



**DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES**

**MATHEMATICAL MODELS AND ECONOMIC
FORECASTING: SOME USES AND MIS-USES OF
MATHEMATICS IN ECONOMICS**

David F. Hendry

Number 530
February 2011

Manor Road Building, Oxford OX1 3UQ

Mathematical Models and Economic Forecasting: Some Uses and Mis-Uses of Mathematics in Economics

David F. Hendry*
Economics Department and Institute for Economic Modelling,
Oxford Martin School, University of Oxford, UK

February 4, 2011

Abstract

We consider three ‘cases studies’ of the uses and mis-uses of mathematics in economics and econometrics. The first concerns economic forecasting, where a mathematical analysis is essential, and is independent of the specific forecasting model and how the process being forecast behaves. The second concerns model selection with more candidate variables than the number of observations. Again, an understanding of the properties of extended general-to-specific procedures is impossible without advanced mathematical analysis. The third concerns inter-temporal optimization and the formation of ‘rational expectations’, where misleading results follow from present mathematical approaches for realistic economies. The appropriate mathematics remains to be developed, and may end ‘problem specific’ rather than generic.

JEL classifications: C02, C22.

KEYWORDS: Economic forecasting; Structural breaks; Model selection; Expectations; Impulse-indicator saturation; Mathematical analyses.

1 Introduction

Mathematics is ubiquitous in modern economics and especially in econometrics. We draw on two examples from the latter where mathematics is essential in order to understand the properties of economic forecasts and the outcomes of empirical model selection exercises respectively. We then use the findings from the first of these to demonstrate important flaws in present approaches to the mathematics of inter-temporal optimization and the formation of expectations, in particular, so-called ‘rational expectations’ as applied to realistic economic time series. The three examples are drawn from work by the author jointly with a number of co-authors.

What prompts the need to discuss the obvious, namely the use of mathematics in economics, since that discipline intrinsically includes econometrics? First, it is a long-standing debate, with historical roots in the 19th century.¹ A particularly amusing complaint about the ‘excessive use of advanced mathematics’ is the discussant of the brilliant analysis of nonsense regressions when first read to the *Royal Statistical Society*,² although most economics undergraduates today would find the mathematics straightforward. Second, recent criticisms have elicited a torrent of supportive responses.³ Third, many non-professional economists seem to suspect that formalization in economics was a partial cause of not

*This research was supported in part by grants from the Open Society Institute and the Oxford Martin School. I am grateful to Jennifer L. Castle and Grayham E. Mizon for helpful comments. Contact details: david.hendry@nuffield.ox.ac.uk. Tel: 44 1865 278587.

foreseeing the financial crisis of 2007–2011. For example, HM Queen Elizabeth II questioned why UK economists had not done so, and the ensuing debate revealed that viewpoint (more precisely, “a failure of the collective imagination of many bright people”).⁴ Fourth, there are formal attacks on our “excessive ambitions”.⁵ Finally, the ESF-PSE Workshop on *The Debate on Mathematical Modeling in the Social Sciences* reflects a widespread desire to reconsider our tools. As ever, there is something to be said on both sides of the debate.

First, my forecasting case study is adapted from research which developed a theory of economic forecasting for settings where the model is mis-specified in unknown ways for an economic process that unexpectedly shifts at unknown times by unknown magnitudes.⁶ That work shows that a general mathematical analysis is both feasible and insightful, radically changing the interpretation of the outcomes of forecasting competitions,⁷ and what can be learned from forecast failures.⁸ Building on such results, later research led to explanations as to why some forecasting methods are ‘robust’ to location shifts after they have occurred,⁹ as well as suggesting possible approaches to forecasting breaks and during breaks.¹⁰

The second case study concerns model selection.¹¹ Since the forms, magnitudes and timings of breaks are usually unknown, a ‘portmanteau’ approach to their detection is required that allows for potential location shifts at every possible point in the sample. Impulse-indicator saturation (IIS) includes an impulse indicator for every observation in the set of candidate regressors, so adds T variables for T observations, then selects significant indicators from that saturating set.¹² Its ability to detect multiple breaks is established,¹³ and IIS allows an automatic test for super exogeneity.¹⁴ The properties of model selection in general, especially when there are more candidate variables, N , for inclusion in the analysis than the number of observations, T —as must occur with IIS—can only be resolved by mathematical analysis and its numerical sister of Monte Carlo simulations. Again, an understanding of the astonishingly good properties of extended general-to-specific procedures would be impossible without advanced mathematical analysis.

The third example concerns the mathematics of inter-temporal optimization and the formation of expectations, in particular, so-called ‘rational expectations’ (RE), where misleading results follow from present approaches applied to realistic economies. When unanticipated location shifts occur, estimated econometric models experience forecast failure, as noted above. However, that finding also entails that conditional expectations formed today of a future period after such a shift will be biased and potentially far from the minimum mean square error predictor ‘proved’ in most textbooks under the unstated assumption that the distributions involved are unchanged. Unfortunately, in economics, location shifts and other forms of structural break are all too common.¹⁵ Conclusions drawn on the ‘as if’ basis that breaks do not occur are inapplicable when they do: on a much grander scale, Euclidean geometry was long believed to be ‘true’, and many theorems, such as “the sum of the angles of a triangle add to 180° ”, were proved on that basis—until Riemann established the existence of non-Euclidean geometries in which the sum can exceed or fall short of 180° depending on the curvature of the space. Thus, the additional assumption was needed for Euclidean geometry that space was flat—an assumption that holds approximately locally, but is violated on the surface of a globe. Similarly, theorems about conditional expectations and the law of iterated expectations require the additional assumption that distributions do not shift, and are inapplicable otherwise.¹⁶ The appropriate mathematics for settings where distributions shift remains to be developed, and may end being ‘problem specific’ rather than generic.

The structure of the paper is as follows. Section 2 explains how mathematics was crucial in developing a theory of economic forecasting relevant to the practical setting where models are mis-specified and the world experiences intermittent unanticipated location shifts, and illustrates some surprising implications that could not have been deduced without a mathematical analysis. Then section 3 considers the formalization of model selection when there are more candidate regressors, $N > T$, than observations, T , although fewer variables, $n < T$, actually matter. Section 4 returns to the implications of forecast failure for inter-temporal optimization theory. Section 5 concludes.

