ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

UNPREDICTABILITY IN ECONOMIC ANALYSIS, ECONOMETRIC MODELING AND FORECASTING

David F. Hendry

Number 551
May 2011

Manor Road Building, Oxford OX1 3UQ

Unpredictability in Economic Analysis, Econometric Modeling and Forecasting

David F. Hendry*
Economics Department and Institute for Economic Modelling,
Oxford Martin School, University of Oxford, UK, and Institute for New Economic Thinking (INET)

Abstract

Unpredictability arises from intrinsic stochastic variation, unexpected instances of outliers, and unanticipated extrinsic shifts of distributions. We analyze their properties, relationships, and different effects on the three arenas in the title, which suggests considering three associated information sets. We note the implications of unanticipated shifts for forecasting, economic analyses of efficient markets, inter-temporal derivations, and general-to-specific model selection, tackling outliers and non-constancy by impulse-indicator saturation, and contrast the potential success in modeling breaks with the major difficulties confronting forecasting.

JEL classifications: C51, C22.
Keywords: Unpredictability; 'Black Swans'; Distributional shifts; Forecasting; Model selection

Contents

*This research was supported in part by grants from the Open Society Institute and the Oxford Martin School. I am indebted to Jennifer L. Castle, Grayham E. Mizon and Bent Nielsen for helpful comments on an earlier version. Contact details: david.hendry@nuffield.ox.ac.uk.

Preface

It is a great pleasure to contribute a paper on econometric modeling and forecasting to a *Festschrift* in honor of Professor Hal White. I have known Hal for more than 30 years, and have not only benefitted greatly from his many important published contributions, but also from his thoughtful comments on my own research, while I visited UCSD, at conferences, and when he was editing. Well known for literally blowing his own trumpet as Dr Jazz, Hal is not prone to do so in the metaphorical sense, and has always been a wonderful colleague, and a delightful golf partner.

One of the most difficult tasks in any science is to overcome earlier prejudices: everyone 'knew' that a general method for valid inference with unknown forms of heteroskedasticity was 'impossible' as there were too many parameters to estimate, but Hal showed that such beliefs were not well founded, in an article that has been cited more than 12000 times. In addition to transforming econometricians' perceptions of how to calculate estimation uncertainty by heteroskedastic-consistent standard errors, despite many features of an error distribution being unknown–and thereby creating a minor industry of generalizations–Hal has thrown light on many issues from general asymptotic distribution theory, networks and non-linearity, empirical modeling and model evaluation, causality and prediction among others. I look forward to many more years of fruitful and incisive output.

1 Introduction

Unpredictability has been formalized as intrinsic stochastic variation in a known distribution, where conditioning on available information does not alter the outcome from the unconditional distribution (see e.g., Clements and Hendry, 2005). Such variation is usually attributed (*inter alia*) to chance distribution sampling, 'random errors', incomplete information, or in economics, many small changes in the choices across individual agents. A variable that is intrinsically unpredictable cannot be modeled or forecast better than its unconditional distribution. However, the converse does not hold: a variable that is not intrinsically unpredictable may still be essentially unpredictable because of two additional aspects of unpredictability. The first concerns independent draws from fat-tailed or heavy-tailed distributions, which leads to a notion we call 'instance unpredictability'. Here the distribution of a variable that is not intrinsically unpredictable is known, as are all conditional and unconditional probabilities, but there is a non-negligible probability of a very discrepant outcome. While that probability is known, it is not known on which draw the discrepant outcome will occur, nor its magnitude, leading to a 'Black Swan' (as in Taleb, 2007), with potentially large costs when it occurs (see Barro, 2009). The third aspect we call 'extrinsic unpredictability', which derives from unanticipated shifts of the distribution itself at unanticipated times–so inherently also involves 'instance unpredictability'–of which location shifts are usually the most pernicious (see, e.g., Castle, Fawcett and Hendry, 2011). The recent financial crisis and ensuing deep recession have brought both instance and extrinsic unpredictability into more salient focus (see Taleb, 2009, and Soros, 2008, 2010, respectively).

There being three aspects of unpredictability suggests considering 3 possible associated information sets that might explain at least a part of their otherwise unaccounted variation. This is well established both theoretically and empirically for putative intrinsic unpredictability, where 'regular' explanatory variables are sought. Empirically, population distributions are never known, so even to calculate the probabilities for instance unpredictability it will always be necessary to estimate the distributional form from available evidence. This applies even more forcefully to extrinsic unpredictability, so we briefly discuss general-to-specific (Gets) modeling when uncertainty about model specification entails a need to select from a candidate set of possible variables, functional-form transformations, lag lengths and multiple outliers and breaks. When an investigator wants a congruent model that variance dominates and encompasses rival models, then the largest available information set should be used, consistent with the implications of intrinsic unpredictability. As shown in Castle, Doornik and Hendry (2011a), although such an information set may comprise more candidate variables, N, than observations, T, automatic model selection algorithms like *Autometrics* can handle $N > T$ (see Doornik, 2009a). Outliers and location shifts may also be corrected in modeling by adding the appropriate variables that cause shifts, by non-linear reaction functions, or if no appropriate variables can be found, by impulse-indicator saturation (denoted IIS: see Hendry, Johansen and Santos, 2008, and Johansen and Nielsen, 2009).

Each type of unpredictability has substantively different effects on economic analyses, econometric modeling, and economic forecasting. Specifically, inter-temporal economic theory, forecasting, and policy analyses could go awry facing extrinsic unpredictability, whereas *ex post*, the outcomes are susceptible to being modeled. The ability to do so highlights the contrast between the possibilities of modeling extrinsic unpredictability *ex post* with the difficulties confronting successful *ex ante* forecasting, where one must either forecast outliers or shifts, or breaks in the variables that entering a model, all of which are demanding tasks. Thus, a congruent model need not outperform in forecasting even when it coincides with the in-sample data generation process (DGP), clarifying a number of findings in the empirical forecasting literature.

The structure of the paper is as follows. Section 2 summarizes the well-known properties of intrinsic unpredictability, its relationship to instance and extrinsic unpredictability in §2.1, with the impact of reduced information discussed in §2.2 to resolve an earlier puzzle. Section 3 discusses instance unpredictability in relation to when a large deviation draw will occur from a distribution known to be fat tailed, and notes its main differences from extrinsic unpredictability, the topic of section 4. Section 5 considers the possibility of three distinct information sets associated respectively with 'normal causality', the timing of outliers, and the occurrence of distributional shifts. Section 6 discusses why there is a fundamental separation in instance and extrinsic unpredictability between modeling and forecasting that does not apply to intrinsic unpredictability: adding a variable which explains a past shift can greatly improve a fitted model without improving its forecasts when shifts in that variable cannot be forecast. Section 7 considers the difficulties involved in determining the properties of breaks from a few data points, an issue confronting both economists and economic agents, with implications for economic analyses of efficient markets and inter-temporal derivations, where we draw on several results on conditional expectations and the law of iterated expectations from Hendry and Mizon (2010). Section 8 investigates the relationships between the three aspects of unpredictability and model selection in processes with unanticipated breaks. That leads to a reconsideration of the role of congruent modeling for forecasting, addressed in §8.1. Section 9 concludes.

