Discussion of
Elusive Return Predictability, by Allan Timmermann

David F. Hendry and James J. Reade*
Economics Department, Oxford University, UK

March 16, 2008

Get real on investment charlatans

'Economic forecasters....make forecasts not because they know, but because they are asked....providing an illusory comfort blanket.... Fulfilling a demand for bogus precision about an unknowable future is a service of sorts, even if the results rarely stand up to intensive scrutiny.... Those who think that regressions and correlations based on past data are a guide to the future....are guilty of wishful thinking.'

Jonathan Davis: *Last Word, Financial Times, June 11, 2007*.

1 Introduction

Is Allan Timmermann guilty as charged by Jonathan Davis? Are we? The above quote, drawn from a full-page article in the 'Money Section', is a strong indictment of econometric modelling for forecasting. The implicit assumption is that such forecasts are for financial markets, rather than say unemployment, GNP or inflation, where there is well-established value added–compare the Bank of England's ex ante fan chart bands against their ex post mean square forecast errors to see their accuracy up to 4 quarters ahead. On that count, **not guilty** for at least some macroeconomic time series forecasts.

However, in financial markets is the accusation also unfair? How predictable are returns on risky assets when the underlying data generating process (DGP) is evolving? Timmermann (2007) emphasizes that return predictability is elusive, so is hardly an overstated claim; and he argues that a forecasting model's best hope is 'local predictability'. Local predictability relates to the situation, where for short episodes, some model predicts well, but once that method becomes common knowledge, and perhaps even before that, any predictability is lost. Such a use of 'regressions and correlations based on past data' is quite distinct from naive 'constant model' forecasts, which may be the butt of Davis's charge.

Some impact of public forecasts on outcomes is inevitable in financial markets, and has been well known since Oscar Morgenstern's famous critique in Morgenstern (1928), as reviewed by Arthur Marget (1929), albeit that publicly credible financial forecasts are rare. 'Elusive' is indeed the key qualifier. That granted, can returns be forecast with useful accuracy by the approaches he discusses, even using alternative evaluation criteria for forecast accuracy? Empirically, none of the methods Timmermann evaluates achieves great gains, so we first reconsider the concept of an unpredictable process and note its implications. Then we address Timmermann's approach in general, rather than specifically for financial returns, as forecasting in non-stationary DGPs raises generic issues, to many of which he has made valuable contributions–see in particular Sullivan, Timmermann and White (1999), Pesaran, Pettenuzzo and Timmermann (2006), Paye and Timmermann (2006). We share the common objective of improving

*Financial support from the UK Economic and Social Research Council under grant RES-062-23-0061 is gratefully acknowledged. We are indebted to Jennie Castle and Nicholas Fawcett for helpful comments.

forecasting by confronting non-stationarity, both unit-roots and breaks, and hence concur with much of his analysis; however, our comment perforce focuses on areas where we adopt different, albeit related, approaches.

1.1 Unpredictability

A non-degenerate random variable ϵ_t is unpredictable with respect to an information set $\mathcal{I}_{t-1}$ over a time interval $\mathcal{T}$ if its distribution conditional on $\mathcal{I}_{t-1}$ is equal to its unconditional density (see Clements and Hendry, 2005):

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

Given the information set $\mathcal{I}_{t-1}$, unpredictability is intrinsic to the variable ϵ_t. Notwithstanding (1), when $\mathcal{I}_{t-1} \subset \mathcal{J}_{t-1}$, it is possible that:

$$\mathsf{D}_{\epsilon_t}(\epsilon_t \mid \mathcal{J}_{t-1}) \neq \mathsf{D}_{\epsilon_t}(\epsilon_t) \quad \forall t \in \mathcal{T}. \tag{2}$$

However, incorporating $\mathcal{J}_{t-1}$ in a model only ensures that ϵ_t is predictable when the conditioning relation is that in (2).

