



Research Program on Forecasting

Economic Forecasting in Theory and Practice: An Interview with David F. Hendry

Neil R. Ericsson

Division of International Finance
Board of Governors of the Federal Reserve System
Washington, DC 20551 USA
ericsson@frb.gov

and

Department of Economics
The George Washington University
Washington, DC 20052 USA
ericsson@gwu.edu

RPF Working Paper No. 2016-012
<http://www2.gwu.edu/~forcpgm/2016-012.pdf>

November 13, 2016

RESEARCH PROGRAM ON FORECASTING
Center of Economic Research
Department of Economics
The George Washington University
Washington, DC 20052
<http://www2.gwu.edu/~forcpgm>

**ECONOMIC FORECASTING IN THEORY AND PRACTICE:
AN INTERVIEW WITH DAVID F. HENDRY**

Neil R. Ericsson*

November 13, 2016

Abstract: David Hendry has made major contributions to many areas of economic forecasting. He has developed a taxonomy of forecast errors and a theory of unpredictability that have yielded valuable insights into the nature of forecasting. He has also provided new perspectives on many existing forecast techniques, including mean square forecast errors, add factors, leading indicators, pooling of forecasts, and multi-step estimation. In addition, David has developed new forecast tools, such as forecast encompassing; and he has improved existing ones, such as nowcasting and robustification to breaks. This interview for the *International Journal of Forecasting* explores David Hendry's research on forecasting.

Keywords: encompassing, equilibrium correction models, error correction, evaluation, exogeneity, forecasting, modeling, nowcasting, parameter constancy, robustification, structural breaks.

JEL classifications: C53.

*Forthcoming in the *International Journal of Forecasting*. The interviewer ("NRE") is a principal economist, Division of International Finance, Board of Governors of the Federal Reserve System, Washington, DC 20551 USA, and a Research Professor, Economics Department, The George Washington University, Washington, DC 20052 USA. The interviewee ("DFH") is a Professor of Economics and a co-director of the programs *Economic Modeling in a Rapidly Changing World* and *Climate Econometrics*, University of Oxford, Oxford, England. They may be reached on the Internet at ericsson@frb.gov and david.hendry@nuffield.ox.ac.uk respectively. The views in this interview are solely the responsibility of the author and the interviewee and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. The interviewee is pleased to acknowledge financial support from the Institute for New Economic Thinking (INET grant #20029822), the Robertson Foundation (grant #9907422), and Statistics Norway through the Research Council of Norway (grant #236935). We are grateful to Julia Campos, Jennifer Castle, Mike Clements, Vivien Hendry, Rob Hyndman, Andrew Kane, Aaron Markiewitz, Jaime Marquez, Andrew Martinez, Bent Nielsen, Felix Pretis, Angela Wenham, and an anonymous referee for helpful comments and discussion, and to Aaron Markiewitz for research assistance. Empirical results and graphics were obtained using 64-bit Ox-Metrics 7.1; see Doornik and Hendry (2013).

1 Early Work on Forecasting

NRE: David, you’ve made major contributions to many areas of economics and econometrics. These include econometric methodology, general-to-specific modeling, Monte Carlo techniques, software implementation, the history of econometric thought, policy analysis, and empirical investigations of consumer expenditure, money demand, inflation, and the housing market. We discussed these topics at length in Ericsson (2004), so let’s focus on another important topic—forecasting. Over the last couple of decades, you’ve made significant contributions to our understanding of economic forecasting. When did you first become interested in forecasting?

1.1 The University of Aberdeen

DFH: It was in 1964. I was an undergraduate at the University of Aberdeen, and I was very much influenced by the empirical economic models of Lawrie Klein (1950) and Jan Tinbergen (1951), who suggested that we might be able to forecast future outcomes. In my undergraduate thesis, I estimated a regression model for annual UK consumers’ expenditure given current income and lagged expenditure—painstakingly worked out on a mechanical calculator. Using the whole-sample parameter estimates, I calculated a “forecast” of the last observation to see how close it was to the outcome.

NRE: In effect, you were evaluating the last residual of your estimation period. What did you find?

DFH: The forecast and the outcome were reasonably close. That’s unsurprising, given how the “forecast” was calculated. Because the forecast was within the estimation period, the corresponding forecast error was included in the sum of squared errors that OLS minimized.

1.2 Macroeconometric Models and Predictive Failure

NRE: When you were writing your PhD thesis under Denis Sargan at the London School of Economics (LSE), you developed a small macro-model of the UK economy that included an equation for consumers’ expenditure. How did your forecasts fare?

DFH: Not well! In late 1967, I calculated *ex ante* forecasts of consumers’ expenditure for the next two quarters: 1968Q1 and 1968Q2. When actual expenditure was later reported by the Central Statistical Office, I found that my model had massive forecast failure. The parameter constancy test rejected, and its *p*-value had so many zeros that it was embarrassing. It took me years to understand why such forecast failure is commonplace.

That particular forecast failure arose from a change in economic policy. During 1968Q1, the Chancellor of the Exchequer (that is, the UK finance minister) threatened to increase Purchase Tax—essentially, a sales tax—if consumers didn’t “behave themselves” and spend less. Consumers responded by spending more, especially on durable goods. So, in the next quarter, the Chancellor duly increased Purchase Tax, and consumers’ expenditure fell. My model did not account for the policy threat, the

policy's implementation, or consumers' responses to both. Consequently, my model's forecasts failed badly.

NRE: Your UK model was subsequently published as Hendry (1974), which included a new test for predictive failure. It generalized Gregory Chow's (1960) single-equation predictive failure test to systems, albeit in a χ^2 version rather than the F version that Jan Kiviet (1986) later developed. How did that experience with your small macro-model influence your work on forecasting?

DFH: It motivated me to investigate the nature of predictive failure. Why did models built from the best available economics using the latest econometrics and fairly good data not produce useful forecasts? In Hendry (1979b), I linked predictive failure to poor model formulation, but that explanation subsequently turned out to be unhelpful, or at least incomplete.

NRE: Other economists were also evaluating forecasts from macro-models. In particular, Charles Nelson wrote two influential papers on *ex ante* forecasts: Nelson (1972) and Cooper and Nelson (1975).

DFH: Charles showed that forecasts from univariate time-series models could beat forecasts from large empirical economic models such as the FRB–MIT–PENN model. From an LSE perspective, such large models treated dynamics inadequately, often simply as autocorrelated errors in static equations. Because of that dynamic misspecification, we suspected that models that included only dynamics could forecast better. I found that simple dynamic models did indeed forecast better than static economic models, even though the latter embedded economic theory whereas the former did not. However, I had misinterpreted the implications of Nelson and Cooper's results. I had not realized that models in differences—such as those in Nelson (1972)—almost invariably forecast better than models in levels if the means of the variables being forecast altered. We now refer to such changes as *location shifts*.

NRE: Nelson and Cooper's forecasts used methods that were proposed by Box and Jenkins (1970). Those methods are robust to location shifts for reasons that we did not appreciate at the time. However, those methods omit information about the long run because they include only variables in their differences.

DFH: Indeed. At a Minneapolis Fed conference in 1975, I criticized Clive Granger for differencing: see Hendry (1977) on Granger and Newbold (1977).

NRE: In his Nobel prize lecture, Clive gives an amusing account of that discussion: “A colleague, David Hendry, stated that the difference between a pair of integrated series could be stationary. My response was that it could be proved that he was wrong, but in attempting to do so, I showed that he was correct, and generalized it to cointegration, and proved the consequences such as the error-correction representation.” [Granger (2004, p. 363)].

DFH: Clive's development of cointegration also resolved the debate between modeling in levels and modeling in differences, as I discuss in Hendry (2004).

NRE: We already knew something about working in differences and in levels from the equilibrium correction models in Denis Sargan's (1964) chapter of the *Colston*

Papers. A decade prior to Denis’s paper, Bill Phillips (1954) had analyzed integral, proportional, and derivative control in formulating policy—also an equilibrium correction framework. An even earlier precedent is Bradford Bixby Smith (1926), a paper re-discovered by Terry Mills (2011).

2 Development of Pertinent Econometric Tools

NRE: In addition to analyzing predictive failure, you developed new econometric tools, focusing on exogeneity, mis-specification analysis, and encompassing. Because these tools helped clarify issues on forecasting, let’s look at these tools, starting with exogeneity.

2.1 Exogeneity

NRE: In the 1970s, you, Rob Engle, and Jean-François Richard reinterpreted the concept of exogeneity, later published in Engle, Hendry, and Richard (1983). You subsequently applied that framework to feedback versus feedforward models in Hendry (1988) and Engle and Hendry (1993), with extensions in Hendry and Santos (2010). How did this work bear on forecasting?

DFH: Exogeneity entails being able to treat a variable as given for a specific purpose, without needing to know how that variable is being generated. Engle, Hendry, and Richard (1983) distinguished between three purposes: inference, such as estimation and testing; forecasting; and policy. We clarified that conditional inference requires weak exogeneity, that conditional forecasting requires strong exogeneity, and that economic policy scenarios require super exogeneity. Our analysis also highlighted practical and theoretical problems with other definitions of exogeneity, such as those based on orthogonality between variables and errors.

NRE: What does super exogeneity imply for forecasting when a regime changes?

DFH: That takes some explaining. Super exogeneity arose naturally as a condition where a variable can be treated as given in a model despite a change in the process generating that variable—a regime change. Thus, in a relationship between variables y and z with a slope coefficient β , super exogeneity ensured the invariance of β to changes in the distribution of z . Engle, Hendry, and Richard (1983) assumed that, when super exogeneity did not hold, regime changes would mainly affect the slope coefficient. However, Favero and Hendry (1992) showed that changes in β were hard to detect if z had a mean of zero. By contrast, changes in β were easy to detect if z had a nonzero mean. The mean of y would then shift, relative to the past, entailing a location shift.

This insight resolved a puzzle that I had encountered while serving on HM Treasury’s Academic Panel. Treasury economists would present a number of empirical models of (say) investment, claiming that it was difficult to distinguish between the models, yet admitting that the next observation on investment rejected all the models on a test of predictive failure. Such failures could be explained by location shifts,

without entailing that the Lucas critique was the cause. Location shifts could occur from structural breaks and need not arise from mis-specified expectations.

NRE: While many people associate the Lucas (1976) critique with policy implications, Bob Lucas actually motivated his paper with examples of predictive failure. As we just discussed, valid estimation, accurate forecasting, and sound policy are separate issues. Lucas conflated the last two.

DFH: Precisely. A model might mis-forecast because of a location shift, which did not affect the parameters of interest in a policy context. That model could still be used successfully for policy. Moreover, intercept correction as such does not affect a model's policy implications, whereas it does affect a model's forecast accuracy; and a change in policy parameters need not imply forecast failure.

NRE: When did you first suspect that different purposes might require different models?

DFH: Regrettably, quite late on! For instance, only in the late 1990s did I realize that a naive model could dominate an estimate of the in-sample data generation process (DGP) for purposes of forecasting. That result surprised me.

2.2 Analysis of Mis-specification

NRE: Mis-specification analysis has been central to understanding economic forecasts in a changing world. In the 1970s, you wrote several theoretical papers that analyzed mis-specified models, including Hendry (1975) on omitted variables and simultaneity bias, and Hendry (1979a) on inconsistent estimators.

DFH: Actually, my analysis of mis-specification began earlier—in Hendry (1973)—and led to two discoveries. First, mis-specified models need not forecast badly. That result was already known for special cases, as Nelson had illustrated. Second, it was possible to develop a general theory of economic forecasting in which the forecasting model was mis-specified for a DGP that itself was nonconstant over time.

2.3 Encompassing, Constancy, and Invariance

NRE: Your research on encompassing arose directly from mis-specification analysis.

DFH: It did. My early work on mis-specification often involved running Monte Carlo simulations. They highlighted two important issues. First, it was essential to formulate a DGP in order to generate simulated data. Claims that economic data have no DGP just baffle me! Some process *must* be generating the data. Second, a DGP necessarily encompasses all its reductions, and hence all mis-specified models. So, if a model closely approximates the DGP, that model should be able to explain the results of everybody else's (mis-specified) models. Davidson, Hendry, Srba, and Yeo (1978) are explicit about the need to explain the results of rival empirical models. Bontemps and Mizon (2008) provide an excellent overview of encompassing.

NRE: How was encompassing relevant in your analysis of forecasting?

DFH: The encompassing principle led to designing tests of forecast encompassing in Chong and Hendry (1986). These tests examine whether the forecasts of one model could explain the forecast errors of another model.

NRE: What other concepts have been important in your analysis of forecasting?

DFH: Constancy, for one; and constancy is distinct from the concept of invariance. For example, a sine wave is invariant but not constant. Relatedly, a given model can be written in many different ways, and how it is written can affect whether it appears to be constant or not. Aggregates such as price indexes can hide a great deal of nonconstancy. An unresolved issue is whether parsimony matters for forecasting.

3 Evaluation of Existing Forecasting Techniques

NRE: Until the early 1990s, you had viewed forecasting as an activity paralleling model design. That perspective arose naturally from the framework for exogeneity in Engle, Hendry, and Richard (1983). However, that perspective also hampered your understanding of forecasting as a separate discipline.