2 Some properties of intrinsic unpredictability

A non-degenerate n-dimensional vector random variable $\boldsymbol{\epsilon}_t$ is an unpredictable process with respect to an information set $\mathcal{I}_{t-1}$ (which always includes the sigma-field generated by the past of $\boldsymbol{\epsilon}_t$, denoted $\sigma[\mathbf{E}_{t-1}]$) over a time period $\mathcal{T}$ if the conditional distribution $\mathsf{D}_{\boldsymbol{\epsilon}_t}(\boldsymbol{\epsilon}_t|\mathcal{I}_{t-1})$ equals the unconditional distribution $\mathsf{D}_{\boldsymbol{\epsilon}_t}(\boldsymbol{\epsilon}_t)$ so that:

$$\mathsf{D}_{\boldsymbol{\epsilon}_t}(\boldsymbol{\epsilon}_t \mid \mathcal{I}_{t-1}) = \mathsf{D}_{\boldsymbol{\epsilon}_t}(\boldsymbol{\epsilon}_t) \quad \forall t \in \mathcal{T}. \tag{1}$$

Such unpredictability is an intrinsic property of $\boldsymbol{\epsilon}_t$ in relation to $\mathcal{I}_{t-1}$, not dependent on knowledge about $\mathsf{D}_{\boldsymbol{\epsilon}_t}(\cdot)$, tantamount to independence between $\boldsymbol{\epsilon}_t$ and $\mathcal{I}_{t-1}$. When $\mathcal{I}_{t-1} = \sigma[\mathbf{X}_{t-1}]$ (say) is the 'universal' information set, (1) clarifies that $\boldsymbol{\epsilon}_t$ is *intrinsically* unpredictable. Intrinsic unpredictability applies equally to explaining the past (i.e., modeling $\{\boldsymbol{\epsilon}_t, t = 1, \ldots, T\}$) and forecasting the future from

T (i.e., of $\{\boldsymbol{\epsilon}_t, t = T+1, \ldots, T+h\}$): the best that can be achieved in both settings is the unconditional distribution, and $\mathcal{I}_{t-1}$ is of no help in reducing either uncertainty. Expectations formed at time t are denoted $\mathsf{E}_t[\cdot]$, and variance is denoted $\mathsf{V}_t[\cdot]$ for each point in $\mathcal{T}$, so assuming the relevant moments exist, intrinsic unpredictability in distribution entails unpredictability in mean and variance:

$$\mathsf{E}_t[\boldsymbol{\epsilon}_t \mid \mathcal{I}_{t-1}] = \mathsf{E}_t[\boldsymbol{\epsilon}_t] \text{ and } \mathsf{V}_t[\boldsymbol{\epsilon}_t \mid \mathcal{I}_{t-1}] = \mathsf{V}_t[\boldsymbol{\epsilon}_t]. \tag{2}$$

where neither the former nor the latter alone need imply the other. As a well known example, $\boldsymbol{\epsilon}_t \sim \mathsf{IN}_n[\boldsymbol{\mu}, \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}]$, denoting an independently distributed Gaussian variable with expected value $\mathsf{E}_t[\boldsymbol{\epsilon}_t] = \boldsymbol{\mu}$ and variance $\mathsf{V}_t[\boldsymbol{\epsilon}_t] = \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}$, is an unpredictable process.

Intrinsic unpredictability is only invariant under non-singular contemporaneous transformations, as inter-temporal transforms must affect (1), implying that no unique measure of forecast accuracy exists (see e.g., Leitch and Tanner, 1991, Clements and Hendry, 2005, and Granger and Pesaran, 2000a, 2000b). Thus, predictability requires combinations with $\mathcal{I}_{t-1}$, as in, for example:

$$\mathbf{y}_t = \boldsymbol{\psi}_t(\mathbf{X}_{t-1}) + \boldsymbol{\epsilon}_t \text{ where } \boldsymbol{\epsilon}_t \sim \mathsf{ID}_n[\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}] \tag{3}$$

so $\mathbf{y}_t$ depends on both the information set and the innovation component. Then:

$$\mathsf{D}_{\mathsf{y}_t}(\mathbf{y}_t \mid \mathcal{I}_{t-1}) \neq \mathsf{D}_{\mathsf{y}_t}(\mathbf{y}_t) \quad \forall t \in \mathcal{T}. \tag{4}$$

In (3), $\mathbf{y}_t$ is predictable in mean even if $\boldsymbol{\epsilon}_t$ is unpredictable as:

$$\mathsf{E}_t[\mathbf{y}_t \mid \mathcal{I}_{t-1}] = \boldsymbol{\psi}_t(\mathbf{X}_{t-1}) \neq \mathsf{E}_t[\mathbf{y}_t],$$

in general. Since:

$$\mathsf{V}_t[\mathbf{y}_t \mid \mathcal{I}_{t-1}] < \mathsf{V}_t[\mathbf{y}_t] \text{ when } \mathsf{V}_t[\boldsymbol{\psi}_t(\mathbf{X}_{t-1})] \neq \mathbf{0} \tag{5}$$

predictability ensures a variance reduction, consistent with its nomenclature, since unpredictability entails equality in (5), and the 'smaller' the conditional variance matrix, the less uncertain is the prediction of $\mathbf{y}_t$ from $\mathcal{I}_{t-1}$.

2.1 Intrinsic, instance and extrinsic unpredictability

When (3) holds, $\mathbf{y}_{T+1}$ is not intrinsically unpredictable, but there are four reasons why $\mathbf{y}_{T+1}$ may not be easily predicted from time T using:

$$\widehat{\mathbf{y}}_{T+1|T} = \boldsymbol{\psi}_{T+1}(\mathbf{X}_T) \tag{6}$$

The first is that in practice, (6) is never available, so instead forecasters must use:

$$\widetilde{\mathbf{y}}_{T+1|T} = \widetilde{\boldsymbol{\psi}}_{T+1}(\mathbf{X}_T) \tag{7}$$

where $\widetilde{\boldsymbol{\psi}}_{T+1}(\cdot)$ is an estimate of what $\boldsymbol{\psi}_{T+1}(\cdot)$ will be. The second is that $\mathbf{X}_T$ may not be known at T, or may be incorrectly measured by a flash or nowcast estimate.

The next reason is instance unpredictability, which arises when the draw of $\boldsymbol{\epsilon}_{T+1}$ induces outcomes $\mathbf{y}_{T+1}$ that are far from the forecast $\widehat{\mathbf{y}}_{T+1|T}$ in the metric of $\boldsymbol{\Omega}_{\boldsymbol{\epsilon}}$, so that:

$$\widehat{\boldsymbol{\epsilon}}_{T+1|T} = \mathbf{y}_{T+1} - \widehat{\mathbf{y}}_{T+1|T}$$

is unexpectedly large (as in Taleb, 2009). That problem can occur even when $\boldsymbol{\psi}_{T+1}(\cdot)$ is known (or constant).

The fourth reason is that, even if $\mathbf{y}_{T+1}$ was predictable according to (4) and $\mathbf{X}_T$ was known, to use such predictability would require either knowledge of $\boldsymbol{\psi}_{T+1}(\cdot)$ in (6) or more realistically, an accurate value $\widetilde{\boldsymbol{\psi}}_{T+1}(\cdot)$. When (4) holds but:

$$\mathsf{D}_{\mathsf{y}_{T+1}}(\cdot) \neq \mathsf{D}_{\mathsf{y}_T}(\cdot) \tag{8}$$

so the distribution itself is not constant but shifts in unanticipated ways at unexpected time points, then from information $\mathcal{I}_T$ available at time T on $\mathsf{D}_{\mathsf{y}_T}(\cdot)$, $\mathbf{y}_{T+1}$ becomes extrinsically unpredictable, irrespective of (1) or (4) holding, a problem exacerbated by the non-constancy of the in-sample distribution making empirical modeling difficult. To successfully forecast from (7) not only entails accurate data on $\mathbf{X}_T$, but also requires both a 'normal' draw $\boldsymbol{\epsilon}_{T+1}$ (or forecasting any outliers) and that $\widetilde{\boldsymbol{\psi}}_{T+1}(\cdot)$ be close to $\boldsymbol{\psi}_{T+1}(\cdot)$ even though shifts occur, together essentially needing a crystal ball. A process is doubly unpredictable when it is both intrinsically and extrinsically unpredictable (and hence instance unpredictable), so the pre-existing unconditional distribution does not match that holding in the next period.[1] For example, $\boldsymbol{\epsilon}_t \sim \mathsf{IN}_n[\boldsymbol{\mu}_t, \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}]$ will be less predictable than expected from probabilities calculated using $\boldsymbol{\Omega}_{\boldsymbol{\epsilon}}$ when future changes in $\boldsymbol{\mu}_t$ cannot be determined in advance.