2 Formalizing forecasting theory

There is a well-developed theory of economic forecasting based on the assumption that the econometric model coincides with a stationary economic data generation process (DGP).¹⁷ Consider an $n \times 1$ vector of variables to be forecast denoted $\mathbf{x}_t \sim D_{\mathbf{x}_t}(\mathbf{x}_t | \mathbf{X}_{t-1}, \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$, where $\mathbf{X}_{t-1} = (\dots \mathbf{x}_1 \dots \mathbf{x}_{t-1})$. A statistical forecast $\tilde{\mathbf{x}}_{T+h|T} = \mathbf{f}_h(\mathbf{X}_T^1)$ is made at time T (the forecast origin) for a future date $T+h$ (the forecast horizon). The key question in this setting is how to select \mathbf{f}_h .

The answer was ‘well known’: the conditional expectation $\hat{\mathbf{x}}_{T+h|T} = E[\mathbf{x}_{T+h} | \mathbf{X}_T^1]$ is unbiased, with $E[(\mathbf{x}_{T+h} - \hat{\mathbf{x}}_{T+h|T}) | \mathbf{X}_T^1] = 0$. Further, $\hat{\mathbf{x}}_{T+h|T}$ has the smallest mean-square forecast-error matrix:

$$M[\hat{\mathbf{x}}_{T+h|T} | \mathbf{X}_T^1] = E\left[(\mathbf{x}_{T+h} - \hat{\mathbf{x}}_{T+h|T})(\mathbf{x}_{T+h} - \hat{\mathbf{x}}_{T+h|T})' | \mathbf{X}_T^1\right].$$

However, that analysis finesses ten distinct problems. The first six concern problems learning about $D_{\mathbf{X}_T^1}(\cdot)$ and $\boldsymbol{\theta}$ from the available sample information, and the last four relate to the forecast period:¹⁸

- (1) *specification* of the set of relevant variables $\{\mathbf{x}_t\}$;
 - (2) *measurement* of the \mathbf{x} s;
 - (3) *formulation* of $D_{\mathbf{X}_T^1}(\cdot)$;
 - (4) *modelling* of the relationships;
 - (5) *estimation* of $\boldsymbol{\theta}$, and;
 - (6) *properties* of $D_{\mathbf{X}_T^1}(\cdot)$, which determine the ‘intrinsic’ uncertainty.
- All of these introduce in-sample uncertainties. Next, over the forecast horizon:
- (7) *properties* of $D_{\mathbf{X}_{T+H}^{T+1}}(\cdot)$ determine forecast uncertainty;
 - (8) *which grows* as H increases;
 - (9) especially for *integrated* data;
 - (10) increased by *changes* in $D_{\mathbf{X}_{T+H}^{T+1}}(\cdot)$ or $\boldsymbol{\theta}$.

These ten issues structure the analysis of forecasting. We now illustrate with a simple example, although its implications are generic, and hold for all forecasting models and DGPs, irrespective of the correctness (or otherwise) of specification of the model, and the properties of the DGP (stationary or integrated, with or without breaks of unknown timing, magnitude and form), and the data accuracy.¹⁹

2.1 Stationary scalar example

Consider a simple first-order autoregressive DGP with a known exogenous variable $\{z_t\}$:

$$y_t = \rho y_{t-1} + \gamma z_t + \epsilon_t \quad \text{where } \epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2] \quad \text{with } |\rho| < 1, \quad (1)$$

and $\text{IN}[0, \sigma_\epsilon^2]$ denotes an independent normal distribution with mean, $E[\epsilon_t] = 0$, and variance $V[\epsilon_t] = \sigma_\epsilon^2$. When ρ and γ are known and constant, the optimal forecast for $T+1$ from y_T for known z_{T+1} is:

$$\bar{y}_{T+1|T} = \rho y_T + \gamma z_{T+1} \quad (2)$$

In terms of the general analysis above, $D_{\mathbf{X}_T^1}(\cdot)$ implies $D_{\mathbf{X}_{T+1}^{T+1}}(\cdot)$, producing an unbiased forecast:

$$E\left[\left(y_{T+1} - \bar{y}_{T+1|T}\right) | y_T, z_{T+1}\right] = (\rho - \rho) y_T + (\gamma - \gamma) z_{T+1} + E[\epsilon_{T+1}] = 0,$$

with the smallest possible variance determined by $D_{\mathbf{X}_T^1}(\cdot)$:

$$V\left[\left(y_{T+1} - \bar{y}_{T+1|T}\right)\right] = \sigma_\epsilon^2.$$

Thus, in this specific case, $D_{\mathbf{x}_{T+1}}(\cdot)$ implies $y_{T+1} \sim \text{IN}[\rho y_T + \gamma z_{T+1}, \sigma_\epsilon^2]$. There will indeed be no forecasting problems, as issues (1)–(10) are ‘assumed away’. However, the ten potential problems return when omniscience is unavailable, even if z_{T+1} is known:

- (1) *Specification is incomplete* if (e.g.) \mathbf{x}_t is a vector not a scalar.
- (2) *Measurement is incorrect* if (e.g.) observe $\tilde{\mathbf{x}}_t$ not \mathbf{x}_t .
- (3) *Formulation is inadequate* if (e.g.) an intercept is needed.
- (4) *Modelling is wrong* if (e.g.) the wrong variables or lags are selected.
- (5) *Estimating ρ and γ* may add biases ($(\rho - \text{E}[\hat{\rho}])$, and $(\gamma - \text{E}[\hat{\gamma}])$), and variances $\text{V}[\hat{\rho}, \hat{\gamma}]$.
- (6) *Properties of $D(\epsilon_t) = \text{IN}[0, \sigma_\epsilon^2]$* determine $\text{V}[y_t]$.
- (7) *Assumed $\epsilon_{T+1} \sim \text{IN}[0, \sigma_\epsilon^2]$* but $\text{V}[\epsilon_{T+1}]$ could differ.
- (8) *Multi-step forecast errors cumulate* $\sum_{h=1}^H \rho^{h-1} v_{T+h}$ with $\text{V} = \frac{1-\rho^{2H}}{1-\rho^2} \sigma_\epsilon^2$.
- (9) $\rho = 1$ induces a *trending forecast variance*, $H\sigma_\epsilon^2$.
- (10) *If ρ changes*, forecast failure could occur.

A forecaster must be prepared for risks from all of (1)–(10), but some matter more.

