ICMA Centre

Henley Business School
UNIVERSITY OF READING

Discussion Paper

Sir Clive W J Granger's Contributions to Forecasting

October 2016

Michael P Clements
ICMA Centre, Henley Business School, University of Reading

Discussion Paper Number:
ICM-2016-09

The aim of this discussion paper series is to disseminate new research of academic distinction. Papers are preliminary drafts, circulated to stimulate discussion and critical comment. Henley Business School is triple accredited and home to over 100 academic faculty, who undertake research in a wide range of fields from ethics and finance to international business and marketing.

admin@icmacentre.ac.uk

www.icmacentre.ac.uk

Sir Clive W.J. Granger's Contributions to Forecasting

Michael P. Clements*
ICMA Centre,
Henley Business School,
University of Reading,
Reading RG6 6BA

October 18, 2016

Abstract

Some of Clive Granger's many and varied contributions to economic forecasting are reviewed. These include contributions to forecast combination and forecast efficiency, to improving forecast practice, to forecast evaluation, and to the theory of forecasting. We also discuss some of the subsequent research and developments in these areas, which have sought to generalize the applicability of Granger's work. We also consider research in related areas motivated at least in part by Granger's work.

*Michael Clements is Professor of Econometrics at the ICMA Centre, Henley Business School, University of Reading, and an associate member of the Institute for New Economic Thinking at the Oxford Martin School, University of Oxford. Email: m.p.clements@reading.ac.uk.

I am grateful to David Hendry for helpful comments and suggestions.

1 Introduction

Professor Clive Granger published extensively on forecasting. This paper is a personal reflection on some of that work, and on the influence it has had on research on forecasting and the subsequent development of the subject. The selection of topics in part reflects the author's own interests. It does not set out to provide a comprehensive account of Clive Granger's contributions to forecasting, or to catalogue his research output on this topic. Some of these contributions were truly ground-breaking. It covers forecast combination in section 2, improving forecast practice in section 3, forecast evaluation in section 4, and forecasting white noise in section 5.

2 Forecast Combination

Barnard (1963) pre-dates the seminal contribution of Bates and Granger (1969), but the latter proposed a more general way of combining different forecasts of the same quantity to improve predictive accuracy. Barnard (1963) had considered taking a simple average of a Box-Jenkins model forecast, and an exponential smoothing model forecast, and showing that the resulting 1-step ahead forecasts of world airline passenger miles per month has a smaller error variance than either set of forecasts alone. Bates and Granger (1969) generalized the simple average to allow convex combinations of two (or more) forecasts, and to allow the relative weights given to each to be chosen optimally based on the past performance of the forecasts.

To illustrate, suppose there are two h-steps-ahead forecasts, f_{1t} and f_{2t}, of the quantity y_t. Assuming the forecasts to be unbiased, i.e. that the forecast errors $e_{it} = y_t - f_{it}$ $(i = 1, 2)$ have zero mean, Bates and Granger (1969) suggest the use of a combined forecast, f_{ct}, of the form:

$$f_{ct} = (1 - \lambda) f_{1t} + \lambda f_{2t}. \tag{1}$$

When $0 \leq \lambda \leq 1$, f_{ct} comprises a simple weighted average of the two individual forecasts, and the weighting parameter λ can be selected based on the relative accuracy of the individual forecasts f_{1t} and f_{2t}.

If the forecast error associated with f_{ct} is denoted by $\varepsilon_t = y_t - f_{ct}$, then the expected squared forecast error of the combined forecast is given by:

$$E(\varepsilon_t^2) = (1 - \lambda)^2 \sigma_1^2 + \lambda^2 \sigma_2^2 + 2\lambda(1 - \lambda)\rho\sigma_1\sigma_2 \tag{2}$$

where σ_1^2 and σ_2^2 denote, respectively, the expected squared errors of f_{1t} and f_{2t}, and ρ denotes the correlation between the forecast errors e_{1t} and e_{2t}. The optimal combination weight associated with a squared error loss function is then derived by choosing λ to minimize (2), i.e.:

$$\lambda_{opt} = \arg\min_{\lambda} \left\{ E(\varepsilon_t^2) \right\} = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}. \tag{3}$$

The expected squared error associated with the optimal combination weight λ_{opt} is given by:

$$E(\varepsilon_t^2\,(\lambda_{opt})) = \frac{\sigma_1^2\sigma_2^2\,(1 - \rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

where, of necessity, $E(\varepsilon_t^2\,(\lambda_{opt})) \leq \min\left\{\sigma_1^2, \sigma_2^2\right\}$. Suppose that f_1 and f_2 are equally accurate, i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Then:

$$E(\varepsilon_t^2\,(\lambda_{opt})) = \sigma^2 \frac{(1 + \rho)}{2}.$$

Given that $|\rho| \leq 1$, then the expected squared error associated with the optimal forecast is less than either individual forecast for all values of ρ other than $\rho = 1$. So there are diversification gains in general, and in particular when the forecasts are equally accurate, unless the forecasts are perfectly correlated.

In practice the optimal weight parameter, and its constituent parameters ρ, σ_1^2 and σ_2^2, will be estimated from the past record of forecasts and outcomes. Denoting the time series of past h-steps-ahead forecast errors by $e_{1t}, e_{2t},\ t = 1, ..., n$, the obvious sample estimator of the population weight parameter (3) is given by:

$$\hat{\lambda}_{opt} = \frac{\sum_{t=1}^{n} e_{1t}^2 - \sum_{t=1}^{n} e_{1t}e_{2t}}{\sum_{t=1}^{n} e_{1t}^2 + \sum_{t=1}^{n} e_{2t}^2 - 2\sum_{t=1}^{n} e_{1t}e_{2t}}. \tag{4}$$

This estimated weight can then be used in the future to produce out-of-sample combined forecasts. Bates and Granger (1969) are aware that the relative performance of the component forecasts may not be constant over time, and suggest a number of ways of allowing time dependence in $\hat{\lambda}$ based on the past forecast errors.

An alternative but equivalent way of estimating λ is via the regression method of Granger and Ramanathan (1984), that is, from ordinary least squares estimation of:

$$e_{1t} = \lambda(e_{1t} - e_{2t}) + \varepsilon_t \tag{5}$$

or, equivalently,

$$y_t = (1 - \lambda)f_{1t} + \lambda f_{2t} + \varepsilon_t. \tag{6}$$

From (6) it is immediately apparent that ε_t is the forecast error of the forecast combination. Hence (6) (or equivalently (5)) implies that the forecast error is uncorrelated with the forecast combination (the explanatory variable) by construction. However, ε_t need not be uncorrelated with either f_{1t} or f_{2t} individually, and it may be optimal to allow non-convex combinations.

The implicit assumption behind taking convex combinations is that the forecasts are efficient in the sense of Mincer and Zarnowitz (1969). Efficiency requires that $\alpha = 0$ and $\beta = 1$ in the realization–forecast regression $y_t = \alpha + \beta f_t + \varepsilon_t$ (see, e.g., Clements and Hendry (1998, Ch. 3) for a discussion), which also serves as sufficient condition for unbiasedness. When this holds for each forecast individually, i.e., $y_t = f_{it} + \varepsilon_{it}$, then (6) is warranted. But otherwise, bias can be easily accommodated by including an intercept in (5) or (6):

$$e_{1t} = \alpha + \lambda(e_{1t} - e_{2t}) + \varepsilon_t$$

resulting in the unbiased combination:

$$f_{ct} = \alpha + (1 - \lambda)f_{1t} + \lambda f_{2t}. \tag{7}$$

More generally, forecast inefficiency suggests relaxing the assumption that the combination weights sum to one (as advocated by Granger and Ramanathan (1984)):

$$f_{ct} = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t} \tag{8}$$

with weights calculated from:

$$y_t = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t} + \varepsilon_t. \tag{9}$$

Clearly (7) and (1) are special cases of (8), where the restrictions $\beta_1 + \beta_2 = 1$, and $\alpha = 0, \beta_1 + \beta_2 = 1$ are imposed, respectively. When the actuals and forecasts are non-stationary integrated processes, (9) could be specified using actual and predicted changes, rather than levels, so that the variables are $I(0)$, and so tests will have their standard distributions, but the outcome of the test for forecast encompassing may not be invariant to this change. An alternative, which may be preferable when the data and forecasts are $I(1)$, is suggested by Ericsson (1992). This amounts to regressing the forecast error for one of the two competing forecasts on the difference between the two forecasts. Provided the forecasting models are reasonably well specified, in the sense that both forecasts are cointegrated with the actuals (such that the forecast errors are $I(0)$), both the dependent variable and the regressor will be $I(0)$.