Due to shifts in the underlying distributions, all expectations operators must be three-way time dated, to denote the relevant random variables, the distributions being integrated over, and the available information at the time the expectations are formed, as in $\mathsf{E}_t[\boldsymbol{\epsilon}_{t+1}|\mathcal{I}_{t-1}]$, which denotes the conditional expectation of $\boldsymbol{\epsilon}_{t+1}$ formed at time t given the information set $\mathcal{I}_{t-1}$:

$$\mathsf{E}_t[\boldsymbol{\epsilon}_{t+1} \mid \mathcal{I}_{t-1}] = \int \boldsymbol{\epsilon}_{t+1} \mathsf{D}_{\boldsymbol{\epsilon}_t}(\boldsymbol{\epsilon}_{t+1}|\mathcal{I}_{t-1})\, \mathrm{d}\boldsymbol{\epsilon}_{t+1} \tag{9}$$

This allows for a random variable being unpredictable in mean or variance because its conditional distribution shifts in unanticipated ways relative to the conditioning information. When $\boldsymbol{\epsilon}_t \sim \mathsf{IN}_n[\boldsymbol{\mu}_t, \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}]$ is extrinsically unpredictable, because future changes in $\boldsymbol{\mu}_t$ cannot be established in advance, which is perhaps the most relevant state of nature for economics, from (9):

$$\mathsf{E}_t[\boldsymbol{\epsilon}_{t+1} \mid \mathcal{I}_{t-1}] = \int \boldsymbol{\epsilon}_{t+1} \mathsf{D}_{\boldsymbol{\epsilon}_t}(\boldsymbol{\epsilon}_{t+1}|\mathcal{I}_{t-1})\, \mathrm{d}\boldsymbol{\epsilon}_{t+1} = \int \boldsymbol{\epsilon}_{t+1} \mathsf{D}_{\boldsymbol{\epsilon}_t}(\boldsymbol{\epsilon}_{t+1})\, \mathrm{d}\boldsymbol{\epsilon}_{t+1} = \boldsymbol{\mu}_t \tag{10}$$

whereas:

$$\mathsf{E}_{t+1}[\boldsymbol{\epsilon}_{t+1} \mid \mathcal{I}_{t-1}] = \int \boldsymbol{\epsilon}_{t+1} \mathsf{D}_{\boldsymbol{\epsilon}_{t+1}}(\boldsymbol{\epsilon}_{t+1}|\mathcal{I}_{t-1})\, \mathrm{d}\boldsymbol{\epsilon}_{t+1} = \int \boldsymbol{\epsilon}_{t+1} \mathsf{D}_{\boldsymbol{\epsilon}_{t+1}}(\boldsymbol{\epsilon}_{t+1})\, \mathrm{d}\boldsymbol{\epsilon}_{t+1} = \boldsymbol{\mu}_{t+1} \tag{11}$$

so that $\mathsf{E}_t[\boldsymbol{\epsilon}_{t+1}|\mathcal{I}_{t-1}]$ does not correctly predict $\boldsymbol{\mu}_{t+1}$. Thus, the conditional expectation $\mathsf{E}_t[\boldsymbol{\epsilon}_{t+1}|\mathcal{I}_{t-1}]$ formed at t is not an unbiased predictor of the outcome $\boldsymbol{\mu}_{t+1}$ at $t+1$, although the crystal-ball predictor $\mathsf{E}_{t+1}[\boldsymbol{\epsilon}_{t+1}|\mathcal{I}_{t-1}]$ remains unbiased. Thus, some well established results on conditional expectations need no longer hold, as Hendry and Mizon (2010) show, an issue to which we return in section 7.

Nevertheless, there may exist additional information sets, denoted $\mathcal{L}_T$ and $\mathcal{K}_T$, which could respectively help predict the outliers or the shifts in (8), as discussed in section 5. Importantly, once a shift has happened, it may be explicable (at worst by indicator variables), so there is a potentially major difference between modeling and forecasting when (8) holds, an aspect addressed in section 8.

[1] Intrinsic and instance unpredictability are close to 'known unknowns' in that the probabilities of various outcomes can be correctly pre-calculated, as in rolling dice, whereas extrinsic unpredictability is more like 'unknown unknowns' in that the probabilities of outcomes cannot be accurately calculated in advance, a dichotomy for forecasting presaged in the first quote of Clements and Hendry (1998).

2.2 Prediction from a reduced information set

Given a subset of information, $\mathcal{J}_{t-1} \subset \mathcal{I}_{t-1}$ where $\mathcal{J}_{t-1} = \sigma[\mathbf{Z}_{t-1}]$ (say) when the data generating process is (3) (so instance and extrinsic effects are absent), predictions will remain unbiased, although less accurate. Since $\mathsf{E}_t[\boldsymbol{\epsilon}_t|\mathcal{I}_{t-1}] = \mathbf{0}$, so $\{\boldsymbol{\epsilon}_t\}$ is unpredictable given all the information, it must be unpredictable from a subset so that:

$$\mathsf{E}_t[\boldsymbol{\epsilon}_t \mid \mathcal{J}_{t-1}] = \mathbf{0} \tag{12}$$

and hence from (3):

$$\mathsf{E}_t[\mathbf{y}_t \mid \mathcal{J}_{t-1}] = \mathsf{E}_t[\boldsymbol{\psi}_t(\mathcal{I}_{t-1}) \mid \mathcal{J}_{t-1}] = \boldsymbol{\phi}_t(\mathbf{Z}_{t-1}) \tag{13}$$

Letting $\mathbf{e}_t = \mathbf{y}_t - \boldsymbol{\phi}_t(\mathbf{Z}_{t-1})$ be the unexplained component from (13), then:

$$\mathsf{E}_t[\mathbf{e}_t \mid \mathcal{J}_{t-1}] = \mathsf{E}_t[\mathbf{y}_t \mid \mathcal{J}_{t-1}] - \boldsymbol{\phi}_t(\mathbf{Z}_{t-1}) = \mathbf{0} \tag{14}$$

so $\mathbf{e}_t$ is a mean innovation with respect to $\mathcal{J}_{t-1}$.

However, since:

$$\mathbf{e}_t = \boldsymbol{\epsilon}_t + \boldsymbol{\psi}_t(\mathbf{X}_{t-1}) - \boldsymbol{\phi}_t(\mathbf{Z}_{t-1}) \tag{15}$$

then taking expectations with respect to the complete information set $\mathcal{I}_{t-1}$:

$$\mathsf{E}_t[\mathbf{e}_t \mid \mathcal{I}_{t-1}] = \boldsymbol{\psi}_t(\mathbf{X}_{t-1}) - \mathsf{E}_t[\boldsymbol{\phi}_t(\mathcal{J}_{t-1}) \mid \mathcal{I}_{t-1}] = \boldsymbol{\psi}_t(\mathbf{X}_{t-1}) - \boldsymbol{\phi}_t(\mathbf{Z}_{t-1}) \neq \mathbf{0} \tag{16}$$

Thus, $\mathbf{e}_t$ is not an innovation relative to $\mathcal{I}_{t-1}$ so from (15) and (16):

$$\mathsf{V}_t[\mathbf{e}_t] = \mathsf{V}_t[\boldsymbol{\epsilon}_t] + \mathsf{V}_t[\boldsymbol{\psi}_t(\mathbf{X}_{t-1}) - \boldsymbol{\phi}_t(\mathbf{Z}_{t-1})] \geq \mathsf{V}_t[\boldsymbol{\epsilon}_t] \tag{17}$$

so larger variance predictions will usually result, again consistent with the concept of predictability.