Inter-temporal transforms also affect predictability: if $\mathcal{I}_{t-1}$ depends on $\mathbf{X}_{t-1} = (\mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-r})$, where $\mathbf{x}'_{t-1} = (x_{1,t-1}, \ldots, x_{k,t-1})$, combinations such as:

$$y_t = g_t(\mathbf{X}_{t-1}, \boldsymbol{\beta}_t, \epsilon_t) \quad \forall t \in \mathcal{T} \tag{3}$$

are predictable. Here, y_t is the variable of interest, and $\boldsymbol{\beta}_t$ is the set of parameters characterizing the relationship $g_t(\cdot)$, which may change over time. Consequently, predictability depends on model formulations–the focus in Timmermann–to the extent that one uses:
(i) different information sets or $\mathbf{X}_{t-1}$;
(ii) different choices of the conditioning relation;
(iii) different formulations of the relationship $g_t(\cdot)$;
(iv) different specifications of the time variation in $\{\boldsymbol{\beta}_t\}$.
Even if one commenced from (3) as the initial model, there are four additional problems in forecasting y_{T+h} for $h > 0$:
(v) selecting the relevant variables in $\mathbf{X}_{t-1}$;
(vi) estimating their associated parameters;
(vii) measuring the forecast origin $\mathbf{x}_T$;
(viii) confronting further changes in $g_{T+1}(\cdot)$.

Since the stock market is intrinsically nearly unpredictable, the issue is whether returns precisely satisfy (1)–so all information about prices is immediately acted upon, generating a 'no-free-arbitrage' martingale difference–or is there a relation like (3) which could be found by an assiduous investigator even when $\mathbf{X}_{t-1}$ is uncertain and $g_t(\cdot)$ shifts over time? Assuming (3) exists, the central problems can therefore be regrouped as:
[1] the specification of $\mathcal{I}_{t-1}$ (via $\mathbf{X}_{t-1}$) at each t, leading to his considering moving sample-period windows: (i);
[2] the formulation of $g_t(\cdot)$ and $\{\boldsymbol{\beta}_t\}$, needing the discovery of and accounting for past shifts, and possibly non-linear relations: (ii)+(iii)+(iv);
[3] insurance against major losses from adopting the wrong model, leading to robust methods and attempts to forecast future breaks: (vii)+(viii);
[4] the selection of the relevant relationship from the evidence: (v); and
[5] the impact of estimation on the accuracy of the resulting forecasts: (vi).
All five issues are considered by Timmermann, but with different emphases than those we highlight below.

1.2 Forecasting in non-stationary DGPs

Timmermann refers to *forecasting methods* rather than *models*. The former explicitly comprises not only the model specification as in (3), but also the estimation technique and choice of estimation window. However, it only implicitly includes the method of selection–a key lacuna discussed below.

Given that construct, he lists a number of 'forecasting methods' for dealing with 'model instability':
[1] use rolling windows–which gets a high profile, despite his noting that there is 'no theory for how to select the length of the rolling window, nor DGPs for which this is the optimal strategy to use';
[2] (a) track evolving estimates; (b) postulate regime-shifting models; and (c) identify historical breaks;
[3] (a) model averaging is part of his 'model set', as are (b) non-linear models (such as STAR, neural nets) and (c) forecasting the best forecaster.

We focus on four areas which we consider to be under-represented, drawing on a range of recent research reported in :
[1]/[5] moving windows and estimation updating–we propose one DGP which partly supports both, namely when there are changes in collinearity;
[2] using impulse saturation to detect breaks;
]3] we note a critique of some 'popular' approaches to model averaging, the absence of other robust forecasting methods, and of attempts to forecast breaks in mean and/or variance; and
[4] model selection, where there has been immense recent progress, neither referenced nor used in his paper.

2 Rolling windows and changing collinearity

The basic analysis is well known (see Castle, Fawcett and Hendry, 2007), but the implications are new. Consider a simple conditional regression as the DGP:

$$y_t = \boldsymbol{\beta}' \mathbf{z}_t + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{4}$$

where $\mathbf{z}_t$ is independent of $\{\epsilon_t\}$ and generated by:

$$\mathbf{z}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \mathbf{\Sigma}\right] \tag{5}$$

for $\mathbf{\Sigma} = \mathbf{H}'\mathbf{\Lambda}\mathbf{H}$ with $\mathbf{H}'\mathbf{H} = \mathbf{I}_n$. The sample estimate is denoted:

$$\widehat{\boldsymbol{\beta}}_{(1,T)} = \left(\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t'\right)^{-1} \sum_{t=1}^{T} \mathbf{z}_t y_t. \tag{6}$$

We allow a break in collinearity that alters $\mathbf{\Sigma}$ at $T+1$ onwards. The resulting 1-step ahead unconditional mean-square forecast error (MSFE) for known regressors from (4) for $\widehat{\boldsymbol{\beta}}_{(1,T)}$ is (see Clements and Hendry, 2005):

$$\mathsf{E}\left[\widehat{\epsilon}_{T+1|T}^2\right] = \sigma_\epsilon^2 \left(1 + \sum_{i=1}^{n} \frac{\lambda_i^*}{T\lambda_i}\right) \tag{7}$$

where $\mathsf{E}\left[\mathbf{z}_{T+1}\mathbf{z}_{T+1}'\right] = \mathbf{\Sigma}^* = \mathbf{H}'\mathbf{\Lambda}^*\mathbf{H}$. If $\widehat{\boldsymbol{\beta}}_{(1,T)}$ is retained, (7) continues to hold at $T+2$ onwards.