There have been a large number of extensions to the analysis discussed here, including: allowing for autocorrelation in ε_t when estimating combination weights (e.g., Diebold (1988) and Coulson and Robins (1993)); allowing weights to vary over time (e.g., Diebold and Pauly (1987), LeSage and Magura (1992), and Deutsch, Granger and Teräsvirta (1994)); simple averaging versus estimating weights (e.g., Makridakis (1983), Stock and Watson (1999), Fildes and Ord (2002) and Genre, Kenny, Meyler and Timmermann (2013)); Bayesian combination methods (e.g., Clemen and Winkler (1986), Diebold and Pauly (1990) and Min and Zellner (1993)); combination for interval, density and probability forecasts (e.g., Wallis (2005), Mitchell and Hall (2005) and Clements and Harvey (2010)) and nonlinear combinations (e.g., via Artificial Neural Networks, Donaldson and Kamstra (1996)). Many Handbooks provide chapters reviewing this literature, including Newbold and Harvey (2002), Timmermann (2006), Clements and Harvey (2009) and Aiolfi, Capistrán and Timmermann (2011).

The notion of conditional efficiency was developed by Nelson (1972) and Granger and Newbold (1973) to denote forecasts which could not be made more accurate by combination with another forecast. More precisely, a forecast f_1 is said to be conditionally efficient with respect to f_2 if the optimal weight on f_2 in a combination with f_1 is zero. Chong and Hendry (1986) interpreted conditional efficiency in terms of the wider concept of encompassing (see, *inter alia*, Mizon (1984), Mizon and Richard (1986) and Hendry and Richard (1989)) as forecast encompassing (see, e.g., Ericsson and Marquez (1993), Andrews, Minford and Riley (1996) and Harvey, Leybourne and Newbold (1998) for further developments and applications).

Finally, Granger (1989) considers the relationship between pooling forecasts and pooling information. The Bates and Granger (1969) example considered below of combining linear and exponential trend models of an output index assumes both models have access to the same information set (just the variable in question, the output index), and the potential for combination arises through the use of different model specifications. Alternatively, the more generic case is perhaps where different models (or forecasters) have access to different information sets. Granger supposes each forecaster $j = 1, \ldots, J$ has access to a specific information set, $\mathcal{I}_{j,t}$, and all share the common information, $\mathcal{I}_{o,t}$. We assume each information set consists of a single variable, and its lags, and all these variables are uncorrelated with each other (at all leads and lags), and with the common variable known to all: $\mathcal{I}_{j,t} = \{x_{j,t}, x_{j,t-1}, \ldots\}$, $\mathcal{I}_{o,t} = (x_t, x_{t-1}, \ldots)$, and $Cov\left(x_{j,t}, x_{i,s}\right) = 0$ for all t and s when $i \neq j$, and $Cov\left(x_t, x_{i,s}\right) = 0$ for all i, s and t.

Given all the information, $\Omega_t = \{\mathcal{I}_{1,t}, \mathcal{I}_{2,t}, \ldots, \mathcal{I}_{J,t}, \mathcal{I}_{o,t}\}$, suppose the optimal 1-step ahead forecast takes the form $E\left(y_{t+1}|\Omega_t\right) = \alpha\left(L\right) x_t + \sum_{j=1}^{J} \beta_j\left(L\right) x_{j,t}$. Each forecaster will report their conditional expectation, given by: $E\left(y_{t+1}|\mathcal{I}_{j,t}\right) = \alpha\left(L\right) x_t + \beta_j\left(L\right) x_{j,t}$. Faced with the J-forecasts, the equally-weighted forecast combination is given by:

$$f_{c,t+1} = \alpha\left(L\right) x_t + \frac{1}{J} \sum_{j=1}^{J} \beta_j\left(L\right) x_{j,t}.$$

so that the individual $x_{j,t}$ are weighted by J^{-1} (as opposed to 1 in the 'pooling of information' case), and for large J, $f_{c,t+1}$ will approach $f_{o,t+1} = E\left(y_{t+1}|\mathcal{I}_{o,t}\right) = \alpha\left(L\right) x_t$. If $f_{o,t+1}$ were available, then the combination $Jf_{c,t+1} - \left(J-1\right) f_{o,t+1}$ matches the pooling of information.

This shows that aggregating information, $E\left(y_{t+1}|\Omega_t\right)$, is not equivalent to aggregating forecasts, $f_{c,t+1}$. This holds in general and not just for equally-weighted forecasts. The problem is that the forecast combination does not use the information efficiently. In the simple example considered here, a second-level of combination with the new forecast $f_{o,t+1}$ results in efficient combination relative to the pooling of information.

There is now a large literature on forecasting when there are many potential predictors. Stock and Watson (2003) is a good example of this literature, showing that combinations of individually unstable forecasts offer some improvement in forecast accuracy over univariate benchmarks. Clements and Galvão (2006) directly compared combining forecasts and combining information in modelling. Factor models are a way of using the information from many predictor variables in a forecasting model (see, e.g., Forni, Hallin, Lippi and Reichlin (2000) and Stock and Watson (2011)), and De Mol, Giannone and Reichlin (2008) for an alternative approach.

We conclude the discussion of forecast combination with i) a re-analysis of an empirical example of forecast combination in Bates and Granger (1969), based on Hendry and Clements (2004), followed by ii) a brief review of recent work allowing for more general loss functions, that is, going beyond the squared-error loss function used by Bates and Granger (1969) and Granger and Ramanathan (1984).

2.1 Forecast Combination Example

Hendry and Clements (2004) re-visit the forecast combination example of Bates and Granger (1969, Table A1, p. 462), which considers linear and exponential trend models of an output index for the gas, electricity and water sector. Table 1 records the output index for the years 1948 to 1965, along with forecast errors from linear and exponential trend models of output $\{y_t\}$, given by:

$$\begin{aligned} y_t &= \alpha + \beta t + error_t, \\ ln\left(y_t\right) &= a + bt + error_t \end{aligned}$$

where t is a linear time trend. The forecast errors in each period t ($t = 1950, \ldots, 1965$) are for forecasts based only on models estimated up to $t-1$. The exponential model forecasts are clearly superior on the sum of squared errors (SSE), and therefore on commonly-used forecast accuracy measures such as the (root) mean squared forecast error. Nevertheless, a combination of the exponential and linear trend models has a smaller SSE. A weight of 0.16 on the linear forecasts results in a combined forecast SSE of 78.8, compared to an SSE for the exponential trend model of 84.4. [1]

Hendry and Clements (2004) consider forecast combination from the perspective of the theory of forecasting enunciated in Clements and Hendry (1998) and Clements and Hendry (1999) (and subsequently refined in, e.g., Castle, Clements and Hendry (2016)). This theory replaces the traditional assumptions of i) the model being essentially complete and correctly-specified for the variables of interest and ii) that model remaining constant over the period being forecast, with their obverses: a) models are incomplete and incorrect in many ways, and b) the processes being modelled will typically exhibit evolution over time as well as abrupt shifts and changes. Their paper suggests a rationale for forecast combination in general terms of the component models being differentially susceptible to structural breaks.

In terms of the output index illustration, it is evident that the 'constant absolute increase' implication of the linear trend model is inappropriate - the forecast errors become large and positive from around 1961 onwards. On average, the exponential model generates negative errors, and combination is seen to work by averaging the over-predictions of the more-accurate exponential model with the under-predictions of the linear trend model over the 1955 − 61 period. Suppose the component forecasts are first bias corrected. The results of bias-correcting the individual forecasts and their SSEs are shown in the last

[1] The numbers are the calculations of Hendry and Clements (2004), who calculate the forecasts and statistics reported in the table from the actual series. Small differences relative to Bates and Granger's figures were attributed to improved precision.

Table 1: Forecast errors of output indices, 1950–65

	Actual	1-step forecast errors				
		Linear	Exponential	Combination	Linear Bias-corrected	Exponential Bias-corrected
1948	58.0					
1949	62.0					
1950	67.0	1.0	0.7	0.77	1.0	0.7
1951	72.0	0.7	0.1	0.21	-0.3	-0.6
1952	74.0	-2.5	-3.4	-3.24	-3.3	-3.8
1953	77.0	-2.2	-3.3	-3.11	-1.9	-2.4
1954	84.0	2.1	0.8	0.99	2.8	2.2
1955	88.0	1.0	-0.6	-0.37	1.2	0.4
1956	92.0	0.4	-1.7	-1.33	0.4	-0.7
1957	96.0	0.0	-2.5	-2.08	-0.0	-1.4
1958	100.0	-0.2	-3.2	-2.71	-0.3	-2.0
1959	103.0	-1.3	-4.8	-4.28	-1.4	-3.4
1960	110.0	1.9	-2.1	-1.47	2.0	-0.3
1961	116.0	3.2	-1.4	-0.71	3.1	0.4
1962	125.0	7.0	1.8	2.60	6.7	3.5
1963	133.0	8.8	2.8	3.74	8.0	4.3
1964	137.0	6.1	-0.9	0.26	4.7	0.3
1965	145.0	8.0	-0.0	1.26	6.3	1.1
Sample bias		2.1	-1.1	-0.6	1.8	-0.1
Sum of squared errors		263.3	84.4	78.8	211.9	77.0

The output series is the output index for the gas, electricity and water sector, given in Bates and Granger (1969, Table A1, p. 462). The combination forecast has fixed weights of 0.16 and 0.84 on the (uncorrected) linear and exponential forecasts

two columns of the table. The bias-correction is calculated in real time, in the sense that the forecast of period t is calculated by adding the sample mean of the forecast errors up to period $t-1$. This is a feasible correction, in that it uses only information available at each forecast origin, but it results in slow adaptation of the forecasts to past systematic errors. Nevertheless, the SSE of the bias-corrected exponential forecasts is 77, less than the combined forecast SSE of 78.8 (with a weight of 0.16). Any fixed-weight combination of the bias-corrected forecasts, with weights in the interval $(0, 1)$, has a larger SSE than that of the exponential model forecasts. The optimal weight on the linear forecasts (after bias-correcting, and imposing the constraint that they sum to unity) was -0.22, with an SSE of 72.61. Negative weights may appear anomalous, but see Timmermann (2006). The fixed-weight combination forecasts are not feasible, as they are calculated based on the full set of forecast errors, and as noted above, Bates and Granger (1969) suggest time-varying weight schemes.