To illustrate, we will first ‘undo’ problem (5), so the specification is correct, but (ρ, γ) have to be estimated from sample data, $t = 1, \dots, T$. Next, we will also violate (1) by omitting z_t , then (10) by changing ρ . Figure 1 illustrates for Monte Carlo simulated data from (1) when $z_t \sim \text{IN}[0, 1]$ with $\rho = 0.8$, $\gamma = 1$ and $\sigma_\epsilon^2 = 1$ when $T = 40$ and $H = 5$. We consider the six panels in turn.

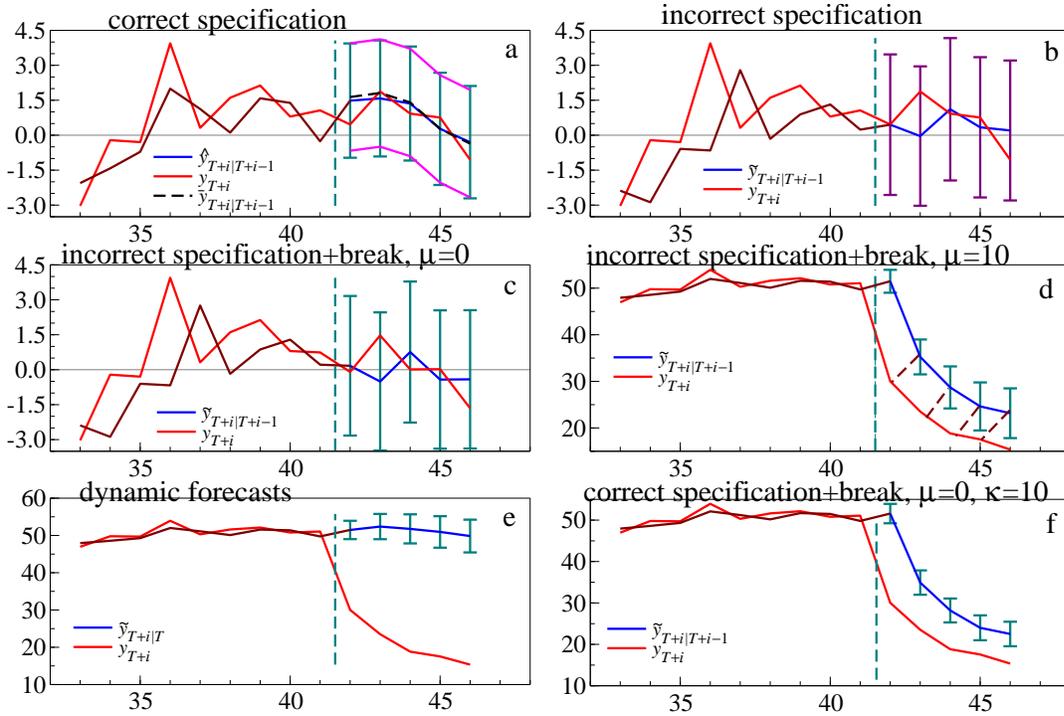


Figure 1: Forecasts under different scenarios

Panel a records forecasts from a single draw of the process in (1), both when (ρ, γ) are known ($\bar{y}_{T+i|T+i-1}$ from (2) with error bands of $\pm 2\hat{\sigma}$) and when estimating them ($\hat{y}_{T+i|T+i-1}$ with bars). The forecasts are almost identical, and there is only a small increase in uncertainty from estimation relative to knowing true parameter values. So not problem (5).

Panel b reports forecasts when z_t is omitted both in estimation and forecasting: the forecasts are poorer, but remain well within their *ex ante* forecast intervals. So not problem (1).

Panel c adds a shift in ρ at $T = 41$ to 0.4, so all of (1), (5) and (10) are violated, yet there is little noticeable impact from *halving* ρ : the forecasts are close to those in panel b, and well within the forecast intervals. In fact, a parameter constancy test barely rejects the false null more often for a halved ρ than for a constant one. Such changes hardly seem disastrous: moreover, similar results will be found if white noise measurement errors are added; or model selection is undertaken when the precise specification is not known. Is forecasting really that resilient in the face of estimation, mis-specification, selection and breaks?

Consider a slight change to the DGP in (1), namely introducing a non-zero intercept $\mu = 10$:

$$y_t = \mu + \rho y_{t-1} + \gamma z_t + \epsilon_t \text{ where } v_t \sim \text{IN}[0, \sigma_\epsilon^2] \text{ and } |\rho| < 1. \quad (3)$$

when everything else remains the same, including the change in ρ of the same magnitude, sign and timing. Since economic data are often indices or have arbitrary units (millions versus billions), μ is relatively arbitrary.

As *Panel d* shows, the forecasts are now catastrophically bad, emphasized by the dreadful 5-step ahead (dynamic) forecasts in *Panel e*: every forecast lies well outside the 95% forecast interval. Parameter constancy tests now reject 100% of the time. Such an outcome is called forecast failure. The data are identical to those in *Panel a* till observation 40, but the sharp fall from observation 41 onwards is obviously different. The dashed lines in *Panel d* show that the 1-step forecasts are systematically too high: at every point, the data are falling yet the forecasts are above the previous outcome. Finally, *Panel f* shows the forecasts for the same break when $\mu = 0$, the model is correctly specified by including z_t , but $\text{E}[z_t] = \kappa = 10$: forecast failure is manifest and similar to *Panel d*.

Without a mathematical analysis of the properties of forecasts, such a dramatic change for the same magnitude, form and timing of a break between no failure in mis-specified zero-intercept processes and massive failure when there are non-zero intercepts, would simply be an unexplained surprise. In fact, it is due to the impact of the non-constant ρ on the pre-existing mean, $\text{E}[y_t] = \theta$. In (1) when $\mu = \kappa = 0$, then $\text{E}[y_t] = 0$ before and after the shift in ρ . But in the second case:

$$\text{E}[y_t] = \theta = \frac{\mu + \gamma \kappa}{(1 - \rho)}$$

shifts markedly from $\theta = 50$ before the break in ρ to $\theta^* = 17$ after. Writing the model in (1) as:

$$\Delta y_t = (\rho - 1)(y_{t-1} - \theta) + \gamma(z_t - \kappa) + \epsilon_t \quad (4)$$

reveals it is an equilibrium-correction model (EqCM), where the equilibrium built-in to the model is θ , so the forecasts will converge back to θ *irrespective of what the data do*. Thus, if $\theta > \theta^*$, the data will fall, but the forecasts will continually return towards θ . This *location shift*, is clearly pernicious for forecasting, and explains *Panel f* as θ shifts when $\kappa \neq 0$. Perhaps more surprising, location shifts are the main problem likely to induce forecast failure, as we now describe, another result that cannot be established without mathematical analysis.