However, with estimation updating, so the next forecast uses $\widehat{\boldsymbol{\beta}}_{(1,T+1)}$:

$$\mathsf{E}\left[\widehat{\epsilon}_{T+2|T+1}^2\right] = \sigma_\epsilon^2 \left(1 + \sum_{i=1}^{n} \frac{\lambda_i^*}{T\lambda_i + \lambda_i^*}\right). \tag{8}$$

The relative reduction in MSFE depends most on the smallest eigenvalue ratio, which can shift radically when there is a break in collinearity so (8) can be considerably smaller than (7).

If instead of $\widehat{\boldsymbol{\beta}}_{(1,T)}$ as in (6), $\widetilde{\boldsymbol{\beta}}_{(m,T)}$ is used, then (7) becomes:

$$\mathsf{E}\left[\widetilde{\epsilon}^2_{T+1|T}\right] = \sigma^2_\epsilon \left(1 + \sum_{i=1}^{n} \frac{\lambda^*_i}{(T-m+1)\,\lambda_i}\right) \tag{9}$$

which must be worse than (7) when $m > 1$. However at $T+1$ using $\widetilde{\boldsymbol{\beta}}_{(m+1,T+1)}$ so some of the break period is in-sample:

$$\mathsf{E}\left[\widetilde{\epsilon}^2_{T+2|T+1}\right] = \sigma^2_\epsilon \left(1 + \sum_{i=1}^{n} \frac{\lambda^*_i}{(T-m+1)\,\lambda_i + \lambda^*_i}\right). \tag{10}$$

The relative weight attributed to the second regime is clearly larger than for 'full-sample' estimates, and will grow faster as the horizon grows, although the loss of earlier information is a counter weight. This example reinforces the advantages of updating, and provides some support to rolling windows, although calculations for small $(T-m+1)$ from (10) show that a short 'window' is dreadful initially (see Castle *et al.*, 2007).

2.1 Model mis-specification

Allowing for model mis-specification makes that picture more interesting, albeit more complicated. Partition $\mathbf{x}'_t = (\mathbf{x}'_{1,t} : \mathbf{x}'_{2,t})$, where $n_1 + n_2 = n$ where $\boldsymbol{\beta}' = (\boldsymbol{\beta}'_1 : \boldsymbol{\beta}'_2)$, and the forecast only uses the former:

$$\overline{y}_{T+1|T} = \mathbf{x}'_{1,T+1}\overline{\boldsymbol{\beta}}_1, \tag{11}$$

so the forecast error is $\overline{\omega}_{T+1|T} = y_{T+1} - \overline{y}_{T+1|T}$. In an obvious notation, the unconditional MSFE is now:

$$\mathsf{E}\left[\overline{\omega}^2_{T+1|T}\right] = \sigma^2_\omega \left(1 + T^{-1}\sum_{i=1}^{n_1} \frac{\lambda^*_i}{\lambda_i}\right) + \boldsymbol{\beta}'_2 \boldsymbol{\Lambda}^*_{22} \boldsymbol{\beta}_2. \tag{12}$$

This trade-off is key to forecast-model selection. Since $\boldsymbol{\beta}'_2\boldsymbol{\Lambda}_{22}\boldsymbol{\beta}_2 = \sum_{i=n_1+1}^{n} \beta^2_i \lambda_i$, defining the non-centrality of the squared t-test of H_0: $\beta_j = 0$ by:

$$\tau^2_{\beta_j} \simeq \frac{T\beta^2_j \lambda_j}{\sigma^2_\omega}, \tag{13}$$

then from (12) the unconditional MSFE is (see Clements and Hendry, 2005):

$$\mathsf{E}\left[\overline{\omega}^2_{T+1|T}\right] \simeq \mathsf{E}\left[\widehat{\epsilon}^2_{T+1|T}\right] + \frac{\sigma^2_\omega}{T} \sum_{j=n_1+1}^{n} \left(\tau^2_{\beta_j} - 1\right) \frac{\lambda^*_j}{\lambda_j}. \tag{14}$$