The Hendry and Clements (2004) extension to the forecast combination example serves to illustrate that gains from combination may result from models with manifestly mis-specified deterministic components. This is consistent with the primacy of breaks or shifts in deterministic factors in causing forecast failure, and the closely-related effects of the mis-specification of such components (see, e.g., the forecast-error taxonomy in Clements and Hendry (2006)). We consider this example again in section 3 in the context of intercept correction.

2.2 Loss Functions and Forecast Combination

The early work on forecast combination of Bates and Granger (1969) and Granger and Ramanathan (1984) assumed combination weights chosen to minimize a symmetric, squared-error loss function, and out-of-sample combinations of forecasts would typically be assessed in terms of the (R)MSFE, the empirical counterpart of expected squared-error loss. The earlier analysis has subsequently been generalized to allow for asymmetric loss by Elliott and Timmermann (2004). For general loss functions and forecast error distributions, the optimal combination weights can be shown to depend on higher-order moments of the forecast error distribution, although under certain restrictions on the form of the forecast error distribution, the optimal combination weights on the individual forecasts are identical to the squared-error loss weights for almost all loss functions - only the value of the constant term in the combination will differ. Beginning with the joint distribution of the actual and forecasts $(y_t, \mathbf{f}_t)'$, Elliott and Timmermann (2004) show that the restrictions require that the marginal distribution of the forecast errors depends only on the first two moments of the forecast errors. This holds when the joint distribution is elliptically symmetric (which includes the multivariate normal and t-distributions). Under these assumptions, the squared-error loss weights of Bates and Granger (1969) remain relevant, and only the constant term differs.

To understand this result, we sketch the analysis in Elliott and Timmermann (2004). The notation for the first two moments of the actuals and forecasts are given by:

$$E\begin{pmatrix} y_t \\ \mathbf{f}_t \end{pmatrix} = \begin{pmatrix} \mu_y \\ \boldsymbol{\mu} \end{pmatrix}, \quad Cov\begin{pmatrix} y_t \\ \mathbf{f}_t \end{pmatrix} = \begin{pmatrix} \sigma_y^2 & \boldsymbol{\sigma}_{21}' \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Write the forecast combination error as $e_t = y_t - \beta_0 - \boldsymbol{\beta}'\mathbf{f}_t$ to define the combination weight vector as $\boldsymbol{\beta}$, and the intercept in the combination by β_0. Then the mean and variance of the combination forecast error are given by:

$$\mu_e = \mu_y - \beta_0 - \boldsymbol{\beta}'\boldsymbol{\mu} \tag{10}$$

$$\sigma_e^2 = \sigma_y^2 + \boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{\sigma}_{21}. \tag{11}$$

Suppose the loss function is $L(e_t)$, where $L(e_t) = e_t^2$ gives the standard squared-error loss. The forecast combination is defined by $(\beta_0, \boldsymbol{\beta})$, and the optimal combination minimizes the expected loss, $E[L(e_t)]$, i.e.,:

$$\min_{\beta_0, \boldsymbol{\beta}} \int L(e_t)\, dF(e_t).$$

Under elliptical symmetry, we can write $E[L(e_t)] = g(\mu_e, \sigma_e^2)$. From (10) and (11), only μ_e depends on β_0. Thus the first order condition for minimizing $E[L(e_t)]$ with respect to β_0 is:

$$\frac{\partial g(\mu_e, \sigma_e^2)}{\partial \beta_0} = \frac{\partial g(\mu_e, \sigma_e^2)}{\partial \mu_e}\frac{\partial \mu_e}{\partial \beta_0} = 0.$$

As $\frac{\partial \mu_e}{\partial \beta_0} = -1$, the optimal value for β_0, β_0^*, solves $\frac{\partial g(\mu_e, \sigma_e^2)}{\partial \mu_e} = 0$. β_0^* depends on $L(.)$, and is set to generate the optimal amount of bias (μ_e^*) given the form of $L(.)$. For squared-error loss, $E[L(e_t)] = \mu_e^2 + \sigma_e^2$, and $\frac{\partial g(\mu_e, \sigma_e^2)}{\partial \beta_0} = -2\mu_e$, so that the optimal amount of bias is of course zero $(\mu_e^* = 0)$.

Consider the first order condition with respect to $\boldsymbol{\beta}$:

$$\frac{\partial g(\mu_e, \sigma_e^2)}{\partial \boldsymbol{\beta}} = \frac{\partial g(\mu_e, \sigma_e^2)}{\partial \sigma_e^2}\frac{\partial \sigma_e^2}{\partial \boldsymbol{\beta}} = 0.$$

Provided $\frac{\partial g(\mu_e, \sigma_e^2)}{\partial \sigma_e^2} \neq 0$, $\frac{\partial g(\mu_e, \sigma_e^2)}{\partial \boldsymbol{\beta}} = 0$ implies that $\frac{\partial \sigma_e^2}{\partial \boldsymbol{\beta}} = 0$, so from (11), $2\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}^* = \mathbf{2}\boldsymbol{\sigma}_{21}$, and $\boldsymbol{\beta}^* = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$ irrespective of the form of $L(.)$, matching the expression for squared-error loss.

As Elliott and Timmermann (2004) remark, if an element of $\boldsymbol{\beta}^*$ is zero under squared-error loss, then the corresponding forecast will also receive zero weight under any other loss function, assuming that the stated properties of the forecast error distribution hold.

3 Improving Forecasting Practice

Granger wrote or co-authored a number of papers on forecasting practice, indicating how the then current approaches might be improved, both from a technical perspective, and in terms of how forecasts are presented. A number of recommendations were made which helped foster marked improvements in the practice of forecasting, and forecasting research.[2] For example, Granger and Newbold (1973) are critical of statistical forecast evaluation criteria which are not monotonic functions of squared-error loss, such as the 'inequality coefficient' (or first U-statistic) of Theil (1958). They also criticise the practice of comparing the distributional and time-series properties of the actual and forecast series, and suggest instead a consideration of the forecast error, and especially the one-step ahead error, as this has readily-testable properties under forecast optimality (e.g., unbiasedness, serially uncorrelatedness: but see section 4.2). When feasible, forecasts should be evaluated in terms of the 'expected utility' resulting from actions or decisions taken on the basis of those forecasts, as in Granger and Pesaran (2000b, 2000a) (discussed in section 4.3).

Granger and Newbold (1973, 1975) suggested econometricians should pay more attention to time-series models, arguing that 'Econometricians rarely, if ever, consider the problem of forecasting a time series in terms of its current and past values....if they did so they might learn a number of valuable lessons applicable to the more sophisticated model building exercises they attempt' (Granger and Newbold (1973)). It is not sufficient to simply show that 'econometric forecasts' outperform 'extrapolative forecasts' (i.e. Box-Jenkin forecasts), but the econometric forecasts should be conditionally efficient (as defined in section 2) with respect to multivariate Box-Jenkins forecasts. There is a call for more stringent comparators than 'no change' or 'same change' predictors, and an emphasis on the forecasts embodying all the useful information in the 'purely statistical' forecasting devices. It is argued that econometric model forecasts would typically contain extraneous, non-numerical information which could not be easily accommodated in the Box-Jenkins models, so that such models would necessarily be at an advantage, and simply being more accurate would not be sufficiently demanding.

The call for greater emphasis on the dynamic relationships between variables would appear to have been answered in Sims (1980), and the subsequent popularity of vector autoregressions in macro-econometric modelling and forecasting (see, e.g., Doan, Litterman and Sims (1984)).

Some 20 years later, Granger (1996) argued that it was usually the case that i) forecasts were published with no indication of the level of uncertainty, ii) there was no recognition of the differing degrees of difficulty in forecasting different variables, iii) or of whether the forecasts were conditional or unconditional, iv) or of the role of data revisions and seasonal adjustment, v) or of the techniques used, the assumptions made, or the information set, vi) that there was a tendency of forecasters to 'all swing together', and vii) and a tendency to under-estimate change.

Since then, there has been substantial progress on many of these. For example, there has been much work on the calculation and representation of forecast uncertainty (i), both in academia and amongst policy makers (see, e.g., Reifschneider and Tulip (2007), Knüppel (2014) and Haddow, Hare, Hooley and Shakir (2013)), and in terms of the effects of data revisions (iv), facilitated by the availability of ready-made real-time datasets (Croushore and Stark (2001)): see, for example, the review chapters by Croushore (2011a, 2011b). There has also been research on the intersection of these two - looking at the impact of data revisions on forecast uncertainty: Clements (2015). There has also been considerable progress on other entries in Granger's list.

