Expectations and Economic Policy in the Presence of Unanticipated Changes

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

Grayham E. Mizon
University of Southampton,
and Institute for New Economic Thinkling, Oxford Martin School, University of Oxford, UK. *

January 2012

Abstract

In many areas of life - economics, meterology, sociology, politics to name a few - there are events that surprise us. However, much behaviour is dependent on expectations of future events. We show that it cannot be proved that conditional expectations based on the current distribution are minimum mean-square error 1-step ahead predictors when unanticipated breaks occur, and consequentially, the law of iterated expectations and Bellman's optimality principle then fail inter-temporally. The former makes the formulation of models for forecasting and economic policy precarious, while the latter two cause problems for models of inter-temporal optimizing behaviour.

1 Introduction

A difficulty faced by many disciplines that are concerned with future behaviour is not simply that uncertainties occur, but also that there are often changes to the underlying relationships that are unanticipated. Such changes, or more precisely structural breaks, not only lead to difficulties in forecasting (see Clements and Hendry, 2001), but also in the formulation of theories and models of the underlying behavioural relationships. The latter is not simply a matter of modeling in the face of structural breaks, but confronts a deeper problem. The mathematical derivations involved in the implementation of theories in empirical models fail to recognize that when there are unanticipated changes, conditional expectations are neither unbiased nor minimum mean-squared error (MMSE) predictors, and that better predictors can be provided by robust devices. As a consequence, the law of iterated expectations then does not hold as an inter-temporal relation unless all distributional shifts are perfectly anticipated. Further, the Bellman optimality principle when applied to stochastic variables that are subject to unanticipated changes no longer holds. Given the prevalence of such changes, learning about the post-change scenario is both difficult, and itself generates further non-stationarities.

The paper is organized as follows. Section 2 notes the important role that expectations play in many areas of decision making and forecasting, including economic policy. The relationship between the

*We are grateful to Jennifer L. Castle, Gunnar Bärdsen, Søren Johansen, Bent Nielsen and Ragnar Nymoen for their valuable comments. We have also benefitted from the comments of two referees on a previous version of the paper.

process that generates the data we observe (DGP) and models thereof is discussed briefly in section 3, as well as the effects of unanticipated changes in the DGP. In the context of the DGP the fact that the conditional expectation is not an unbiased predictor and need not have minimum mean-squared error when there are location shifts is proved in section 4. It is shown in section 5 that when there are such shifts, the law of iterated expectations does not hold inter-temporally, and in section 6 that the Bellman optimality principle does not apply in these circumstances. The implications of these results for economic analysis and modeling are then discussed in section 7. Alternative approaches are therefore required, and as an illustration the value of having a modeling methodology that can produce relevant, reliable, and robust models generally but especially for policy analysis, is described in section 8, emphasizing the important role that automatic model selection can play. Reference is made to the impressive results that this approach to modeling has achieved in economics. Conclusions are provided in section 9.

2 Expectations

Expectations play an important role in many areas of human activity including economic, social, political, and trading (e.g. in commodity and financial markets) behaviour, as well as in the forecasting of natural phenomena such as weather. Thus they are critical components of the theories of the associated disciplines: economics, sociology, psephology and meterology. Forecasting variables as part of systems that are subject to unanticipated changes is difficult as the recent experience of floods in Brazil and Australia as well as early heavy snow in Europe in December 2010 illustrated for meteorology. Vulcanologists and seismologists also experience difficulty in predictiong the magnitude and precise timing of volcanic erruptions and earthquakes, as in Japan and New Zealand recently. Equally, the recent financial crisis illustrated the problem for economics. Central banks use interest rates for inflation 'targets' based on expected, or forecast, levels of inflation and related variables one or two years ahead. Nevertheless, it is unclear how accurate agents' expectations of future variables are in practice, including even sophisticated agents. For example, despite a substantive investment in modeling and forecasting – and a committee of experts to advise it (the Monetary Policy Committee) – the Bank of England still significantly mis-forecasts CPI inflation (see Bank of England, 2008, with inflation later rising well outside the range in its 'fan chart'). Equally, almost no oil price forecasts for 2008 included a price near the $147 high, nor below the $40 per barrel that eventuated. Although exchange rates are a key financial price, Nickell (2009) shows the consensus forecast systematically mis-forecasting by a large margin over a long time period. Equally, the near collapse of many of the world's largest financial institutions in 2008–2009 reveals how inaccurate their expectations of asset values have proved to be. These examples show that it is very difficult to form accurate expectations about future events, with the primary cause of such failures being location shifts, when the means of future distributions differ from those of the current distribution (see Clements and Hendry, 1998).

The converse is also true: it can be difficult to discern some breaks and harder still to determine their source. Figure 1 illustrates six different types of break which all occur in practice, alter future distributions when they do so, and require careful modeling to capture their effects. A trend break (panel a) can take some time to detect, despite its immense long-run impact, partly because of the 'noisiness' of economic time series from cycles and shocks, but also because such breaks must perforce be relatively small. A shift from economic growth at a quarterly rate of 0.5 to one of 1.0 would double living standards in 18 rather than 36 years, yet corresponds to a coefficient change from 0.005 to 0.01 on a linear trend in a log-linear process, or in the intercept of a model expressed in $\mathsf{I}(0)$ variables. A step shift (panel b) is the first difference of a trend break, and would be equally undetectable for such a small effect as a change in