Three conclusions follow from this analysis of changing collinearity when using mis-specified models, both aspects of which seem realistic. First, one minimizes $\mathsf{E}[\overline{\omega}^2_{T+1|T}]$ by eliminating all regressors where $\tau^2_{\beta_j} < 1$. Secondly $\mathsf{E}[\overline{\omega}^2_{T+1|T}]$ is possibly smaller than the MSFE of the estimated DGP equation. Thirdly, one cannot forecast better simply by dropping collinear variables if they are relevant with $\tau^2_{\beta_j} > 1$. Hence it is invaluable to correctly eliminate or retain variables when λ^*_j/λ_j is large. The conclusions of the previous section are harder to re-establish under mis-specification, but seem likely to remain relevant, namely updating helps, and a rolling window may help.

We note that discrete and intermittent shifts in $\boldsymbol{\beta}$ when $\mathsf{E}[y_t]$ stays constant would provide the strongest case for a rolling window estimator.

3 Robust strategies

There are two different reasons for seeking robustness:
(I) against measurement errors, an example of which would be the 'fundamental price' versus the observed price. Then a Holt–Winters (see e.g., Holt, 1957) or exponentially weighted moving average would be a more robust procedure.
(II) changing DGPs as is the focus here, in which case intercept corrections (ICs) and differencing are more robust (see Clements and Hendry, 1998, and Hendry, 2006).

Unfortunately, solutions for the first generally conflict with those for the second: for example, an EWMA correction for measurement errors is oppositely signed to an IC which offsets a location shift. If breaks are mainly location shifts, then relatively robust solutions exist to mitigate forecast failure (see Clements and Hendry, 2001). A 'real-world' application of such devices to forecasting discounted net TV advertising revenues 10 years ahead for pricing ITV3's licence fee renewal illustrates their potential: see http://www.ofcom.org.uk/research/tv/reports/tvadvmarket.pdf. The primary econometric system for forecasting was developed by PcGets 'quick modeller' discussed in Hendry and Krolzig (2005), but forecasting by the differenced system, which is a robust device that nevertheless retains causal effects: see Hendry (2006). The finally selected forecasts were based on an average across a small group of related methods, an issue to which we now turn.

4 Model averaging

Pooled forecasts often improve over any single method in the pool, but there are circumstances when pooling can go seriously wrong. First, if a subset of variables in a large candidate group is important, but another larger subset is irrelevant, then the weight ascribed to irrelevant components in the pooled outcome biases the coefficients of the relevant–in essence, poor models are accorded too much 'importance' overall even though individually each receives little weight: see Hendry and Reade (2007a, 2007b). Secondly, when there are location shifts, in-sample causal relations need not help in forecasting, although there may be offsetting effects that help an average (see Hendry and Clements, 2004). However, Clements and Hendry (1999) propose a number of simple forecasting models that are robust in the presence of past breaks, such as differenced models, but these are often not considered in the model pool. Further, if a past break has been detected and modelled by an indicator as one of the candidate variables, then averaging can be much worse than the general model, again by according too much weight to the half of the models without the indicator.

To avoid such mis-performance when forecasting, it seems essential to select a small subset of 'good' models for the average, retaining some robust predictors in 'averaged set'. Our approach to selection is explained in section 6.

5 Detecting breaks by impulse saturation

The 'impulse saturation' strategy proposed by Hendry, Johansen and Santos (2004) adds T indicator variables to a regression for an estimation sample of size T by first adding half the indicators (equivalent to just 'dummying out' $T/2$ observations), recording any significant ones (see Salkever, 1976), then omitting these, and adding the other half, recording significance again. Finally, the two sets of significant indicators are combined to remove outliers and location shifts. At significance level α, then αT 'null' indicators will be selected on average: for $\alpha = 0.01$, and $T = 100$, then just 1 indicator will be retained by chance. A feasible algorithm is discussed in Hendry and Krolzig (2005), formalized in Hendry *et al.*

(2004), and implemented in *Autometrics* (see Doornik, 2007). Alternative splits, such as $T/3$, $T/4$, etc. do not affect the null retention rate, nor do unequal splits.

There is an important difference between outlier detection and impulse saturation, illustrated using *Autometrics* in figure 1. The top-left panel shows the absence of outliers at $\alpha = 0.025$ when the fitted model is just the mean and the DGP is standardized white noise with a location shift of 10 error standard deviations (σ) for 30 periods. The top-right panel confirms that all the residuals lie within $\pm 2\widehat{\sigma}$. By way of contrast, the lower-left panel shows the detection of the shift after impulse saturation, again at $\alpha = 0.025$ confirmed in the lower-right panel as capturing all the shifts. This technique is part of an automatic general-to-specific (Gets) modelling methodology which also has implications for forecasting.