Granger applauded the use of past forecast errors as a method of intercept-correction, or putting the forecast 'back-on-track', as advocated in Clements and Hendry (1996) and discussed more fully in Clements and Hendry (1999, ch. 6), as a way of improving forecasting practice.

[2] We consider one of his recommendations separately in section 5

The role of intercept correction in mitigating the effects of structural breaks and dynamic mis-specification are explained by Clements and Hendry (1999) as follows. Suppose the forecaster assumes y_t is generated by

$$y_t = \mu + u_t \text{ where } u_t \sim N\left(0, \sigma_u^2\right) \tag{12}$$

and μ is estimated from a sample of size T by least squares, and used to forecast:

$$y_{T+1|T} = \widehat{\mu} = T^{-1} \sum_{t=1}^{T} y_t. \tag{13}$$

When (12) is the actual process that generates the data, $y_{T+1|T}$ is the estimated conditional expectation, $E\left(y_{T+1} \mid y_T\right) = \mu$. But suppose there is error autocorrelation, as in the linear trend model of section 2, and in addition a shift in $E\left(y_t\right)$ at time T_1, since the actual data generating mechanism is:

$$y_t = \mu + \delta 1_{\{t \geq T_1\}} + \rho y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim IN\left(0, \sigma_\epsilon^2\right), \tag{14}$$

where $|\rho| < 1$, and $1_{\{t \geq T_1\}} = 0$ for $t < T_1$, and otherwise $1_{\{t \geq T_1\}} = 1$. Hence $E\left(y_{T+1}\right) = \left(\mu + \delta\right) / \left(1 - \rho\right)$ and $E\left(y_{T+1} \mid y_T\right) = \mu + \delta + \rho y_T$, but the mis-specified model has a forecast mean of:

$$E\left(\widehat{\mu}\right) = T^{-1} \left(\sum_{t=1}^{T_1} E\left(y_t\right) + \sum_{t=T_1+1}^{T} E\left(y_t\right) \right) = \frac{\mu + \delta}{1 - \rho} - \kappa \frac{\delta}{1 - \rho},$$

where $\kappa = T^{-1} T_1$, and $T_1 \leq T$. When $T_1 = T$, $\kappa = 1$, and $E\left(\widehat{\mu}\right) = \mu \left(1 - \rho\right)^{-1}$, and the in-sample estimate reflects none of the changed intercept.

Setting the model 'back on track' adds to the forecast the forecast-origin error, $\widehat{u}_T = y_T - \widehat{\mu}$ (and so fits the last observation perfectly), yielding:

$$\widehat{y}_{\iota,T+1} = \widehat{\mu} + \widehat{u}_T = y_T. \tag{15}$$

Note that:

$$E\left(\widehat{y}_{\iota,T+1}\right) = E\left(y_T\right) = \frac{\mu + \delta}{1 - \rho},$$

which is unconditionally unbiased, despite the model being mis-specified. Further:

$$\begin{aligned} y_{T+1} - \widehat{y}_{\iota,T+1} &= \mu + \delta + \left(\rho - 1\right) y_T + \epsilon_{T+1} \\ &= \left(\rho - 1\right) \left(y_T - \frac{\mu + \delta}{1 - \rho} \right) + \epsilon_{T+1}, \end{aligned}$$

so that the unconditional expected squared-error is:

$$E\left[\left(y_{T+1} - \widehat{y}_{\iota,T+1}\right)^2\right] = \sigma_\epsilon^2 + \left(1 - \rho\right)^2 Var\left(y_T\right) = \frac{2\sigma_\epsilon^2}{1 + \rho}, \tag{16}$$

as against the minimum obtainable (under correct specification, a known break, and known parameters) of σ_ϵ^2.

Intercept correction in this example yields unbiased forecasts, with an increase in the forecast-error variance which is decreasing in ρ, provided $\rho > 0$, that is, the variance cost is decreasing in the degree of dynamic mis-specification. In general, for models which are not dynamically-mis-specified, the overall efficacy of intercept correction in terms of expected squared error will rely on a favourable tradeoff between an inflated forecast-error variance and reduced (squared) bias.

In terms of the output index illustration recorded in table 1, it is apparent that the linear trend model generates a sequence of positive forecast errors over the second half of the period (even after bias-correcting), which is consistent with an inappropriate specification of the deterministic term andor a shift in the trend function. Either way, the above analysis suggests the use of an intercept correction. Adding in the last error reduces the linear model (bias-corrected) sum of squares from the value of 211.9 shown in the table to 121.1. This strategy is not successful for the non-linear trend model - the sum of squares is increased from 77.0 to 111.4. The forecasts from this model do not exhibit systematic bias, and there is therefore no bias offset to the higher forecast-error variance.

4 Forecast Evaluation

4.1 Generalized Cost of Error

Granger (1969) was a key paper in the development of a prediction theory which generalizes the quadratic forecast loss functions of classical prediction theory. Granger argued that in practice - at least in the fields of economic and management - the 'cost of error' may not be proportional to the squared forecast error, and cited two motivating examples.[3] Granger comes up with the practical recommendation that when the loss function is symmetric (but not quadratic), one may use the least-squares predictor (that is, proceed as one would under squared-error loss), and that when the loss function is non-symmetric, a simple constant bias term added to the predictor will be a reasonable solution. More precisely, assuming Gaussianity, and a symmetric loss function, the least squares predictor will not be optimal but will constitute an 'efficient and computationally simple method'. If the loss function is non-symmetric, the least squares predictor can be adjusted by a constant amount at a second stage. The constant is chosen to minimize the sum of the losses over a sample of actual values and (least-squares) forecasts. Granger suggests these prescriptions may be sensible even when the data are non-Gaussian. When the data are non-Gaussian, the optimal predictor will not necessarily be a linear function of the past data, and so the use of the linear least squares predictor will be non-optimal. Building on the earlier work of Whittle (1963), Granger argues the use of linear predictors may yield reasonable approximations, particularly if there is no information on the form of the non-linearity (which we interpret as absent knowledge of the relevant non-linear model of the sort surveyed in Granger and Teräsvirta (1993)).

Granger (1969) details the derivation of the optimal predictor for a non-symmetric linear loss function. Firstly, in terms of the notation of that paper, the optimal predictor h, where h is some function of the data $X_t, X_{t-1}, \ldots$, is given by:

$$\begin{aligned} E\left[g\left(X_{t+k}-h\right) \mid X_t, X_{t-1}, \ldots\right] &= E_c\left[\left(g\left(X_{t+k}-h\right)\right)\right] && (17) \\ &= \int_{-\infty}^{\infty} g\left(x-h\right) f_c\left(x\right) dx && (18) \end{aligned}$$

where $g\left(\right)$ is the loss function, and f_c is the conditional density function of X_{t+k}. For squared-error loss, $g\left(x\right)=x^2$, and writing $\left(X_{t+k}-h\right)=\left(X_{t+k}-M\right)+\left(M-h\right)$, where $M=E_c\left(X_{t+k}\right)$ is the conditional

[3] In the first a bank decides how large a computer to buy to handle its current accounts, with an over-prediction of the bank's future business resulting in too large and expensive a computer, and an under-prediction a computer unable to handle all the accounts, and Granger suggests there is no reason to expect under- and over-predictions of the same magnitude would be equally costly. The second example concerns a bank acting as an issuing agent for a share issue, where too high a price leaves the bank liable to purchase the shares at a high price and too low a price will reduce the amount of money made by the issuing firm. Again, there is no reason to think the costs are symmetric. These examples suggest in real-world situations costs are unlikely to be symmetric (or, presumably, necessarily quadratic), but Granger was also aware that 'real-world cost functions are rarely available'. Granger (1993, p.651)

expectation:

$$
\begin{aligned}
E_c\left[\left(X_{t+k}-h\right)^2\right] &= E_c\left[\left(X_{t+k}-M\right)^2\right]+E_c\left[\left(M-h\right)^2\right]+2E_c\left[\left(X_{t+k}-M\right)\left(M-h\right)\right] \\
&= E_c\left[\left(X_{t+k}-M\right)^2\right]+E_c\left[\left(M-h\right)^2\right]
\end{aligned}
$$

(the last term in the first line is necessarily zero - the conditional-expectation forecast error is uncorrelated with forecast origin functions M and h). The expression is minimized by $h = M$, showing that the conditional expectation is the optimal predictor for squared-error loss. When X_t is Gaussian, M is a linear function of $X_t, X_{t-1}, \ldots$, and $f_c(x)$ is normal with mean M.

When loss is non-symmetric but linear, e.g.,

$$
\begin{aligned}
g(x) &= ax, \quad x>0, \quad a>0 \\
&= 0, \quad x=0 \\
&= bx, \quad x<0 \quad b<0
\end{aligned}
$$

from (18) Granger obtains:

$$
E_c\left[\left(g\left(X_{t+k}-h\right)\right)\right]=a\int_h^{\infty}(x-h)\,f_c(x)\,dx+b\int_{-\infty}^{h}(x-h)\,f_c(x)\,dx.
$$

Setting the derivative with respect to h to zero implicitly defines h via $F_c(h) = a(a-b)^{-1}$, where $F_c(x)$ is the c.d.f. of X_{t+k}. Assuming Gaussianity, $h = M + \alpha$ where α does not depend on $X_t, X_{t-1}, \ldots$. Granger's suggestion is that in practice it may generally be reasonable to assume Gaussianity holds, in which case the optimal predictor is a simple constant adjustment to the conditional expectation, as here.