2.2 Forecast-error taxonomy

We now change the DGP to involve lagged rather than current z :

$$y_t = \theta + \rho(y_{t-1} - \theta) + \gamma(z_{t-1} - \kappa) + \epsilon_t \text{ for } t = 1, \dots, T \quad (5)$$

where $\epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2]$, $\text{E}[y_t] = \theta$ and $\text{E}[z_t] = \kappa$ with $\gamma \neq 0$, but z_{t-1} is omitted from the model:

$$y_t = \mu + \rho y_{t-1} + v_t$$

The break occurs at T , which leads to the post-break DGP:

$$y_t = \theta^* + \rho^* (y_{t-1} - \theta^*) + \gamma^* (z_{t-1} - \kappa^*) + \epsilon_t \text{ for } t = T + 1, \dots \quad (6)$$

The forecasting model:

$$\hat{y}_{T+1|T} = \hat{\theta} + \hat{\rho} (\hat{y}_T - \hat{\theta}) \quad (7)$$

is estimated over $t = 1, \dots, T$ delivering parameter estimates $(\hat{\theta}, \hat{\rho})$. The omitted variable and the dynamics induce biases, so $\mathbf{E}[\hat{\theta}] = \theta_e$ and $\mathbf{E}[\hat{\rho}] = \rho_e$. The forecast from an estimated \hat{y}_T at the forecast origin yields a forecast error of $\hat{\epsilon}_{T+1|T} = y_{T+1} - \hat{y}_{T+1|T}$. Ignoring interaction terms (corresponding to estimation covariances of $\mathbf{O}_p(T^{-1})$), the forecast error can be decomposed into the following taxonomy.

$\hat{\epsilon}_{T+1 T} \simeq$	Component	Expectation	Variance
$(1 - \rho^*) (\theta^* - \theta)$	(ia) equilibrium-mean change	$(1 - \rho^*) (\theta^* - \theta)$	0
$+ (\rho^* - \rho) (y_T - \theta)$	(ib) slope change	0	$(\rho^* - \rho)^2 \mathbf{V}[y_T]$
$+ (1 - \rho) (\theta - \theta_e)$	(iia) equilibrium-mean mis-specification	$(1 - \rho) (\theta - \theta_e)$	0
$+ (\rho - \rho_e) (y_T - \theta)$	(iib) slope mis-specification	0	$(\rho - \rho_e)^2 \mathbf{V}[y_T]$
$-\rho (\hat{y}_T - y_T)$	(iii) forecast-origin uncertainty	$-\rho (\mathbf{E}[\hat{y}_T] - y_T)$	$\rho^2 \mathbf{V}[\hat{y}_T - y_T]$
$-(1 - \rho) (\hat{\theta} - \theta_e)$	(iva) equilibrium-mean estimation	0	$\mathbf{O}_p(T^{-1})$
$-(\hat{\rho} - \rho_e) (y_T - \theta)$	(ivb) slope estimation	$\simeq 0$	$\mathbf{O}_p(T^{-1})$
$+\gamma^* (z_T - \kappa^*)$	(v) omitted variable	0	$(\gamma^*)^2 \mathbf{V}[z_T]$
$+\epsilon_{T+1}$	(vi) innovation error.	0	σ_ϵ^2

(8)

The third and fourth columns give the game away, but starting at the foot of the table:

(vi): $\mathbf{E}[\epsilon_{T+1}] = 0$ and $\mathbf{V}[\epsilon_{T+1}] = \sigma_\epsilon^2$ so there is no bias, but an $\mathbf{O}_p(1)$ variance component that is irreducible when $\{\epsilon_t\}$ is, and remains, an innovation error;

(v): again, $\mathbf{E}[\gamma^* (z_T - \kappa^*)] = 0$ and $\mathbf{V}[\gamma^* (z_T - \kappa^*)] = \sigma_z^2$, so there is also no bias despite the omission and the change in parameters, but an $\mathbf{O}_p(1)$ variance component (reducible if $\{z_{t-1}\}$ is included as a regressor with an offsetting estimation variance effect of $\mathbf{O}_p(T^{-1})$);

(ivb): slope estimation has $\mathbf{E}[(\hat{\rho} - \rho_e)(y_T - \theta)] \simeq 0$ as $\mathbf{E}[\hat{\rho} - \rho_e] = 0$ and $\mathbf{E}[y_T - \theta] = 0$, with a variance from estimation of $\mathbf{O}_p(T^{-1})$;

(iva): equilibrium-mean estimation has $\mathbf{E}[(1 - \rho)(\hat{\theta} - \theta_e)] = 0$ with an estimation variance of $\mathbf{O}_p(T^{-1})$;

(iii): forecast-origin uncertainty only has $\mathbf{E}[\rho(\hat{y}_T - y_T)] = 0$ if the forecast origin is unbiasedly estimated, but that can be achieved using modern methods of model selection applied to ‘nowcasting’²⁰ and has a variance component, probably of $\mathbf{O}_p(1)$;

(iib) slope mis-specification again has $\mathbf{E}[(\rho - \rho_e)(y_T - \theta)] = 0$ and an $\mathbf{O}_p(1)$ variance component unconditionally;

(iia) equilibrium-mean mis-specification is the first potentially serious component as $\theta \neq \theta_e$ is possible if there have been earlier in-sample location shifts that were not modelled, but IIS could resolve that difficulty;

(ib) slope change surprisingly has $\mathbf{E}[(\rho^* - \rho) (y_T - \theta)] = 0$ as $\mathbf{E}[y_T - \theta] = 0$ irrespective of $\rho^* \neq \rho$, a point illustrated above;

(ia) equilibrium-mean change is the fundamental problem: $\theta^* \neq \theta$ induces forecast failure.