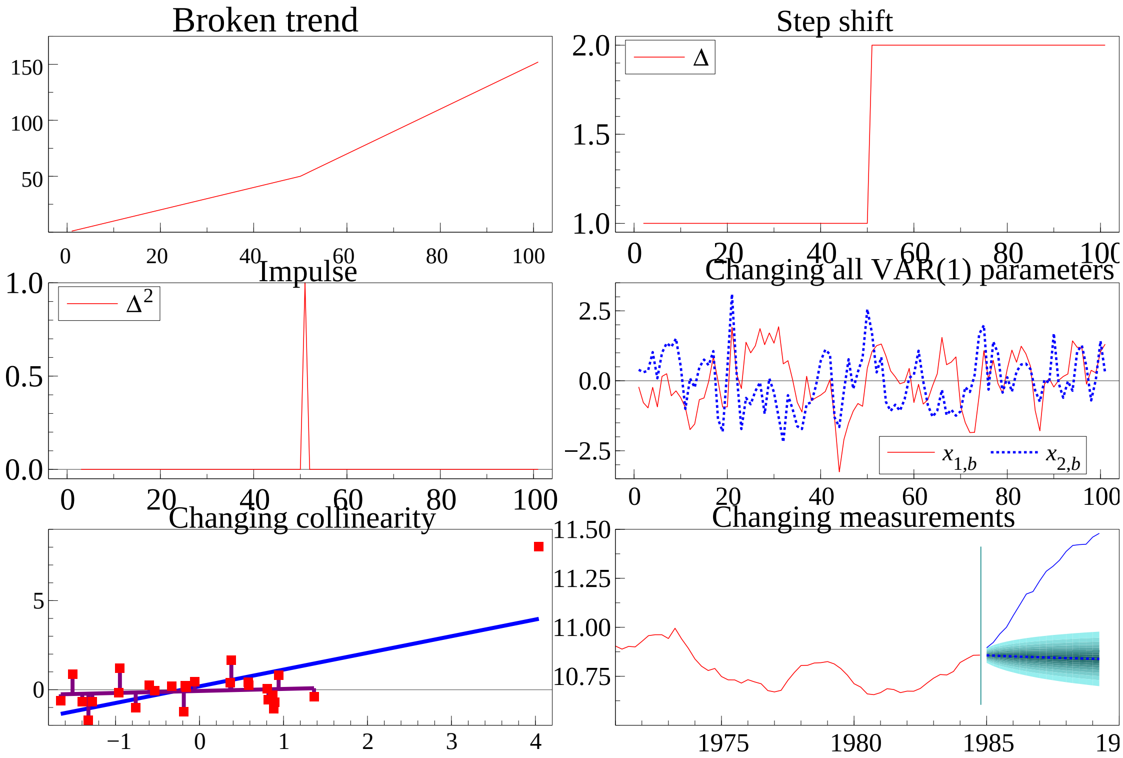

Figure 1: **Six different breaks**

growth, but could correspond to a general location shift of any magnitude, so is usually detectable, and in Clements and Hendry (1998) is analytically derived as the main cause of forecast failure. Panel c shows the first difference of a location shift, which can be conflated with a large shock, albeit that these have very different implications for impulse-response analyses in the absence of a correct weak exogeneity specification (as shown in Hendry and Mizon, 2000), and integrates to a location shift in I(1) processes. Panel d often surprises, as there is no obviously visible break in the data shown, which was generated by a first-order bivariate vector autoregression (VAR(1)) where every coefficient was changed by 30-40 error standard deviations (σ) , and the intercepts by more than 100σ. Thus, some breaks can be very difficult to detect, even when they are massive (see e.g., Hendry, 2000). Conversely, false perceptions of breaks can also be induced: panel e shows an apparent break associated with forecast failure when in fact the model in question is constant, and the break is in the collinearity between the conditioning variables–see Castle, Fawcett and Hendry (2010). The final panel, f, is a much-studied data series where the measurement of the opportunity cost of holding money was altered by legislative fiat, and induced dramatic forecast failure in models that failed to use the new measure (as shown), whereas models which shifted to the new measure maintained constant parameters: see e.g., Hendry (2006).

Thus, after a shift in the probability distribution needed to calculate future expectations, agents cannot immediately 'know' the new form. Rather they have a complicated learning task to undertake, involving a signal extraction problem as to what, if anything, has shifted, when it shifted, what aspects shifted, and by how much they have shifted, requiring many observations after the break to ascertain. The difficulties even of learning in a relatively constant environment are well known (see e.g., Evans and Honkapohja, 2001, and Young, 2004). Yet in the time taken to learn, the distribution could well have shifted again, further complicating an already difficult task. Since 'crises' occur with impressive frequency and are rarely anticipated, any empirical modeling and forecasting methods that do not explicitly address breaks are bound to be inadequate. A powerful justification for using expectations from models based on theory is that conditional expectations minimize the forecast mean-squared error. However, in the presence of unanticipated location shifts, among others, it is no longer the case that conditional expectations are unbiased, nor MMSE 1-step ahead predictors, as we prove in section 4 after addressing the formulation of conditional expectations in both models and DGPs.[1]

3 Conditional expectations in models and DGPs

Conditional expectations are the mean of the corresponding conditional distribution of one set of variables $\mathbf{y}_t$ conditional on another set of variables $\mathbf{z}_t$, defined relative to the joint distribution of all these variables $\mathbf{x}_t = (\mathbf{y}_t, \mathbf{z}_t)'$. In addition, for practical applications of conditional distributions, a distinction has to be made between calculations in the data generation process (DGP) and those in models thereof. This latter distinction has been discussed by various authors (see *inter alia* Hendry, 1995, Mizon, 1995, Spanos, 1986) and is relevant in the present context since the conditional expectations of interest are those of the DGP, which is unknown and so analyses have to be conducted using models as approximations. In the absence of a meta-DGP that explains all changes, the existence of structural changes entails that the problems analyzed in this paper occur in the DGP as well as models of aspects of it. Hence, even in the DGP, conditional expectations, despite remaining constant for periods of time, will change and thus not provide unbiased or MMSE predictors. A further problem arises with models that do not provide a good description of the economy. One of the potential contributors to the latter problem lies in the necessity of omitting some relevant variables (marginalization) and conditioning on others that may not be weakly exogenous. Further, an economic or econometric model may provide a poor description of the data we observe and so be non-congruent (see *inter alia* Hendry, 1995, Mizon, 1995, Bontemps and Mizon, 2003). Even a model that does characterize the data well can be subject to structural change, and that is the focus of this paper.

Adopting the notation that $\mathsf{E}_t[\cdot]$ is the expectation operator in the DGP at time t, and $\mathcal{E}_t[\cdot]$ is an expectation with respect to the model at the same time t, a simple example illustrates the issue. Consider the DGP in which y_t is generated by $y_t = \mu + \alpha y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim \mathsf{IN}[0, 1]$ and $|\alpha| < 1$, where a theory model also asserts that $\{y_t\}$ is generated by $y_t = \mu + \alpha y_{t-1} + u_t$ with $u_t \sim \mathsf{IN}[0, \sigma^2]$. This theory model would describe a data sample $Y_T^1 = (y_1, \ldots, y_T)$ well, and its conditional expectation $\mathcal{E}_T[y_{T+1}|Y_T^1] = \mu + \alpha y_T$ as a predictor of y_{T+1} would perform well when the DGP remained constant, since $\mathsf{E}_T[y_{T+1}|Y_T^1] = \mu + \alpha y_T$ also. However, if the DGP unexpectedly changed at time $T+1$ such that $y_t = \nu + \alpha y_{t-1} + \varepsilon_t$ for $t > T$ with $\nu \neq \mu$, then $\mathsf{E}_{T+1}[y_{T+1}|Y_T^1] = \nu + \alpha y_T \neq \mu + \alpha y_T = \mathcal{E}_T[y_{T+1}|Y_T^1]$, so the model now predicts badly.

In the following three sections 4, 5 and 6 the analysis is for DGPs, for which it is seen unanticipated distribution shifts cause problems. Although there are serious problems for analysis in the context of DGPs when there are unanticipated changes, there are even more difficulties for empirical modeling in which the DGP is unknown and so the resulting models are incomplete and maybe badly mis-specified. However, there are modeling strategies, such as that described briefly in section 8, which can address these difficulties with success. We note though that even DGPs are incomplete since they do not capture

[1] Given that we all live in a very large world, it is highly likely that there will be a few individuals who claim to foresee any change: e.g., Nooriel Roubini, who gained the epithet Dr Doom for his views. Some also foresee changes that never eventuate. Hence 'unanticipated' refers to the views of the vast majority of individuals, not necessarily all.

the “unknown unknowns" that distinguish the meta-DGP which explains all changes from the DGP.

4 Conditional expectations are not necessarily MMSE predictors

Since the primary causes of forecast failures are location shifts (see Clements and Hendry, 1998, 1999), we prove that the usual claim that the conditional expectation is the unbiased minimum mean-squared error predictor (MMSEP) is false for the case where the means of future distributions differ from the current because unanticipated breaks occur. More precisely, given an information set, $\mathbf{X}^1_{t-1}$, available at time $t-1$, the conditional expectation about a variable x_t formed at time $t-1$ for time t is denoted $\mathsf{E}_{t-1}[x_t|\mathbf{X}^1_{t-1}]$, and $\mathsf{V}_{t-1}[e_t|\mathbf{X}^1_{t-1}]$ denotes the corresponding conditional variance when e_t is the prediction error defined in (1). The first subscript denotes the date of the distribution over which expectations are calculated, the $|$ denotes conditioning, the subscript on x_t denotes the period for which the relevant expectation is formed, and $\mathbf{X}^1_{t-1}$ denotes the conditioning information. Thus, $\mathsf{E}_t[x_t|\mathbf{X}^1_{t-1}]$ is a potentially different expectation, as is $\mathsf{E}_t[x_{t+1}|\mathbf{X}^1_{t-1}]$, showing that three time subscripts are clearly needed. The conditional distribution at time s of x_t is denoted $\mathsf{f}_s(x_t|\mathbf{X}^1_{t-1})$ with $s \leqslant t$ in the context of forecasting.

Let:

$$e_t = x_t - \mathsf{E}_{t-1}\left[x_t \mid \mathbf{X}^1_{t-1}\right] \tag{1}$$

be the error from predicting x_t by the conditional expectation of x_t given $\mathbf{X}^1_{t-1}$ formed at $t-1$. Then:

$$\mathsf{E}_{t-1}\left[e_t \mid \mathbf{X}^1_{t-1}\right] = \mathsf{E}_{t-1}\left[x_t \mid \mathbf{X}^1_{t-1}\right] - \mathsf{E}_{t-1}\left[x_t \mid \mathbf{X}^1_{t-1}\right] = 0 \tag{2}$$

and:

$$\mathsf{E}_{t-1}\left[e_t^2 \mid \mathbf{X}^1_{t-1}\right] = \mathsf{V}_{t-1}\left[x_t \mid \mathbf{X}^1_{t-1}\right].$$

Thus, the usual claim that the conditional expectation is MMSEP seems correct.