The year 1991 marked a turning point in your thinking about economic forecasting, and the catalyst was the UK Parliament's Treasury and Civil Service Committee. You submitted evidence on economic forecasting to that Committee. Preparation of your report to the Committee—detailed in Hendry (1991)—led you to a broader understanding of the subject. You subsequently produced a torrent of insightful evaluations of many existing forecast techniques, including error correction models and cointegration, mean square forecast errors, add factors, leading indicators, pooling of forecasts, multi-step estimation for forecasting, and forecast competitions. You and Mike Clements also developed a theory of forecasting, which included a taxonomy of forecast errors and a theory of unpredictability, and which had implications for parsimony and aggregation. From your theory of forecasting, you were able to improve forecasts themselves, as through robustification and nowcasting.

These clarifications and improvements set quite an agenda for our discussion. Let's start in 1991. How did you become involved in this parliamentary enquiry?

3.1 Error Correction Models

DFH: A misunderstanding of error correction models was responsible, and some background clarifies why. The debacle of Treasury mis-forecasts in the previous several years led to the parliamentary enquiry, and I was asked to be an adviser in the enquiry. The Treasury, in its empirical macro-modeling, had adopted what Davidson, Hendry, Srba, and Yeo (1978) called error correction models. These models incorporated feedbacks that were based on long-run relationships between variables, and those feedbacks were intended to capture how economic agents corrected previous mistakes. By the mid-1980s, research by Rob Engle, Clive Granger, Søren Johansen, Katarina Juselius, Peter Phillips, myself, and others had clarified the isomorphism between cointegration and error correction. However, we had not yet realized that

the so-called error correction models were actually equilibrium correcting, not error correcting. It took some time to grasp why error correction and equilibrium correction had distinct and different implications for forecasting.

Mike and I give an example in Clements and Hendry (1999, Section 2.5). If the equilibrium mean shifts, the model's forecasts will converge back to the *old* equilibrium mean—not the new one—and so the model will systematically mis-forecast. For example, if the equilibrium mean shifts up, the new data will tend to exceed the model's previous equilibrium. That measured disequilibrium will induce the model to predict a decline, hence the forecasts will move *opposite* to the outcomes.

NRE: The Treasury's forecasts exhibited this phenomenon. Why?

DFH: Legislation shifted the equilibrium mean. The Building Societies Act of 1986 increased credit availability in the British economy, thereby inducing a location shift. Consumers borrowed heavily and spent that money, with expenditure exceeding income for about three years and overheating the economy. In 1989, the government sharply raised income taxes and interest rates, and the economy crashed. On the resulting downswing, consumers' expenditure fell by much more than did personal disposable income. That behavior was incompatible with most economic theories of consumption, and incompatible with "error correction" models correcting previous errors. Thus, the poor performance of the Treasury forecasts was partly my fault.

Evidence submitted to the parliamentary Committee included many forecasts from many forecasters, and also dozens of *ex post* forecast evaluations that tried to sort out why forecasts had gone wrong. Surprisingly, there was almost no theory of *economic* forecasting. Most theories of forecasting were from the physical sciences or statistical sciences. Those theories were not relevant to economic forecasting, where intermittent structural breaks are a key data feature.

3.2 Mean Square Forecast Errors and Cointegration

NRE: Discovering this lack of theory prompted your initial collaboration with Mike Clements on mean square forecast errors and cointegration. Mean square forecast errors (MSFEs) are a standard tool for comparing forecasts from different models. Clements and Hendry (1993, 1995) questioned their value, generating considerable controversy. In fact, the discussants' published comments on your 1993 paper are longer than the paper itself. What was the origin of these papers?

DFH: Cointegration. Engle and Yoo (1987) had shown that imposing cointegration significantly improved forecasts in terms of MSFEs. This result seemed to demonstrate real advantages to cointegration—not just in modeling, understanding, and interpretation—but also in forecasting. We replicated Engle and Yoo's Monte Carlo experiments and found that imposing cointegration did *not* appear to reduce MSFEs. Our analysis of Engle and Yoo's results appeared as Clements and Hendry (1995).

The discrepancy between Engle and Yoo's results and ours arose because Engle and Yoo had calculated MSFEs for the variables' levels whereas we had calculated MSFEs for the cointegrating combination. Inadvertently, we had discovered that data transformations affected MFSEs. Additionally, we found that rankings across models

often depended more on the choice of data transformation, and less on whether or not cointegration was imposed, or even whether the model included the equilibrium correction term. We formalized algebraically these properties of MSFEs in Clements and Hendry (1993).

NRE: So, the ranking of different models' forecasts could alter, depending upon how you transformed the variables being forecast. In Ericsson (2008), I illustrated this problem by comparing forecasts in levels and forecasts in differences for two models of crude oil spot prices. For forecasts of the level of oil prices, the MSFE for the first model was more than four times that for the second model. However, for forecasts of the change of oil prices, the MSFE for the first model was less than half that for the second model. Thus, a simple transformation of the variable being forecast altered the MSFE ranking of the models, *with no change to any of the forecasts or to the underlying data*. Furthermore, the oil-price example illustrated that, for a *given* model, the MSFE was not invariant to the transformation from levels to differences. As you show in Clements and Hendry (1993), MSFEs lack robustness when the data are transformed, when forecasts are multivariate, and when forecasts are multi-step ahead. All three situations are common in economics.

DFH: Yes. Rankings by MSFEs often lack invariance. Even so, the discussion of Clements and Hendry (1993) was energetic. Many forecasters saw MSFEs as central to forecast evaluation, yet Mike and I had shown that model rankings by MSFEs depended upon the choice of transformation. Also, some discussants were critical of our results because we had assumed that the models being compared were congruent—namely, that the models had constant parameters and that their errors were homoscedastic white noise and unpredictable from the available information. In fact, congruence is not required to show the lack of invariance of MFSEs.

NRE: Indeed. Clements and Hendry (1993) showed that useful comparison of MSFEs required highly restrictive assumptions about the forecasts—that the forecasts must be of a single specific variable just one step ahead. Data transformations, multivariate forecasts, and multi-step-ahead forecasts are all outside that limited structure because they imply a vector of forecasts. As you and Mike discussed, a generalization of the MSFE does exist for a vector of forecasts—the predictive likelihood.

DFH: Predictive likelihood is the only direction-invariant measure, as it doesn't depend on nonsingular linear scale-preserving transformations of the system.

NRE: Has the predictive likelihood been used for forecast evaluation?

DFH: Not much, so far. Wallis (1993) pioneered its use, but its practical implementation was hindered because its calculation seemed to require having sufficient observations on all the multi-step-ahead forecast errors in order to estimate their variance-covariance matrix. However, results in Abadir, Distaso, and Žikeš (2014) encouraged me to revisit predictive likelihood in Hendry and Martinez (2016), where we show that one can evaluate multi-step-ahead system forecasts with relatively few forecast errors. Explicit loss functions also have come back into favor, as in Granger (2001).

3.3 Add Factors and Intercept Correction

NRE: Mike became a frequent co-author of yours. In addition to looking at MSFEs and cointegration, you and Mike re-examined the ubiquitous forecast tool known as “add factors”. Today we interpret add factors as a form of intercept correction (IC) and hence as a potentially useful method for robustifying forecasts against the effects of structural breaks.

This contrasts with your earlier harsh views on add factors. I remember a discussion that you had with Fed economists on forecasting. Peter Hooper was presenting forecast results on the Fed’s Multi-country Model (MCM) to a small workshop in 1985. You were critical of his adjustment of the forecasts with add factors: Why adjust forecasts if the model is good? How have your views on add factors evolved?

DFH: We need to go back at least to Klein (1971). Lawrie discussed *that* add factors might improve economic forecasting, but he gave no theory explaining *why* they might do so. There was no such theory at the time. My initial insight came from work with Mike Clements. We realized that some types of add factors might mitigate forecast failure that was caused by location shifts at the start of the forecast period. In Clements and Hendry (1996a), we showed analytically and in practice how intercept correction could improve forecasts in the face of location shifts. Intercept correction differences the forecast error that would have occurred otherwise. If the original error is ε , IC delivers an error $\Delta\varepsilon$ —where Δ denotes the first difference—thereby removing the forecast error’s systematic component. Consequently, IC is a valuable tool in the face of location shifts.

3.4 Leading Indicators

NRE: Leading indicators are another tool aimed at improving forecasts. Do they work?

DFH: That’s a good question. Rebecca Emerson and I found that the variables selected as leading indicators changed all too often, suggesting that they didn’t lead for very long. Also, picking leading indicators by maximum in-sample correlation was unreliable. In Emerson and Hendry (1996), we concluded that using only leading indicators for economic forecasting was not a fruitful route to pursue.

That said, leading indicators could have *some* role in forecasting. For instance, a cointegrated system can be written as a set of differenced variables that are explained by lagged cointegrating combinations and lagged differenced variables. That system is interpretable as a system of leading indicators by being based on effects that happened in the past. Also, higher frequency information such as from surveys may improve forecasting performance, with that information acting as a leading indicator. And, leading indicators may help predict turning points and breaks, as in Birchenhall, Jessen, Osborn, and Simpson (1999). These areas remain to be explored analytically.

3.5 Pooling of Forecasts

NRE: Bates and Granger (1969) proposed combining or “pooling” forecasts as a mechanism for improving forecast performance. Chong and Hendry (1986) later showed that pooling is unnecessary under the null of forecast encompassing. So, why should we pool forecasts?

DFH: In Bates and Granger’s framework, each model has information that the other model doesn’t. Pooling combines the information in the models’ forecasts.

Bates and Granger didn’t address the question of whether pooling *forecasts* was better than utilizing all the information in the *models* generating the forecasts. Hendry and Clements (2004) showed that there wasn’t a unique answer. It can pay to pool forecasts in some situations and not in others. Also, as implied by Chong and Hendry (1986), you don’t need to pool a model’s forecasts with any other forecasts if that model is congruent and encompassing.

NRE: Pooling is often viewed as being benign at worst, serving as insurance against bad forecasts by averaging across a range of forecasts. Is that view valid?

DFH: No. Years ago, Charles Nelson gave a humorous counter-example to that argument for non-selective pooling. Suppose that averaging always helped, and that averaging across more forecasts helped more. Then we could just ask the first twenty-five beggars on the street what their forecasts were, and average their forecasts with existing forecasts to improve the latter. Such averaging is unlikely to help!

More dramatically, imagine that you have a set of good models and also a set of poisonous models. Averaging the forecasts of the poisonous models with those of the good models can poison the pooled forecasts. To eliminate the poisonous models, we’ve got to select good models, whence averaging over just those good models’ forecasts may reduce the risk a little. However, in the literature, model averaging is often over all 2^N possible models for N explanatory variables. Most of those 2^N models are poisonous in our sense because they are distorted by omitted variables, unmodeled nonlinearities, intermittent location shifts, etc. One has to be careful which forecasts one averages across, and how that averaging is carried out.

NRE: Forecasts from different models may be of value in themselves. Divergence of different models’ forecasts can indicate breaks that are occurring and hence can serve as “canaries in the mine”. The Bank of England has used a suite of models in this manner, as Hatch (2001) discusses.

DFH: Agreed. When models are sufficiently different, they are not all affected in the same way by a major unanticipated shift. Including robust forecasting devices in the suite of models can help too. Robust devices are not affected systematically once the break point is past, although they will still mis-forecast as the break hits.

Also, *models* only need be robustified once a forecast failure occurs. Institutions that make forecasts should not be blamed for bad forecasts that were made in advance of an unanticipated break, but they should be blamed for systematically mis-forecasting after a break. Robust methods can help avoid such errors.

3.6 Multi-step Estimation for Forecasting

NRE: Can multi-step estimation or “dynamic estimation” improve forecasts?

DFH: Sometimes. Multi-step estimation has a long history in economics and has been proposed as a better approach for forecasting. There are two main approaches for forecasting multiple steps ahead: “standard”, and multi-step estimation. In the standard approach, one-step-ahead forecasts are generated from a conventionally estimated model. Those forecasts are then repeatedly substituted into the model ($h-1$) times, thereby generating forecasts h steps ahead ($h > 1$). In multi-step estimation, the first ($h-1$) lags in the model are substituted out. That derived model is then estimated, and it immediately produces h -step-ahead forecasts. Clements and Hendry (1996b) sought to mathematize and formalize multi-step estimation to see when it might work well and, if so, why.

Various theorems can be proved about multi-step estimation. In particular, the superiority of multi-step estimation requires nonstationarity of the data. If the DGP is stationary, h -step-ahead forecasts are close to the model’s equilibrium mean, so the choice of approach doesn’t matter much. Conversely, multi-step estimation can be useful if the model is mis-specified and there are unit roots, or near-unit roots, or slow trends, or breaks. Even then, multi-step estimation need not dominate the standard forecasting method: see Chevillon and Hendry (2005). Also, multi-step estimates for a given model can be worthwhile to include in pooling, provided that such a model is known to be non-poisonous.

3.7 Forecast Competitions

NRE: Many comparisons of forecasts have sought to determine which techniques are best for forecasting. Makridakis and Hibon (2000) report the well-known M3 competition. What’s the background?

DFH: The *International Journal of Forecasting* has hosted a series of forecasting competitions organized by Spyros Makridakis, hence the “M” in M3. Many different time series were divided into subperiods, each of which was then forecast by many methods, albeit usually only one step ahead. Various evaluation criteria were applied to each forecasting device on each dataset to find which methods had the best *ex post* forecast performance as measured by the chosen criteria. Those methods then “won” the competition. Because parsimonious methods such as damped trend often did well, whereas less parsimonious methods such as econometric models did poorly, Makridakis and Hibon (2000) concluded that parsimony was key to good forecast performance.