In summary, once in-sample breaks are removed, from good forecast origin estimates:

$$\mathbf{E}[\hat{\epsilon}_{T+1|T}] \simeq (1 - \rho^*) (\theta^* - \theta) \quad (9)$$

and that bias persists at $\hat{\epsilon}_{T+2|T+1}$ etc., so long as (7) is used, even though no further breaks ensue. Keeping μ constant while shifting ρ to ρ^* induces a shift in θ to θ^* . The power of that insight is exemplified by

(a) changing *both* μ and ρ by large magnitudes, such that $\theta = \theta^*$, then demonstrating that the outcome is isomorphic to $\mu = \mu^* = 0$ (and hence $\theta = \theta^*$) as above, so no break is detected;²¹ and (b) when $\mu = \mu^* = 0$ and z_{t-1} is correctly included, then $\kappa \neq \kappa^*$ induces forecast failure by shifting θ .²²

The specificity of the example is irrelevant to the entailed result, which applies to all models in the equilibrium-correction class: they fail systematically when $E[y]$ changes as the models' forecasts are forced to converge back to θ irrespective of the value of θ^* . The class of EqCMs is huge and comprises all regression models; autoregressions; dynamic systems; vector autoregressions (VARs); dynamic-stochastic general equilibrium systems (DSGEs); autoregressive conditional heteroscedastic (ARCH) models; and generalized ARCH (GARCH) among others. Shifts in means are a pervasive and pernicious problem affecting forecasts from all such models.

2.3 Empirically-relevant theory

Such a theory needs to allow for the model being mis-specified for the DGP, with parameters estimated from inaccurate observations, on an integrated-cointegrated system, intermittently altering unexpectedly from structural breaks. That theory has achieved some success as it explains the prevalence of forecast failure, accounts for the results of forecasting competitions, and explains much of the good performance of 'consensus' forecasts. Of equal importance, it corrects some 'folklore' of forecasting, namely that forecast failure is *not* due to 'poor econometric methods', 'inaccurate data', 'incorrect estimation', or 'data-based model selection'.²³

Location shifts are the key to break detection: if there were no such shifts, forecast failure at 1% would be a 1 in 100 event. A crucial feature of (8) is that forecast errors persist unless the model is revised or abandoned. The former is difficult, as the cause of the forecast failure needs to be rapidly diagnosed and treated, and unfortunately, previous findings on forecasting breaks and during breaks show the large uncertainty attached to such attempts.²⁴ The latter requires a new model, which is even harder after a large unanticipated location shift. Fortunately, there is another approach—transform the initial model to avoid systematic forecast failure after location shifts.²⁵

To illustrate that result, reconsider the forecasting model in (7), but instead of using the level, which depends on θ , difference the model, retaining the original estimated parameter values:

$$\Delta \tilde{y}_{T+1|T} = \hat{\rho} \Delta (\hat{y}_T - \hat{\theta})$$

written as:

$$\tilde{y}_{T+1|T} = \hat{y}_T + \hat{\rho} \Delta \hat{y}_T \quad (10)$$

At the break point at time T , (10) makes the same magnitude forecast error as (7) precisely because the break is unpredicted. But one period later:

$$\tilde{y}_{T+2|T+1} = \hat{y}_{T+1} + \hat{\rho} \Delta \hat{y}_{T+1} = y_{T+1} + \hat{\rho} \Delta \hat{y}_{T+1} + (\hat{y}_{T+1} - y_{T+1}) \quad (11)$$

where (11) distinguishes an incorrect estimate of the forecast origin from the consequences of a break. When unbiased forecast origin estimates are available, so $E[\hat{y}_{T+1}] = y_{T+1}$ and the 'noise term' $\hat{\rho} \Delta \hat{y}_{T+1}$ is omitted to highlight the key point, then:

$$\tilde{y}_{T+2|T+1} = y_{T+1} = \theta^* + \rho^* (y_T - \theta^*) + \gamma^* (z_T - \kappa^*) + \epsilon_{T+1}.$$

Consequently, the forecast error is:

$$\begin{aligned} y_{T+2} - \tilde{y}_{T+2|T+1} &= \theta^* + \rho^* (y_{T+1} - \theta^*) + \gamma^* (z_T - \kappa^*) + \epsilon_{T+2} \\ &\quad - (\theta^* + \rho^* (y_T - \theta^*) + \gamma^* (z_T - \kappa^*) + \epsilon_{T+1}) \\ &= \rho^* \Delta y_{T+1} + \gamma^* \Delta z_{T+1} + \Delta \epsilon_{T+2} \end{aligned} \quad (12)$$

which is noisy, but not systematic, and delivers near unbiased forecasts because:

$$y_{T+1} = \theta^* + \rho^* (y_T - \theta^*) + \gamma^* (z_T - \kappa^*) + \epsilon_{T+1}$$

despite the omission of z_T .

Whereas the estimated in-sample DGP suffers from all the main sources of forecast error, namely stochastic and deterministic breaks, omitted variables, inconsistent parameters, estimation uncertainty and innovation errors, the ‘differenced’ transform reflects all the effects needed—parameter changes, differences of omitted variables, with no estimation components. There are two drawbacks, namely the unwanted presence of ϵ_{T+1} in (12), which doubles the innovation error variance; and all variables are lagged one extra period, which adds the ‘noise’ of $I(-1)$ effects. Nevertheless, there is a clear trade-off between avoiding systematic forecast failure and adding somewhat to the forecast-error variance when no location shifts occur. After the unanticipated occurrence of a location shift, as with the recent financial crisis, forecast failure is ubiquitous in EqCMs, but not in differenced variants thereof, so there need be no connection between the in-sample ‘quality’ (or verisimilitude) of a model and that of its later forecasts.

2.4 Designing Monte Carlo simulations

Simulation evidence is complementary to mathematical analysis in that, while mathematics is fundamental to understanding the analytically-tractable cases, simulation analysis helps examine empirically-relevant cases that may be intractable analytically.²⁶ The insights from mathematical analysis remain essential when designing Monte Carlo studies to focus on ‘canonical’ cases, isolating aspects that are invariant across the simulations. For example, in a mean-zero autoregressive process, the units of the error standard deviation are irrelevant, but cease to be so if there is a non-zero intercept in the data generating process. When all parameters shift but leave the equilibrium mean constant is isomorphic to a zero mean, so allows a specific simulation to entail general results.

3 Selecting econometric models from a mass of candidate variables

There are many critical analyses of model selection, almost all of which assume ‘correct’ models with constant parameters where simply fitting the given specification dominates selection. This is not a realistic characterization of the situation confronting empirical investigators of economic time series. Data processes are complicated and evolving, so models derived from economic theory provide only a guide to some of the main variables, and rarely address breaks or outliers which vitiate any *ceteris paribus* assumptions. Thus, model selection is inevitable in practice, where only some substantively relevant aspects are correctly included, some are omitted, and some irrelevant aspects are also included, usually correlated with omitted variables.