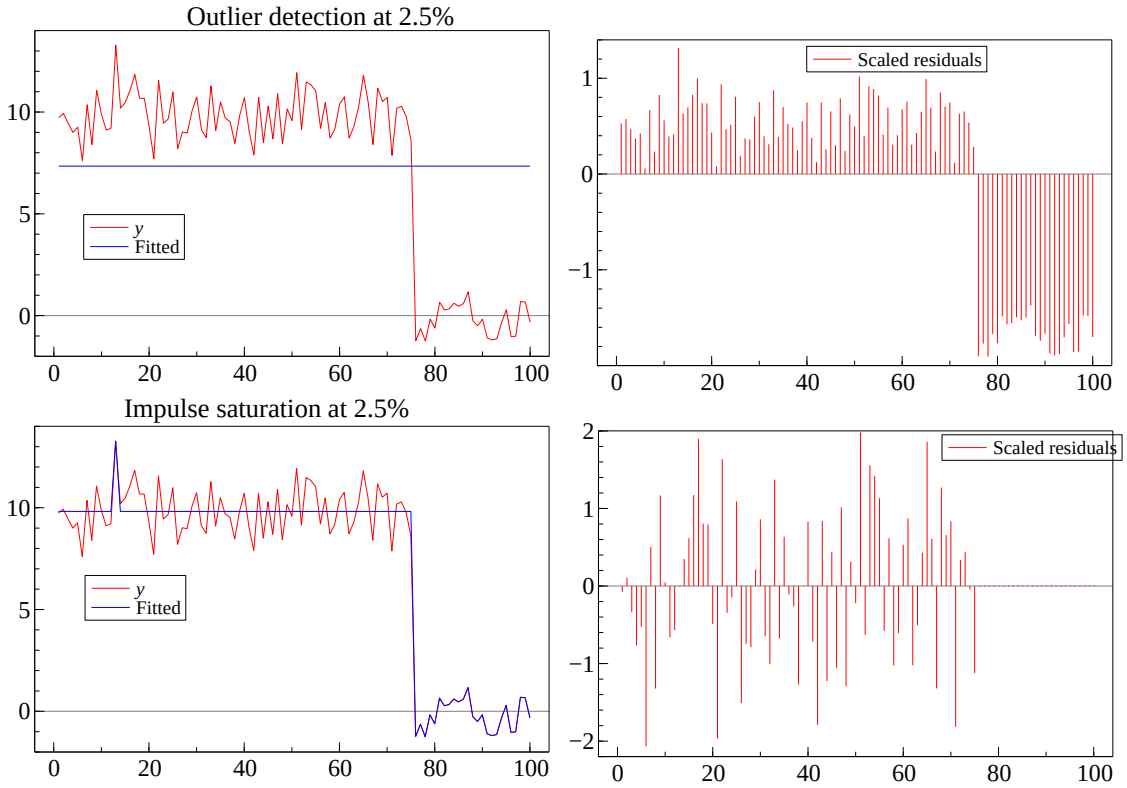

Figure 1: Absence of outliers despite a break

6 Model selection

Several references to BIC in his paper suggest that Timmermann believes BIC is a sensible selection criterion in the general DGPs he considers. We are unaware of evidence that BIC is a good selection method in the face of non-constancy of means, variances, and other features of distributions such as shape, systematic nature of the time series, volatility clustering etc. all of which could apply in the present setting. We believe that automatic general-to-specific Gets has much to offer, especially when impulse saturation is part of the procedure, as there is a crucial difference between selecting *models* and selecting *variables*. Given an initial set of k candidate regressors, there are 2^k possible models. Many selection methods such as Raffalovich, Deane, Armstrong and Tsao (2001) involve selecting models, as do the basic selection algorithms in the model averaging applications of Raftery, Madigan and Hoeting (1997). If $k = 1000$, that induces 10^{301} possible models; and it is hard to believe that any selection technique will choose accurately between these (not to mention it being a prohibitive task). Conversely, when $\alpha = 0.001$, only one irrelevant variable will be retained by chance even if none actually mattered;

yet any relevant variables with sample t-statistics in excess of about 3.3 in absolute value will be retained. Selecting variables by a process of reduction has seen a large amount of research since the selection algorithms of Hoover and Perez (1999) and Hendry and Krolzig (2005), culminating in automated model selection packages such as *PcGets* (Hendry and Krolzig, 2001) and *Autometrics*. These algorithms can be analyzed theoretically and checked by Monte Carlo simulation to calibrate significance levels for the selection rules and mis-specification tests to ensure they result in minimal undominated congruent selections. Hendry and Krolzig (2005) also discuss how to bias correct the parameter estimates after selection, which could be valuable when forecasting.