NRE: What led you to evaluate the M3 competition?

DFH: I could not understand why parsimonious models *per se* should do so well. For instance, the sample mean of the level is parsimonious, but it often is a dreadful forecast.

To understand the empirical results in the M3 competition, Mike Clements and I developed a general analytical framework describing a taxonomy for forecast errors.

That taxonomy revealed which sources of forecast error most affected each forecasting method, thus clarifying why some methods out- or under-performed others, and when. For intermittent location shifts, all methods mis-forecast at the break. However, after the break-point, methods that are not robust to such breaks tend to make systematic forecast errors, whereas robust methods get the forecasts back on track. Hendry and Doornik (1997) demonstrated these implications in Monte Carlo simulations.

The taxonomy showed that rankings of forecasts should not depend particularly on the number of parameters in either the model or the DGP, whereas the rankings do depend on the robustness of the forecasting devices. The design of forecast competitions such as M3 happened to favor robust devices by having many short forecasting subperiods with intermittent location shifts in the data, thus giving the impression that parsimony *per se* was advantageous in forecasting. In Clements and Hendry (2001), Mike and I showed that many of the key empirical results in the M3 competition were derivable from the taxonomy.

4 A Theory of Economic Forecasting

NRE: That insight brings us to a theory of economic forecasting. Let's focus on congruence, parsimony, the forecast error taxonomy, and the nature of unpredictability. What are the criteria for building a good forecasting model? Do congruence and parsimony matter?

4.1 Congruence and Parsimony

DFH: I initially assumed that congruence was essential to good forecasting—that the best in-sample model should be the best forecasting model. There were counterexamples, however. Allen and Fildes (2005) found little relation between the forecast performance of a model and how well-specified it was empirically. Consider a double-differenced predictor of the form $\widehat{\Delta x}_{T+1|T} = \Delta x_T$. The difference Δx is being predicted one period ahead for the observation $T + 1$, conditional on a sample ending with observation T , the “forecast origin”. The empirical model corresponding to this double-differenced predictor is often non-congruent and hence mis-specified, but it also often performs well in forecasting.

Surprisingly, good out-of-sample forecast performance does not reliably indicate whether an empirical model adequately describes the phenomenon being modeled, nor need forecast performance entail anything about the theory on which that model is based. IC highlighted the disconnect between a model's forecast accuracy and its verisimilitude. The same model could forecast poorly without IC and well with IC.

NRE: What *is* the role for congruence in forecasting?

DFH: A congruent model avoids in-sample location shifts, which could bias the forecasts and would tend to increase forecast errors. Congruence also ensures the smallest innovation variance for the model's error on the information used, and it delivers valid in-sample inferences, an especially important consideration when selecting variables.

NRE: Was parsimony also up for debate as a criterion for choosing a forecasting model?

DFH: Absolutely. Parsimony could exclude irrelevant variables that might shift and thereby cause forecast failure. Most “real-world” forecasters that I knew claimed that non-parsimonious models forecasted badly and so one should always use a parsimonious model. That claim was sustained by empirical experience, but not by analysis, as it confounded parsimony with other features of forecasting models. For example, differenced devices such as Box–Jenkins models and Bayesian vector autoregressions with a Minnesota prior appeared to out-forecast other methods. Our research on MSFEs and the M3 competition revealed that evaluating forecasts of growth rates delivered smaller MSFEs than the corresponding forecasts implied for levels or differentials. Apparent superior performance might reflect the evaluation criterion and also the differencing of the data that achieved robustness after location shifts, rather than model parsimony.

NRE: What sort of analysis would help us understand such forecast behavior?

DFH: We needed an empirically relevant theory of economic forecasting to ascertain what contributed to a good forecasting model. That theory would allow for the model being mis-specified and hence being distinct from the DGP. That theory would also allow for estimation of parameters from inaccurate observations in an integrated-cointegrated system that intermittently and unexpectedly altered from structural breaks. It was not obvious that such a theory could be developed! The implied research agenda was highly speculative and I had difficulty getting funding—until the Leverhulme Foundation generously offered me a Research Professorship.

4.2 A Taxonomy of Forecast Errors

NRE: To help interpret the problems that arise in economic forecasting, you developed a taxonomy of the sources of forecast error. Initially, you solved the taxonomy for vector autoregressive models and simple time-series models. More recently, you’ve considered open dynamic simultaneous systems and nonlinear formulations. One key insight from the taxonomy is that location shifts are a common source of predictive failure in empirical forecasting. Why are location shifts so central to predictive failure, and what are the other sources of forecast error?

DFH: The taxonomy delineates all possible sources of forecast error—nine sources in total. These sources derive from the three components of a model:

- i.* unobserved terms,
- ii.* observed stochastic variables, and
- iii.* deterministic terms.

The first component is what the model fails to explain, and it thus includes mis-measurement of the data at the forecast origin, omitted variables, and the innovation errors in the DGP. The second and third components characterize what *is* modeled

and often correspond to the slope parameter and the equilibrium mean. So, for concreteness, I'll focus on the slope parameter and the equilibrium mean when discussing these two components.

Each of the model's three components is itself subject to three potential problems:

- a. estimation uncertainty,
- b. mis-specification, and
- c. change in the DGP's parameter values.

That leads to a 3×3 decomposition, implying nine sources of forecast error.

NRE: What are the consequences of forecast error?

DFH: The consequences depend on the source of forecast error. The taxonomy allows us to derive the effects of each source, so let's consider each source in turn. As for the consequences, I'll focus on the forecast's bias and the forecast error's variance, which are often of primary concern to forecasters.

- i(a) Estimation of unobserved terms.* What's not modeled—that is, the unobserved terms—can affect forecast performance in many ways. For example, mis-estimation of the most recent observed data (the forecast origin) can induce systematic predictive failure. This problem can be resolved in part by nowcasting. EuroStat (2016) provides a recent overview. Estimation of the forecast origin also adds to the forecast error variance.
- i(b) Mis-specification of unobserved terms.* Omission of a relevant variable *per se* does not induce forecast bias. However, such an omission increases the forecast error variance. Including the omitted variable in the model can reduce the forecast error variance, but with an offsetting effect on the estimation variance. Examples of omitted variables include mis-measurement of in-sample data and ignored nonlinearity.
- i(c) Change of unobserved terms.* By definition, the DGP's innovation error is unpredictable from all possible information. The innovation error has a zero mean, and its variance is whatever it is. That said, changes in the distribution of the innovation error will affect forecast properties.
- ii(a) Estimation involving stochastic variables.* Biased estimation of the slope coefficients—as from incorrect exogeneity assumptions—need not bias forecasts, but estimation does increase the forecast error variance.
- ii(b) Mis-specification involving stochastic variables.* Mis-specification also need not bias forecasts, but it does increase the forecast error variance.
- ii(c) Change involving stochastic variables.* Surprisingly, an unmodeled change need not bias forecasts, provided the model's equilibrium mean remains constant.
- iii(a) Estimation involving deterministic terms.* Estimation involving deterministic terms need not bias forecasts, but it increases the forecast error variance.

iii(b) Mis-specification involving deterministic terms. Mis-specification involving deterministic terms is potentially serious, as when that mis-specification is due to unmodeled in-sample location shifts or changes in trend.

iii(c) Change involving deterministic terms. An out-of-sample change involving deterministic terms—as with a change in the equilibrium mean—is the fundamental problem in economic forecasting. Such a change implies a location shift, which induces systematic mis-forecasting.

[Appendix A portrays the taxonomy tabularly.] What I’ve just described treats each source of forecast error in isolation. If forecast errors arise from multiple sources, interactions between sources may also matter. Comprehensive derivations and analyses of the taxonomy appear in Clements and Hendry (1994, 1998, 1999, 2006).

NRE: How did your thinking about the taxonomy evolve over time?

DFH: One key insight came during a seminar in which I was explaining a very early version of the taxonomy. I noticed that the change in the slope coefficient [source *ii(c)* above] was multiplied by the deviation of the forecast-origin data from its equilibrium mean. Consequently, if forecasting happened to start when the data were in equilibrium, changes in the slope parameter would not affect the forecast errors. Indeed, if the mean of the data stayed constant and the forecast origin were accurately measured, forecasts would not be systematically biased—even if all the other problems were present. Conversely, out-of-sample location shifts would systematically bias the forecasts, even if the forecast model were the in-sample DGP itself. I suddenly realized this result in the middle of the seminar and expressed my astonishment!

NRE: Predictive failure is thus easy to detect when there are location shifts, and it’s more difficult to detect otherwise.

DFH: Exactly. Predictive failure due to a location shift is easily detected because the MSFE includes the squared shift in the mean. Relatedly, the magnitude of the location shift depends on the units of measurement, even when the model is in logs. The means of variables often have units dependent on using \$millions rather than \$billions (say), or on the base year of a price index. For example, the equilibrium mean of a model may change from 1 to 2, or from 100 to 200. It’s surprising that such a choice is important, but we’ve confirmed it in Monte Carlo simulations.

NRE: Would that alter the units of the error variance as well?

DFH: It would do so for the original levels of the variables, but not for their logs.

NRE: Starting with Clements and Hendry (1994), the taxonomy led to the question: Can a general theory of economic forecasting be developed?

DFH: I think that the answer is “yes”, and we’ve made considerable progress in developing a general theory. For instance, we now have a theory of macroeconomic forecast failure that doesn’t assume how the model is estimated, how badly mis-specified it is, or what changes occur in the economy. Many aspects still need more research, though, including how to forecast breaks, and how to best select forecasting models for realistic economic processes.

4.3 Unpredictability and Aggregation

NRE: The concept of unpredictability has provided a number of insights in forecasting. How did this concept develop?

DFH: From a mistake. I was aware of the Doob (1953) decomposition of the likelihood by a sequential factorization, which delivers a martingale difference sequence that is unpredictable from all the available information. However, I had falsely concluded that all variables were reducible to a function of unpredictable components.

I realized my mistake when visiting the Bank of Portugal. I was staying at Hotel Palácio de Seteais in Sintra and sitting poolside—about the only time in my life sitting by a swimming pool, since I’m not much into that sort of inactivity—when a theory of unpredictability suddenly dawned on me. In that theory, unpredictability is an *intrinsic* property of a non-degenerate random variable in relation to an information set. With a sufficiently large information set, the unexplained component of a variable would be unpredictable on *any* information set.

NRE: That theory proved limited in its applicability. Reducing the information set that a forecasting model uses need not worsen the model’s forecasts, and it might even improve them. For instance, with an unanticipated location shift, structural models could forecast worse than a random walk, even though the latter uses a minimal information set. This theoretical result parallels Cooper and Nelson’s (1975) empirical findings.

DFH: Still, less information *could* worsen the model’s forecasts. I kept puzzling as to why additional information would sometimes help and othertimes hinder. Then, from working on Clements and Hendry (2005), I realized that unpredictability has three distinct senses, with each sense defined by an information set. The implications of using more or less information could differ, depending upon the type of information being considered and so upon the sense of unpredictability. That resolved the previous apparently contradictory results.

NRE: The “knowns” imply what’s unknown, and the “unknowns” are what matter for forecast performance.

DFH: Agreed. The three senses of unpredictability are called intrinsic, instance, and extrinsic.

- *Intrinsic unpredictability.* This is the conventional sense of unpredictability. The information set is the sigma field generated by the history of all the variables, although the variables’ particular realizations are unknown *ex ante*.
- *Instance unpredictability.* Here, “outliers” are expected to occur—as with a (known) thick-tailed distribution—but their magnitudes, signs, and timings are unknown.
- *Extrinsic unpredictability.* Extrinsic unpredictability arises from unanticipated changes in the distribution. Even if the variables’ initial distribution is known, their distribution can alter unexpectedly, as with an unanticipated location shift. Such shifts can have pernicious effects on modeling, forecasting, and policy analysis; see Hendry and Mizon (2014).

NRE: What other benefits did the theory of unpredictability bring?

DFH: It helped clarify the methodology of forecasting aggregated data. Kirstin Hubrich—then at the ECB—was forecasting aggregated series by aggregating the forecasts of the disaggregates. She was also forecasting the aggregate directly from its own lags. As we showed in Hendry and Hubrich (2011), the theory of unpredictability suggested a new approach: use the disaggregates to forecast the aggregate directly. Because an aggregated series is a linear combination of its disaggregated components, forecasting the disaggregates individually and adding them up should not be better than forecasting the aggregate directly from the disaggregates, provided that the information set explaining the disaggregates is the same as the information set explaining the aggregate.

The research with Kirstin focused on disaggregation of a variable. In Clements and Hendry (2011a), Mike and I considered *time* disaggregation, which can improve forecasts after a break because the break can be detected more quickly.

4.4 The Role of Mathematical Analysis

NRE: How important was *mathematical* formalization in developing the theory of economic forecasting?

DFH: It was fundamental. Readers of Clements and Hendry (1998) and Clements and Hendry (1999) often expressed surprise that the books were so mathematical. However, a general theory of economic forecasting required mathematical formalization to clarify how and why important aspects of empirical forecasting resulted from specific features of the forecasts' implementation. That formalization did not require details about the forecast model, DGP, or data properties, such as multi-collinearity and shifts in the DGP. Nevertheless, we derived useful theorems about forecasting.