However, when distributions shift, so that $\mathsf{f}_t(\cdot) \neq \mathsf{f}_{t-1}(\cdot)$, then $\mathsf{E}_t[\cdot] \neq \mathsf{E}_{t-1}[\cdot]$ since:

$$\mathsf{E}_{t-1}\left[x_t \mid \mathbf{X}^1_{t-1}\right] = \int x_t \mathsf{f}_{t-1}\left(x_t|\mathbf{X}^1_{t-1}\right) \mathrm{d}x_t.$$

but:

$$\mathsf{E}_t\left[x_t \mid \mathbf{X}^1_{t-1}\right] = \int x_t \mathsf{f}_t\left(x_t|\mathbf{X}^1_{t-1}\right) \mathrm{d}x_t$$

Although (2) is true, that is unhelpful *ex post* as the realized average error will be:

$$\begin{aligned}
\mathsf{E}_t\left[e_t \mid \mathbf{X}^1_{t-1}\right] &= \mathsf{E}_t\left[\left(x_t - \mathsf{E}_{t-1}\left[x_t|\mathbf{X}^1_{t-1}\right]\right) \mid \mathbf{X}^1_{t-1}\right] \\
&= \mathsf{E}_t\left[x_t \mid \mathbf{X}^1_{t-1}\right] - \mathsf{E}_t\left[\mathsf{E}_{t-1}\left[x_t|\mathbf{X}^1_{t-1}\right] \mid \mathbf{X}^1_{t-1}\right] \\
&= \int x_t \left[\mathsf{f}_t\left(x_t|\mathbf{X}^1_{t-1}\right) - \mathsf{f}_{t-1}\left(x_t|\mathbf{X}^1_{t-1}\right)\right] \mathrm{d}x_t \neq 0
\end{aligned} \tag{3}$$

when $\mathsf{f}_t(\cdot) \neq \mathsf{f}_{t-1}(\cdot)$. Thus, the conditional expectation $\mathsf{E}_{t-1}\left[x_t|\mathbf{X}^1_{t-1}\right]$ need not be unbiased for $\mathsf{E}_t\left[x_t|\mathbf{X}^1_{t-1}\right]$, which is the *relevant* conditional mean at time t. Also:

$$\sigma^2_{e_t} = \mathsf{E}_t\left[e_t^2 \mid \mathbf{X}^1_{t-1}\right] = \mathsf{E}_t\left[\left(x_t - \mathsf{E}_{t-1}\left[x_t \mid \mathbf{X}^1_{t-1}\right]\right)^2 \mid \mathbf{X}^1_{t-1}\right]. \tag{4}$$

Hence the conditional expectation is the MMSEP of x_t at $t-1$, but need not be at t where it can be biased and may not have the minimum variance.

4.1 Static illustration

Even in the simplest setting with no dynamics, if:

$$x_t \sim \mathsf{IN}\left[\mu_t, \sigma_x^2\right] \tag{5}$$

where $x_t = \mu_t + \epsilon_t$ then:

$$\begin{aligned}\mathsf{E}_t\left[x_t \mid \mathbf{X}_{t-1}^1\right] &= \mu_t \\ \mathsf{E}_{t-1}\left[x_t \mid \mathbf{X}_{t-1}^1\right] &= \mu_{t-1}\end{aligned}$$

so that as in (3) when the mean changes, $e_t = \epsilon_t + \mu_t - \mu_{t-1}$, so:

$$\mathsf{E}_t\left[e_t \mid \mathbf{X}_{t-1}^1\right] = \mu_t - \mu_{t-1} = \nabla\mu_t \neq 0 \tag{6}$$

and from (4):

$$\sigma_{e_t}^2 = \mathsf{E}_t\left[(\mu_t + \epsilon_t - \mu_{t-1})^2 \mid \mathbf{X}_{t-1}^1\right] = \sigma_x^2 + (\nabla\mu_t)^2 > \sigma_x^2 \tag{7}$$

Consequently, if the underlying process is wide-sense non-stationary, the conditional expectation based on the current distribution is not an unbiased predictor of the next period mean, and could have a large variance relative to the variance of the process.

As an alternative predictor, consider another function $\mathsf{G}_{t-1}\left[x_t|\mathbf{X}_{t-1}^1\right]$, and analogously to (1) let:

$$\eta_t = x_t - \mathsf{G}_{t-1}\left[x_t \mid \mathbf{X}_{t-1}^1\right]. \tag{8}$$

Then, for $\mathsf{H}_{t-1}\left[x_t|\mathbf{X}_{t-1}^1\right] = \mathsf{G}_{t-1}\left[x_t|\mathbf{X}_{t-1}^1\right] - \mathsf{E}_{t-1}\left[x_t|\mathbf{X}_{t-1}^1\right]$:

$$\begin{aligned}\sigma_{\eta_t}^2 &= \mathsf{E}_t\left[\eta_t^2 \mid \mathbf{X}_{t-1}^1\right] = \mathsf{E}_t\left[\left(x_t - \mathsf{G}_{t-1}[x_t|\mathbf{X}_{t-1}^1]\right)^2 \mid \mathbf{X}_{t-1}^1\right] \\ &= \mathsf{E}_t\left[\left(x_t - \mathsf{E}_{t-1}[x_t|\mathbf{X}_{t-1}^1] - \left\{\mathsf{G}_{t-1}[x_t|\mathbf{X}_{t-1}^1] - \mathsf{E}_{t-1}[x_t \mid \mathbf{X}_{t-1}^1]\right\}\right)^2 \mid \mathbf{X}_{t-1}^1\right] \\ &= \mathsf{E}_t\left[\left(e_t - \mathsf{H}_{t-1}[x_t|\mathbf{X}_{t-1}^1]\right)^2 \mid \mathbf{X}_{t-1}^1\right] \\ &= \mathsf{E}_t\left[e_t^2 \mid \mathbf{X}_{t-1}^1\right] + \left(\mathsf{H}_{t-1}[x_t|\mathbf{X}_{t-1}^1]\right)^2 - 2\mathsf{E}_t\left[e_t\mathsf{H}_{t-1}[x_t|\mathbf{X}_{t-1}^1]\right] \\ &= \sigma_{e_t}^2 + \left(\mathsf{H}_{t-1}[x_t|\mathbf{X}_{t-1}^1]\right)^2 - 2\mathsf{E}_t\left[e_t\mathsf{H}_{t-1}[x_t|\mathbf{X}_{t-1}^1]\right]\end{aligned}$$

When $\mathsf{E}_t \neq \mathsf{E}_{t-1}$, the best *ex post* predictor of x_t in MSE terms need not be $\mathsf{E}_{t-1}[x_t|\mathbf{X}_{t-1}^1]$ as it is possible for $\sigma_{\eta_t}^2 < \sigma_{e_t}^2$. That cannot occur when $\mathsf{E}_t = \mathsf{E}_{t-1}$ as then $e_t = \epsilon_t$ and $\mathsf{E}_t[\epsilon_t\mathsf{H}_{t-1}[x_t|\mathbf{X}_{t-1}^1]] = 0$, whereas more generally (6) shows:

$$\begin{aligned}\mathsf{E}_t\left[e_t\mathsf{H}_{t-1}[x_t|\mathbf{X}_{t-1}^1] \mid \mathbf{X}_{t-1}\right] &= \mathsf{E}_t\left[(\nabla\mu_t + \epsilon_t)\left(\mathsf{G}_{t-1}[x_t|\mathbf{X}_{t-1}^1] - \mu_{t-1}\right) \mid \mathbf{X}_{t-1}^1\right] \\ &= \nabla\mu_t\left(\mathsf{G}_{t-1}[x_t|\mathbf{X}_{t-1}^1] - \mu_{t-1}\right)\end{aligned} \tag{9}$$

