–121.
- Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, **51**, 277–304.
- Ericsson, N. R. (2008). Comment on ‘Economic forecasting in a changing world’ (by Michael Clements and David Hendry). *Capitalism and Society*, **3**, 19–36.
- Ericsson, N. R., and Hendry, D. F. (1999). Encompassing and rational expectations: How sequential corroboration can imply refutation. *Empirical Economics*, **24**, 1–21.
- Friedman, M. (1953). *Essays in Positive Economics*. Chicago: University of Chicago Press.
- Frith, C. D., and Frith, U. (2008). The self and its reputation in autism. *Neuron*, **57**, 331–332.
- Frydman, R., and Goldberg, M. D.(eds.)(2007). *Imperfect Knowledge Economics: Exchange Rates and Risk*. Princeton, New Jersey: Princeton University Press.
- Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, **16**, 121–130.
- Hendry, D. F. (1980). Econometrics: Alchemy or science?. *Economica*, **47**, 387–406.
- (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.
- (2000). Epilogue: The success of general-to-specific model selection. In *Econometrics: Alchemy or Science?*, pp. 467–490. Oxford: Oxford University Press. New Edition.
- (2004). Causality and exogeneity in non-stationary economic time series. In Welfe, A. (ed.), *New Directions in Macromodelling*, pp. 21–48. Amsterdam: North Holland.
- (2005). Bridging the gap: Linking economics and econometrics. In Diebolt, C., and Kyrtsou, C.(eds.), *New Trends in Macroeconomics*, pp. 53–77. Berlin: Springer Verlag.
- (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, **135**, 399–426.
- (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In Mills, T. C., and Patterson,

- K. D.(eds.), *Palgrave Handbook of Econometrics*, pp. 3–67. Basingstoke: Palgrave MacMillan.
- Hendry, D. F., and Ericsson, N. R. (1991). An econometric analysis of UK money demand in ‘Monetary Trends in the United States and the United Kingdom’ by Milton Friedman and Anna J. Schwartz. *American Economic Review*, **81**, 8–38.
- Hendry, D. F., Johansen, S., and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, **33**, 317–335. Erratum, 337–339.
- Hendry, D. F., and Krolzig, H.-M. (2001). *Automatic Econometric Model Selection*. London: Timberlake Consultants Press.
- and — (2004). Automatic model selection: A new instrument for social science. *Electoral Studies*, **23**, 525–544.
- and — (2005). The properties of automatic Gets modelling. *Economic Journal*, **115**, C32–C61.
- Hendry, D. F., Marcellino, M., and Mizon, G. E.(eds.)(2008). *Encompassing*. Special Issue: Oxford Bulletin of Economics and Statistics.
- Hendry, D. F., and Mizon, G. E. (2000). On selecting policy analysis models by forecast accuracy. In Atkinson, A. B., Glennerster, H., and Stern, N.(eds.), *Putting Economics to Work: Volume in Honour of Michio Morishima*, pp. 71–113. London School of Economics: STICERD.
- and — (2005). Forecasting in the presence of structural breaks and policy regime shifts. In Andrews, D. W. K., and Stock, J. H.(eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 480–502. Cambridge: Cambridge University Press.
- Hendry, D. F., and Nielsen, B. (2007). *Econometric Modeling: A Likelihood Approach*. Princeton: Princeton University Press.
- Hendry, D. F., and Santos, C. (2009). An automatic test of super exogeneity. In Watson, M., Bollerslev, T., and Russell, J.(eds.), *Volatility and Time Series Econometrics*. Oxford: Oxford University Press.
- Hoover, K. D. (1995a). Facts and artifacts: Calibration and the empirical assessment of real business cycle models. *Oxford Economic Papers*, **47**, 24–44.
- (1995b). In defense of data mining: Some preliminary thoughts. In Hoover, K. D., and Sheffrin, S. M.(eds.), *Monetarism and the Methodology of Economics: Essays in Honor of Thomas Mayer*, pp. 242–257. Aldershot:

Edward Elgar.

- Hoover, K. D., Johansen, S., and Juselius, K. (2008). Allowing the data to speak freely: The macroeconometrics of the cointegrated vector autoregression. *American Economic Review: Papers and Proceedings*, **98**, 251–255.
- 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.
- Ireland, P. (2004). A method for taking models to the data. *Journal of Economic Dynamics and Control*, **28**(6), 1205–1226.
- Johansen, S. (2006). Confronting the economic model with the data?. In Colander, D. (ed.), *Post-Walrasian Macroeconomics*, pp. 287–300. Cambridge: Cambridge University Press.
- Johansen, S., and Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. in Castle, and Shephard (2009), pp. 1–36.
- Juselius, K. (1993). VAR modelling and Haavelmo's probability approach to econometrics. *Empirical Economics*, **18**, 595–622.
- Juselius, K., and Franchi, M. (2007). Taking a DSGE model to the data meaningfully. *Economics-The Open-Access, Open-Assessment E-Journal*, **2007-4**.
- Keynes, J. M. (1939). Professor Tinbergen's method. *Economic Journal*, **44**, 558–568.
- Kirman, A. (1989). The intrinsic limits of economic theory: The emperor has no clothes. *Economic Journal*, **99**, 126–139.
- (1995). The evolution of economic theory. In d'Autume, A., and Cartelier, J.(eds.), *L'Economie Devient-elle une Science Dure?*, pp. 92–107. Paris: Economica. Reprinted in English as: *Is Economics Becoming a Hard Science?* Edward Elgar, 1997.
- Klemperer, P. (2002). What really matters in auction design. *The Journal of Economic Perspectives*, **16**, 169–189.
- 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.
- Kydland, F. E., and Prescott, E. C. (1990). Business cycles: Real facts and a monetary myth. *Federal Reserve Bank of Minneapolis, Quarterly Review*, **14**, 3–18.

- and — (1991). The econometrics of the general equilibrium approach to business cycles. *Scandinavian Journal of Economics*, **93**, 161–178.
- Lux, T. (2008). Dahlem report: The financial crisis and the systemic failure of academic economics. Unpublished report, Department of Economics, University of Kiel.
- Mizon, G. E., and Richard, J.-F. (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica*, **54**, 657–678.
- Roth, A. E. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, **70**, 1341–1378.
- Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.
- Stock, J. H., and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, **14**, 11–30.
- Summers, L. H. (1991). The scientific illusion in empirical macroeconomics. *Scandinavian Journal of Economics*, **93**, 129–148.
- Taleb, N. N. (2007). *The Black Swan*. New York: Random House.
- Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time series (with discussion). *Journal of the Royal Statistical Society*, **89**, 1–64. Reprinted in Hendry, D. F. and Morgan, M. S. (1995), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.