NRE: In economics, there is often a divergence between theoretical models of the economy and the economy itself.

DFH: Unfortunately so. Many economists emphasize a theory about the world in which agents are intertemporal optimizers using all available sources of information, subject to various constraints. Macroeconomic theory models themselves have improved over time, partly by adding *ad hoc* features such as smoothing. However, Hendry and Mizon (2014) show that standard intertemporal optimization theory is not applicable when unanticipated location shifts occur, as with the recent financial crisis. As the typology highlights, virtually all models suffer forecast failure when unanticipated location shifts occur. In Hendry and Johansen (2015) and Hendry (2016), Søren and I present a systematic encompassing framework for deciding between alternative empirical macroeconomic models.

NRE: Do you see a similar divergence between theory and practice in the forecasting literature?

DFH: No. Most researchers who work on the theory of forecasting keep a close eye on empirical findings, and they also often undertake pseudo-forecasting of historical outcomes to see how well various forecasting devices would have performed.

NRE: While mathematics and statistics help understand empirical findings, the literature on forecasting is primarily focused on improving forecasts.

DFH: Indeed! I, for one, take very seriously the folklore and rules-of-thumb in the empirical forecasting literature. When forecasters show that add factors improve forecasts, I want to know why. When they argue that parsimony is advantageous in forecasting, I want to know why. Add factors (as intercept corrections) turn out to be an ingenious robustification device, whereas parsimony *per se* does not seem to be advantageous. Parsimony may appear to help in forecasting, but only coincidentally so. Some parsimonious models such as random walks are formulations that are also robust forecasting devices. The mathematical formulation of economic forecasting has been central to achieving such an understanding.

5 Improvements to Forecasting Techniques

NRE: The discipline has moved forward dramatically since you reported to the 1991 parliamentary Committee. There’s now a much better understanding of the art and science of economic forecasting. The taxonomy clarified the sources of predictive failure. It also led to new techniques that robustify forecasts after breaks and that augment robust devices with information from economic models. Robustification has led to research on nowcasting and, from a completely different route, impulse indicator saturation. Let’s start with robustifying forecasts, as occurs with the differencing of vector equilibrium correction models, aka DVEqCMs or differenced VEqCMs.

5.1 Robustification

NRE: Virtually all standard economic models are equilibrium correction models. That includes dynamic stochastic general equilibrium (DSGE) models, New Keynesian Phillips Curve models, structural vector autoregressions, and so-called error correction models. When the equilibrium mean alters, the model’s equilibrium correction term pushes the model’s forecasts towards the old equilibrium, not the new one, inducing the sort of systematic predictive failure that we’ve often seen in practice.

Intercept correction and differencing can robustify the forecast of an equilibrium correction model because they serve as good proxies for such shifts in equilibrium. Hendry (2006) formalizes this; and Castle, Clements, and Hendry (2013) illustrate empirically with an assessment of robustified UK GDP forecasts. Does the taxonomy provide insights on forecast robustification?

DFH: Very much so. The taxonomy shows that few things can go wrong in forecasting a variable if the forecasting model for the second difference of that variable has no parameters and no deterministic terms. If the data do not accelerate, the second difference of the variable being forecast (x , say) has a mean of zero. Letting t be the time subscript, then $E(\Delta^2 x_t) = 0$, which implies that Δx_T is an unconditionally unbiased forecast for Δx_{T+1} . However, Δx_T is the current growth of x , not its future growth, so such a “forecast” device never really forecasts; but Δx_T will be close to Δx_{T+1} in the absence of acceleration.

NRE: Δx_T may be an unbiased forecast. Does Δx_T have any other interpretation?

DFH: Yes, and a surprising one. The first difference Δx_T is a single measure that aggregates almost all the information needed in forecasting. This needs explaining.

As a reference point, we build congruent, encompassing, cointegrated models to test theories, understand the economy, and conduct policy analysis. These models also need to account for breaks and other nonstationarities. For forecasting, though, we can difference these models (as with DVEqCMs) to eliminate deterministic terms such as intercepts and location shifts. Doing so introduces the growth rate Δx_T in the model for forecasting Δx_{T+1} ; and Δx_T depends on the cointegrating relationship. This new system thus retains the economics and the policy-relevant causal information that underlie the model. Differencing the model also introduces the first difference of the model's other economic variables. For instance, when forecasting Δx_{T+1} with a DVEqCM, the primary right-hand side variable becomes Δx_T .

NRE: Aha! Because Δx_T is generated by the DGP, it necessarily includes relevant variables for forecasting Δx_{T+1} , whereas a model of Δx_T is a simplification of the DGP and need not include the relevant variables. Furthermore, there's no need to disentangle the individual components of a model when forecasting, unlike when modeling or in policy analysis. The data themselves provide the basis for forecasting. What practical implications does differencing have for forecasting?

DFH: Differencing creates a system that is robust after location shifts because Δx_T includes all stochastic and deterministic shifts and any variables omitted from the model. Moreover, use of Δx_T to forecast Δx_{T+1} obviates the need to estimate model parameters. The current growth rate Δx_T thus captures everything you always wanted to know about forecasting, but were afraid to ask!

NRE: Are there yet other interpretations of Δx_T ?

DFH: Remarkably, yes. Let me backtrack to VEqCMs to explain. In the simplest VEqCM, we forecast Δx_{T+1} by its mean growth rate and the current disequilibrium, which is the deviation of the cointegration vector from the equilibrium mean. Both the mean growth rate and the current disequilibrium employ full-sample estimates of the parameters. In the DVEqCM, however, the mean growth rate is estimated by the current growth rate Δx_T , and the disequilibrium is estimated by the deviation in the cointegrating relation from its previous value. Both terms in the DVEqCM are estimates that use only *current-dated* Δx , although the cointegrating coefficients themselves need to be estimated with a longer sample.

The VEqCM thus uses fixed values of its two key components, shifts in which can cause forecast failure. By contrast, the DVEqCM uses estimates based on just the current observation, which may be more relevant for forecasting than the full historical sample. This approach generates a class of “data-based” forecasting devices, which could utilize a single observation (as in the DVEqCM), a subset of observations, or the full sample (as in the VEqCM). The observations actually used imply a trade-off between rapid adaptation and imprecision in estimation. Other approaches, such as in Phillips (1995), adapt the forecasts to location shifts through automated variable reselection and parameter estimate updating.

Eitrheim, Husebø, and Nymoen (1999) empirically documented these implications of the taxonomy by comparing real-world forecasts from Norges Bank’s macro-model RIMINI with forecasts from simple robust devices such as Δx_T . RIMINI produced 12-quarter-ahead forecasts, and also three sequential 4-quarter-ahead forecasts over the same 12-quarter period. The robust methods lost out at the longer forecast horizon, but quite often they won at 4 quarters ahead, despite using very little information. Bårdsen, Eitrheim, Jansen, and Nymoen (2005) further examined robustification devices in forecasting comparisons with RIMINI.

NRE: You mentioned that parsimony is not the key to robust forecasts. A simple example compares two one-parameter models that have very different forecasts. One parsimonious forecast is the full-sample mean through period T , namely, $\hat{x}_{T+1} = \sum_{t=1}^T x_t/T$. Another parsimonious forecast is the data’s previous value, namely, $\tilde{x}_{T+1} = x_T$. Those two models generate vastly different forecasts after location shifts, yet both models are equally parsimonious.

DFH: In an important sense, the second model is actually more parsimonious than the first because the second model has no parameters to estimate, whereas the first model needs to estimate the mean. That said, a DVEqCM can deliver robust forecasts after location shifts, even though the DVEqCM might be highly non-parsimonious.

Returning to your two simple models, the ranking of their forecasts may alter if (say) the problem is an IID measurement error rather than an unanticipated location shift. The first model’s forecast is $\sum_{t=1}^T x_t/T$ and so averages the measurement errors, with a weight of $1/T$ on any individual measurement error. The second model’s forecast is x_T itself and so gives a unit weight to the T^{th} -period measurement error and a zero weight to all other measurement errors, hence increasing its forecast error variance relative to that of the first model.

NRE: The choice of forecasting model and its form of robustification involves trade-offs between the consequences of measurement errors at the forecast origin and the consequences of location shifts. As source $i(a)$ in the taxonomy highlights, mis-measurement of the forecast origin is a serious problem.

DFH: Location shifts are not always bad news. For near-integrated measurement errors, Duffy and Hendry (2015) show that location shifts that co-break across variables can reveal the underlying relationships. The specifics of the particular data measurement system should be considered carefully.

NRE: At a practical level, the consequences of measurement errors may be especially pronounced in nowcasting, so let’s turn to that topic.

5.2 Nowcasting

DFH: I started thinking about nowcasting in a more structured way when Mike Clements and I (2003) were consulting for the UK Statistics Commission, evaluating how the UK’s Office for National Statistics (ONS) calculated its flash estimates of the national accounts.

NRE: Nowcasting can imply measurement errors of the forecast origin. Sometimes, those errors are systematic and large, as with recent British and US economic statis-

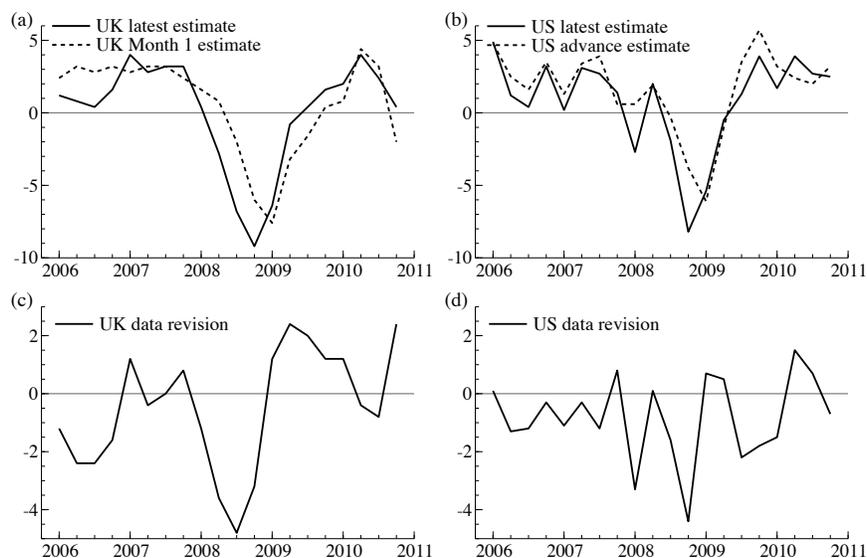


Figure 1: Initial estimates and latest estimates of UK and US GDP growth (quarterly rates, measured in percent per annum), and the corresponding data revisions.

tics. Can improved methods of nowcasting help reduce real-time forecast problems that arise from mis-measuring the forecast origin?

DFH: Definitely. For the United Kingdom, large data revisions during the financial crisis were not surprising in light of the methods that the ONS used to produce their flash estimates. The ONS releases its flash estimate (or “Month 1 estimate”) of quarterly GDP growth about a month after the quarter’s end, and that flash estimate is derived in part from many disaggregate observations. Some disaggregate observations become available too late for inclusion in the flash estimate, so those missing observations are “infilled”, based on interpolation models such as Holt–Winters. Sudden changes in data behavior—as occurred during the financial crisis—can make interpolation methods inappropriate; and they led to flash estimates of aggregate economic growth that were systematically above the final data in the downturn, and systematically below the data in the upturn.

NRE: The large systematic data revisions during the financial crisis are visually striking. In Figure 1a, I have graphed the Month 1 estimate of UK GDP growth, along with the “latest estimate” of UK GDP growth—that is, as measured today for values in the past. Directly below this graph, Figure 1c plots the implied data revisions, constructed as the difference between the two growth rate series. All quarterly GDP growth rates for 2008 were revised downward, and markedly so, with data revisions of 3%–5% per annum for Q2, Q3, and Q4.

Large systematic data revisions are also apparent in US data. The US Bureau of Economic Analysis releases its “advance estimate” approximately 30 days after the end of the quarter, and it made errors similar to those made by the ONS. I have

graphed that advance estimate and corresponding latest estimate in Figure 1b, and the implied data revisions in Figure 1d. The latter show that the BEA systematically over-estimated US GDP growth during the financial crisis, including by 3.3% per annum for 2008Q1 and by 4.4% for 2008Q4. These mis-measurements made it difficult for policymakers to ascertain the timing and extent of the crisis, as Stekler and Symington (2016) and Ericsson (2016) discuss. [Appendix B documents Figure 1's data.]

DFH: Systematic errors such as these have led to proposed improvements to nowcasting, as documented in EuroStat (2016). The taxonomy delineates what does and does not cause forecast failure and so has implications for nowcasting. When a statistical agency estimates (say) GDP growth from a set of disaggregate components, the agency should check whether previous forecasts of those components are close to their now known outcomes. If they aren't, a location shift is probably responsible, so any missing disaggregates should be infilled, taking into account information about recent breaks. Considerable contemporaneous information is available for nowcasting, including surveys, Google Trends, mobile phone data, and prediction markets. All could be used to improve the accuracy of forecast-origin estimates. Also, nowcasts made this way could be created before the end of the reference period, thereby reducing the delay with which flash estimates appear.

NRE: Your recent papers on nowcasting also employ automatic model selection with Jurgen Doornik's and your econometrics package *Autometrics*, which is documented in Doornik and Hendry (2013). How can automatic model selection help nowcasting?