Again in the special case of (5) let:

$$\mathsf{G}_{t-1}\left[x_t \mid \mathbf{X}_{t-1}\right] = \mu_{t-1} + \delta$$

which might be an intercept-corrected forecast, then:

$$\mathsf{H}_{t-1}\left[x_t \mid \mathbf{X}_{t-1}\right] = \delta$$

and so:

$$\mathsf{E}_t\left[e_t \mathsf{H}_{t-1}\left[x_t \mid \mathbf{X}_{t-1}\right]\right] = \delta\left(\mu_t - \mu_{t-1}\right) = \delta\nabla\mu_t.$$

Consequently:

$$\sigma^2_{\eta_t} = \sigma^2_{e_t} + \delta^2 - 2\delta\nabla\mu_t \tag{10}$$

so $\sigma^2_{\eta_t} < \sigma^2_{e_t}$ if (say) $\delta > 0$ and:

$$\delta - 2\nabla\mu_t < 0. \tag{11}$$

Hence, when $\nabla\mu_t > 0$ then $\sigma^2_{\eta_t} < \sigma^2_{e_t}$, provided that $\delta > 0$ and $\delta < 2\nabla\mu_t$. Therefore, if the modification to the conditional mean is in the correct direction, but does not seriously overshoot, then the it results in a lower MSE than the conditional mean predictor. Note that when $\nabla\mu_t > 0$, then $\sigma^2_{\eta_t} > \sigma^2_{\epsilon_t}$ whenever the mean adjustment is in the wrong direction, i.e., $\delta < 0$. Alternatively, when $\nabla\mu_t < 0$, then $\sigma^2_{\eta_t} < \sigma^2_{e_t}$ provided that $\delta < 0$ and $|\delta| < 2|\nabla\mu_t|$. In summary, it follows that $\sigma^2_{\eta_t} < \sigma^2_{e_t}$ whenever $\nabla\mu_t$ and δ have the same sign (i.e., the modification is in the correct direction) and $|\delta| < 2|\nabla\mu_t|$ (i.e., the modification is not too large).

4.2 Dynamic illustration

As a more realistic illustration of these formulae, consider a stationary first-order autoregressive DGP:

$$y_t = \gamma + \rho y_{t-1} + \epsilon_t \;\text{ where }\; \epsilon_t \sim \mathsf{IN}\left[0, \sigma^2_\epsilon\right] \tag{12}$$

with $|\rho| < 1$ that holds for $t = 1, 2, \ldots, T-1$. Then expectations are constant over that period, so that:

$$\mathsf{E}\left[y_t\right] = \mu = \gamma + \rho\mathsf{E}\left[y_{t-1}\right] + \mathsf{E}\left[\epsilon_t\right] = \gamma + \rho\mu \tag{13}$$

and hence $\mu = \gamma/(1-\rho)$ is the equilibrium mean of $\{y_t\}$ over that sample. The conditional expectation, given the history of the process is:

$$\mathsf{E}\left[y_t \mid y_{t-1}\right] = \gamma + \rho y_{t-1} + \mathsf{E}\left[\epsilon_t \mid y_{t-1}\right] = \gamma + \rho y_{t-1} \tag{14}$$

and in this setting, $\mathsf{E}\left[y_t|y_{t-1}\right]$ is an unbiased, MMSE predictor of y_t:

$$y_t - \mathsf{E}\left[y_t \mid y_{t-1}\right] = \epsilon_t$$

with:

$$\mathsf{E}\left[\epsilon_t\right] = 0 \;\text{ and }\; \mathsf{V}\left[\epsilon_t\right] = \sigma^2_\epsilon$$

which is the smallest obtainable.

Next, for $t = T, T+1, \ldots$ the structural change is denoted:

$$y_t = \gamma^* + \rho^* y_{t-1} + \epsilon_t \tag{15}$$

where $\epsilon_t \sim \mathsf{IN}\left[0, \sigma^2_\epsilon\right]$ as before (changing that distribution adds to the conclusion) and $|\rho^*| < 1$ still. Now expectations must be dated to avoid incorrect calculations, so we write $\mathsf{E}_{T-1}\left[\cdot\right]$, $\mathsf{E}_T\left[\cdot\right]$ etc., where the subscripts denote the pre-break and post-break distributions determined by (12) and (15) respectively. From (15):

$$\mathsf{E}_T\left[y_T\right] = \gamma^* + \rho^*\mathsf{E}_T\left[y_{T-1}\right] + \mathsf{E}_T\left[\epsilon_T\right] = \gamma^* + \rho^*\mu \tag{16}$$

which in general is not equal to μ if either parameter differs between (12) and (15).[2] Moreover:

$$\mathsf{E}_{T+1}\left[y_{T+1}\right] = \gamma^* + \rho^*\mathsf{E}_{T+1}\left[y_T\right] + \mathsf{E}_{T+1}\left[\epsilon_{T+1}\right] = \gamma^* + \rho^*\left(\gamma^* + \rho^*\mu\right) = \gamma^*\left(1+\rho^*\right) + \left(\rho^*\right)^2\mu \quad (17)$$

which keeps changing, and although it converges on $\mu^* = \gamma^*/(1-\rho^*)$, does not equal μ^* for a number of periods.

However, at $T-1$, it is not known that the break will occur, so agents forming conditional expectations about y_T given y_{T-1} must perforce use the distribution at that time, leading to:

$$\mathsf{E}_{T-1}\left[y_T \mid y_{T-1}\right] = \gamma + \rho y_{T-1} + \mathsf{E}_{T-1}\left[\epsilon_T \mid y_{T-1}\right] = \gamma + \rho y_{T-1} \quad (18)$$

Thus, their conditional expectations error is:

$$\begin{aligned} y_T - \mathsf{E}_{T-1}\left[y_T \mid y_{T-1}\right] &= \gamma^* + \rho^* y_{T-1} + \epsilon_T - \gamma - \rho y_{T-1} \\ &= \left(\gamma^* - \gamma\right) + \left(\rho^* - \rho\right) y_{T-1} + \epsilon_T \\ &= \nabla\gamma + \nabla\rho\, y_{T-1} + \epsilon_T. \end{aligned}$$

On average (i.e., unconditionally), that error will transpire to be:

$$\mathsf{E}_T\left[y_T - \mathsf{E}_{T-1}[y_T|y_{T-1}]\right] = \left(\gamma^* - \gamma\right) + \left(\rho^* - \rho\right)\mathsf{E}_T\left[y_{T-1}\right] = \nabla\gamma + \nabla\rho\,\mu$$

so the prediction is biased. Moreover, unless the agents are omniscient and instantly discover their mistake (somehow 'learning' two parameters from the one error), then they will make a similar mistake in the next period, so the bias persists. For example, if agents keep the in-sample parameter values, but update the data, so use:

$$\mathsf{E}_{T-1}\left[y_{T+1} \mid y_T\right] = \gamma + \rho y_T + \mathsf{E}_{T-1}\left[\epsilon_{T+1} \mid y_T\right] = \gamma + \rho y_T$$

this leads to the average error:

$$\begin{aligned} \mathsf{E}_{T+1}\left[y_{T+1} - \mathsf{E}_{T-1}\left[y_{T+1} \mid y_T\right]\right] &= \mathsf{E}_{T+1}\left[\left(\gamma^* - \gamma\right) + \left(\rho^* - \rho\right) y_T + \epsilon_{T+1}\right] \\ &= \nabla\gamma + \nabla\rho\,\left(\gamma^* + \rho^*\mu\right). \end{aligned}$$

If expectations were undated, then it is unclear what $\mathsf{E}\left[y_{T+1}\right]$ might be, but if any aspect of the in-sample model's parameters has shifted, the correct unconditional expectation is never:

$$\mathsf{E}\left[y_{T+1}\right] = \frac{\gamma^*}{1-\rho^*} \quad \text{nor} \quad \mathsf{E}\left[y_{T+1}\right] = \frac{\gamma}{1-\rho}.$$