Thus, in the context of intrinsic unpredictability, more (relevant) information improves the accuracy of prediction, but less information by itself does not lead to biased outcomes. Although such a result can be proved as above (see e.g., Clements and Hendry, 2005), it conflicts with the intuition that a loss of information about what causes shifts can lead to seriously biased forecasts. The resolution of this apparent paradox lies in the assumed knowledge of what the conditioning symbol (|) entails in equations like (12), which hides a key difficulty: the mapping between random variables and information sets will never be known, so implicitly, (13) is a crystal-ball predictor, the absence of which is precisely the empirical problem. Neither of the functions $\boldsymbol{\psi}_t(\mathbf{X}_{t-1})$ nor $\boldsymbol{\phi}_t(\mathbf{Z}_{t-1})$ will ever be known, especially if not constant. Specifically, $\boldsymbol{\psi}(\mathbf{X}_{t-1})$ could be constant when $\boldsymbol{\phi}_t(\mathbf{Z}_{t-1})$ is not, so such an information reduction would have major costs in the loss of constancy when the relevant function is unknown, as will occur with extrinsic unpredictability.

3 Instance unpredictability

As Taleb (2007, 2009) has stressed, rare large-magnitude events, or 'Black Swans', do unexpectedly occur. One formulation of that insight is to interpret 'Black Swans' as highly discrepant draws from fat-tailed distributions, where there is a constant, but small, probability of such an occurrence each period. Both the timing and the magnitude of the discrepant events are then unpredictable, even when the form of the distribution is known and constant over time. When a large outcome materializes unexpectedly at time τ, say, substantial costs or benefits can result, sometimes both, but for different groups of agents. Barro (2009) estimates very high costs from 'consumption disasters', finding 84 events over approximately the last 150 years with falls of more than 10% per capita, cumulating to a total duration of almost 300 'bad' years across his sample of 21 countries, mainly due to wars. However, there is a marked reduction in their frequency after World War II.

Recent research on many financial variables has revealed a proliferation of 'jumps', as measured by bipower variation (see e.g., Barndorff-Nielsen and Shephard, 2004, 2006). These seem to be *ex ante* instance unpredictable events, superimposed on the underlying Ornstein–Uhlenbeck processes. To mitigate possibly large costs from financial outliers, Taleb (2009) argues for more robust systems that avoid dependence on the non-occurrence of very large draws: for example, systems with more nodes and less interdependence.

Empirically, the distribution of future events is not likely to be known, with tail properties especially difficult to estimate from available data, and distributions may also shift over time. Some of the considerations discussed below for forecasting breaks also apply to forecasting 'Black Swans', although if they were indeed genuinely independent large draws from fat-tailed distributions, no information would make their timing possible so success in doing so is unlikely, especially when the underlying distributions are neither known nor constant over time.

A key difference between instance unpredictability and extrinsic unpredictability arises under independent sampling. In the former, a 'Black Swan' relative to the usual outcomes is unlikely to also occur on the next draw, and even less likely in several successive outcomes: flocks of 'Black Swans' are improbable.[2] However, when the mean of a distribution changes, as in some of the examples cited below, as well as in Barro (2009), successive outcomes are likely to be around the new mean, so a cluster appears.

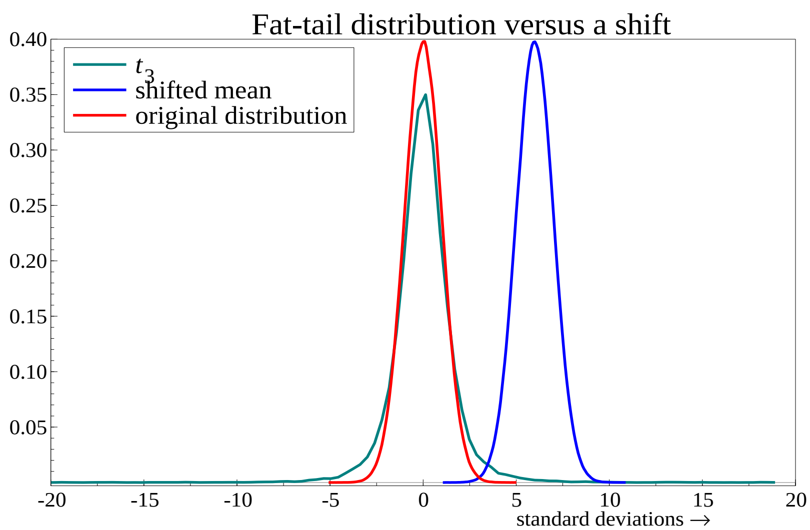


Figure 1 illustrates this result, contrasting a t_3 with a normal distribution and also with a mean-shifted normal distribution. There are potentially extreme draws from the t_3, but they should occur rarely, and would be equally likely in either tail, although that would not be true for, say, a log-normal or some of the extreme-value distributions. Conversely, while the first outcome after a mean shift would initially appear to be a 'Black Swan', even with independent draws it would be followed by many more outcomes that were discrepant relative to the original distribution, but not relative to its mean-shifted replacement.

A 'Black Swan' is often viewed as a large deviation in a differenced variable, as in asset market returns or price changes. If not rapidly reversed, such a jump, or collapse, entails a location shift in the corresponding level. Thus, instance unpredictability in the differences of variables entails extrinsic unpredictability in the levels (and *vice versa*), the topic to which we now turn.

[2]This only applies to draws from a given distribution. As Gillian Tett, *Financial Times Magazine*, March 26/27, p54, remarks 'Indeed, black swans have suddenly arrived in such a flock...', as can happen when looking across many distributions of economic, political, natural and environmental phenomena.

4 Extrinsic unpredictability

The world has seen many major shifts since 1400 (or any recent epoch longer than 50 years), in demography (average age of death changing from around 40 to around 80, with the average number of births per female falling dramatically), real variables such as incomes per capita (increasing 6-8 fold), and all nominal variables (some 1000 fold). Current outcomes in the Western world today are not highly discrepant draws from the distributions relevant in the Middle Ages, but 'normal' draws from distributions with very different means. Such shifts can be precipitated by changes in legislation, science and technology, medicine, financial innovation, climate, and political and economic regimes among other sources. While these examples are of major shifts over long periods, the recent financial crisis, and the many similar examples in the last quarter of the 20th Century, demonstrate that sudden large shifts still occur (see the cases documented by Barrell, Dury, Holland, Pain and te Velde, 1998, and Barrell, 2001, *inter alia*). In asset markets, endogenous changes in agents' behavior can alter the underlying 'reality', as argued by Soros (2008, 2010) in his concept of reflexivity, inducing feedbacks that can lead to dramatic changes as agents' views move together, changing the system, which thereby ends in a different state.

Moreover, the distributions of the differences of many economic variables have also changed radically, and are also not stationary. For example, real growth per capita was a fraction of one percent per annum till the Industrial Revolution (see Apostolides, Broadberry, Campbell, Overton and van Leeuwen, 2008, for the evidence, and Allen, 2009, for an insightful analysis of the Industrial Revolution), remained low for the next couple of hundred years, but is now around 2% pa in OECD countries, and much higher in emerging economies borrowing modern Western technology.

Shifts of distributions remain common, and would be unproblematic *per se* if they could be modeled and predicted. Co-breaking, where location shifts cancel, would enable some aspects to be forecast even when breaks themselves could not (see Hendry and Massmann, 2007), analogously to cointegration reducing stochastic trends ($\mathsf{I}(1)$ variables) to stationarity. Differencing plays a similar role, removing unit roots and converting location shifts to impulses. Many of the examples of demographic shifts noted earlier have such a property: annual changes in average age of death in OECD countries have been remarkably constant at about a weekend a week since around 1860, other than major shifts during the First World War and the ensuing flu' epidemic. The changes in many economic distributions also shift, especially those of nominal variables, but the change in the change may be more constant as few variables permanently accelerate (think 'inflation surprises', where $\Delta^2 p_t$, say, is sometimes treated as intrinsically unpredictable).