DFH: One way is by building forecasting models of the disaggregated series. That requires effective methods of automatic model selection. Model selection is a surprisingly contentious issue, possibly because there are so many poor selection methods. Some nowcasting approaches seek to avoid model selection altogether by summarizing the information from large numbers of variables by using principal components or factors: see Forni, Hallin, Lippi, and Reichlin (2001), Artis, Banerjee, and Marcellino (2005), Stock and Watson (2011), and Castle, Clements, and Hendry (2013).

NRE: Is *Autometrics* better at model building than manual model selection?

DFH: We think so. Hendry and Doornik (2014) document how automated approaches such as *Autometrics* avoid the pernicious properties of many earlier approaches, which employed poor algorithms and inappropriate selection and evaluation criteria. Whether starting from a large model that nests the DGP or from a model that is the DGP itself, model search à l'*Autometrics* retains roughly the same relevant variables, and it obtains a controlled average number of irrelevant variables.

5.3 Impulse Indicator Saturation

NRE: How general are such automated model selection methods?

DFH: Amazingly general. For instance, the recent theory of model selection can handle more potential variables than there are observations.

NRE: That sounds impossible.

DFH: Even my co-author Søren Johansen initially thought so! I remember our discussion vividly. Søren and I were in Stockholm for the 2003 Nobel prize award ceremonies for Clive Granger and Rob Engle, and I tried to persuade Søren to take the idea seriously. We were walking across a bridge in freezing December weather—not the most opportune setting.

The canonical case for this problem in model selection is impulse indicator saturation (IIS), in which the set of candidate explanatory variables includes a dummy variable for each observation. The solution to this canonical case is implicit in several existing techniques. For instance, as Salkever (1976) shows, the Chow (1960) statistic can be calculated by including zero-one indicator variables for all observations in the forecast period and then testing those indicators' joint significance. Recursive estimation is another example. Its “forward” version can be calculated by estimating the model, including an indicator variable for every observation in the latter part of the sample, and then sequentially removing the indicators, one indicator at a time. Both forward and backward versions of recursive estimation can be calculated in this fashion. Together, they require indicators for all observations in the sample and thus analyze as many potential variables as there are observations.

NRE: Andrews's (1993) unknown breakpoint test and Bai and Perron's (1998) generalization thereon are interpretable in this way. How did you come up with IIS?

DFH: I discovered it by a fluke. I was a late entrant to Jan Magnus and Mary Morgan's (1999) econometric modeling competition, with my analysis published as Hendry (1999) in their book *Two Experiments in Econometrics*. Researchers in the competition were given data from 1929 to 1989 on the US demand for food, thus building on Tobin's (1950) empirical analysis through 1948. Most investigators discarded the data for the interwar period and for World War II as being too difficult to model. For example, a standard demand model fitted over the whole sample delivered positive price elasticities.

NRE: Shouldn't more observations be better than fewer?

DFH: Exactly. A larger sample is better—if it is used in the right way. I wanted to replicate other researchers' findings for the postwar subsample while actually using the whole sample. To do so, I estimated a given model over the whole sample, including indicator variables for all observations up to the beginning of the postwar period. Several of those indicator variables were highly significant. Three were associated with a food program in the United States during the Great Depression. Unsurprisingly, the food program affected the demand for food. The other significant indicator variables were for years during World War II.

I then reversed the whole procedure, estimating the model over the whole sample but including indicators for the postwar period. That was equivalent to estimating the model over the first part of the sample. A few postwar indicators were marginally significant, as the corresponding Chow test revealed. Between these two estimations, I had included an indicator for every observation, albeit in two large blocks. All indicators could be considered—just not all at once.

Finally, I estimated the model over the whole sample, including the indicators selected in the two subsample estimations. Of those indicators, only those for the food program and World War II were significant, and they had clear economic explanations. By including just those indicators, the whole sample could be adequately captured by a single model. The large data variability during the interwar period and World War II also greatly reduced the estimated economic parameters' standard errors, relative to those in the same model estimated on the postwar period alone.

NRE: That explains IIS as a procedure. What's the distributional theory for IIS?

DFH: To understand IIS's properties, Hendry, Johansen, and Santos (2008) considered a stylized version of IIS with a split-half sample. Under the null hypothesis that there are no outliers or breaks in the DGP, IIS incurs only a small loss of efficiency. For example, for a sample size of 100, on average one impulse indicator out of the 100 total would be significant at the 1% significance level. Because an impulse indicator merely removes one observation from the sample, the method is 99% efficient under the null hypothesis. IIS is almost costless, despite searching across 100 indicators.

Under the alternative hypothesis, IIS can detect multiple outliers and location shifts. Castle, Doornik, and Hendry (2012) demonstrate high power for multiple location shifts that are "large enough". Importantly, IIS can detect breaks that are near the ends of the sample. Johansen and Nielsen (2009) generalized the theory of IIS to include autoregressive distributed-lag models with or without unit roots, and they proved that IIS did not affect the rate of convergence of other parameter estimates to their population values.

NRE: IIS adds blocks of dummies to estimation and model selection. IIS can consider many blocks, thereby allowing many different alternatives to be considered.

DFH: That feature of IIS has remarkable implications. Under the null, an indicator for a given observation is significant only if it is discrepant. Its significance doesn't depend particularly on how or how often you split the indicators into blocks, provided that the blocks are large and that multiple search paths are explored.

Let's now consider the alternative of multiple unmodeled breaks or outliers. For ease of discussion, let's assume two outliers. Detection of one outlier (the first, say) can be difficult unless the *other* outlier is accounted for. Failing to account for the second outlier in the model induces a larger estimated error variance, making the first outlier appear less significant than it actually is. Hence, there's a need to include sufficient indicators to capture all actual outliers.

Single-step algorithms such as step-wise regression and Lasso can fail badly in such situations because they lack a mechanism to ensure capturing all relevant outliers and breaks. For instance, in a Monte Carlo simulation with a ten standard deviation break that persists for 25% of the sample, step-wise regression and Lasso can easily fail to detect even one outlier, whereas IIS picks up all of them. Many early model selection algorithms were poor because they used single-step expanding searches that explored only one search path. Hoover and Perez (1999) showed the advantages of multiple-path contracting searches that are guided by encompassing evaluations.

NRE: The block-search algorithm can be generalized to include candidate variables

such as standard economic variables, and not just impulse indicators. Even if the number of candidate variables N is greater than the number of observations T , block searches can still be implemented, so long as the number of variables in each block is smaller than T . To consider lots of alternatives, we can iterate the block search and vary the composition of the blocks as we iterate. What are the properties of the algorithm in such a situation?

DFH: Proofs of the properties of such an algorithm are similar to those for “standard” IIS. What matters appears to be the number of degrees of freedom used up from examining a set of candidate variables, rather than the number of observations corresponding to a set of indicators. Purely contracting searches are not possible for $N > T$, but the principle of examining many large blocks remains. Blocks help avoid inadvertently eliminating variables that are correlated with already selected variables, and blocks help detect effects that are camouflaged by breaks.

Surprisingly, we can select jointly across lag length, functional form, relevant variables, and breaks, even when doing so implies $N > T$. Under the null, estimates of the parameters of interest are still relatively efficient. Under the alternative, it is particularly important to consider all of these complications jointly because they are likely to be connected.

Other procedures tend to address just one or a few issues, rather than all of them at once. Nonparametric statistics can determine functional form but, in so doing, assume constant parameters, accurate measurements, and inclusion of all relevant variables. Robust statistics can tackle contaminated data but assume an otherwise correct specification. The block-search approach aims at considering all complications together. As Hendry and Johansen (2015) show, it can do so without distorting the distribution of the parameter estimates of a correct theory-specified model. Søren and I serendipitously discovered that result while trying to prove something else.

5.4 Forecasting Breaks

NRE: By dealing with location shifts, IIS links model selection to forecasting. Unmodeled location shifts have adverse effects on forecasting, so how should we deal with them? I actually have three questions, rather than just that one.

- How can we forecast a break before it occurs?
- How can we recognize a break once it does occur?
- How can we improve forecasts once a break has begun?

Let’s consider these questions in reverse order, starting with the last one.

DFH: Once a break has started, differencing and intercept correction offer effective immediate ways for avoiding systematic forecast failure. Alternatively, we might estimate the break’s functional form to help predict the next period’s outcome. That function might be as simple as a step shift. Castle, Doornik, Hendry, and Pretis (2015) have proposed a generalization of IIS called step indicator saturation as an algorithm for detecting step shifts. An in-sample detected step shift generates an intercept correction for out-of-sample forecasts, thus linking saturation techniques to robustification of forecasts in the presence of breaks.

Unfortunately, the initial subsample for a break period is typically small, making estimation of the break’s functional form difficult. Castle, Fawcett, and Hendry (2010) find that estimating the parameters of an ogive functional form costs as much in terms of MSFE as intercept correction or differencing, even if the form and timing of the break are known. Repeated breaks offer more promise, as with repeated volcanic eruptions, where the functional form of an eruption’s effect on air temperature is known. For instance, Pretis, Schneider, Smerdon, and Hendry (2016) show that the return to normal temperatures after a new eruption can be forecast relatively accurately, once the lowering of temperature from that eruption has been observed.

NRE: We already have an answer to the second question: saturation techniques such as IIS can help detect a break, once it occurs. So, let’s turn to the first question: How can we forecast a break before it occurs? As an example, very few people foresaw the recent financial crisis.

DFH: In retrospect, there was evidence about increasing stress in the financial system, prior to the crisis. That evidence is akin to vulcanologists’ measurements of increased pressure inside a volcano, prior to a volcanic eruption. Such volcanic data has helped forecast eruptions, as with Mount St. Helens in 1980. Eruptions were once viewed as “beyond science”, but vulcanologists have made great progress in understanding what causes eruptions, and therefore what to measure so as to forecast eruptions.

Such progress is promising for economic forecasting too. Proper monitoring systems of the financial sector might have flagged a looming problem through the build-up of sub-prime loans, their high default rates, and excessive leveraging of the institutions making those loans, all interacting with house prices that were out of line with average incomes; see Eichner, Kohn, and Palumbo (2010). That said, I have applied many times for funding to investigate how to forecast breaks, but the applications have almost always been rejected, and often with extremely sceptical reactions.

NRE: Returning to the first question, can one reliably forecast an economic break?

DFH: Perhaps; and that’s an age-old question. To quote Smith (1927, p. 457) from nearly a century ago, “[t]his possibility of change in relationship is perhaps the most fundamental weakness in all statistical forecasting”. Smith (1929) followed up on this view in a paper “Judging the Forecast for 1929”—an interesting year historically! I wish I had been aware of Smith’s work decades ago.

Returning to the question, it’s helpful to distinguish between two kinds of information: regular “causal” information, and information reflecting a break in relationship. Take UK narrow money demand as an example. The first type of information includes incomes, prices, and interest rates, whereas the second type includes the UK Finance Act of 1984. That Act required commercial banks to start withholding income tax from interest payments on individuals’ deposit accounts (that is, savings accounts) and to remit that withheld tax to the Inland Revenue. This legislative change precipitated an outflow of funds from deposit accounts, led to the introduction of interest-bearing checking accounts, and hence radically shifted the previous relationship between narrow money, prices, incomes, and interest rates.

NRE: Does the second type of information bear on shifts in the DGP?

DFH: Yes. If we had known that the 1984 Finance Act would lead to the introduction of interest-bearing checking accounts, and if we had known how agents would respond to their introduction, we could have produced fairly accurate money-demand forecasts. Without that knowledge, massive forecast failure resulted. By the late 1980s, forecasts of the narrow money stock from a previously well-specified model were systematically off by about half of the stock of money.

NRE: Does information exist that would enable one to forecast breaks?

DFH: I think so, but much more research is needed. If such information could be observed in advance, then it should be feasible to develop models for forecasting breaks, an issue addressed in Castle, Fawcett, and Hendry (2011). Policy-determined events—such as allowing Lehman Brothers to go bankrupt—seem likely to remain unpredictable. We lack information that would help forecast the outcome, prior to the policy decision. Ironically, if such a bankruptcy were correctly forecast, and if the government acted on that forecast to prevent the bankruptcy, no bankruptcy would occur, leading to an apparent forecast failure.

NRE: At a conceptual level, additional conditioning information—such as that policy decision—may remove time dependence in the distribution.

DFH: Agreed, but you need “breaks in” to get “breaks out”. That echoes “No Causes In, No Causes Out”, which is the title to Chapter 2 in Nancy Cartwright’s remarkable (1989) book. An observed break in a relationship may well be due to a variable that is excluded from the model. If so, that excluded variable itself must have experienced a location shift, which may be difficult to predict. For instance, if inflation depends on oil prices, and oil prices shift, the change in oil prices could explain a shift in inflation—but shifts in oil prices may be nearly unpredictable.

NRE: Could co-breaking help sort this out? What’s the background on co-breaking?

DFH: I was sketching out the idea of co-breaking during a conference in Florence when an unanticipated break occurred—a small earthquake. That earthquake had not been forecast!

I thought that co-breaking would rescue forecasting from its most pernicious problem—systematic forecast failure. Hendry and Massmann (2007) show that co-breaking cancels out multiple location shifts through linear combinations of variables, just as Johansen (1988) shows that cointegration cancels out multiple unit roots. However, while the co-breaking solution works for the co-breaking combinations, it doesn’t work for the remaining variables, which still require forecasts of their shifts.