To illustrate, consider the perfectly orthogonal regression model:

$$y_t = \boldsymbol{\beta}' \mathbf{z}_t + \epsilon_t \tag{15}$$

where $\mathsf{E}[\mathbf{z}_t \mathbf{z}_t'] = \mathbf{\Lambda}$ is an $k \times k$ diagonal matrix and $T >> k$. Order the k sample t^2-statistics testing H_0: $\beta_j = 0$:

$$\mathsf{t}^2_{(k)} \geq \mathsf{t}^2_{(k+1)} \geq \cdots \geq \mathsf{t}^2_{(1)}.$$

The cut-off n between included and excluded variables is:

$$\mathsf{t}^2_{(n)} \geq c_\alpha > \mathsf{t}^2_{(n+1)},$$

so all larger t^2 values are retained and smaller are eliminated. Thus, only one decision is needed even for $k = 1000$, and 'repeated testing' does not occur although path searches give the impression of 'repeated testing'. Despite the potential for 10^{301} possible models, the false null retention rate can be maintained at r variables by setting $\alpha = r/k$, so $1/1000$ for $r = 1$.

7 Conclusions

Allan Timmermann has made a brave effort to forecast the near unpredictable. His research takes a number of useful steps down the road to understanding the general problem of forecasting in non-stationary time series, especially where time invariance is unlikely. We hope our analysis adds to that understanding. Practitioners who seek to employ his ideas for other potentially forecastable economic time series have much to draw on, but, in selecting a viable forecasting methodology, might also consider the other research avenues sketched in our response.

References

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2007). Forecasting, structural breaks and nonlinearities. Mimeo, 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). 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.

Doornik, J. A. (2007). Autometrics. Working paper, Economics Department, University of Oxford.

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

Hendry, D. F., and Clements, M. P. (2004). Pooling of forecasts. *Econometrics Journal*, **7**, 1–31.

Hendry, D. F., Johansen, S., and Santos, C. (2004). Selecting a regression saturated by indicators. Unpublished paper, Economics Department, University of Oxford.

Hendry, D. F., and Krolzig, H.-M. (2001). *Automatic Econometric Model Selection with PcGets*. London: Timberlake Consultants Press.

Hendry, D. F., and Krolzig, H.-M. (2005). The properties of automatic Gets modelling. *Economic Journal*, **115**, C32–C61.

Hendry, D. F., and Reade, J. J. (2007a). Model selection and robust forecast combinations. Working paper, Economics Department, Oxford University.

Hendry, D. F., and Reade, J. J. (2007b). Modelling and forecasting using model averaging. Working paper, Economics Department, Oxford University.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. Onr research memorandum 52, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

Hoover, K. D., and Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, **2**, 167–191.

Marget, A. W. (1929). Morgenstern on the methodology of economic forecasting. *Journal of Political Economy*, **37**, 312–339. Reprinted in Hendry, D. F. and Morgan, M. S. (1995), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press; and in T.C. Mills (ed.), *Economic Forecasting*. Edward Elgar, 1999.

Morgenstern, O. (1928). *Wirtschaftsprognose: eine Untersuchung ihrer Voraussetzungen und Möglichkeiten*. Vienna: Julius Springer.

Paye, B., and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, **13**, 274–315.

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

Raffalovich, L., Deane, D., Armstrong, D., and Tsao, H.-S. (2001). Model selection procedures in social science research: Monte-Carlo simulation results. Working paper 2005/16, center for social and demographic analysis, State University of New York, Albany.

Raftery, A., Madigan, D., and Hoeting, J. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, **92**, 179–191.

Salkever, D. S. (1976). The use of dummy variables to compute predictions, prediction errors and confidence intervals. *Journal of Econometrics*, **4**, 393–397.

Sullivan, R., Timmermann, A., and White, H. (1999). Data-snooping, technical trading rules and the bootstrap. *Journal of Finance*, **54**, 1647–1692.

Timmermann, A. (2007). Elusive return predictability. *International Journal of Forecasting*. Forthcoming.