Granger's ideas have been extended to allow for processes which are conditionally-Gaussian, which allow for time-varying forecast-error variances instead of a constant conditional variance, allowing for ARCH and GARCH processes (e.g., Engle (1982) and Bollerslev (1986)). Christoffersen and Diebold (1997) show that allowing for time-varying forecast-error variances gives rise to an adjustment which is no longer constant, but instead depends on the forecast variance of the process. Consider the 'linex' loss function of Varian (1975), which is a popular choice in the literature as it permits a closed-form solution for the optimal predictor:

$$
C\left(e_{t+k|t}\right)=b\left[\exp\left(ae_{t+k|t}\right)-ae_{t+k|t}-1\right], \quad a\neq 0,\, b\geq 0
$$

where $e_{t+k|t}$ is the forecast error, and the notation makes explicit the target period and forecast origin as $t+k$ and t, respectively. For $a > 0$, the loss function is approximately *lin*ear for $e_{t+k|t} < 0$ ('over-predictions'), and *ex*ponential for $e_{t+k|t} > 0$, ('under-predictions'), and conversely for $a < 0$. Assume the process being forecast is conditionally Gaussian:

$$
y_{t+k} \mid \mathcal{I}_t \sim N\left(y_{t+k|t}, \sigma^2_{t+k|t}\right),
$$

then the optimal predictor $\hat{y}_{t+k|t}$ can be shown to be:

$$
\hat{y}_{t+k|t}=y_{t+k|t}+\frac{a}{2}\sigma^2_{t+k|t}. \tag{19}
$$

where $y_{t+k|t} = E\left(y_{t+k} \mid \mathcal{I}_t\right)$ is the conditional expectation, and the 'adjustment term' $\frac{a}{2}\sigma^2_{t+k|t}$ depends on the the degree of asymmetry a, the forecast horizon k, and the past data (the last two through the forecast of the variance, $\sigma^2_{t+k|t}$). In Granger's setup the assumption that $\sigma^2_{t+k|t} = \sigma^2_k$ gives rise to a constant adjustment (given a and k). For small a, the loss function is approximately quadratic (from a

Taylor-series expansion of $C(e)$ about $e = 0$) , $C(e) \simeq \frac{ba^2}{2}e^2$, and so $\hat{y}_{t+k|t} \to y_{t+k|t}$ as the degree of asymmetry lessens.

Recently Patton and Timmermann (2007) have shown that essentially the same results hold for general asymmetric loss functions, under relatively weak conditions on the form of the loss function and the data generating process. The condition on the data generating process is that the variable of interest is conditionally location-scale distributed, whereas the loss function must be homogeneous in the forecast error.

Formally, we need to assume:

$$y_{t+k} \mid \mathcal{I}_t \sim D\left(y_{t+k|t}, \sigma^2_{t+k|t}\right),$$

for some constant distribution function D, where, as above, $y_{t+k|t} = E(y_{t+k} \mid \mathcal{I}_t)$ and $\sigma^2_{t+k|t} = Var(y_{t+k} \mid \mathcal{I}_t)$ and the loss function satisfies:

$$L\left(a \cdot e_{t+k|t}\right) = g(a)\, L\left(e_{t+k|t}\right),$$

for some positive function g, and all $a \neq 0$. Patton and Timmermann (2007, Proposition 2) show that the optimal forecast is given by:

$$\hat{y}_{t+k|t} = y_{t+k|t} + \phi_h \sigma_{t+k|t} \tag{20}$$

where ϕ_h is a constant that depends on the form of D and L.[4]

From (20) it follows immediately that:

$$E\left(y_{t+k} - \hat{y}_{t+k|t} \mid \mathcal{I}_t\right) = E\left[y_{t+k} - \left(y_{t+k|t} + \phi_h \sigma_{t+k|t}\right) \mid \mathcal{I}_t\right] = -\phi_h \sigma_{t+k|t}$$

so that optimal forecasts are (conditionally) biased. But although the bias of a rational forecaster should depend on the forecast standard deviation, it should not depend on other variables known at time t. This suggests testing for rational expectations with asymmetric losses by running a regression such as:

$$e_{t+k|t} \equiv y_{t+k} - \hat{y}_{t+k|t} = \zeta_1 \sigma_{t+k|t} + \boldsymbol{\zeta}_2' \mathbf{Z}_t + \varepsilon_{t+k} \tag{21}$$

where $\mathbf{Z}_t$ is a vector of variables known at time t, $\mathbf{Z}_t \subset \mathcal{I}_t$. Under the null of efficient use of information - in the sense that forecasts cannot be systematically improved using forecast-origin information - we would expect to find $\boldsymbol{\zeta}_2 = 0$, but $\zeta_1 \neq 0$ if loss is asymmetric: see e.g., Pesaran and Weale (2006).

Asymmetric loss has been widely used in a number of contexts: to consider whether apparently biased and inefficient forecasts are consistent with rational behaviour (see, e.g., Elliott, Komunjer and Timmermann (2005, 2008)); as an explanation of the observed dispersion of inflation expectations (resulting from heterogeneity in the degree of asymmetry of individuals' loss functions) (Capistrán and Timmermann (2009)); and as a possible explanation of apparent inconsistencies between survey respondents' point predictions and probability distributions (see, e.g., Clements (2009, 2014)); amongst many others.

Hendry and Mizon (2014) stress that the calculation of the optimal predictor requires knowledge of the conditional distribution over which the integration is assumed to be carried out. For example, the calculation of $y_{t+k|t} = E(y_{t+k} \mid \mathcal{I}_t)$ requires knowledge of the conditional distribution of y_{t+k} given $\mathcal{I}_t$. This requires stationarity, so that the distribution can be calculated from past data. In the event of shifts in the underlying distributions, expectations based on now outdated distributions will not be 'optimal', and may create a role for the robust forecasting devices recently reviewed by Castle, Clements and Hendry (2015) and Castle *et al.* (2016).

[4] Homogeneity of the loss function rules out linex loss. A practical implication of adopting the Patton and Timmermann (2007) framework rather than linex loss (together with the assumption that the data generating process is conditionally normal, in order to obtain an expression for the optimal predictor) is that the optimal predictor should depend linearly on the conditional standard deviation, rather than the conditional variance.

4.2 Properties of Optimal Forecasts

Granger (1999) recognized that the standard properties of optimal forecasts only applied for a squared-error loss function, and outlined a forecast theory for generalized cost functions. Under squared-error loss, it is straight forward to show that optimal forecasts are unbiased, that the forecast-error variance is monotonically non-decreasing in the forecast horizon (at least in population), and that k-step ahead forecast errors can be written as a moving-average process which is at most of order $k-1$ (and thus one-step forecast errors are serially-uncorrelated, as alluded to in section 3). These theoretical properties can be used to test the optimality of empirical forecasts. However, under more general loss functions than squared-error loss, these properties may no longer apply. As an example, the section 4.1 shows that optimal forecasts will typically be biased under asymmetric loss: equation (19) shows that the optimal predictor differs from the conditional expectation - the unbiased optimal predictor under squared-error loss.

Given a general cost function $C(e)$ defined on the forecast error e, Granger suggests taking the derivative of the loss function with respect to the forecast, evaluated at the forecast error corresponding to the optimal forecast. It is straightforward to derive some of the properties of this 'generalized forecast error' from the first-order condition which defines the optimal predictor. The optimal predictor is defined by:

$$\underset{\hat{y}_{t+1|t}}{\arg\min} \int C\left(y_{t+1} - \hat{y}_{t+1|t}\right) dP_{t+1|t}\left(y_{t+1}\right)$$

where $P_{t+1|t}$ is the conditional c.d.f. The first-order condition is:

$$\int C' dP_{t+1|t}\left(y_{t+1}\right)$$

Granger sets $Z_{t+1|t} = C'\left(y_{t+1} - \hat{y}_{t+1|t}\right)$, and it follows immediately that $Z_{t+1|t}$ is conditionally unbiased $E\left(Z_{t+1|t} \mid \mathcal{I}_t\right) = 0$, and therefore unconditionally unbiased, $E\left(Z_{t+1|t}\right) = 0$. Moreover, $E\left(Z_{t+1|t} \mid \mathcal{I}_t\right) = 0$ implies that:

$$E\left(Z_{t+1|t} W_t \mid \mathcal{I}_t\right) = 0$$

where $W_t \subset \mathcal{I}_t$ (and more generally $E\left(Z_{t+1|t} g(W_t) \mid \mathcal{I}_t\right) = 0$) for any finite function $g(.)$), which suggests that in the test regression:

$$Z_{t+1|t} = \alpha + \beta W_t + \varepsilon_{t+1}$$

under optimality both $\alpha = 0$ and $\beta = 0$. A valid choice for (an element of) W_t would be the explanatory variable $\sigma_{t+k|t}$ used in the test regression (21), for example, given that the forecast standard deviation is based on information in $\mathcal{I}_t$. Hence the switch from the forecast error to the derivative of the cost function allows simple tests of whether the forecast efficiently uses forecast-origin information. Since $Z_{t|t-1} \subset \mathcal{I}_t$, and $e_{t|t-1} = y_t - \hat{y}_{t|t-1} \subset \mathcal{I}_t$, $Z_{t+1|t}$ should be uncorrelated with both $Z_{t|t-1}$ and $e_{t|t-1}$.