NRE: Relatedly, model mis-specification *per se* cannot explain predictive failure. Suppose the world were ergodic and we ran a regression of y on z , and we also ran a regression of z on y . Both regression models would be constant, and both would predict as well out of sample as they would fit in sample: see Miller (1978) and Hendry (1979a). Changes in data correlations or changes in data means are needed for a model to systematically mis-forecast or become nonconstant.

DFH: That result is key to understanding predictive failure. With data from a stationary stochastic process, a model that is congruent and encompassing in-sample

will usually dominate in forecasting. Its interval forecasts will also be relatively accurate, and its forecasts will approximate the conditional expectations and so will be close to the corresponding minimum MSFE predictors.

However, when the DGP has unanticipated shifts, these implications no longer hold. The conditional expectation may be biased, and it need not be a minimum MSFE predictor. Changes in collinearity can induce explosions in MSFEs, and misspecification can cause forecast failure. In fact, in forecasting, a non-causal model may dominate a model that includes every causal variable. For example, the lagged first difference will be a robust predictor, even if the lagged first difference is not in the DGP. Conversely, VEqCMs and DSGE models are non-robust predictors and can suffer systematic forecast failure, even if correctly specified in-sample. That result entails the need to rethink how expectations are formulated in economic models.

NRE: Are fat-tailed distributions also problematic when forecasting?

DFH: Yes. With fat-tailed distributions, apparent outliers are more prevalent than would be expected from the normal distribution. Empirically, though, outliers are often clustered in time. The probability of observing several successive similar outliers is very small with independent draws—akin to the sudden appearance of a flock of “black swans”.

NRE: Can clustering arise from a shift in the distribution?

DFH: Absolutely. When a distribution shifts, observations that would have been outliers under the previous distribution can be commonplace under the new distribution. Hence the “outliers” will be clustered. Also, if the distribution’s mean does not shift, Chebyshev’s inequality limits the number of outliers likely to occur because the density must integrate to unity.

6 Applications and Implications

NRE: Having focused on the theory of economic forecasting, let’s turn to applications.

6.1 UK Office of Communications

NRE: You worked with Ofcom—the UK government’s Office of Communications—on forecasts of net advertising revenue for the TV broadcasting network ITV. These forecasts had policy consequences and were interesting in their own right. What is the background?

DFH: Ofcom is the British government agency responsible for UK radio and TV broadcasting licenses, and ITV is the biggest commercial TV network in the United Kingdom. In 2004, Ofcom needed to price the renewal of the license for ITV to advertise on TV. The license fee had been specified to be calculated from forecasts of discounted net advertising revenue (NAR) over the subsequent decade.

In Hendry (1992), I had developed a VEqCM for key variables in forecasting NAR—hours broadcast, audience reach, and the price of advertising time. I subsequently improved that VEqCM using a precursor to *Autometrics* called *PcGets*,

which I had developed with Hans-Martin Krolzig. Ofcom then augmented that new VEqCM by forecasts from a macro-model for variables such as GDP, company profits, interest rates, and inflation.

In forecasting NAR, Ofcom initially preferred to forecast from that augmented VEqCM, rather than from the corresponding DVEqCM. Ofcom was concerned with how the differencing in the DVEqCM would eliminate long-run relationships from the VEqCM. However, representatives from the advertising industry described recent breaks in TV advertising that arose from innovations such as video-recorders, personalized and Internet advertising, and satellite and cable TV channels. Those breaks would be difficult to model with available data, yet they could cause systematic forecast failure by the VEqCM. I persuaded Ofcom that differencing the VEqCM would remove those location-shift components but retain long-run information.

Ofcom published forecasts for NAR over 2004–2014 in Raven, Hoehn, Lancefield, and Robinson (2004, Figure 6.5). Forecasts were calculated for three models: a “long-run trend” model, the VEqCM, and the corresponding DVEqCM. The respective forecasts were increasing, relatively flat, and slightly declining over time.

NRE: So, Ofcom robustified the forecasts by differencing the VEqCM; and robustification was important because there had been recent unmodeled shifts.

DFH: Indeed. Differencing removed location shifts in excluded variables such as the introduction of personal video-recorders, which had reduced TV advertising revenue.

NRE: Were these forecasts important in policy?

DFH: Very much so. Ofcom set a lower license fee because the DVEqCM forecasts showed NAR declining, rather than increasing. However, while the DVEqCM did perform the best of the three models *ex post*, even its forecasts proved too optimistic. Many of the variables included in the DVEqCM themselves experienced unanticipated location shifts during the forecast period. For instance, in the wake of the financial crisis, actual GDP and profits were much lower than forecast, poignantly illustrating that unanticipated location shifts can induce systematic forecast errors.

6.2 The UK Housing Market

NRE: In the 1970s, you began modeling the UK housing and mortgage markets. You were unable to sell your models’ forecasts to any major UK company involved in construction, construction materials, or mortgages. Why?

DFH: Therein lies a story. In the late 1970s, I gave a presentation on my models to representatives from many of those companies. At the end of the presentation, someone queried the model’s ability to forecast the number of housing starts in April five years ahead. I replied that “no one needs to know that today”. The questioner said “your model can’t do it” and walked out. Then, to my shock, so did most of the rest of the audience. Given how many representatives thought that that was a sensible question, it didn’t surprise me that some of those companies collapsed in the sharp recession of the early 1980s.

NRE: What did you learn from forecasting the housing market?

DFH: What insight I gained started with a puzzle. During 1972, UK house prices rose dramatically in response to a major increase in mortgage lending by building societies. Later, I checked how well my model would have forecast through that period. When forecasting a few quarters after the then largest-ever increase in UK house prices, the model forecast a fall in prices, while prices actually continued to rise substantially. Nevertheless, coefficients estimated over the pre-forecast period were almost identical to those estimated over the whole sample; and the whole-sample residuals were homoscedastic.

NRE: When did you come up with an explanation?

DFH: Not until over a decade later, when Mike Clements and I were developing the general theory of forecasting. That theory distinguishes between “internal breaks”, which are shifts in the model’s parameters, and “external breaks”, which are shifts in the unmodeled included variables. Mike and I showed that a change in multi-collinearity among the model’s variables leaves estimated coefficients almost unchanged but can greatly increase MSFEs, contrasting with the irrelevance of multi-collinearity to forecast uncertainty when multi-collinearity is constant. This problem with multi-collinearity cannot be solved by orthogonalizing the model’s variables or by eliminating relevant multi-collinear variables. The latter can lead to even worse forecasts. However, updating parameter estimates with new data can reduce MSFEs.

For UK house prices, the correlations of mortgage lending with disposable income, interest rates, and inflation altered markedly when mortgage lending itself increased. Despite the accrual of more information from changes in multi-collinearity, the MSFE also increased, in line with our general theory of forecasting.

NRE: Model nonlinearities proved central to explaining house-price “bubbles”. What inspired you to include nonlinearities in your model?

DFH: I had learned that Van der Pol’s cubic differential equation could describe heartbeats, and that heartbeats could manifest sudden surges. Changes in UK house prices seemed rather like heartbeats so, in my model, I included the cube of the excess demand for housing, as represented by the cube of lagged house-price inflation. The cube was significant, as I showed in Hendry (1984).

My formulation had difficulties, though. It predicted some large jumps in house prices that didn’t materialize. It also implied that large changes in house prices were explosive. In practice, once the market was far from equilibrium, excessively high or low house-price-to-income ratios drove the market back towards equilibrium, as followed after the UK housing bubble in the late 1980s. Richard and Zhang (1996) improved on my nonlinear formulation by using a cubic in the observed deviation from the long-run equilibrium rather than the cubic of house-price inflation.

6.3 Forecasting and Policy Analysis

NRE: You have undertaken policy research with various coauthors, including Graham Mizon and myself. Forecasting has also been an important aspect of your policy advice to government, as with Ofcom and HM Treasury. What role does forecasting play in policy, and what’s the relationship between forecasting and policy analysis?

DFH: Forecasting is integral to policy, and some formalization helps explain why. Policy analysis depends intrinsically on partial derivatives between causally related variables. If y depends on z and the policymaker wants to change z to affect y , then the partial derivative $\partial y/\partial z$ must coincide with the model’s coefficient of z (β , say). Otherwise, policy won’t have the anticipated effect. Unfortunately, changes in the model’s slope coefficient β can be hard to detect, contrasting with changes in the equilibrium mean. That said, changes in slope coefficients may have become easier to detect now by using multiplicative indicator saturation, as proposed by yourself in Ericsson (2011) and analyzed further in Kitov and Tabor (2015).

Relatedly, impulse response functions are ubiquitous for calculating partial derivatives but may be misleading when structural breaks occur. As Hendry and Mizon (2000) show, the model’s coefficients can alter dramatically—even with their signs switching in the DGP—yet the model’s estimated impulse responses may be relatively unaffected by those parameter changes. Furthermore, robustification methods such as IC may markedly alter a model’s forecasts without changing the model’s policy implications. So, don’t judge a policy model by its forecast performance.

NRE: We’ve discussed how the Lucas critique conflates forecasting and policy. It treats forecast failure as indicative of a model failing to isolate structural parameters—parameters such as partial derivatives. What’s the importance of expectations formation in forecasting?

DFH: Much economic analysis is predicated on agents forming conditional (“rational”) expectations about future values of economic variables. However, *rational* agents would not in general use in-sample conditional expectations when facing unanticipated intermittent location shifts. As I showed in Hendry (1997) and Hendry and Mizon (2014), those in-sample conditional expectations would correspond to a systematically biased predictor and so would not be “rational”.

Even if agents cannot forecast a break, they can benefit from adjusting their expectations rapidly after the break occurs, as with robust forecasting devices. For instance, in Hendry and Ericsson (1991), we interpret current inflation Δp_T in a money-demand equation as a robust forecast of next period’s inflation Δp_{T+1} . Models of rational expectations have been extended to incorporate learning, which could better describe agents’ behavior in the face of unanticipated breaks. However, as the forecast error taxonomy shows, vastly different values of the DGP’s parameters may imply almost identical data paths. Agents may not be able to discern changes in those parameters from just a few observations after the changes occurred.

NRE: Is IIS a robust device for isolating partial derivatives of policy interest?

DFH: Not necessarily. IIS can detect and account for unanticipated unmodeled location shifts, once they occur. However, even with IIS, partial derivatives are inconsistently estimated if the included variables are correlated with omitted but relevant variables.

An IIS-based test of super exogeneity can help here by checking the underlying assumptions and specification of the policy model. That test is constructed as follows. Start by modeling z_t with IIS to ascertain the breaks in the process for z . Then

estimate the conditional model of y_t given z_t , and test whether the break dummies that are significant in the model of the z process are also significant in the conditional model. If those dummies are *not* significant in the conditional model, then y co-breaks with z . That is, shifts in z generate corresponding shifts in y . Co-breaking helps establish the super exogeneity of z , and the super exogeneity of z should be established before implementing policy. If super exogeneity is rejected, a policy's actual consequences may differ markedly from its anticipated consequences.

7 Looking Back and Looking Forward

NRE: How has an academic environment influenced your research on forecasting?

DFH: Teaching is a very important catalyst. By explaining an idea to others, I often come to understand the idea better myself. For instance, when I was teaching a course for the Oxford MPhil, a student asked me to explain the DVEqCM. I had just spent considerable time explaining the DVEqCM in class and, in a somewhat frustrated response, I literally read out-loud the equation for the DVEqCM that I had written on the board. From that act of verbalization, I suddenly realized several implications of the DVEqCM for the theory of robust forecasting. What a surprise! That theory is now described in Hendry (2006). As we discussed in Ericsson (2004), I have had several other such insights when teaching.

NRE: What roles do seminars and conferences play for you in research?

DFH: Presentations at seminars and conferences force me to clarify my research, and they generate important feedback. Seminars and conferences also are useful forums for finding out what other researchers are doing. The *International Symposium on Forecasting* conferences have been particularly valuable in this regard. They've stimulated many insightful discussions, including with Geoff Allen, Robert Fildes, Rob Hyndman, Fred Joutz, Dilek Onkal, and Herman Stekler.

NRE: How useful has research funding been?

DFH: Extremely useful, when I received it! However, many of my research grant applications for forecasting were rejected. Fortunately, though, I was awarded two personal research fellowships: one from the Leverhulme Trust for five years, and one from the Economic and Social Research Council for three years. These fellowships bought out some of my teaching responsibilities, enabling me to develop the general theory of forecasting.

NRE: Perhaps funding difficulties arose because forecasting has been regarded by many as the orphan of economics. Those who could do economics, did it; those who couldn't do economics, forecasted. That said, you have been awarded sizable grants for "new economic thinking".

DFH: Yes, and thankfully so. James Martin and George Soros generously funded my program *Economic Modeling in a Rapidly Changing World (EMoD)*—James Martin through the Oxford Martin School, and George Soros through the Open Society

Foundations and the Institute for New Economic Thinking (INET). That initial five-year grant for *EMoD* supported Oxford economics faculty and post-doctoral research fellows in analyzing difficulties that empirical modeling, economic analysis, policy, and forecasting confront when there are rapid unanticipated changes. INET recently extended our *EMoD* grant for three more years jointly with John Muellbauer, and the Robertson Foundation has awarded a grant for our program *Climate Econometrics*.

NRE: Is this funding indicative of a change in attitude about economic forecasting?