Selection is essential when there are large numbers of potential explanatory variables. But can model selection work well in that setting? The canonical case of more variables than observations, $N > T$, is including an impulse indicator for every observation in the candidate regressor set. In the simplest analysis (the ‘split-half’ case), one regression only includes the first $T/2$ of these indicators initially. By dummifying out that first subset of observations, estimates are based on the remaining data, and any observations in the first half that are discrepant will result in significant indicators.²⁷ The location of the significant indicators is recorded, then the first $T/2$ are replaced by the second half and the procedure repeated. The two sets of significant indicators are then added to the general model for selection of those that remain significant together with selecting over the non-dummy variables. This is the approach called impulse-indicator saturation (IIS) above.²⁸ IIS is an efficient method: under the null of no breaks, outliers or data contamination, the cost of applying IIS at a significance level α is the loss of αT of the

sample, so at $\alpha = 0.01$ and $T = 100$, IIS is 99% efficient. This follows because under the null, αT indicators will be retained by chance sampling, and each merely ‘dummies out’ an observation. Thus, despite adding as many indicator variables as observations to the set of candidate variables to be selected from, when IIS is not needed the costs are almost negligible; and if IIS is required, the most pernicious effects of induced location shifts on non-constant intercepts, slopes and equation standard errors can be corrected.

The mathematical analyses supporting these claims are in the references, and are consistent with a wide range of Monte Carlo simulations. No amount of non-mathematical thinking could have delivered such an astonishing insight: indeed most reactions are that adding $N > T$ candidate variables to the model search cannot be done, and if it could, it would produce garbage. But in fact it is easy to do, and almost costless.

3.1 As many candidate variables as observations

The analytic approach to understanding IIS can be applied when there are $N = T$ IID mutually orthogonal candidate regressors $z_{i,t}$, where none matters under the null. Formally, the DGP is:

$$y_t = \epsilon_t \tag{13}$$

and the general, but inestimable, model can be expressed as:

$$y_t = \sum_{j=1}^N \delta_j z_{j,t} + \epsilon_t \tag{14}$$

where $\delta_j = 0 \forall j = 1, \dots, N$. We consider the analogue of the ‘split-half’ analysis from IIS. Thus, add the first $N/2$, and select those with $|t_{\delta_{j=0}}| > c_\alpha$ at significance level $\alpha = 1/T = 1/N$. Record which were significant, and drop them all. Now add the second block of $N/2$, again select those with $|t_{\delta_{j=0}}| > c_\alpha$ at significance level $\alpha = 1/N$, and record which are significant. Finally, combine the recorded variables from the two stages (if any), and select again at significance level $\alpha = 1/N$. At both sub-steps, on average $\alpha N/2 = 1/2$ of a variable will be retained by chance under the null, so on average $\alpha N = 1$ will be retained from the combined stage. Again, despite examining the relevance of $N = T$ additional irrelevant variables, almost none is retained and the statistical analysis is 99% efficient under the null at eliminating irrelevant variables, merely costing one degree of freedom on average.

3.2 More candidate variables than observations

These results can be extended to having $N > T$ general candidate variables in the search, where $n < N$ are relevant.²⁹ The k theory-determined variables are not selected over, so are forced to be retained by the search. When the theory is correctly specified, the costs of searching over the remaining $N - k$ candidates is trivial for small α , as now $\alpha(N - k)$ irrelevant variables will be retained by chance and each merely costs a ‘degree of freedom’. The real surprise is that the distribution of the estimates of the coefficients of the relevant variables are exactly the same as if no search was undertaken at all. This is because the relevant and irrelevant variables can be orthogonalized without loss of generality, and as the latter are irrelevant, orthogonalizing does not alter the parameters of the relevant variables—and it is well known that estimator distributions are unaffected by the omission or inclusion of orthogonal variables. Without an advanced mathematical analysis, such a result is unimaginable.

Most economists and econometricians believe model selection is a pernicious but necessary activity—as shown above, it is in fact almost costless despite $N > T$, and invaluable when needed. Their beliefs were not based on sound mathematics, and that signals the dangers of not using powerful analytical

tools. The practical difficulty is to be sure the tool is correctly based, and relevant to the target situation, a problem to which we now turn.

4 Models of expectations

The very notation used for the mathematics of expectations in economics is inadvertently designed to mislead. Instead of $E[\mathbf{x}_{T+h}|\mathbf{X}_T^1]$ as in §2, one must write conditional expectations as:

$$E_{T+h}[\mathbf{x}_{T+h}|\mathbf{X}_T^1].$$

Thus *three* time subscripts are needed: that for the date of the conditioning information (here \mathbf{X}_T^1); that for the date of the variable being expected (here \mathbf{x}_{T+h}); and that for the distribution over which the expectation is formed (here E_{T+h}). If the distribution is stationary, then $E_{T+h} = E_T$, where the latter is the only feasible distribution at the time the expectation is formed. Otherwise, we have a paradox if $D_{x_t}(\cdot)$ is not constant as one needs to know the whole future distribution to derive the forecast. Worse, one cannot prove that $\tilde{\mathbf{x}}_{T+h|T} = E_T[\mathbf{x}_{T+h}|\mathbf{X}_T^1]$ is a useful forecast if $D_{x_{T+h}}(\cdot) \neq D_{x_T}(\cdot)$.

Theories of expectations must face the realities of forecasting discussed above. ‘Rational’ expectations (RE) correspond to the conditional expectation given available information (denoted \mathcal{I}_t):

$$y_{t+1}^{re} = E[y_{t+1} | \mathcal{I}_t]. \quad (15)$$

RE assumes free information, unlimited computing power, and the discovery of the form of $E[y_{t+1}|\mathcal{I}_t]$ by economic agents. If (15) is to be useful, it should be written as:

$$y_{t+1}^e = E_{t+1}[y_{t+1} | \mathcal{I}_t] = \int y_{t+1} f_{t+1}(y_{t+1} | \mathcal{I}_t) dy_{t+1}. \quad (16)$$

Only then is y_{t+1}^e even unbiased for y_{t+1} . But (16) requires a crystal ball for *future* $f_{t+1}(y_{t+1}|\mathcal{I}_t)$. The best an agent can do is to form a ‘sensible expectation’, y_{t+1}^{se} , forecasting $f_{t+1}(\cdot)$ by $\hat{f}_{t+1}(\cdot)$:

$$y_{t+1}^{se} = \int y_{t+1} \hat{f}_{t+1}(y_{t+1} | \mathcal{I}_t) dy_{t+1}. \quad (17)$$

If the moments of $f_{t+1}(y_{t+1}|\mathcal{I}_t)$ alter, there are no good rules for $\hat{f}_{t+1}(\cdot)$, but $\hat{f}_{t+1}(y_{t+1}|\mathcal{I}_t) = f_t(\cdot)$ is not a good choice. Agents cannot know how \mathcal{I}_t will enter $f_{t+1}(\cdot)$ if there is no time invariance.