The key feature of extrinsic unpredictability is that the distributional shift is unanticipated, even if the variable is otherwise partly predictable:

$$\mathsf{D}_{\mathsf{y}_{t+1}}\left(\mathbf{y}_{t+1} \mid \cdot\right) \neq \mathsf{D}_{\mathsf{y}_t}\left(\mathbf{y}_t \mid \cdot\right) \quad \forall t \in \mathcal{T} \tag{18}$$

where that change was not expected. From the viewpoint of forecasting, location shifts (changes in the means of distributions) are the most pernicious as they induce systematic forecast failure, so will be the focus below, but in other settings, unanticipated shifts in unconditional variances can also be problematic (see e.g., Clements and Hendry, 2006, for a taxonomy of forecast errors in GARCH processes).

As discussed in section 2, when distributions shift, care is required in defining expectations to reflect the probability distribution integrated over, $\mathsf{D}_{\mathsf{y}_{t+1}}(\mathbf{y}_{t+1}|\cdot)$, the timing of the expectation (e.g., t), and the timing and contents of the available information that is conditioned on (e.g., $\mathcal{J}_{t-1}$), such that, for a scalar:

$$\mathsf{E}_t\left[y_{t+1} \mid \mathcal{J}_{t-1}\right] = \int y_{t+1}\mathsf{D}_{\mathsf{y}_t}\left(y_{t+1}|\mathcal{J}_{t-1}\right)\mathrm{d}y_{t+1} \tag{19}$$

When $\mathsf{D}_{\mathsf{y}_t}(y_t|\cdot) \neq \mathsf{D}_{\mathsf{y}_{t+1}}(y_{t+1}|\cdot)$ one period later:

$$\mathsf{E}_{t+1}[y_{t+1} \mid \mathcal{J}_{t-1}] = \int y_{t+1} \mathsf{D}_{\mathsf{y}_{t+1}}(y_{t+1}|\mathcal{J}_{t-1})\,\mathrm{d}y_{t+1} \tag{20}$$

so the mean calculated from (20) will be unbiased for $\mathsf{E}_{t+1}[y_{t+1}]$, but that from (19) will not be. Unfortunately, forecasting y_{t+1} from time t by (20) would require a crystal ball to know the complete future distribution $\mathsf{D}_{\mathsf{y}_{t+1}}(y_{t+1}|\cdot)$. Thus, only (19) is available to economic agents or forecasters, and after a location shift, the resulting bias can cause forecast failure, considered in section 6. First, we consider the possibility that there may be other sources of information to account for shifts, then address the impact of shifts on econometric modeling, before turning to forecasting and economic analyses.

5 Three information sets

If the universal information set $\mathcal{I}_{t-1}$ does not enable predictability, then the random variable in question is intrinsically unpredictable. That result applies equally to modeling and forecasting: nothing useful can be said beyond the unconditional distribution. Few observable economic variables are intrinsically unpredictable, although modelers often seek an error that is, an issue addressed in section 8.1. In practice, investigators use available information $\mathcal{J}_{t-1} \subset \mathcal{I}_{t-1}$ that facilitates some predictability. However, even if a variable $\mathbf{y}_t$ is not intrinsically unpredictable, neither the conditional nor unconditional distribution need be constant, so despite $\mathsf{D}_{\mathsf{y}_t}(\mathbf{y}_t|\mathcal{J}_{t-1}) \neq \mathsf{D}_{\mathsf{y}_t}(\mathbf{y}_t)$, $\mathbf{y}_{T+1}$ may still be essentially unpredictable at T because of instance or extrinsic unpredictability. To overcome that problem, we consider two additional information sets, denoted $\mathcal{L}$ and $\mathcal{K}$ above, that are not subsets of $\mathcal{J}$.

The possible information set $\mathcal{L}$ is one that would help predict the timing of a bad draw from a known fat-tailed distribution. The considerations involved seem close to those for predicting shifts of distributions discussed below, although instance unpredictability is a known unknown when the *ex ante* probability can be calculated, whereas extrinsic unpredictability is an unknown unknown relative to the information set in use: see Castle *et al.* (2011) on the necessary conditions for predicting location shifts. Empirically, however, instance unpredictability is usually an unknown unknown as well, because the relevant distribution is not known in advance, although characteristics, such as its having a fat tail, may be.

The third possible source of information, $\mathcal{K}$, is one that might help reduce extrinsic unpredictability. For simplicity, we assume that $\mathcal{L}_{t-1} \subset \mathcal{K}_{t-1}$ since timing predictability is needed to predict location shifts. Then, let the universal information set determining the data generating process (DGP) of $\{\mathbf{y}_t\}$ be $\mathcal{I}_{t-1} = (\mathcal{J}_{t-1}, \mathcal{K}_{t-1}, \mathcal{M}_{t-1})$ where $\mathcal{J}_{t-1} = \sigma[\mathbf{Z}_{t-1}]$ denotes known 'standard economic forces' (e.g., for money demand, these would be incomes, prices, interest rates and lags thereof as in, e.g., Hendry and Ericsson, 1991), $\mathcal{K}_{t-\delta} = \sigma[\mathbf{W}_{t-\delta}]$ (say) which if known would explain some of the changes from $\mathsf{D}_{\mathsf{y}_{t-1}}(\cdot)$ to $\mathsf{D}_{\mathsf{y}_t}(\cdot)$ where $0 \leq \delta \leq 1$, and $\mathcal{M}_{t-1}$ is a relevant source of information but is completely unknown.

Write the DGP of $\mathbf{y}_t$ as:

$$\mathbf{y}_t = \mathbf{\Gamma}\mathbf{X}^v_{t-1} + \boldsymbol{\epsilon}_t \tag{21}$$

where ^{v} denotes column vectoring. Using only $\mathcal{J}_{t-1}$ entails marginalizing $\mathcal{I}_{t-1}$ with respect to $\mathcal{M}_{t-1}$ and $\mathcal{K}_{t-1}$, so $\mathcal{J}_{t-1}$ no longer fully characterizes the DGP of $\{\mathbf{y}_t\}$, only its local data generating process (LDGP):

$$\mathbf{y}_t = \mathbf{\Psi}_t\mathbf{Z}^v_{t-1} + \mathbf{e}_t \tag{22}$$

When the resulting parameter $\mathbf{\Psi}_t$ is non-constant, but $\mathbf{\Gamma}$ is constant, then the information in $\mathcal{K}$ and $\mathcal{M}$ must 'explain' the breaks in (22). Shifts then alter the 'regular relationship' (e.g., in inflation relative to its usual determinants, as occurs when wars start: see Hendry, 2001). When $(\mathcal{J}_{t-1}, \mathcal{K}_{t-\delta})$ are both known, this leads to a different LDGP, marginalized with respect to $\mathcal{M}_{t-1}$ only. When $\mathcal{K}_{t-\delta}$ is indeed the information set that induces a constant LDGP, it must do so by accounting for shifts in $\mathbf{\Psi}_t$, so assuming the simplest linear setting to illustrate:

$$\mathbf{y}_t = \mathbf{\Phi}_1 \mathbf{Z}_{t-1}^{v} + \mathbf{\Phi}_2 \mathbf{W}_{t-\delta}^{v} + \mathbf{v}_t \tag{23}$$