DFH: I hope so. Many top econometricians are now involved in the theory of forecasting, including Frank Diebold, Hashem Pesaran, Peter Phillips, Lucrezia Reichlin, Jim Stock, Timo Teräsvirta, Ken Wallis, and Mark Watson. Their technical expertise as well as their practical forecasting experience is invaluable in furthering the field. A mathematical treatment can help understand economic forecasts, as the taxonomy illustrated. Recent developments are summarized in the books by Hendry and Ericsson (2001), Clements and Hendry (2002), Elliott, Granger, and Timmermann (2006), and Clements and Hendry (2011b). Forecasting is no longer an orphan of the profession.

NRE: What are some recent research topics at *EMoD* and *Climate Econometrics*?

DFH: We are analyzing the mathematical and statistical bases for expectations formation and intertemporal optimization when economic agents face unanticipated breaks, and we are developing methods of empirical model discovery that can handle multiple intermittent shifts. We are also investigating inequality in wealth and income, modeling immigration into Norway, establishing a database of civilization's progress (www.ourworldindata.org), and formulating alternative macro-models with financial channels. In climate econometrics, our new methods for detecting breaks are isolating the effects of volcanic eruptions on temperature, detecting when earthquakes occurred, and helping model increases in sea level. The overriding theme is to develop approaches appropriate to a world undergoing rapid unanticipated changes, and to improve forecasting methods in such a setting.

NRE: An ambitious agenda!

References

- Abadir, K. M., W. Distaso, and F. Žikeš (2014) “Design-free Estimation of Variance Matrices”, *Journal of Econometrics*, 181, 2, 165–180.
- Allen, P. G., and R. Fildes (2005) “Levels, Differences and ECMs—Principles for Improved Econometric Forecasting”, *Oxford Bulletin of Economics and Statistics*, 67, Supplement, 881–904.
- Andrews, D. W. K. (1993) “Tests for Parameter Instability and Structural Change with Unknown Change Point”, *Econometrica*, 61, 4, 821–856.
- Artis, M. J., A. Banerjee, and M. Marcellino (2005) “Factor Forecasts for the UK”, *Journal of Forecasting*, 24, 4, 279–298.
- Bai, J., and P. Perron (1998) “Estimating and Testing Linear Models with Multiple Structural Changes”, *Econometrica*, 66, 1, 47–78.

- Bårdsen, G., Ø. Eitrheim, E. S. Jansen, and R. Nymoen (2005) *The Econometrics of Macroeconomic Modelling*, Oxford University Press, Oxford.
- Bates, J. M., and C. W. J. Granger (1969) “The Combination of Forecasts”, *Operational Research Quarterly*, 20, 451–468.
- Birchenhall, C. R., H. Jessen, D. R. Osborn, and P. Simpson (1999) “Predicting U.S. Business-cycle Regimes”, *Journal of Business and Economic Statistics*, 17, 3, 313–323.
- Bontemps, C., and G. E. Mizon (2008) “Encompassing: Concepts and Implementation”, *Oxford Bulletin of Economics and Statistics*, 70, supplement, 721–750.
- Box, G. E. P., and G. M. Jenkins (1970) *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
- Cartwright, N. (1989) *Nature’s Capacities and Their Measurement*, Clarendon Press, Oxford.
- Castle, J. L., M. P. Clements, and D. F. Hendry (2013) “Forecasting by Factors, by Variables, by Both or Neither?”, *Journal of Econometrics*, 177, 2, 305–319.
- Castle, J. L., J. A. Doornik, and D. F. Hendry (2012) “Model Selection When There Are Multiple Breaks”, *Journal of Econometrics*, 169, 2, 239–246.
- Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2015) “Detecting Location Shifts During Model Selection by Step-indicator Saturation”, *Econometrics*, 3, 2, 240–264.
- Castle, J. L., N. W. P. Fawcett, and D. F. Hendry (2010) “Forecasting with Equilibrium-correction Models During Structural Breaks”, *Journal of Econometrics*, 158, 1, 25–36.
- Castle, J. L., N. W. P. Fawcett, and D. F. Hendry (2011) “Forecasting Breaks and Forecasting During Breaks”, Chapter 11 in M. P. Clements and D. F. Hendry (eds.) *Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford, 315–353.
- Chevillon, G., and D. F. Hendry (2005) “Non-parametric Direct Multi-step Estimation for Forecasting Economic Processes”, *International Journal of Forecasting*, 21, 2, 201–218.
- Chong, Y. Y., and D. F. Hendry (1986) “Econometric Evaluation of Linear Macro-economic Models”, *Review of Economic Studies*, 53, 4, 671–690.
- Chow, G. C. (1960) “Tests of Equality Between Sets of Coefficients in Two Linear Regressions”, *Econometrica*, 28, 3, 591–605.
- Clements, M. P., and D. F. Hendry (1993) “On the Limitations of Comparing Mean Square Forecast Errors”, *Journal of Forecasting*, 12, 8, 617–637 (with discussion).
- Clements, M. P., and D. F. Hendry (1994) “Towards a Theory of Economic Forecasting”, Chapter 2 in C. P. Hargreaves (ed.) *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press, Oxford, 9–52.
- Clements, M. P., and D. F. Hendry (1995) “Forecasting in Cointegrated Systems”, *Journal of Applied Econometrics*, 10, 2, 127–146.
- Clements, M. P., and D. F. Hendry (1996a) “Intercept Corrections and Structural Change”, *Journal of Applied Econometrics*, 11, 5, 475–494.
- Clements, M. P., and D. F. Hendry (1996b) “Multi-step Estimation for Forecasting”, *Oxford Bulletin of Economics and Statistics*, 58, 4, 657–684.
- Clements, M. P., and D. F. Hendry (1998) *Forecasting Economic Time Series*, Cambridge University Press, Cambridge.

- Clements, M. P., and D. F. Hendry (1999) *Forecasting Non-stationary Economic Time Series*, MIT Press, Cambridge.
- Clements, M. P., and D. F. Hendry (2001) “Explaining the Results of the M3 Forecasting Competition”, *International Journal of Forecasting*, 17, 4, 550–554.
- Clements, M. P., and D. F. Hendry (eds.) (2002) *A Companion to Economic Forecasting*, Blackwell Publishers, Oxford.
- Clements, M. P., and D. F. Hendry (2003) “Report of a Scoping Study of Forecasting in the National Accounts at the Office for National Statistics”, Annex A in *Forecasting in the National Accounts at the Office for National Statistics*, Statistics Commission Report No. 12, Statistics Commission, London, December.
- Clements, M. P., and D. F. Hendry (2005) “Guest Editors’ Introduction: Information in Economic Forecasting”, *Oxford Bulletin of Economics and Statistics*, 67, Supplement, 713–753.
- Clements, M. P., and D. F. Hendry (2006) “Forecasting with Breaks”, Chapter 12 in G. Elliott, C. W. J. Granger, and A. Timmermann (eds.) *Handbook of Economic Forecasting*, Volume 1, Elsevier, Amsterdam, 605–657.
- Clements, M. P., and D. F. Hendry (2011a) “Forecasting From Misspecified Models in the Presence of Unanticipated Location Shifts”, Chapter 10 in M. P. Clements and D. F. Hendry (eds.) *Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford, 271–314.
- Clements, M. P., and D. F. Hendry (eds.) (2011b) *Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford.
- Cooper, J. P., and C. R. Nelson (1975) “The Ex Ante Prediction Performance of the St. Louis and FRB–MIT–PENN Econometric Models and Some Results on Composite Predictors”, *Journal of Money, Credit, and Banking*, 7, 1, 1–32.
- Davidson, J. E. H., D. F. Hendry, F. Srba, and S. Yeo (1978) “Econometric Modelling of the Aggregate Time-series Relationship Between Consumers’ Expenditure and Income in the United Kingdom”, *Economic Journal*, 88, 352, 661–692.
- Doob, J. L. (1953) *Stochastic Processes*, John Wiley, New York (republished 1990).
- Doornik, J. A., and D. F. Hendry (2013) *PcGive 14*, Timberlake Consultants Press, London (3 volumes).
- Duffy, J. A., and D. F. Hendry (2015) “The Impact of Near-Integrated Measurement Errors on Modelling Long-run Macroeconomic Time Series”, draft, Department of Economics, University of Oxford, Oxford.
- Eichner, M. J., D. L. Kohn, and M. G. Palumbo (2010) “Financial Statistics for the United States and the Crisis: What Did They Get Right, What Did They Miss, and How Should They Change?”, FEDS Discussion Paper No. 2010-20, Board of Governors of the Federal Reserve System, Washington, D.C., April; available at www.federalreserve.gov/pubs/feds/2010/201020/201020abs.html.
- Eitrheim, Ø., T. A. Husebø, and R. Nymoen (1999) “Equilibrium-correction Versus Differencing in Macroeconometric Forecasting”, *Economic Modeling*, 16, 4, 515–554.
- Elliott, G., C. W. J. Granger, and A. Timmermann (eds.) (2006) *Handbook of Economic Forecasting*, Volume 1, Elsevier, Amsterdam.
- Emerson, R. A., and D. F. Hendry (1996) “An Evaluation of Forecasting Using Leading Indicators”, *Journal of Forecasting*, 15, 4, 271–291.

- Engle, R. F., and D. F. Hendry (1993) “Testing Super Exogeneity and Invariance in Regression Models”, *Journal of Econometrics*, 56, 1/2, 119–139.
- Engle, R. F., D. F. Hendry, and J.-F. Richard (1983) “Exogeneity”, *Econometrica*, 51, 2, 277–304.
- Engle, R. F., and B. S. Yoo (1987) “Forecasting and Testing in Co-integrated Systems”, *Journal of Econometrics*, 35, 1, 143–159.
- Ericsson, N. R. (2004) “The *ET* Interview: Professor David F. Hendry”, *Econometric Theory*, 20, 4, 743–804.
- Ericsson, N. R. (2008) “Comment on ‘Economic Forecasting in a Changing World’ (by Michael Clements and David Hendry)”, *Capitalism and Society*, 3, 2, 2, 1–16.
- Ericsson, N. R. (2011) “Justifying Empirical Macro-econometric Evidence in Practice”, invited presentation, online conference *Communications with Economists: Current and Future Trends* commemorating the 25th anniversary of the *Journal of Economic Surveys*, November.
- Ericsson, N. R. (2016) “Eliciting GDP Forecasts from the FOMC’s Minutes Around the Financial Crisis”, *International Journal of Forecasting*, 32, 2, 571–583.
- EuroStat (ed.) (2016) *Handbook on Rapid Estimates*, UN/EuroStat, Brussels (forthcoming).
- Favero, C., and D. F. Hendry (1992) “Testing the Lucas Critique: A Review”, *Econometric Reviews*, 11, 3, 265–306 (with discussion).
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2001) “Coincident and Leading Indicators for the Euro Area”, *Economic Journal*, 111, 471, C62–C85.
- Granger, C. W. J. (2001) “Evaluation of Forecasts”, Chapter 6 in D. F. Hendry and N. R. Ericsson (eds.) *Understanding Economic Forecasts*, MIT Press, Cambridge, Massachusetts, 93–103.
- Granger, C. W. J. (2004) “Time Series Analysis, Cointegration, and Applications”, in T. Frängsmyr (ed.) *The Nobel Prizes 2003*, Almqvist and Wiksell International, Stockholm, 360–366.
- Granger, C. W. J., and P. Newbold (1977) “The Time Series Approach to Econometric Model Building”, in C. A. Sims (ed.) *New Methods in Business Cycle Research: Proceedings from a Conference*, Federal Reserve Bank of Minneapolis, Minneapolis, Minnesota, 7–21 (with discussion).
- Hatch, N. (2001) “Modeling and Forecasting at the Bank of England”, Chapter 8 in D. F. Hendry and N. R. Ericsson (eds.) *Understanding Economic Forecasts*, MIT Press, Cambridge, Massachusetts, 124–148.
- Hendry, D. F. (1973) “On Asymptotic Theory and Finite Sample Experiments”, *Economica*, 40, 158, 210–217.
- Hendry, D. F. (1974) “Stochastic Specification in an Aggregate Demand Model of the United Kingdom”, *Econometrica*, 42, 3, 559–578.
- Hendry, D. F. (1975) “The Consequences of Mis-specification of Dynamic Structure, Autocorrelation, and Simultaneity in a Simple Model with an Application to the Demand for Imports”, Chapter 11 in G. A. Renton (ed.) *Modelling the Economy*, Heinemann Educational Books, London, 286–320 (with discussion).