When $f_{t+1}(\cdot) \neq f_t(\cdot)$, forecasting devices robust to location shifts avoid systematic mis-forecasting after breaks, as illustrated above. But if agents use robust predictors, and are not endowed with prescience that sustains an unbiased RE, then one needs to re-specify expectations in economic-theory models. But the problem is unfortunately even worse. Consider a very simple example—if $x_t \sim \text{IN}[\mu_t, \sigma_x^2]$, then:

$$\begin{aligned} E_t[x_t | \mathbf{X}_{t-1}] &= \mu_t \\ E_{t-1}[x_t | \mathbf{X}_{t-1}] &= \mu_{t-1} \end{aligned}$$

when the mean changes, so letting $\epsilon_t = x_t - E_{t-1}[x_t|\mathbf{X}_{t-1}]$:

$$E_t[\epsilon_t] = \mu_t - \mu_{t-1} \neq 0 \quad (18)$$

shows that the conditional expectation is biased. But such a result also entails that the law of iterated expectations does not hold inter-temporally without the additional assumption that the distribution does not shift, and is inapplicable otherwise.³⁰ If the distribution shifts, many of the ‘mathematical derivations of inter-temporal optimization’ are invalid in the same way that Euclidean calculations are invalid on a

sphere. And as with non-Euclidean geometry, a different mathematics is needed depending on the shape of the relevant space, so here, different calculations will be required depending on the unanticipated breaks experienced by economies, the abilities of economic agents to learn what those breaks entail, and the speeds with which they reform their plans and expectations about the future. Thus, more powerful mathematical tools are urgently required to enable such analyses.

5 Conclusion

The paper has considered three possible situations of the use or mis-use of mathematics in economics and econometrics. The first concerned the properties of economic forecasts and forecast failure in particular, where a mathematical analysis was both essential and highly revealing. While only a specific example was given, the analysis holds independently of how well or badly specified the forecasting model is, and how the process being forecast actually behaves. Location shifts were isolated as the primary cause of forecast failure, with the myriad of other possible model mis-specifications and data mis-measurements playing a secondary role, despite prior intuitions to the contrary.

The second situation concerned model selection when there are more candidate variables N than the number of observations T . Again, an understanding of the astonishingly good properties of extended general-to-specific based procedures would be impossible without advanced mathematical analysis. That is particularly true of the finding that the distributions of the estimated coefficients of a correct theory model's forced variables are not affected by selecting over any number of irrelevant candidate variables. Yet there are innumerable assertions in the econometrics literature (and beyond) that selection is pernicious 'data mining', leads to 'over-fitting', etc., all without substantive mathematical proofs.

The third concerned the mathematics of inter-temporal optimization and the formation of expectations, in particular, so-called 'rational expectations', where misleading results followed from present approaches applied to realistic economies. Conventional notation fails to address the three different times relevant to expectations formation, namely that of the available conditioning information, of the target variable to be forecast and of the time the expectation is formed. Consequently, the effects of shifts in distributions over which expectations are calculated have been hidden. Conditional expectations formed today for an outcome tomorrow need not be unbiased nor minimum variance. The appropriate mathematics remains to be developed, and may end being 'problem specific' rather than generic. Nevertheless, the conclusion is inexorable: the solution is more powerful and more general mathematical techniques, with assumptions that more closely match 'economic reality'.

Notes

¹Discussed by John N. Keynes, *The Scope and Method of Political Economy*, New York: Kelley and Millman, 1891, and Joseph Schumpeter, "The common sense of econometrics", *Econometrica*, **1**, 1933, 5–12, among others (see David F. Hendry and Mary S. Morgan, *The Foundations of Econometric Analysis*, Cambridge: Cambridge University Press, 1995).

²By G. Udney Yule, "Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time series (with discussion)", *Journal of the Royal Statistical Society*, **89**, 1926, 1–64.

³See e.g., John LLewellyn, "It's possible to subtract mathematics from economics" *The Observer*, 16 August, 2009 <http://www.guardian.co.uk/business/2009/aug/16/economics-economics>.

⁴See e.g., <http://www.guardian.co.uk/uk/2009/jul/26/monarchy-credit-crunch>.

⁵Like that to which I responded in "Comment on 'Excessive ambitions' (by Jon Elster)", *Capitalism and Society*, **4**, 2009, DOI: 10.2202/1932-0213.1056.

⁶Taken from David F. Hendry and Bent Nielsen, *Econometric Modeling: A Likelihood Approach*, Princeton: Princeton University Press, 2007, building on Michael P. Clements and David F. Hendry, *Forecasting Non-stationary Economic Time Series*, Cambridge, Mass.: MIT Press, 1999.

⁷See Spyros Makridakis and Michelle Hibon "The M3-competition: Results, conclusions and implications", *International Journal of Forecasting*, **16**, 451–476, and Robert Fildes and Keith Ord "Forecasting competitions—their role in improving

forecasting practice and research”, in Michael P. Clements and David F. Hendry (eds.) *A Companion to Economic Forecasting*, Oxford: Blackwells, 2002, and compare Michael P. Clements and David F. Hendry, 2001, “Explaining the results of the M3 forecasting competition”, *International Journal of Forecasting*, **17**, 550–554.

⁸See David F. Hendry and Jurgen A. Doornik “The implications for econometric modelling of forecast failure”, *Scottish Journal of Political Economy*, **44**, 1997, 437–461, and Michael P. Clements and David F. Hendry, “Explaining forecast failure in macroeconomics”, in Clements and Hendry (eds.), 2002, *op. cit.*

⁹see David F. Hendry, “Robustifying forecasts from equilibrium-correction models”, *Journal of Econometrics*, **135**, 2006, 399–426.

¹⁰See Jennifer L. Castle, Nicholas W. P. Fawcett, and David F. Hendry, “Forecasting Breaks and During Breaks”, in Michael P. Clements and David F. Hendry, (eds.), *Oxford Handbook of Economic Forecasting*, Oxford: Oxford University Press, forthcoming.