Here, $\mathbf{W}_{t-\delta}^{v}$ is dated almost contemporaneously with $\mathbf{\Psi}_t$ (and therefore $\mathbf{y}_t$) to produce a constant relationship, so an investigator may still need to forecast shifts in $\mathbf{W}_{t-\delta}$ at time t (hence very high frequency information may help). Thus, $\mathbf{W}_{t-1}$ may not help in forecasting even if $\mathbf{W}_t$ would capture the shift in the relationship between $t-1$ and t due to changes in (*inter alia*) legislation, financial innovation, technology, and policy regimes. Castle *et al.* (2011) discuss a variety of information sets that may improve the predictability of shifts, including leading indicators and survey data, disaggregation over time and variables, *Google Trends* (see Choi and Varian, 2009) and prediction markets data (like the Iowa electronic market, http://www.biz.uiowa.edu/iem/index.cfm), although none need help in any given instance.[3]

When $\mathcal{M}$ accounts for the distributional shifts, such that knowing $\mathcal{J}$ and $\mathcal{K}$ does not lead to a constant LDGP, the only viable modeling approach seems to be to remove past location shifts in-sample, although that still leaves forecasting hazardous. To induce (22) from the LDGP in (23) requires a relation like:

$$\mathbf{W}_t^{v} = \mathbf{\Pi}_t \mathbf{Z}_{t-1}^{v} + \boldsymbol{\eta}_t \tag{24}$$

To paraphrase Cartwright (1989), 'Breaks out (as in (22)) need breaks in (as in (24))': see Hendry (1988) for a derivation. Note that (24) is a projection, so does not entail an ability to forecast $\mathbf{W}_t$ from $\mathbf{Z}_{t-1}$ because $\mathbf{\Pi}_t$ is unknown (and perhaps unknowable) at $t-1$.

6 Instance and extrinsic unpredictability in forecasting

As noted above, no information would make the timing and magnitude of genuinely independent large draws predictable. Once an outlier or a shift has occurred, however, even if the 'causes' are unknown, appropriate indicator variables can be entered in a model to remove the problem. If the causes are known, the relevant variable(s) can be added. Nevertheless, knowing and adding such variables need not help in forecasting, unless one can also forecast future shifts in them. For example, oil price changes can have an important impact on inflation, and are often significant in empirical models of price inflation, but remain at least as difficult to forecast as the inflation change itself. Thus, both instance and extrinsic unpredictability may be susceptible to *ex post* modeling, yet not improve *ex ante* forecasting when future shifts cannot be accurately forecast.

When $\boldsymbol{\psi}_{T+1}(\cdot)$ in (6) changes with $\mathcal{K}_{T+1}$, and thereby induces location shifts, we can write that dependence as:

$$\mathsf{E}_{T+1}\left[\mathbf{y}_{T+1} \mid \mathcal{J}_T, \mathcal{K}_{T+1}\right] = \boldsymbol{\psi}_0\left(\mathbf{Z}_T\right) + \boldsymbol{\psi}_1\left(\mathbf{W}_{T+1}\right) \tag{25}$$

where $\boldsymbol{\psi}_0(\cdot)$ is constant. The second term is generally zero, and accounts for shifts by a step function or non-linear response. At $T+1-\delta$, however, an investigator at best knows ($\mathbf{Z}_T$, $\mathbf{W}_{T+1-\delta}$), in which case:

$$\mathsf{E}_{T+1-\delta}\left[\mathbf{y}_{T+1} \mid \mathcal{J}_T, \mathcal{K}_{T+1-\delta}\right] = \boldsymbol{\psi}_0\left(\mathbf{Z}_T\right) + \mathsf{E}_{T+1-\delta}\left[\boldsymbol{\psi}_1\left(\mathcal{K}_{T+1}\right) \mid \mathcal{K}_{T+1-\delta}\right] \tag{26}$$

If a location shift is unpredictable, so $\mathsf{E}_{T+1-\delta}\left[\boldsymbol{\psi}_1\left(\mathcal{K}_{T+1}\right) \middle| \mathcal{K}_{T+1-\delta}\right] = \mathbf{0}$, there will be no perceptible difference from the information set that enters $\boldsymbol{\psi}_{T+1}\left(\mathcal{J}_T\right)$. The aim, therefore, must be to ascertain

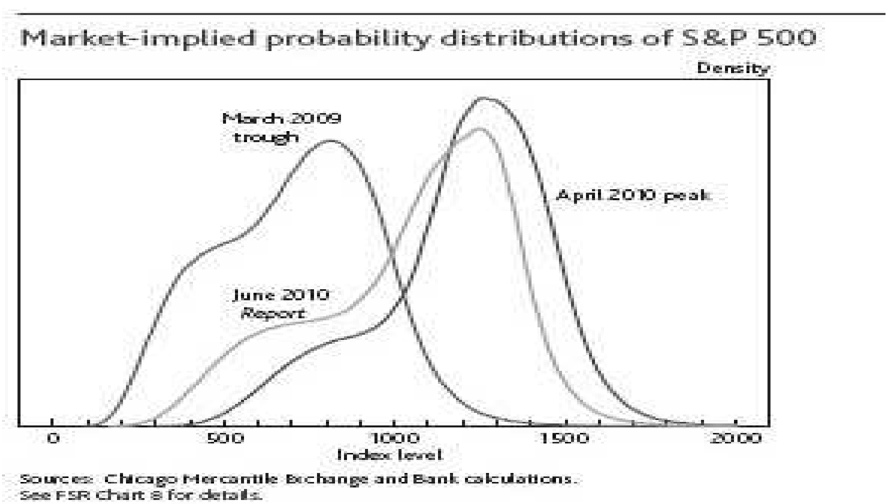
[3] They discuss possible information about the 2004 Indian Ocean tsunami as an example. The recent disastrous Japanese tsunami highlights the unpredictability of the causal earthquake and its impact on nearby land, but the later predictability of the resulting tsunami for other Pacific Ocean inhabitants.

available information $\mathcal{K}_{T+1-\delta}$ such that $\mathsf{E}_{T+1-\delta}\left[\psi_1\left(\mathcal{K}_{T+1}\right)|\mathcal{K}_{T+1-\delta}\right] \simeq \psi_1\left(\mathbf{W}_{T+1}\right)$, a daunting, but not impossible task, dependent on the specifics of the problem, although it is likely that some outliers and breaks will not be anticipated, so forecasts that are robust after breaks will often be needed (Castle, Fawcett and Hendry, 2010, offer some suggestions).

7 Extrinsic unpredictability and economic analyses

An apparent 'break' can be anything from a large measurement error that will be corrected at a future date, or an independent fat-tail draw, through to a location shift in either a level or a difference, or trend shifts. From just one or two observations following any unexpected draw, it can be almost impossible to discriminate its 'type' or future progress, and hence it is often not known what action to take (compare Granger and Pesaran, 2000a, and Pesaran and Skouras, 2002). As Castle *et al.* (2011) show, even when the break form and the date of its occurrence are known, the uncertainty in estimating the parameters of any break function from a few data points can vitiate the advantages of that knowledge, and hence not dominate a robust device. Thus, the key difficulty is determining how persistent any 'shift' will be, and that is a fundamental problem *ex ante* for both econometricians and economic agents.

It seems unlikely that economic agents are any more successful than professional economists in foreseeing when breaks will occur, or divining their type from one or two observations after they have happened. That link with forecast failure has important implications for economic theories about agents' expectations formation in a world with extrinsic unpredictability. General equilibrium theories rely heavily on *ceteris paribus* assumptions, especially that equilibria do not shift unexpectedly. As discussed in section 2, the mean of an unconditional distribution is the minimum mean-square error predictor in an intrinsically unpredictable process (more generally, the conditional expectation) only when the distribution is constant or a crystal ball is available, and fails under extrinsic unpredictability when it can no longer be established that the conditional expectation today of tomorrow's outcome is a useful predictor (see Hendry and Mizon, 2010). Moreover, those authors show that the law of iterated expectations does not apply inter-temporally when the distributions that enter the formulation change over time. Thus, it would not be rational to use the conditional expectation under extrinsic unpredictability, which may lead agents to use forecasting devices that are robust after location shifts.