Granger shows that the results generalize beyond the one-step horizon, so that:

$$E\left(Z_{t+k|t} g(W_t) \mid \mathcal{I}_t\right) = 0$$

showing unbiasedness (by setting $g(W_t) = 1$) and forecast efficiency (by setting $g(W_t) = \hat{y}_{t+k|t}$). In addition, $Z_{t+k|t}$ will be autocorrelated at most up to $k-1$, matching the standard result for the forecast error from an optimal forecast under squared-error loss.

Patton and Timmermann (2003) provide some extensions, and Artis and Marcellino (2001) an empirical application.

4.3 A Decision-theoretic Approach to Forecast Evaluation

Granger and Pesaran (2000b, 2000a) suggest, where possible, evaluating forecasts in terms of the expected economic value emanating from actions taken based on those forecasts. The key ideas can be explained in terms of a simple two state - two action model. The two possible states in period $t+1$ are "Bad" ($s_{t+1}=1$) and "Good" ($s_{t+1}=0$), and there are two possible actions open to the decision-maker in period t. To take action, "Yes", indicated by $y_t=1$, or to decline to take action, "No", $y_t=0$. Thus, actions are taken in advance. The payoff matrix associated with this decision problem is given in table 2. In the Payoff Matrix, $U_{t+1,by}$ represents the decision maker's utility if the b̲ad state occurs after the y̲es decision is taken, and so on.

Table 2: Payoff Matrix for a Two-State, Two-Action Decision Problem

		States (s_{t+1})	
		Bad $(s_{t+1}=1)$	Good $(s_{t+1}=0)$
Decisions (y_t)	Yes $(y_t=1)$	$U_{t+1,by}$	$U_{t+1,gy}$
	No $(y_t=0)$	$U_{t+1,bn}$	$U_{t+1,gn}$

Table 2 reports the payoffs (U) from combining outcomes (or realized states) and (prior) actions to economic values. A forecast probability at period t of the Bad event occurring in period $t+1$ can be used by the decision maker of a particular forecast probability to determine their action. Let π_{t+1} be the actual probability that $s_{t+1}=1$, $\pi_{t+1}=\Pr(s_{t+1}=1)$, and $\hat{\pi}_{t+1}$ the forecast probability. Probabilities of states are assumed independent of actions. Then, the expected utility of taking action ($y_t=1$) based on the forecast probabilities is given by:

$$U_{t+1,by}\hat{\pi}_{t+1}+U_{t+1,gy}\left(1-\hat{\pi}_{t+1}\right) \tag{22}$$

and of not acting:

$$U_{t+1,bn}\hat{\pi}_{t+1}+U_{t+1,gn}\left(1-\hat{\pi}_{t+1}\right). \tag{23}$$

The decision maker will taken the action ($y_t=1$) if (22) exceeds (23), that is, if:

$$\hat{\pi}_{t+1}>q_{t+1}$$

where:

$$q_{t+1}=\frac{\delta_{t+1,g}}{\delta_{t+1,g}+\delta_{t+1,b}},\text{ and } \delta_{t+1,g}=U_{t+1,gn}-U_{t+1,gy},\ \ \delta_{t+1,b}=U_{t+1,by}-U_{t+1,bn}.$$

It is assumed that $q_{t+1}>0$, because $\delta_{t+1,b}>0$ and $\delta_{t+1,g}>0$ by assumption. $U_{by}>U_{bn}$ because 'action' will alleviate the costs incurred in the bad state (having gritted the roads when there is a frost will reduce the number of road traffic accidents). $U_{gn}>U_{gy}$ because gritting is costly and unnecessary when there is not a frost.

Hence the decision rule is $y_t^*=1(\hat{\pi}_{t+1}>q_{t+1})$. The economic benefit that accrues at period $t+1$ will depend on which state materialises and the action taken at period t:

$$\begin{aligned} v_{t+1}\left(y_t;s_{t+1}\right) &= U_{t+1,by}s_{t+1}y_t+U_{t+1,gy}\left(1-s_{t+1}\right)y_t \\ &\quad +U_{t+1,bn}s_{t+1}\left(1-y_t\right)+U_{t+1,gn}\left(1-s_{t+1}\right)\left(1-y_t\right) \end{aligned}$$

Using the optimal decision rule:

$$
\begin{aligned}
\upsilon_{t+1}\left(y_t^*; s_{t+1}\right) &= U_{t+1,by} s_{t+1} y_t^* + U_{t+1,gy}\left(1 - s_{t+1}\right) y_t^* \\
&\quad + U_{t+1,bn} s_{t+1}\left(1 - y_t^*\right) + U_{t+1,gn}\left(1 - s_{t+1}\right)\left(1 - y_t^*\right). \qquad (24)
\end{aligned}
$$

If we substitute $y_t^* = 1(\hat{\pi}_{t+1} > q_{t+1})$, we obtain υ_{t+1} as a function of the forecast probability, $\upsilon_{t+1}\left(\hat{\pi}_{t+1}; s_{t+1}\right)$:

$$
\upsilon_{t+1}(\hat{\pi}_{t+1}; s_{t+1}) = a_{t+1} + b_{t+1}(s_{t+1} - q_{t+1})1\left(\hat{\pi}_{t+1} > q_{t+1}\right), \qquad (25)
$$

where $a_{t+1} = s_{t+1}U_{t+1,bn} + (1 - s_{t+1})U_{t+1,gn}$ and $b_{t+1} = U_{t+1,by} - U_{t+1,bn} + U_{t+1,gn} - U_{t+1,gy}$. Since the part of the economic value given by a_{t+1} does not depend on the probability forecast estimate, $\hat{\pi}_{t+1}$, it can be ignored when comparing two or more rival forecast probabilities (say, $\widehat{\pi}_{t+1}$ and $\widetilde{\pi}_{t+1}$).

The expected economic value of using the probability forecast $\hat{\pi}_{t+1}$ is given by

$$
E\left[\upsilon_{t+1}(\hat{\pi}_{t+1}; s_{t+1}) \mid \Omega_t\right] = E\left(a_{t+1} \mid \Omega_t\right) + b_{t+1}\left(\pi_{t+1} - q_{t+1}\right)1(\hat{\pi}_{t+1} > q_{t+1}) \qquad (26)
$$

where expectations are taken with respect to the true conditional probability distribution of s_{t+1}, denoted by $E\left(\cdot \mid \Omega_t\right)$, and $\pi_{t+1} = E\left(s_{t+1} \mid \Omega_t\right) = \Pr\left(s_{t+1} = 1 \mid \Omega_t\right)$, $1 - \pi_{t+1} = \Pr\left(s_{t+1} = 0 \mid \Omega_t\right)$.

Ignoring the dependence of probabilities on actions, from (26) the part of the expected economic value that depends on the probability forecast $\hat{\pi}_{t+1}$ is given by:

$$
E\left[\upsilon_{t+1}(\hat{\pi}_{t+1}, s_{t+1}) \mid \Omega_t\right] = b_{t+1}\left(\pi_{t+1} - q_{t+1}\right)1(\hat{\pi}_{t+1} > q_{t+1}). \qquad (27)
$$

Suppose we have a set of probability forecasts and states for $t = 1, \ldots, T$, then the expectation in (27) can be evaluated by averaging over the observations to give the average realized economic value:

$$
\upsilon = \frac{1}{T}\sum_{t=1}^{T} b_t\left(s_t - q_t\right)1\left(\widehat{\pi}_t > q_t\right), \qquad (28)
$$

where π_{t+1} is replaced by the binary indicator s_t, given that the true probabilities are not observed.

The Kuipers score (Ks) is defined as:

$$
Ks = H - F
$$

where H is the 'hit rate', the proportion of the total number of Bad states that were correctly forecast, and F is the 'false alarm' rate, defined as the proportion of the total number of Good states that were incorrectly forecast as being Bad states. The advantage of the Ks statistic over measures such as the quadratic and log probability scores (QPS and LPS) is that always forecasting the Bad state to occur (or always forecasting the Good state) will score zero, whereas such strategies may fare well on QPS and LPS. The Ks evaluates forecasts of events rather than forecasts of the probabilities of events. The Bad state is forecast to occur when $1\left(\hat{\pi}_{t+1} > q_{t+1}\right) = 1$. We can express H and F as:

$$
H = \frac{\sum_{t=1}^{T} s_t 1\left(\hat{\pi}_t > q_t\right)}{\sum_{t=1}^{T} s_t}, \quad F = \frac{\sum_{t=1}^{T}\left(1 - s_t\right)1\left(\hat{\pi}_t > q_t\right)}{\sum_{t=1}^{T}\left(1 - s_t\right)}.
$$

Granger and Pesaran (2000b) show that in some circumstances KS, a nominally purely statistical evaluation criterion, and the economic value criterion, are proportional to each other, that is:

$$
\upsilon = b\overline{s}\left(1 - \overline{s}\right)Ks
$$

where $\overline{s} = T^{-1}\sum_{t=1}^{T} s_t$, the estimate of the (unconditional) probability of the Bad state. To obtain this expression, the decision problem has to be simplified by assuming that $b_t = b$, all t, $q_t = q = \overline{s}$, all t.