¹¹This draws on research by David F. Hendry and Hans-Martin Krolzig, “The properties of automatic Gets modelling”, *Economic Journal*, **115**, 2005, C32–C61; Jennifer L. Castle, Jurgen A. Doornik and David F. Hendry, “Evaluating automatic model selection”, *Journal of Time Series Econometrics*, **3** (1), DOI: 10.2202/1941-1928.1097; Jennifer L. Castle and David F. Hendry, “Automatic selection of non-linear models”, in Liuping Wang, Hugues Garnier and T. Jackman (eds.) *System Identification, Environmental Modelling and Control*, New York: Springer, forthcoming; Jurgen A. Doornik, “Autometrics” in Jennifer L. Castle and Neil Shephard (eds.) *The Methodology and Practice of Econometrics*, Oxford: Oxford University Press, 88–121, 2009; David F. Hendry and Grayham E. Mizon, “Econometric modelling of time series with outlying observations”, *Journal of Time Series Econometrics*, **3** (1), DOI: 10.2202/1941-1928.1100.

¹²Its properties in a simple setting are analyzed in David F. Hendry, Søren Johansen and Carlos Santos, “Automatic selection of indicators in a fully saturated regression”, *Computational Statistics*, **33**, 2008, 317–335, Erratum, 337–339, and extended by Søren Johansen and Bent Nielsen, “An analysis of the indicator saturation estimator as a robust regression estimator”, 1–36 in Castle and Shephard (eds.), 2009, *op.cit.*, to both stationary and unit-root autoregressive models.

¹³See Jennifer L. Castle, Jurgen A. Doornik and David F. Hendry, 2011, “Model selection when there are multiple breaks”, *Journal of Econometrics*, forthcoming

¹⁴See David F. Hendry and Carlos Santos, “An automatic test of super exogeneity”, 164–193 in Mark W. Watson, Tim Bollerslev, and Jeff Russell (eds.) *Volatility and Time Series Econometrics*, Oxford: Oxford University Press, 2010, extending earlier research by Robert F. Engle, David F. Hendry and Jean-Francois Richard, “Exogeneity”, *Econometrica*, **51**, 1983, 277–304, and Robert F. Engle and David F. Hendry, “Testing super exogeneity and invariance in regression models”, *Journal of Econometrics*, **56**, 1993, 119–139.

¹⁵See e.g., James H. Stock and Mark W. Watson, “Evidence on structural instability in macroeconomic time series relations”, *Journal of Business and Economic Statistics*, **14**, 1996, 11–30, and Ray Barrell, “Forecasting the world economy”, 149–169 in David F. Hendry and Neil R. Ericsson (eds.), *Understanding Economic Forecasts*, Cambridge, Mass.: MIT Press, 2001.

¹⁶See David F. Hendry and Grayham E. Mizon, “On the mathematical basis of inter-temporal optimization”, Working paper 497, Economics Department, Oxford, 2010.

¹⁷See the famous treatise of Trygve Haavelmo, “The probability approach in econometrics”, *Econometrica*, **12**, 1944, 1–118, extended by Lawrence R. Klein, *An Essay on the Theory of Economic Prediction*, Chicago: Markham Publishing Company, 1971.

¹⁸See Hendry and Nielsen, 2007, *op. cit.*

¹⁹Michael P. Clements and David F. Hendry, “Forecasting with breaks in data processes”, 605–657 in Graham Elliott, Clive W. J. Granger and Allan Timmermann (eds.), *Handbook of Econometrics on Forecasting*, Amsterdam: Elsevier, 2006, provide a general ‘non-parametric’ analysis.

²⁰See e.g., Jennifer L. Castle, Nicholas W. P. Fawcett, and David F. Hendry, “Nowcasting is not just contemporaneous forecasting”, *National Institute Economic Review*, **210**, 2009, 71–89, and Jennifer L. Castle and David F. Hendry, “Nowcasting from disaggregates in the face of location shifts”, *Journal of Forecasting*, **29**, 2010, 200–214.

²¹Figure 1 Panel a: see e.g., David F. Hendry, “On detectable and non-detectable structural change”, *Structural Change and Economic Dynamics*, **11**, 2000, 45–65.

²²See e.g., Hendry and Mizon, 2010, *op.cit.*

²³See, e.g., the unsubstantiated assertions on what went wrong in economic forecasting at the Bank of Canada by Don Coletti, Ben Hunt, David Rose and Robert J. Tetlow, “The Bank of Canada’s new quarterly projection model, part 3. The dynamic model: QPM”, Technical report 75, 1996, Bank of Canada, Ottawa: “the inability of relatively unstructured, estimated models to predict well for any length of time outside their estimation period seemed to indicate that small-sample econometric problems were perhaps more fundamental than had been appreciated and that too much attention had been paid to capturing the idiosyncrasies of particular samples.”

²⁴See e.g., Jennifer L. Castle, Nicholas W. P. Fawcett and David F. Hendry, “Forecasting with equilibrium-correction models during structural breaks”, *Journal of Econometrics*, **158**, 2010, 25–36.

²⁵See David F. Hendry, “Robustifying forecasts from equilibrium-correction models”, *Journal of Econometrics*, **135**, 2006, 399–426, and for a non-technical account, see Michael P. Clements and David F. Hendry, “Economic forecasting in a changing world”, *Capitalism and Society*, **3**, 2008, 1–18

²⁶A more detailed analysis of efficient Monte Carlo simulation is provided in David F. Hendry, “Monte Carlo experimentation in econometrics”, 937–976 in Zvi Griliches and Michael D. Intriligator (eds.), *Handbook of Econometrics*, **2**, Amsterdam: North-Holland, 1984.

²⁷In essence, that lies behind the approach for testing parameter constancy using indicators in David S. Salkever, “The use of dummy variables to compute predictions, prediction errors and confidence intervals”, *Journal of Econometrics*, **4**, 1976, 393–397.

²⁸see Hendry, Johansen and Santos, *op. cit.* and Johansen and Nielsen, *op. cit.*, who derive the distributions of estimators after IIS when there are no outliers or breaks, and relate IIS to robust estimation.

²⁹ See David F. Hendry and Søren Johansen, “Model selection when forcing retention of theory variables”, Unpublished paper, Economics Department, University of Oxford, 2010.

³⁰As shown in Hendry and Mizon, 2010, *op.cit.*; a non-technical discussion is provided in David F. Hendry and Grayham E. Mizon, 2011, “What needs rethinking in macroeconomics?”, *Global Policy*, forthcoming.