Now consider the alternative predictor to the conditional mean given by $\mathsf{K}_{T-1}[y_T|y_{T-1}]$ and analogously to (1) and (8), define:

$$\psi_T = y_T - \mathsf{K}_{T-1}[y_T|y_{T-1}].$$

Let:

$$\mathsf{J}_{T-1}[y_T|y_{T-1}] = \mathsf{K}_{T-1}[y_T|y_{T-1}] - \mathsf{E}_{T-1}[y_T|y_{T-1}]$$

and

$$u_t = y_T - \mathsf{E}_{T-1}[y_T|y_{T-1}] = \nabla\gamma + \nabla\rho\, y_{T-1} + \epsilon_T$$

[2]If the process remains stationary then it would be possible for the equilibrium to remain constant, but it would need both of γ^* and ρ^* to change with $0 < \gamma^*/\gamma = (1-\rho^*)/(1-\rho) < \infty$.

so that for $\mathsf{E}_T[u_T^2|y_{T-1}] = \mathsf{E}_T[u_T^2] = \sigma^2_{u_T}$:

$$\begin{aligned}
\sigma^2_{\psi_T} &= \mathsf{E}_T\left[\psi_T^2 \mid y_{T-1}\right] = \mathsf{E}_T\left[(y_T - \mathsf{K}_{T-1}[y_T|y_{T-1}])^2 \mid y_{T-1}\right] \\
&= \mathsf{E}_T\left[(y_T - \mathsf{E}_{T-1}[y_T|y_{T-1}] - \{\mathsf{K}_{t-1}[y_T|y_{T-1}] - \mathsf{E}_{T-1}[y_T|y_{T-1}]\})^2 \mid y_{T-1}\right] \\
&= \mathsf{E}_T\left[(u_T - \mathsf{J}_{T-1}[y_T|y_{T-1}])^2 \mid y_{T-1}\right] \\
&= \sigma^2_{u_T} + (\mathsf{J}_{T-1}[y_T|y_{T-1}])^2 - 2\mathsf{E}_T\left[u_T \mathsf{J}_{T-1}[y_T|y_{T-1}]\right].
\end{aligned}$$

When $\mathsf{E}_T \neq \mathsf{E}_{T-1}$, the best *ex post* predictor of y_T in MSE terms need not be $\mathsf{E}_{T-1}[y_T \mid y_{T-1}]$ as it is possible for $\sigma^2_{\psi_T} < \sigma^2_{u_T}$. That cannot occur when $\mathsf{E}_T = \mathsf{E}_{T-1}$ as then $\sigma^2_{u_T} = \sigma^2_\epsilon$ and $\mathsf{E}_T[u_T \mathsf{J}_{T-1}[y_T|y_{T-1}]] = \mathsf{E}_T[\epsilon_T \mathsf{J}_{T-1}[y_T|y_{T-1}]] = 0$, whereas in general:

$$\begin{aligned}
&\mathsf{E}_T\left[(u_T \mathsf{J}_{T-1}[y_T|dy_{T-1}]) \mid y_{T-1}\right] \\
&= \mathsf{E}_T\left[(\nabla\gamma + \nabla\rho\, y_{T-1} + \epsilon_T)\left(\mathsf{K}_{T-1}[y_T|y_{T-1}] - (\gamma + \rho y_{T-1})\right) \mid y_{T-1}\right] \\
&= \mathsf{E}_T\left[u_T \mathsf{K}_{T-1}[y_T|y_{T-1}] \mid y_{T-1}\right] - \gamma\nabla\gamma - (\rho\nabla\gamma + \gamma\nabla\rho)y_{T-1} - \rho\nabla\rho y^2_{T-1} \\
&\neq 0.
\end{aligned}$$

In the case of (12) let:

$$\mathsf{K}_{T-1}\left[y_T \mid y_{T-1}\right] = \gamma + \rho y_{T-1} + \delta$$

which might be an intercept-corrected forecast, then:

$$\mathsf{J}_{T-1}\left[y_T \mid y_{T-1}\right] = \delta$$

and so:

$$\mathsf{E}_T\left[u_T \mathsf{J}_{T-1}\left[y_T \mid y_{T-1}\right]\right] = \delta \mathsf{E}_T\left[u_T \mid y_{T-1}\right] = \delta(\nabla\gamma + \nabla\rho\, y_{T-1}).$$

Hence:

$$\sigma^2_{\psi_T} = \sigma^2_{u_T} + \delta^2 - 2\delta(\nabla\gamma + \nabla\rho\, y_{T-1})$$

which for example illustrates that an intercept adjusted forecast might have a lower forecast error variance than the conditional mean since $\sigma^2_{\psi_T} < \sigma^2_{u_T}$ is possible. Noting that $(\nabla\gamma + \nabla\rho\, y_{T-1})$ is the forecast error of the conditional mean predictor (apart from ϵ_T which has a zero mean) it is clear that this result is analogous to (10) of the static case.

The difficulties described in this section arise in the DGP and do not involve any subjective probabilities that might arise in an empirical model. Hence when attention is turned to the practical problem of forecasting using empirical models the difficulties multiply. For example, the possibility exists that a badly mis-specified model might by chance forecast accurately. However, a biased forecast is unlikely to be the most rational basis for forecasting. We now consider a further implication of these results.

5 The law of iterated expectations and unanticipated change

When expectation distributions are unaltered the law of iterated expectations often is written:

$$\mathsf{E}_z\left[\mathsf{E}_y\left[y \mid z\right]\right] = \mathsf{E}_y\left[y\right] \tag{19}$$

and proved by:

$$\begin{aligned}
\mathsf{E}_z\left[\mathsf{E}_y\left[y \mid z\right]\right] &= \int_{\mathcal{Z}}\left(\int_{\mathcal{Y}} y\mathsf{f}\left(y|z\right)\mathrm{d}y\right)\mathsf{g}\left(z\right)\mathrm{d}z = \int_{\mathcal{Z}}\int_{\mathcal{Y}} y\mathsf{f}\left(y|z\right)\mathsf{g}\left(z\right)\mathrm{d}z\mathrm{d}y \\
&= \int_{\mathcal{Y}} y\left(\int_{\mathcal{Z}}\mathsf{h}\left(y,z\right)\mathrm{d}z\right)\mathrm{d}y = \int_{\mathcal{Y}} y\mathsf{p}\left(y\right)\mathrm{d}y = \mathsf{E}_y\left[y\right]
\end{aligned}$$

where $\mathsf{h}(y,z) = \mathsf{f}(y|z)\mathsf{g}(z) = \mathsf{p}\left(y\right)\psi\left(z|y\right)$ is the joint distribution of (y,z) and:

$$\int_{\mathcal{Z}}\mathsf{h}\left(y,z\right)\mathrm{d}z = \mathsf{p}\left(y\right).$$

When the variables correspond to a common set at different dates drawn from the same distribution, then (19) becomes:

$$\mathsf{E}_{x_t}\left[\mathsf{E}_{x_{t+1}}\left[x_{t+1} \mid x_t\right]\right] = \mathsf{E}_{x_{t+1}}\left[x_{t+1}\right].$$

The formal derivation is close to that in (19), namely:

$$\begin{aligned}
\mathsf{E}_{x_t}\left[\mathsf{E}_{x_{t+1}}\left[x_{t+1} \mid x_t\right]\right] &= \int_{x_t}\left(\int_{x_{t+1}} x_{t+1}\mathsf{f}\left(x_{t+1}|x_t\right)\mathrm{d}x_{t+1}\right)\mathsf{p}\left(x_t\right)\mathrm{d}x_t \\
&= \int_{x_t}\int_{x_{t+1}} x_{t+1}\mathsf{f}\left(x_{t+1}|x_t\right)\mathsf{p}\left(x_t\right)\mathrm{d}x_t\mathrm{d}x_{t+1} \\
&= \int_{x_{t+1}} x_{t+1}\left(\int_{x_t}\mathsf{h}\left(x_{t+1},x_t\right)\mathrm{d}x_t\right)\mathrm{d}x_{t+1} \\
&= \int_{x_{t+1}} x_{t+1}\mathsf{p}\left(x_{t+1}\right)\mathrm{d}x_{t+1} = \mathsf{E}_{x_{t+1}}\left[x_{t+1}\right] \qquad (20)
\end{aligned}$$

Thus, if the distributions remain constant, the law of iterated expectations holds.