Clements (2004) discusses the application of the decision-based approach to the UK Monetary Policy Committee's forecasts of inflation.

5 Forecasting White Noise

Granger (1983) shows that white noise processes can be forecastable. This suggested that the Box-Jenkins (Box and Jenkins (1970)) time-series modelling approach might not fully exploit the predictability of the time series. The Box-Jenkins approach consisted of fitting an autoregressive-moving average (ARMA) model to the series y_t where the AR and MA lag polynomials were of high enough order to result in a white noise error term. That is, an error term ε_t with no discernible structure, such that the conditional expectation is zero, $E\left(\varepsilon_t \mid \mathcal{I}_{t-1}\right) = 0$, where $\mathcal{I}_t = \left(y_t, y_{t-1}, \ldots\right)$ (or, equivalently, $\mathcal{I}_t = \left(\varepsilon_t, \varepsilon_{t-1}, \ldots\right)$). At the time, autocorrelation and partial autocorrelation functions were used to identify the form of the model (that is, the orders of the AR and MA polynomials), and to diagnose any systematic patterns in the model's error terms, so the emphasis was on linear dependence in the time series, and on any non-modelled linear dependence in the error term after fitting the selected ARMA model.

Granger argued that modelling a time-series up to a white noise error may still leave room for improvement in terms of forecasting, and argued for the development of new techniques able to detect non-linear dependence.

We will illustrate with two examples taken from Granger (1983). The first is a simple linear model where the use of a wider information set than that suggested by the Box-Jenkins approach would yield more accurate forecasts, and the second is an example of a non-linear model generating more accurate forecasts.

In the first, the variable y_t is generated by:

$$y_t = x_{t-1} + \varepsilon_t$$

where x_t and ε_t are independent white noise. All the autocovariances of y_t are zero, so that y_t is white noise. For example, consider $Cov(y_t, y_{t-1})$:

$$\begin{aligned} Cov\left(y_t, y_{t-1}\right) &= Cov\left(x_{t-1} + \varepsilon_t, x_{t-2} + \varepsilon_{t-1}\right) \\ &= Cov\left(x_{t-1}, x_{t-2} + \varepsilon_{t-1}\right) + Cov\left(\varepsilon_t, x_{t-2} + \varepsilon_{t-1}\right) \end{aligned}$$

which is zero because $Cov\left(x_t, x_{t-s}\right) = Cov\left(\varepsilon_t, \varepsilon_{t-s}\right) = 0$ for all t and s, other than $s = 0$, by virtue of x and ε being white noise, and because $Cov\left(x_t, \varepsilon_{t-s}\right) = 0$ for all t and s by virtue of x and ε being independent. Moreover, all correlations of y_t are zero, and not just the first.

Next, consider the expected squared forecast error when the information set $\mathcal{I}_t$ includes x_t, so that the optimal forecast is $E\left(y_t \mid \mathcal{I}_{t-1}\right) = x_{t-1}$. Then:

$$E\left[\left(y_t - E\left(y_t \mid \mathcal{I}_{t-1}\right)\right)^2\right] = E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2.$$

Given that y_t is white noise, the appropriate ARMA model is an ARMA$(0,0)$, i.e., $y_t = u_t$, where u_t is a white noise error. For this model, the information set is the null set, $E\left(y_t \mid \emptyset\right) = 0$, and the forecast-error variance is:

$$E\left(u_t^2\right) = \sigma_\varepsilon^2 + E\left(x_{t-1}^2\right).$$

Hence the population ARMA model's forecasts are inferior to the model which conditions on x_{t-1}.

Granger defines the 'usefulness' of a forecasting model as:

$$R^2 = 1 - \frac{Var(e_t)}{Var\,(y_t)}$$

where $e_t = y_t - E\,(y_t \mid \mathcal{I}_{t-1})$, and in this example, $0 < R^2 < 1$, whereas the equivalent R^2 for the ARMA$(0,0)$ is zero.

Clements and Hendry (2005) discuss the role of information in economic forecasting, and establish that 'more information unambiguously does not worsen predictability, even in intrinsically non-stationary processes', where a variable is said to be predictable 'if its distribution conditional on $\mathcal{I}_{t-1}$ differs from its unconditional distribution'. Hence their definition of predictability corresponds to Granger's notion of 'usefulness'. For a stationary process, $Var\,(e_t) = Var\,(y_t)$ when the forecast (the conditional expectation) is equal to the unconditional mean, since $E\,(y_t \mid \mathcal{I}_{t-1}) = E\,(y_t)$ implies $Var\,(e_t) = E\left[(y_t - E\,(y_t \mid \mathcal{I}_{t-1}))^2\right] = E\left[(y_t - E\,(y_t))^2\right] = Var\,(y_t)$. Clements and Hendry (2005) go on to further refine their notions of predictability and forecastability when models are mis-specified and there is non-constancy (that is, given the worldview of Clements and Hendry (1998, 1999)).

The second example is a bilinear model (see Granger and Andersen (1978)):

$$y_t = \beta y_{t-2}\varepsilon_{t-1} + \varepsilon_t$$

One can show that $0 < R^2 < 0.5$ when the model is invertible ($\beta\sigma_\varepsilon < 0.707$), so that although y_t appears to be white noise it is forecastable non-linearly. Related examples are discussed, including:

$$y_t = \beta x_{t-1}y_{t-1} + u_t \tag{29}$$

where x_t and u_t are Gaussian white noise. Then it follows that the autocorrelation function of y_t is everywhere zero, indicating white noise, and moreover $Cov\,(y_t, x_{t-s}) = 0$ for all s, indicating that y_t cannot be forecast linearly from x_t. However, as indicated by (29), y is forecastable from past values of y and x.

Granger and Teräsvirta (1993) provide an extensive treatment of non-linear models and forecasting, and Granger's work on non-linear models is the topic of the contribution by Jennifer L. Castle and David F. Hendry.

6 Further Reading

This review does not aim to be comprehensive. Topics which are not covered include the following. The concept of 'time distance' as an alternative to the standard approach to evaluating forecasts based on the 'vertical distance', see: Granger and Jeon (2003a, 2003b). The forecasting of transformed series, in Granger and Newbold (1976), and of the Box-Cox transformation, in Nelson and Granger (1979). Finally, we have not discussed Granger's work on stock market price predictability and the efficient markets hypothesis (see, Granger and Morgenstern (1970), Timmermann and Granger (2004) and Granger (1992)).

Mills (2013) provides a readable account of Granger's work, not confined to forecasting.

References

Aiolfi, M., Capistrán, C., and Timmermann, A. (2011). Forecast combinations, chapter 11. In Clements, M. P., and Hendry, D. F. (eds.), *The Oxford Handbook of Economic Forecasting*, pp. 355–388: Oxford University Press.

Andrews, M. J., Minford, A. P. L., and Riley, J. (1996). On comparing macroeconomic forecasts using forecast encompassing tests. *Oxford Bulletin of Economics and Statistics*, **58**, 279–305.

Artis, M., and Marcellino, M. (2001). Fiscal forecasting: the track record of the IMF, OECD and EC. *Econometrics Journal*, **4**, S20–S36.

Barnard, G. A. (1963). New methods of quality control. *Journal of the Royal Statistical Society, A*, **126**, 255–259.

Bates, J. M., and Granger, C. W. J. (1969). The combination of forecasts. *Operations Research Quarterly*, **20**, 451–468. Reprinted in T.C. Mills (ed.), *Economic Forecasting*. Edward Elgar, 1999.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **51**, 307–327.

Box, G. E. P., and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

Capistrán, C., and Timmermann, A. (2009). Disagreement and biases in inflation expectations. *Journal of Money, Credit and Banking*, **41**, 365–396.

Castle, J. L., Clements, M. P., and Hendry, D. F. (2015). Robust approaches to forecasting. *International Journal of Forecasting*, **31**, 99–112.

Castle, J. L., Clements, M. P., and Hendry, D. F. (2016). An overview of forecasting facing breaks. *Journal of Business Cycle Research*, **12(1)**, 3–23. DOI 10.1007/s41549-016-0005-2.

Chong, Y. Y., and Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, **53**, 671–690. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.

Christoffersen, P. F., and Diebold, F. X. (1997). Optimal prediction under asymmetric loss. *Econometric Theory*, **13**, 808–817.

Clemen, R. T., and Winkler, R. L. (1986). Combining economic forecasts. *Journal of Business and Economic Statistics*, **4**, 39–46.

Clements, M. P. (2004). Evaluating the Bank of England density forecasts of inflation. *Economic Journal*, **114**, 844 – 866.