- Hendry, D. F. (1977) “Comments on Granger–Newbold’s ‘Time Series Approach to Econometric Model Building’ and Sargent–Sims’ ‘Business Cycle Modeling Without Pretending to Have Too Much *A Priori* Economic Theory’”, in C. A. Sims (ed.) *New Methods in Business Cycle Research: Proceedings from a Conference*, Federal Reserve Bank of Minneapolis, Minneapolis, 183–202.
- Hendry, D. F. (1979a) “The Behaviour of Inconsistent Instrumental Variables Estimators in Dynamic Systems with Autocorrelated Errors”, *Journal of Econometrics*, 9, 3, 295–314.
- Hendry, D. F. (1979b) “Predictive Failure and Econometric Modelling in Macroeconomics: The Transactions Demand for Money”, Chapter 9 in P. Ormerod (ed.) *Economic Modelling: Current Issues and Problems in Macroeconomic Modelling in the UK and the US*, Heinemann Education Books, London, 217–242.
- Hendry, D. F. (1984) “Econometric Modelling of House Prices in the United Kingdom”, Chapter 8 in D. F. Hendry and K. F. Wallis (eds.) *Econometrics and Quantitative Economics*, Basil Blackwell, Oxford, 211–252.
- Hendry, D. F. (1988) “The Encompassing Implications of Feedback Versus Feedforward Mechanisms in Econometrics”, *Oxford Economic Papers*, 40, 1, 132–149.
- Hendry, D. F. (1991) “Economic Forecasting: A Report to the Treasury and Civil Service Committee”, submitted to the House of Commons, *Memoranda on Official Economic Forecasting*, Treasury and Civil Service Committee, Session 1990–91, Her Majesty’s Stationery Office, London, July 12.
- Hendry, D. F. (1992) “An Econometric Analysis of TV Advertising Expenditure in the United Kingdom”, *Journal of Policy Modeling*, 14, 3, 281–311.
- Hendry, D. F. (1997) “The Econometrics of Macroeconomic Forecasting”, *Economic Journal*, 107, 444, 1330–1357.
- Hendry, D. F. (1999) “An Econometric Analysis of US Food Expenditure, 1931–1989”, Chapter 17 in J. R. Magnus and M. S. Morgan (eds.) *Methodology and Tacit Knowledge: Two Experiments in Econometrics*, John Wiley and Sons, Chichester, 341–361.
- Hendry, D. F. (2004) “The Nobel Memorial Prize for Clive W. J. Granger”, *Scandinavian Journal of Economics*, 106, 2, 187–213.
- Hendry, D. F. (2006) “Robustifying Forecasts from Equilibrium-correction Systems”, *Journal of Econometrics*, 135, 1–2, 399–426.
- Hendry, D. F. (2016) “Deciding Between Alternative Approaches in Macroeconomics”, Discussion Paper No. 778, Department of Economics, University of Oxford, Oxford, January.
- Hendry, D. F., and M. P. Clements (2004) “Pooling of Forecasts”, *Econometrics Journal*, 7, 1, 1–31.
- Hendry, D. F., and J. A. Doornik (1997) “The Implications for Econometric Modelling of Forecast Failure”, *Scottish Journal of Political Economy*, 44, 4, 437–461.
- Hendry, D. F., and J. A. Doornik (2014) *Empirical Model Discovery and Theory Evaluation: Automatic Selection Methods in Econometrics*, MIT Press, Cambridge, Massachusetts.
- Hendry, D. F., and N. R. Ericsson (1991) “Modeling the Demand for Narrow Money in the United Kingdom and the United States”, *European Economic Review*, 35, 4, 833–881 (with discussion).
- Hendry, D. F., and N. R. Ericsson (eds.) (2001) *Understanding Economic Forecasts*, MIT Press, Cambridge.

- Hendry, D. F., and K. Hubrich (2011) “Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate”, *Journal of Business and Economic Statistics*, 29, 2, 216–227.
- Hendry, D. F., and S. Johansen (2015) “Model Discovery and Trygve Haavelmo’s Legacy”, *Econometric Theory*, 31, 1, 93–114.
- Hendry, D. F., S. Johansen, and C. Santos (2008) “Automatic Selection of Indicators in a Fully Saturated Regression”, *Computational Statistics*, 23, 2, 317–335, 337–339.
- Hendry, D. F., and A. B. Martinez (2016) “Evaluating Multi-step System Forecasts with Relatively Few Forecast-error Observations”, Discussion Paper No. 784, Department of Economics, University of Oxford, Oxford, March; *International Journal of Forecasting*, forthcoming.
- Hendry, D. F., and M. Massmann (2007) “Co-breaking: Recent Advances and a Synopsis of the Literature”, *Journal of Business and Economic Statistics*, 25, 1, 33–51.
- Hendry, D. F., and G. E. Mizon (2000) “Reformulating Empirical Macroeconometric Modelling”, *Oxford Review of Economic Policy*, 16, 4, 138–159.
- Hendry, D. F., and G. E. Mizon (2014) “Unpredictability in Economic Analysis, Econometric Modeling and Forecasting”, *Journal of Econometrics*, 182, 1, 186–195.
- Hendry, D. F., and C. Santos (2010) “An Automatic Test of Super Exogeneity”, Chapter 12 in T. Bollerslev, J. R. Russell, and M. W. Watson (eds.) *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, Oxford University Press, Oxford, 164–193.
- Hoover, K. D., and S. J. Perez (1999) “Data Mining Reconsidered: Encompassing and the General-to-specific Approach to Specification Search”, *Econometrics Journal*, 2, 2, 167–191 (with discussion).
- Johansen, S. (1988) “Statistical Analysis of Cointegration Vectors”, *Journal of Economic Dynamics and Control*, 12, 2/3, 231–254.
- Johansen, S., and B. Nielsen (2009) “An Analysis of the Indicator Saturation Estimator as a Robust Regression Estimator”, Chapter 1 in J. L. Castle and N. Shephard (eds.) *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, Oxford University Press, Oxford, 1–36.
- Kitov, O. I., and M. N. Tabor (2015) “Detecting Structural Changes in Linear Models: A Variable Selection Approach Using Multiplicative Indicator Saturation”, draft, Department of Economics, University of Oxford, Oxford.
- Kiviet, J. F. (1986) “On the Rigour of Some Misspecification Tests for Modelling Dynamic Relationships”, *Review of Economic Studies*, 53, 2, 241–261.
- Klein, L. R. (1950) *Economic Fluctuations in the United States, 1921–1941* (Cowles Commission Monograph No. 11), John Wiley, New York.
- Klein, L. R. (1971) *An Essay on the Theory of Economic Prediction*, Markham Publishing Company, Chicago.
- Lucas, Jr., R. E. (1976) “Econometric Policy Evaluation: A Critique”, in K. Brunner and A. H. Meltzer (eds.) *The Phillips Curve and Labor Markets*, North-Holland, Amsterdam, Carnegie-Rochester Conference Series on Public Policy, Volume 1, *Journal of Monetary Economics*, Supplement, 19–46 (with discussion).
- Magnus, J. R., and M. S. Morgan (eds.) (1999) *Methodology and Tacit Knowledge: Two Experiments in Econometrics*, John Wiley and Sons, Chichester.

- Makridakis, S., and M. Hibon (2000) “The M3-Competition: Results, Conclusions and Implications”, *International Journal of Forecasting*, 16, 4, 451–476.
- Miller, P. J. (1978) “Forecasting with Econometric Methods: A Comment”, *Journal of Business*, 51, 4, 579–586.
- Mills, T. C. (2011) “Bradford Smith: An Econometrician Decades Ahead of His Time”, *Oxford Bulletin of Economics and Statistics*, 73, 2, 276–285.
- Nelson, C. R. (1972) “The Prediction Performance of the FRB–MIT–PENN Model of the U.S. Economy”, *American Economic Review*, 62, 5, 902–917.
- Phillips, A. W. (1954) “Stabilisation Policy in a Closed Economy”, *Economic Journal*, 64, 254, 290–323.
- Phillips, P. C. B. (1995) “Automated Forecasts of Asia-Pacific Economic Activity”, *Asia-Pacific Economic Review*, 1, 1, 92–102.
- Pretis, F., L. Schneider, J. E. Smerdon, and D. F. Hendry (2016) “Detecting Volcanic Eruptions in Temperature Reconstructions by Designed Break-indicator Saturation”, *Journal of Economic Surveys*, 30, 3, 403–429.
- Raven, J., T. Hoehn, D. Lancefield, and B. Robinson (2004) *Economic Analysis of the TV Advertising Market*, PricewaterhouseCoopers LLP, London, December (available from www.ofcom.org.uk/__data/assets/pdf_file/0018/23913/tvadvmarket.pdf).
- Richard, J.-F., and W. Zhang (1996) “Econometric Modelling of UK House Prices Using Accelerated Importance Sampling”, *Oxford Bulletin of Economics and Statistics*, 58, 4, 601–613.
- Salkever, D. S. (1976) “The Use of Dummy Variables to Compute Predictions, Prediction Errors, and Confidence Intervals”, *Journal of Econometrics*, 4, 4, 393–397.
- Sargan, J. D. (1964) “Wages and Prices in the United Kingdom: A Study in Econometric Methodology”, in P. E. Hart, G. Mills, and J. K. Whitaker (eds.) *Econometric Analysis for National Economic Planning*, Volume 16 of *Colston Papers*, Butterworths, London, 25–54 (with discussion).
- Smith, B. B. (1926) “Combining the Advantages of First-Difference and Deviation-From-Trend Methods of Correlating Time Series”, *Journal of the American Statistical Association*, 21, 153, 55–59.
- Smith, B. B. (1927) “Forecasting the Volume and Value of the Cotton Crop”, *Journal of the American Statistical Association*, 22, 160, 442–459.
- Smith, B. B. (1929) “Judging the Forecast for 1929”, *Journal of the American Statistical Association*, 24, 165, 94–98 (Supplement).
- Stekler, H. O., and H. Symington (2016) “Evaluating Qualitative Forecasts: The FOMC Minutes, 2006–2010”, *International Journal of Forecasting*, 32, 2, 559–570.
- Stock, J. H., and M. W. Watson (2011) “Dynamic Factor Models”, Chapter 2 in M. P. Clements and D. F. Hendry (eds.) *Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford, 35–59.
- Timbergen, J. (1951) *Business Cycles in the United Kingdom, 1870–1914*, North-Holland, Amsterdam.
- Tobin, J. (1950) “A Statistical Demand Function for Food in the U.S.A.”, *Journal of the Royal Statistical Society, Series A*, 113, 2, 113–141.
- Wallis, K. F. (1993) “On the Limitations of Comparing Mean Square Forecast Errors: Comment”, *Journal of Forecasting*, 12, 8, 663–666.

Appendix A. The Taxonomy

Section 4.2 describes a taxonomy for the sources of forecast error. Table A1 portrays that taxonomy in a convenient 3×3 decomposition for the nine sources of forecast error, labeled $i(a)$ – $iii(c)$.

Table A1: A taxonomy of the sources of forecast error.

Component	Problem		
	<i>a.</i> Estimation	<i>b.</i> Mis-specification	<i>c.</i> Change
<i>i.</i> Unobserved terms (what’s not modeled)	$i(a)$ [forecast origin]	$i(b)$ [omitted variable]	$i(c)$ [innovation error]
<i>ii.</i> Observed stochastic variables (q.v. slope parameters)	$ii(a)$	$ii(b)$	$ii(c)$
<i>iii.</i> Deterministic terms (q.v. equilibrium mean)	$iii(a)$	$iii(b)$	$iii(c)$

Appendix B. The Data

This appendix documents the data in Figure 1.

The UK data are produced by the Office for National Statistics at the UK Statistics Authority. The URL for the source file is:

www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/revisionstrianglesforukgdpabmi .

The source file was downloaded on October 14, 2016, and had been released by the ONS on September 30, 2016 and last updated by the ONS on September 19, 2016. The series are the “Month 1 estimate” and “Latest estimate” vintages of quarterly growth rates for real UK GDP, market prices, chained volume measure, as found under the tab “ABMI_triangle” of the file “Quarterly GDP at Market Prices (ABMI).xls” in the zipped file “gdprevisionstrianglesabmi.zip”.

The US data are produced by the Bureau of Economic Analysis at the US Department of Commerce. The URL for the source file is:

www.bea.gov/national/xls/GDP-GDI%20vintage%20history.xls .

The source file was downloaded on October 14, 2016 and had been last updated and released by the BEA on September 29, 2016. The series are the “Advance” and most recently “Revised” vintages of real US GDP in percent change from preceding period, as found under the tab “Vintage History” of the file “GDP-GDI vintage history.xls”.

Table B1 lists the data themselves, as published by the ONS and the BEA and in the units as published.

Table B1: Initial estimates and latest estimates of UK and US GDP growth (quarterly rates), as used in Figure 1.

Date	UK GDP growth rate (percent per quarter)		US GDP growth rate (percent per annum)	
	Month 1 estimate	Latest estimate	Advance estimate	Latest estimate
2006 Q1	0.6	0.3	4.8	4.9
2006 Q2	0.8	0.2	2.5	1.2
2006 Q3	0.7	0.1	1.6	0.4
2006 Q4	0.8	0.4	3.5	3.2
2007 Q1	0.7	1.0	1.3	0.2
2007 Q2	0.8	0.7	3.4	3.1
2007 Q3	0.8	0.8	3.9	2.7
2007 Q4	0.6	0.8	0.6	1.4
2008 Q1	0.4	0.1	0.6	-2.7
2008 Q2	0.2	-0.7	1.9	2.0
2008 Q3	-0.5	-1.7	-0.3	-1.9
2008 Q4	-1.5	-2.3	-3.8	-8.2
2009 Q1	-1.9	-1.6	-6.1	-5.4
2009 Q2	-0.8	-0.2	-1.0	-0.5
2009 Q3	-0.4	0.1	3.5	1.3
2009 Q4	0.1	0.4	5.7	3.9
2010 Q1	0.2	0.5	3.2	1.7
2010 Q2	1.1	1.0	2.4	3.9
2010 Q3	0.8	0.6	2.0	2.7
2010 Q4	-0.5	0.1	3.2	2.5