However, the law of iterated expectations need not hold when distributions shift, as the factorization $\mathsf{h}\left(x_{t+1},x_t\right) = \mathsf{f}\left(x_{t+1}|x_t\right)\mathsf{p}\left(x_t\right)$ of the joint density is not achieved by the law of iterated expectations. This problem arises when the distribution shifts between t and $t+1$ as follows. First, note that:

$$\mathsf{E}_{x_t}\left[x_{t+1} \mid \mathcal{I}_t\right] \text{ and } \mathsf{E}_{x_t}\left[x_{t+1} \mid \mathcal{I}_{t-1}\right]$$

are different entities when $\mathcal{I}_t$ and $\mathcal{I}_{t-1}$ are information sets at t and $t-1$ respectively. Similarly when distributions shift we have:

$$\mathsf{E}_{x_t}\left[x_{t+1} \mid \mathcal{I}_t\right] \neq \mathsf{E}_{x_{t+1}}\left[x_{t+1} \mid \mathcal{I}_t\right]$$

the former of these being needed for an unbiased conditional prediction as shown in the previous section. Now, however:

$$\mathsf{E}_{x_t}\left[\mathsf{E}_{x_{t+1}}\left[x_{t+1} \mid x_t\right]\right] \neq \mathsf{E}_{x_{t+1}}\left[x_{t+1}\right]$$

since:

$$\begin{aligned}
\mathsf{E}_{x_t}\left[\mathsf{E}_{x_{t+1}}\left[x_{t+1} \mid x_t\right]\right] &= \int_{x_t}\left(\int_{x_{t+1}} x_{t+1}\mathsf{f}_{x_{t+1}}\left(x_{t+1}|x_t\right)\mathrm{d}x_{t+1}\right)\mathsf{p}_{x_t}\left(x_t\right)\mathrm{d}x_t \\
&= \int_{x_t}\int_{x_{t+1}} x_{t+1}\mathsf{f}_{x_{t+1}}\left(x_{t+1}|x_t\right)\mathsf{p}_{x_t}\left(x_t\right)\mathrm{d}x_t\mathrm{d}x_{t+1} \\
&= \int_{x_{t+1}} x_{t+1}\left(\int_{x_t}\mathsf{f}_{x_{t+1}}\left(x_{t+1}|x_t\right)\mathsf{p}_{x_t}\left(x_t\right)\mathrm{d}x_t\right)\mathrm{d}x_{t+1} \\
&\neq \int_{x_{t+1}} x_{t+1}\mathsf{p}_{x_{t+1}}\left(x_{t+1}\right)\mathrm{d}x_{t+1} = \mathsf{E}_{x_{t+1}}\left[x_{t+1}\right] \qquad (21)
\end{aligned}$$

The reason the law of iterated expectations does not hold in this case is that $\mathsf{f}_{x_{t+1}}\left(x_{t+1}|x_t\right)\mathsf{p}_{x_t}\left(x_t\right) \neq \mathsf{f}_{x_{t+1}}\left(x_{t+1}|x_t\right)\mathsf{p}_{x_{t+1}}\left(x_t\right) = \mathsf{h}_{t+1}\left(x_{t+1}, x_t\right)$ unlike the situation in (20) where there is no shift in distribution.

Thus, when distributions shift over time as in (5) expectations are affected by their timing:

$$\begin{aligned}
\mathsf{E}_{x_t}\left[x_{t+1}|x_t\right] &= \mu_t \neq \mathsf{E}_{x_{t+1}}\left[x_{t+1}\right] = \mu_{t+1} \\
\mathsf{E}_{x_{t+1}}\left[x_{t+1}|x_t\right] &= \mu_{t+1}
\end{aligned}$$

noting that x_t and x_{t+1} are independent in this example. Thus in this case we have:

$$\mathsf{E}_{x_t}\left[\left(\mathsf{E}_{x_{t+1}}\left[x_{t+1}|x_t\right]\right)\right] = \mathsf{E}_{x_t}\left[\mu_t\right] = \mu_t \neq \mu_{t+1} = \mathsf{E}_{x_{t+1}}\left[x_{t+1}\right].$$

Equally, for the analogous model to (12):

$$\mathsf{E}_{x_t}\left[x_{t+1}|x_t\right] = \gamma + \rho x_t \neq \mathsf{E}_{x_{t+1}}\left[x_{t+1}\right] = \gamma^* + \rho^*\mu^*$$

and

$$\mathsf{E}_{x_t}\left[\left(\mathsf{E}_{x_{t+1}}\left[x_{t+1}|x_t\right]\right)\right] = \mathsf{E}_{x_t}\left[\gamma^* + \rho^* x_t\right] = \gamma^* + \rho^*\mu \neq \mathsf{E}_{x_{t+1}}\left[x_{t+1}\right] = \mu_{t+1}$$

when $\mu = \gamma/(1-\rho)$ and $\mu^* = \gamma^*/(1-\rho^*)$. Finally note that with consistent dating it remains true that:

$$\mathsf{E}_{x_t}\left[\left(\mathsf{E}_{x_t}\left[x_{t+1} \mid x_t\right]\right)\right] = \mathsf{E}_{x_t}\left[x_{t+1}\right] = \mu_t.$$

More generally, there are two sources of updating from, say, $\mathsf{E}_{x_t}\left[x_{t+1}|x_{t-1}\right]$ to $\mathsf{E}_{x_{t+1}}\left[x_{t+1}|x_t\right]$: new information is embodied in x_{t-1} becoming x_t; and shifts in the distribution implied by a change from E_{x_t} to $\mathsf{E}_{x_{t+1}}$. Much of the economics literature (see e.g., Campbell and Shiller, 1987) assumes that the former is an unanticipated change, written as $\mathsf{E}\left[x_{t+1}|x_t\right] - \mathsf{E}\left[x_{t+1}|x_{t-1}\right]$, which is an innovation, ν_t, and the relevant information becomes known one period later. That is not true of the latter, where the new distribution has to be learned over time–and may have shifted again in the meantime. Even if the distribution, denoted $\mathsf{f}_{t+1}\left(x_{t+1}|x_t\right)$, became known one period later:

$$\begin{aligned}
\mathsf{E}_{x_{t+1}}\left[x_{t+1} \mid x_t\right] - \mathsf{E}_{x_t}\left[x_{t+1} \mid x_{t-1}\right] &= \mathsf{E}_{x_{t+1}}[x_{t+1}|x_t] - \mathsf{E}_{x_{t+1}}\left[x_{t+1}|x_{t-1}\right] \\
&\quad + \left(\mathsf{E}_{x_{t+1}}\left[x_{t+1}|x_{t-1}\right] - \mathsf{E}_{x_t}\left[x_{t+1}|x_{t-1}\right]\right) \\
&= \nu_t + \int x_{t+1}\mathsf{f}_{t+1}\left(x_{t+1}|x_{t-1}\right)\mathrm{d}x_{t+1} - \int x_{t+1}\mathsf{f}_t\left(x_{t+1}|x_{t-1}\right)\mathrm{d}x_t \\
&= \nu_t + \left(\mu_{t+1} - \mu_t\right)
\end{aligned}$$

where the last line uses (5). In practice, both means need to be estimated, a nearly intractable task for agents–or statisticians and econometricians–when distributions are shifting.

6 Bellman principle of optimality and unanticipated change

Analogous to the law of iterated expectations not holding, the recursive method of solving stochastic inter-temporal optimization problems, known as the Bellman principle of optimality (see Bellman, 1957), does not apply when there are unanticipated changes. The basic principle is described by Kreps (Kreps, 1990, page 798) in the following way:

If a strategy is optimal for each point in time at that point in time, given that an optimal strategy will be used thereafter, then the strategy is optimal.