The derivation deducing a martingale difference sequence (MDS) from 'no arbitrage' in, e.g., Jensen and Nielsen (1996), also explicitly requires no shifts in the underlying probability distributions to obtain a MDS, which once assumed allows deducing the intrinsic unpredictability of equity price changes and hence market (informational) efficiency. Unanticipated shifts also imply unpredictability, but need not entail efficiency. The unanticipated, and possibly even unpredictable, explosion of the oxygen cylinder on Apollo 13, which wrecked its space mission, was highly inefficient, and is an example of the invalidity of inferring informational efficiency from unpredictability *per se*, when the source is extrinsic rather than intrinsic. That distributional shifts occur even in financial markets is illustrated by the market-implied probability distributions of the S&P500 from the Bank of England *Financial Stability Report*, June 2010 shown above.

In other arenas, location shifts can play a positive role in clarifying both causality, as demonstrated in White and Kennedy (2009), and super exogeneity (see Hendry and Santos, 2010). Also White (2006) considers estimating the effects of natural experiments, many of which involve large location shifts. Thus, while more general theories of the behavior of economic agents and their methods of expectations formation are required under extrinsic unpredictability, and forecasting becomes prone to failure, large shifts could also help reveal the linkages between variables.

8 Unpredictability and model selection

All three forms of unpredictability impinge on econometric modeling and model selection. A process that is intrinsically unpredictable cannot be modeled better than its unconditional distribution. For economic data, the only relevant situation where that might apply is to the error processes in models, and we consider that aspect in §8.1.

Ex post, instance or extrinsic unpredictability in a process can be modeled, because once a shift has occurred, it can be taken into account, just as once there exists a relationship, or an inter-temporal link, that can be modeled. Although unanticipated shifts may be highly detrimental to both economic analysis and forecasting, they need not impugn empirical modeling or model selection when correctly handled. Shifts that are not modeled can be disastrous for estimation and inference, but impulse-indicator saturation (IIS: see Hendry *et al.*, 2008, and Johansen and Nielsen, 2009) can handle multiple location shifts as well as 'remove' most of the outliers from a fat-tailed distribution. As implemented in automatic model selection algorithms like *Autometrics* (see Doornik, 2009a, 2009b, and Hendry and Doornik, 2009), IIS also enables jointly locating breaks with selection over variables, functional forms and lags: see Castle, Doornik and Hendry (2011b).

Care is required in interpreting empirical models with substantively important indicator variables, when the shifts they represent could recur. Future draws are then likely to include either outliers or shifts, so measures of uncertainty need to reflect that possibility, an issue emphasized by Hendry (2001) and Pesaran, Pettenuzzo and Timmermann (2006). On the other hand, Castle *et al.* (2011) show the value of IIS in avoiding unmodeled outliers contaminating the selection of non-linear functions or threshold non-linear models in the context of forecasting shifts and during shifts.

8.1 Congruent modeling for forecasting

The result in (17) underpins general-to-specific (Gets) model selection and the related use of congruence and encompassing as a basis for econometric model selection (see e.g., Hendry and Doornik, 2009). In terms of Gets, less is learned from $\mathcal{J}_{t-1}$ than $\mathcal{I}_{t-1}$, and the variance (when it exists) of the unpredictable component is larger. In terms of encompassing (see Bontemps and Mizon, 2003, 2008, for recent results), a later investigator may discover additional information in $\mathcal{I}_{t-1}$ beyond $\mathcal{J}_{t-1}$ which explains part of a previously unpredictable error. Nevertheless, Allen and Fildes (2001) among others, find no relationship between congruence, as exemplified by rigorous mis-specification testing, and subsequent forecasting performance. Since systematic forecast failure is primarily due to extrinsically unpredictable shifts in the future distribution, such a finding makes sense, but raises the question: what is the role for congruent models in forecasting? There are four potential advantages of using a congruent encompassing model for forecasting despite the omnipresent possibility of unanticipated breaks.

First, such models deliver the smallest variance for the innovation error, a key determinant of forecast-error variances, especially important when breaks do not occur over the forecast horizon. However, in-sample dominance does not ensure the avoidance of forecast failure, and Clements and Hendry (1998) establish the robustness after location shifts of double-differenced forecasting devices, such as $\Delta\widehat{x}_{T+1|T} = \Delta x_T$, which can outperform despite their non-congruence and non-encompassing. Clements and Hendry (2001) use that analysis to explain the outcomes of forecasting competitions, where the simplicity of a model is viewed as essential for success (see e.g., Makridakis and Hibon, 2000), but argue that is due to confounding parsimony with robustness, as such competitions did not include non-parsimonious but robust models. How a congruent encompassing model is used in the forecast period also matters, and Castle *et al.* (2010) illustrate the forecasting advantages of transforming an in-sample model to be robust after breaks, and show that then a non-parsimonious model can outperform $\Delta\widehat{x}_{T+1|T} = \Delta x_T$ after a break at time $T-1$.

Secondly, modeling mean shifts in-sample removes one of the sources of systematic mis-forecasting. All taxonomies of forecast errors show that the equilibrium mean plays a key role therein (see e.g., Clements and Hendry, 1998): shifts, mis-estimation, or mis-specification of the equilibrium mean all induce non-zero mean forecast errors. Shifts can only be avoided by either forecasting their arrival, or using a formulation that is robust after breaks, as just discussed; mis-estimation should be small in a constant well-specified model, where the free intercept represents the mean of the dependent variable; and removing in-sample location shifts and outliers reduces that source of mis-specification.

Thirdly, eliminating from a selected model irrelevant variables that might break in the forecast period avoids a further source of forecast failure. Here, model selection at tight significance levels can achieve that aim jointly with selecting relevant variables, non-linear functions and lags thereof, and removing in-sample location shifts and outliers by IIS: see e.g., Castle *et al.* (2011a). Moreover, the bias corrections after selection noted by those authors also reduce the impact of any adventitiously significant irrelevant variables which might shift.

Fourthly, in congruent models, decisions during model selection can be based on conventional inference procedures, such as t-tests. However, when unmodeled breaks occur, heteroskedastic-consistent standard errors (HCSEs), and autocorrelation-consistent (HACSEs) generalizations thereof (see e.g., White, 1980, and Andrews, 1991) appear to be needed. Because of the unmodeled shifts, residuals will exhibit heteroskedasticity and autocorrelation, even though those features are not present in the errors, so HCSEs and HACSEs will incorrectly attribute the underlying uncertainty to the problems for which they were respectively designed.

A further important issue is the ratio of the largest eigenvalue of the data second-moment matrix (say λ_1) to the smallest (λ_n), albeit that is not just a problem for congruent models. Castle *et al.* (2010) show that location shifts reduce the collinearities between variables, having the greatest impact on the smallest eigenvalues, and since mean square forecast errors (MSFEs) depend most on λ_1/λ_n, changes in collinearity after a break adversely increase forecast uncertainty. This effect cannot be avoided by deleting the collinear variables, nor is the problem mitigated by orthogonalizing the variables, which can transform an external break (one affecting marginal processes) to an internal one (shifting the conditional model of interest). However, eliminating low-significance variables by model selection, and rapidly updating after a shift, especially by a relatively short moving window, can both help alleviate this problem (see e.g., Phillips, 1995, and Pesaran and Timmermann, 2007, respectively).

9 Conclusion

The three distinctions within unpredictability of intrinsic, instance and extrinsic, have different implications for economic analyses, econometric modeling, and forecasting. The first entails that conditioning information does not alter uncertainty, so that the unconditional distribution is the best basis for all three activities, which are therefore on an equal footing of 'uninformativeness'. The second adds the possibility that even when the distributional form is known, either unconditionally or conditionally (so otherwise there would be some predictability), an outcome can be highly discrepant at an unexpected time, and thereby impose substantial costs. Such outliers do not persist, but are usually in dfferenced variables (like rates of return) so can permanently change a level. The third concerns unanticipated shifts in distributions, which do persist, as with location shifts.