However, when the future distribution of the state variables $\mathbf{x}_t$ in a non-stationary context is not known because it is subject to unanticipated changes the optimal strategies in the future are likely to be different from the present perception of them. Hence the existence of an invariant optimal strategy over the whole time horizon is unlikely in a potentially ever changing environment.

Consider a decision problem in which at time t an action a_t has to be taken from a set of feasible actions $A(\mathbf{x}_t)$ that depends on the observed state variable (e.g. state of the economy) $\mathbf{x}_t$. Initially let $\mathbf{x}_t$ be non-stochastic and determined according to $\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}, a_{t-1})$. Then for the value function $V(\mathbf{x}_t)$ over the horizon h with discount factor $0 < \beta < 1$ and return function $r(\mathbf{x}_t, a_t)$ following decision a_t the decision problem can be written as:

$$V(\mathbf{x}_t) = \max_{a_\tau \in A(x_\tau)} \sum_{\tau=t}^{t+h} \beta^\tau r(\mathbf{x}_\tau, a_\tau) \tag{22}$$

Noting that the first decision a_t can be separated from all future decisions (22) can be re-written as:

$$V(\mathbf{x}_t) = \left\{ r(\mathbf{x}_t, a_t) + \max_{a_\tau \in A(x_\tau)} \beta \sum_{\tau=t+1}^{t+h} \beta^{\tau-1} r(\mathbf{x}_\tau, a_\tau) \right\}$$

thus revealing the recursive nature of the decision problem and leading to the Bellman equation for this non-stochastic discrete case:

$$V(\mathbf{x}_t) = \max_{a_t \in A(x_t)} \{ r(\mathbf{x}_t, a_t) + \beta V(\mathbf{x}_{t+1}) \} \tag{23}$$

In many decision problems of this type the state variable $\mathbf{x}_t$ is stochastic and so its future values are unknown and replaced by expectations. Let $F(\lambda) = Prob(\mathbf{x}_{t+i} \leq \lambda \,|\mathbf{x}_t, a_t)$ for $i = 1, 2, ...h$ be the conditional distribution function for $\mathbf{x}_{t+i}$ then when $\mathbf{x}_t$ is stationary

$$V(\mathbf{x}_t) = \max_{a_t \in A(x_t)} \{ r(\mathbf{x}_t, a_t) + \beta \mathsf{E}[V(\mathbf{x}_{t+1}) \mid \mathbf{x}_t, a_t] \}.$$

However, when $F(\lambda)$ is subject to unanticipated changes, as is commonly the case in many areas of application such as economics, the expectation must also be dated yielding:

$$V(\mathbf{x}_t) = \max_{a_t \in A(x_t)} \{ r(\mathbf{x}_t, a_t) + \beta \mathsf{E}_t[V(\mathbf{x}_{t+1}) \mid \mathbf{x}_t, a_t] \}. \tag{24}$$

Note that at time t the future distributions and hence expectations are not known and so the expectation in (24) can only be E_t. Since $\mathsf{E}_{t+h} \neq \mathsf{E}_t$ with unanticipated change the decisions arising from the Bellman equation (24) are not necessarily optimal precisely as with the law of interated expectations.

The main results of Sections 4, 5 and 6, namely that when distributions shift the conditional expectation is not the unbiased MMSE predictor and the law of iterated expectations plus the Bellman equation do not hold, mean that the mathematical derivations commonly underlying inter-temporal optimization theory are invalid if *any* location shifts have occurred.

7 Implications for economic analysis and modeling

The analysis in the previous sections was for DGPs, but when we consider economic analysis and modeling economic theory and empirical models are the available entities since DGPs are unknown. Hence a major objective in practice must be to have the best available theory, model or forecast for the particular problem being analysed. Economic theories, models and forecasts that are relevant to the phenomena of interest, reliable in that they are not sensitive to minor variations (e.g. as a result of not having exploited all available relevant information), and robust in the face of unanticipated changes, are important ingredients in economic analysis and modeling, including economic policy analysis. We now describe some areas of economic analysis and modeling where unanticipated changes cause difficulties. Ultimately though, when the DGP is not known what is required of theoretical and empirical models is that they be the best available for analyzing the phenomenon of interest.

It is notoriously difficult to forecast economic variables with the many forecast failures serving to emphasize that economies are non-stationary and evolving. There is a long history of economic models suffering forecast failure and being out-performed by so-called 'naive devices'. A major reason for this failure is the fact that almost no forecasting models allow for unanticipated location shifts (changes in the previous means of the variables under analysis), although these clearly occur empirically.

Forecast failure is due to unanticipated location shifts (see Clements and Hendry, 1999, 1998). Location shifts induce systematic mis-forecasting in all forms of equilibrium-correction models, which comprise most macro-econometric systems in use. Conversely, every parameter in the data generation process can be shifted without any noticeable effect on the data or a model thereof when there is no location shift as illustrated in section 2. Similarly, any location shift effect can be created by many different combinations of DGP parameters shifting, but which ones changed may not be discernable from the evidence till long after the occurrence of forecast failure. Thus, the verisimilitude of a model cannot be reliably checked by its forecasting success or failure.

Systematic mis-forecasting can be mitigated by using the differences of the econometric system, retaining precisely the same estimates, even when the DGP parameters involved have changed. The costs of unnecessary differencing when there is no location shift are relatively small. In both cases, the policy implications of the structural system are the same, but may or may not be useful depending on the unknown source and form of the location shift. In neither case will the systems considered here, or their differences, forecast future location shifts: a different class of model seems to be needed for that, based on different information (see e.g., Castle *et al.*, 2010, Castle, Fawcett and Hendry, 2011). Nevertheless, at least avoiding systematic forecast failure is crucial if policy is to be well based on what the future might bring forth.

An implication of the results in sections 4 and 5 is that the existence of unanticipated changes leads to difficulties for models based on inter-temporal optimization and conditional expectations. DSGE models have rational expectations (RE), construed as the pre-existing conditional expectation, built into them and this presents a problem. Hall (1978) pointed out an important implication of RE, namely that $e_{t+1} = x_{t+1} - \mathsf{E}_t\left[x_{t+1}|\mathbf{X}_t\right]$ is unpredictable given $\mathbf{X}_t$, and so when there are structural breaks serious forecast errors will arise. This presents a problem for economic theory-led models, such as DSGE models, whenever there is a structural change.

In practice, no agent can possibly know even the current distribution to compute its conditional expectation, which instead has to be estimated in some way from the information available to that agent. That requires a minimum of a sample of observations, formulated in a model, from which the estimated conditional expectation is then calculated–and when distributions are shifting, that task borders on the impossible. Historically, most of the theory of rational expectations was developed for stationary processes, and while learning introduced a form of non-stationarity as in Evans and Honkapohja (2001), the theory has not been updated to a wide-sense non-stationary world, partly because it is not obvious what a rational forecast would be when location shifts occur, as they manifestly do. Since their derivations rely on solving inter-temporal optimization problems, assuming agents form their expectations of the unknown future events using their current conditional expectations, DSGEs must be intrinsically non-structural when the distributions underlying those expectations alter. Thus, the Lucas (1976) critique applies automatically to the basic form of DSGE because their very derivations necessitate that expectations distributions never change. Muellbauer (2009) presents a similar critique of the use of DSGE with rational expectations in the particular context of personal sector consumption and housing.

The Bellman optimality principle has been applied in many areas of economic theory (see e.g. Stokey and Lucas, 1989 for an early overview). Areas of application include: consumption theory, the inter-temporal capital asset pricing model of Merton (1973), resource extraction (see e.g., Gilbert, 1979), public finance (see e.g., Kydland and Prescott, 1980), industrial organization (see e.g., Doraszelski and Pakes, 2007), and many others. The results in section 6 imply that unless these models can account for the unanticipated changes that affect most areas of the economy they too will fail to capture a common feature of observed behaviour.

However, economic analysis and economic policy analysis require more than just capturing a mean shift. Tempting though it may be to identify variables that ‘explain’ such shifts in-sample and include them in the model, this will improve forecasts only if it is possible to accurately forecast their shifts. A structural model is required for reliable analysis, but more realistically, one might seek an ability to quickly: (a) identify a new regime’s characteristics, and (b) develop a model of that regime. Precisely how this can be done within the framework of models similar to DSGEs is unclear, but the modeling strategy outlined in the next section may be more promising in a world of intermittent unanticipated location shifts. Further discussion of these and related points is contained in Hendry and Mizon (2011b).

8 Modeling methodology

Although there are alternative ways of developing empirically well founded and policy-relevant models, there are few that have been able to successfully deal with the problems arising from unanticipated change described above. Given that unanticipated changes are by definition unknown before they occur it is imperative to have methods that react quickly once the changes have taken place. One approach to this problem that has proved to be highly successful is the incorporation of impulse indicator saturation (IIR) (see Hendry, Johansen and Santos, 2008) in a general-to-simple model selection procedure. General models designed to embrace a range of theories, different functional forms, and provide a good characterization of the data, including possible regime changes, are essential – no current theories are structural in the sense of being invariant to all relevant regime change. Attention can then be paid to valid conditioning and marginalization, which is crucial, particularly when models are being developed for policy analysis. Equally, it provides a framework to distinguish behaviourally relevant dynamics from proxy dynamics that often arise to accommodate regime change and expectations. The choice of the general unrestricted model (GUM) is very important, and involves much human input based on experience, relevant theory, institutional knowledge, the purpose for which the modeling is being done, and the known properties of the data, including its quality. Once the GUM has been specified, the major task is that of selecting a model from the large number of possible sub-models that are embedded in the GUM, such that the final selection is coherent with the data characteristics (congruent), and achieves this parsimoniously at least as well as the alternative models within the GUM (encompassing). By focusing on selecting variables rather than models, recent developments in the automation of this selection process have produced remarkable results, extending to handling potentially more candidate variables than observations, and jointly selecting variables, functional forms, multiple breaks, and data contamination. Hendry and Johansen (2011) show that if the theory variables are not selected over when the theory model is a complete and correct representation of the data evidence, then the distributions of the parameter estimates after selection, possibly over more candidate variables than observations, are **identical** to those obtained by direct estimation of the theory model. Thus, the search costs are essentially zero. Conversely, if the theory model is incomplete or incorrect, but a sufficiently general GUM nests the DGP, then a viable representation of that DGP will be retained after selection even when the theory variables are maintained. Finally, if the theory is incomplete and the GUM does not nest the DGP, selection can still deliver a far better model, avoiding serious non-constancies and providing smaller MSEs for the parameters of interest in the correct specification (see Castle and Hendry, 2010). Consequently, selection provides a near Pareto optimal approach for all these realistic settings. For general discussions of the achievements of the new approach to automatic model selection, see *inter alia*, Castle, Doornik and Hendry (2011, 2009). The results of this large body of research are embodied in the software package *Autometrics* (see Doornik, 2009). Hendry and Mizon (2011a) provide an example of this approach to modeling in the context of a re-examination of Tobin's model of the demand for food in the USA (Tobin, 1950) using an extended data set.

9 Conclusions

Expectations of future events are important in many areas of human behaviour and the natural environment. However, almost none of the relevant time series is stationary, either weakly or strictly: distributions shift. This causes problems for the analysis of these series. In particular, it cannot be proved that conditional expectations based on contemporaneous distributions are minimum mean-square error 1-step predictors when unanticipated breaks occur, and consequentially, the law of iterated expectations and the Bellman equation fail inter-temporally. Although no model is perfect, choosing amongst the available models on the basis of theory coherence, no matter how inconsistent the result is with empirical evidence, has little to recommend it for policy analysis and forecasting. Modeling is an evolutionary process, and it is important to have criteria that enable selection to lead to models that will survive challenges from all sources of information, rather than models that become extinct following successive failures to accurately capture the unfolding of events in the economy. To offset the negative results on expectations, we have briefly described a modeling methodology that offers exciting prospects, and has an excellent record to date.

References

Bank of England (2008). *August Inflation Report*. London: Bank of England Monetary Policy Committee.

Bellman, R. E. (1957). *Dynamic Programming*. Princeton, NJ: Princeton University Press. Republished 2003: Dover, ISBN 0486428095.

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

Campbell, J. Y., and Shiller, R. J. (1987). Cointegration and Tests of Present Value Models. *Journal of Political Economy*, *95*(5), 1062–1088.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2009). Model selection when there are multiple breaks. Working paper 472, Economics Department, University of Oxford.

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

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, M. P., and Hendry, D. F.(eds.), *Oxford Handbook of Economic Forecasting*, pp. 315–353. Oxford: Oxford University Press.

Castle, J. L., and Hendry, D. F. (2010). Model selection in under-specified equations with breaks. Discussion paper 509, Economics Department, Oxford University.

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

Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Clements, M. P., and Hendry, D. F. (2001). An historical perspective on forecast errors. *National Institute Economic Review*, **177**, 100–112.

Doornik, J. A. (2009). Autometrics. In Castle, J. L., and Shephard, N.(eds.), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, pp. 88–121. Oxford: Oxford University Press.

Doraszelski, U., and Pakes, A. (2007). A framework for applied dynamic analysis in IO. In Armstrong, M., and Porter, R. H.(eds.), *Handbook of Industrial Organisation: Volume 3*, pp. 1887–1964. Amsterdam: North-Holland, Elsevier. ISBN-13:978-0-444-82435-6.

Evans, G. W., and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton: Princeton University Press.

Gilbert, R. (1979). Optimal depletion of uncertain stock. *Review of Economic Studies*, **46**, 47–57.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: Evidence. *Journal of Political Economy*, **86**, 971–987.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65. Reprinted in *The Economics of Structural Change*, Hagemann, H. Landesman, M. and Scazzieri, R. (eds.), Edward Elgar, Cheltenham, 2002.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, **135**, 399–426.

Hendry, D. F., and Johansen, S. (2011). Model selection when forcing retention of theory variables. Unpublished paper, Economics Department, University of Oxford.

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 Mizon, G. E. (2000). Reformulating empirical macro-econometric modelling. *Oxford Review of Economic Policy*, **16**, 138–159.

Hendry, D. F., and Mizon, G. E. (2011a). Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics*, **3, Issue 1**. DOI: 10.2202/1941-1928.1100.

Hendry, D. F., and Mizon, G. E. (2011b). What needs re-thinking in macroeconomics?. *Global Policy*, **2, Issue 2**, 176–183. DOI: 10.1111/j.1758-5899.2011.00082x.

Kreps, D. M. (1990). *A course in microeconomic theory*. London: Harvester Wheatsheaf.

Kydland, F. E., and Prescott, E. C. (1980). Dynamic optimal taxation, rational expectations and optimal control. *Journal of Economic Dynamics and Control*, **2**, 79–91.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. In Brunner, K., and Meltzer, A.(eds.), *The Phillips Curve and Labor Markets*, vol. 1 of *Carnegie-Rochester Conferences on Public Policy*, pp. 19–46. Amsterdam: North-Holland Publishing Company.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, **41**, 867–887.

Mizon, G. E. (1995). Progressive modelling of macroeconomic time series: the LSE methodology. In Hoover, K. D. (ed.), *Macroeconometrics: Developments, Tensions and Prospects*, pp. 107–169. Dordrecht: Kluwer Academic Press.

Muellbauer, J. (2009). Household Decisions, Credit Markets and the Macroeconomy: Implications for the Design of Central Bank Models. Paper prepared for the Eigth BIS Annual Conference, 25-26 June 2009, on Financial System and Macroeconomic Resilience: Revisited, Bank of International Settlements.

Nickell, S. J. (2009). Monetary policy making in the UK. Mimeo, Nuffield College, Oxford.

Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

Stokey, N., and Lucas, R. (1989). *Recursive Methods in Economic Dynamics*. Cambridge, Massachusetts: Harvard University Press.

Tobin, J. (1950). A statistical demand function for food in the U.S.A.. *Journal of the Royal Statistical Society*, *A*, *113*(2), 113–141.

Young, P. H. (2004). *Strategic Learning and its Limits*. Oxford: Oxford University Press.