After they have happened, the effects from the second and third can be modeled (at a minimum by indicator variables), but if neither can be predicted on the available information, forecasts will be hazardous. Given the nature of innovation and economic behavior, it seeems highly unlikely that a meta-distribution of breaks can be formulated once and for all, precluding parametrizing all possible breaks or relying on intrinsic unprecitability alone: there are too many future unknown unknowns.

One implication is that economic agents may adopt robust methods to avoid systematic forecast failure following shifts. Although sources of information may exist that could help with predicting outliers or shifts, previous research has shown that it may prove difficult to benefit from these in practice. Consequently, methods which can help ascertain the likely persistence of breaks after they have occurred seem to offer the best prospect for mitigating systematic forecast failure for both economists and economic agents.

References

Allen, P. G., and Fildes, R. A. (2001). Econometric forecasting strategies and techniques. In Armstrong, J. S. (ed.), *Principles of Forecasting*, pp. 303–362. Boston: Kluwer Academic Publishers.

Allen, R. C. (2009). *The British Industrial Revolution in Global Perspective*. Cambridge: Cambridge University Press.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59**, 817–858.

Apostolides, A., Broadberry, S., Campbell, B., Overton, M., and van Leeuwen, B. (2008). English Gross Domestic Product, 1300–1700: Some preliminary estimates. Discussion paper, University of Warwick, Coventry.

Barndorff-Nielsen, O. E., and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of financial Econometrics*, **2**, 1–37.

Barndorff-Nielsen, O. E., and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics*, **4**, 1–30.

Barrell, R. (2001). Forecasting the world economy. In Hendry, D. F., and Ericsson, N. R.(eds.), *Understanding Economic Forecasts*, pp. 149–169. Cambridge, Mass.: MIT Press.

Barrell, R., Dury, K., Holland, D., Pain, N., and te Velde, D. (1998). Financial market contagion and the effects of the crises in East Asia, Russia and Latin America. *National Institute Economic Review*, **166**, 57–73.

Barro, R. J. (2009). Rare disasters, asset prices and welfare costs. *American Economic Review*, **99**, 243–264.

Bontemps, C., and Mizon, G. E. (2003). Congruence and encompassing. In Stigum, B. P. (ed.), *Econometrics and the Philosophy of Economics*, pp. 354–378. Princeton: Princeton University Press.

Bontemps, C., and Mizon, G. E. (2008). Encompassing: Concepts and implementation. *Oxford Bulletin of Economics and Statistics*, **70**, 721–750.

Cartwright, N. (1989). *Nature's Capacities and their Measurement*. Oxford: Clarendon Press.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011a). Evaluating automatic model selection. *Journal of Time Series Econometrics*, **3 (1)**, DOI: 10.2202/1941–1928.1097.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011b). Model selection when there are multiple breaks. *Journal of Econometrics*, forthcoming.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2010). Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, **158**, 25–36.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2011). Forecasting breaks and during breaks. In Clements, and Hendry (2011), Ch. 11.

Castle, J. L., and Shephard, N.(eds.)(2009). *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.

Choi, H., and Varian, H. (2009). Predicting the present with Google Trends. Unpublished paper, Economics Research Group, Google.

Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Clements, M. P., and Hendry, D. F. (2001). Explaining the results of the M3 forecasting competition. *International Journal of Forecasting*, **17**, 550–554.

Clements, M. P., and Hendry, D. F. (2005). Guest editors' introduction: Information in economic forecasting. *Oxford Bulletin of Economics and Statistics*, **67**, 713–753.

Clements, M. P., and Hendry, D. F. (2006). Forecasting with breaks. In Elliott, G., Granger, C. W. J., and Timmermann, A.(eds.), *Handbook of Econometrics on Forecasting*, pp. 605–657. Amsterdam: Elsevier.

Clements, M. P., and Hendry, D. F.(eds.)(2011). *Oxford Handbook of Economic Forecasting*. Oxford: Oxford University Press.

Doornik, J. A. (2009a). Autometrics. In Castle, and Shephard (2009), pp. 88–121.

Doornik, J. A. (2009b). Econometric model selection with more variables than observations. Working paper, Economics Department, University of Oxford.

Granger, C. W. J., and Pesaran, M. H. (2000a). A decision-theoretic approach to forecast evaluation. In Chon, W. S., Li, W. K., and Tong, H.(eds.), *Statistics and Finance: An Interface*, pp. 261–278. London: Imperial College Press.

Granger, C. W. J., and Pesaran, M. H. (2000b). Economic and statistical measures of forecasting accuracy. *Journal of Forecasting*, **19**, 537–560.

Hendry, D. F. (1988). The encompassing implications of feedback versus feedforward mechanisms in econometrics. *Oxford Economic Papers*, **40**, 132–149.

Hendry, D. F. (2001). Modelling UK inflation, 1875–1991. *Journal of Applied Econometrics*, **16**, 255–275.

Hendry, D. F., and Doornik, J. A. (2009). *Empirical Econometric Modelling using PcGive: Volume I*. London: Timberlake Consultants Press.

Hendry, D. F., and Ericsson, N. R. (1991). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, **35**, 833–886.

Hendry, D. F., Johansen, S., and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, **33**, 317–335. Erratum, 337–339.

Hendry, D. F., and Massmann, M. (2007). Co-breaking: Recent advances and a synopsis of the literature. *Journal of Business and Economic Statistics*, **25**, 33–51.

Hendry, D. F., and Mizon, G. E. (2010). On the mathematical basis of inter-temporal optimization. Discussion paper 497, Economics Department, Oxford.

Hendry, D. F., and Santos, C. (2010). An automatic test of super exogeneity. In Watson, M. W., Bollerslev, T., and Russell, J.(eds.), *Volatility and Time Series Econometrics*, pp. 164–193. Oxford: Oxford University Press.

Jensen, B. A., and Nielsen, J. A. (1996). Pricing by no arbitrage. In Cox, D. R., Hinkley, D. V., and Barndorff-Nielsen, O. E.(eds.), *Time Series Models: In Econometrics, Finance and other Fields*, pp. 179–225. London: Chapman and Hall.

Johansen, S., and Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, and Shephard (2009), pp. 1–36.

Leitch, G., and Tanner, J. E. (1991). Economic forecast evaluation: Profits versus the conventional error measures. *American Economic Review*, **81**, 580–590.

Makridakis, S., and Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, **16**, 451–476.

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, **73**, 1057–1084.

Pesaran, M. H., and Skouras, S. (2002). Decision-based methods for forecast evaluation. In Clements, M. P., and Hendry, D. F.(eds.), *A Companion to Economic Forecasting*, pp. 241–267. Oxford: Blackwells.

Pesaran, M. H., and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, **137**, 134–161.

Phillips, P. C. B. (1995). Automated forecasts of Asia-Pacific economic activity. *Asia-Pacific Economic Review*, **1**, 92–102.

Soros, G. (2008). *The New Paradigm for Financial Markets*. London: Perseus Books.

Soros, G. (2010). *The Soros Lectures*. Philadelphia: Perseus Books.

Taleb, N. N. (2007). *The Black Swan*. New York: Random House.

Taleb, N. N. (2009). Errors, robustness, and the fourth quadrant. *International Journal of Forecasting*, **25**, 744–759.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

White, H. (2006). Time series estimation of the effects of natural experiments. *Journal of Econometrics*, **135**, 527–566.

White, H., and Kennedy, P. (2009). Retrospective estimation of causal effects through time. In Castle, and Shephard (2009), pp. 59–87.