Clements, M. P. (2009). Internal consistency of survey respondents' forecasts: Evidence based on the Survey of Professional Forecasters. In Castle, J. L., and Shephard, N. (eds.), *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry. Chapter 8*, pp. 206–226. Oxford: Oxford University Press.

Clements, M. P. (2014). US inflation expectations and heterogeneous loss functions, 1968–2010. *Journal of Forecasting*, **33(1)**, 1–14.

Clements, M. P. (2015). Assessing macro uncertainty in real-time when data are subject to revision. *Journal of Business & Economic Statistics*. Forthcoming.

Clements, M. P., and Galvão, A. B. (2006). Combining predictors and combining information in modelling: Forecasting US recession probabilities and output growth. In Milas, C., Rothman, P., and van Dijk, D. (eds.), *Nonlinear Time Series Analysis of Business Cycles. Contributions to Economic Analysis Series*, pp. 55–73: Elsevier.

Clements, M. P., and Harvey, D. I. (2009). Forecasting combination and encompassing. In Mills, T. C., and Patterson, K. (eds.), *Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics*, pp. 169–198: Palgrave MacMillan.

Clements, M. P., and Harvey, D. I. (2010). Forecast encompassing tests and probability forecasts. *Journal of Applied Econometrics*, **25**, 1028–1062.

Clements, M. P., and Hendry, D. F. (1996). Intercept corrections and structural change. *Journal of Applied Econometrics*, **11**, 475–494.

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

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. (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., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 605–657: Elsevier, Horth-Holland.

Coulson, N. F., and Robins, R. P. (1993). Forecast combination in a dynamic setting. *Journal of Forecasting*, **12**, 63–68.

Croushore, D. (2011a). Forecasting with real-time data vintages, chapter 9. In Clements, M. P., and Hendry, D. F. (eds.), *The Oxford Handbook of Economic Forecasting*, pp. 247–267: Oxford University Press.

Croushore, D. (2011b). Frontiers of real-time data analysis. *Journal of Economic Literature*, **49**, 72–100.

Croushore, D., and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, **105(1)**, 111–130.

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components?. *Journal of Econometrics*, *146*(2), 318–328.

Deutsch, M., Granger, C. W. J., and Teräsvirta, T. (1994). The combination of forecasts using changing weights. *International Journal of Forecasting*, **10**, 47–57.

Diebold, F. X. (1988). Serial correlation and the combination of forecasts. *Journal of Business & Economic Statistics*, **6**, 105–111.

Diebold, F. X., and Pauly, R. (1987). Structural change and the combination of forecasts. *Journal of Forecasting*, **6**, 21–40.

Diebold, F. X., and Pauly, R. (1990). The use of prior information in forecast combination. *International Journal of Forecasting*, **6**, 503–508.

Doan, T., Litterman, R., and Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, **3**, 1–100.

Donaldson, R. G., and Kamstra, M. (1996). Forecast combining with neural networks. *Journal of Forecasting*, **15**, 49–61.

Elliott, G., Komunjer, I., and Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies*, **72**, 1107–1125.

Elliott, G., Komunjer, I., and Timmermann, A. (2008). Biases in macroeconomic forecasts: Irrationality or asymmetric loss. *Journal of the European Economic Association*, **6**, 122–157.

Elliott, G., and Timmermann, A. (2004). Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics*, **122**, 47–79.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Ericsson, N. R. (1992). Parameter constancy, mean square forecast errors, and measuring forecast performance: An exposition, extensions, and illustration. *Journal of Policy Modeling*, **14**, 465–495.

Ericsson, N. R., and Marquez, J. (1993). Encompassing the forecasts of U.S. trade balance models. *Review of Economics and Statistics*, **75**, 19–31.

Fildes, R., and Ord, K. (2002). Forecasting competitions – their role in improving forecasting practice and research. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*, pp. 322–353: Oxford: Blackwells.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized factor model: Identification and estimation. *Review of Economics and Statistics*, **82**, 540–554.

Genre, V., Kenny, G., Meyler, A., and Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average?. *International Journal of Forecasting*, **29(1)**, 108–121.

Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Operations Research Quarterly*, **20**, 199–207.

Granger, C. W. J. (1983). Forecasting white noise. In Zellner, A. (ed.), *Proceedings of the Conference on Applied Time Series Analysis of Economic Data*, pp. 308–314. Washington, DC: Bureau of the Census.

Granger, C. W. J. (1989). Combining forecasts - Twenty years later. *Journal of Forecasting*, **8**, 167–173.

Granger, C. W. J. (1993). On the limitations of comparing mean squared forecast errors: Comment. *Journal of Forecasting*, **12**, 651–652.

Granger, C. W. J. (1996). Can we improve the perceived quality of economic forecasts. *Journal of Applied Econometrics*, **11**, 455–473.

Granger, C. W. J. (1999). Outline of forecast theory using generalized cost functions. *Spanish Economic Review*, **1**, 161–173.

Granger, C. W. J., and Andersen, A. P. (1978). *Introduction to Bilinear Time Series Models.* Göttingen: Vandenhoeck & Ruprecht.

Granger, C. W. J., and Morgenstern, O. (1970). *Predictability of Stock Market Prices.* Lexington, Massachusetts: D. C. Heath and Company.

Granger, C. W. J., and Newbold, P. (1973). Some comments on the evaluation of economic forecasts. *Applied Economics*, **5**, 35–47. Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics.* Cheltenham: Edward Elgar.

Granger, C. W. J., and Newbold, P. (1975). Economic forecasting: The atheist's viewpoint. In Renton, G. A. (ed.), *Modelling the Economy.* London: Heinemann Educational Books.

Granger, C. W. J., and Newbold, P. (1976). Forecasting transformed series. *Journal of Royal Statistical Society, Series B*, **38**, 189–203.

Granger, C. W. J., and Pesaran, M. H. (2000a). A decision-based approach to forecast evaluation. In Chan, 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 forecast accuracy. *Journal of Forecasting*, **19**, 537–560.

Granger, C. W. J., and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, **3**, 197–204.

Granger, C. W. J., and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Granger, C. W. J. (1992). Forecasting stock market proces: Lessons for forecasters. *International Journal of Forecasting*, **8**, 3–13.

Granger, C. W. J., and Jeon, Y. (2003a). A time-distance criterion for evaluating forecasting models. *International Journal of Forecasting*, *19*(2), 199–215.

Granger, C. W. J., and Jeon, Y. (2003b). Comparing forecasts of inflation using time distance. *International Journal of Forecasting*, *19*(3), 339–349.

Haddow, A., Hare, C., Hooley, J., and Shakir, T. (2013). Macroeconomic uncertainty: what is it, how can we measure it and why does it matter?. *Bank of England Quarterly Bulletin*, 100–109. 2013 Q2.

Harvey, D. I., Leybourne, S., and Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business and Economic Statistics*, **16**, 254–259. Reprinted in T.C. Mills (ed.), *Economic Forecasting*. Edward Elgar, 1999.

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

Hendry, D. F., and Richard, J.-F. (1989). Recent developments in the theory of encompassing. In Cornet, B., and Tulkens, H. (eds.), *Contributions to Operations Research and Economics. The XXth Anniversary of CORE*, pp. 393–440. Cambridge, MA: MIT Press. Reprinted in J. Campos, N.R. Ericsson and D.F. Hendry (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

Hendry, D. F., and Mizon, G. E. (2014). Unpredictability in economic analysis, econometric modeling and forecasting. *Journal of Econometrics*, *182*(1), 186–195.

Knüppel, M. (2014). Efficient estimation of forecast uncertainty based on recent forecast errors. *International Journal of Forecasting*, *30*(2), 257–267.

LeSage, J. P., and Magura, M. (1992). A mixture-model approach to combining forecasts. *Journal of Business and Economic Statistics*, **10**, 445–452.

Makridakis, S. Winkler, R. L. (1983). Averages of forecasts: some empirical results. *Management Science*, **29**, 987–996.

Mills, T. C. (2013). *A Very British Affair. Six Britons and the Development of Time Series Analysis During the 20th Century*: Palgrave Advanced Tests in Econometrics. Palgrave Macmillan, Basingstoke.

Min, C., and Zellner, A. (1993). Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics*, **56**, 89–118.

Mincer, J., and Zarnowitz, V. (1969). The evaluation of economic forecasts. In Mincer, J. (ed.), *Economic Forecasts and Expectations*, pp. 3–46. New York: National Bureau of Economic Research.

Mitchell, J., and Hall, S. G. (2005). Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR 'Fan' Charts of Inflation. *Oxford Bulletin of Economics and Statistics*, *67*(s1), 995–1033.

Mizon, G. E. (1984). The encompassing approach in econometrics. In Hendry, D. F., and Wallis, K. F. (eds.), *Econometrics and Quantitative Economics*, pp. 135–172. Oxford: Basil Blackwell.

Mizon, G. E., and Richard, J.-F. (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica*, **54**, 657–678.

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the US economy. *American Economic Review*, **62**, 902–917. Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar.

Nelson, H. L., and Granger, C. W. J. (1979). Experience using the Box-Cox transformation when forecasting economic time series. *Journal of Econometrics*, **10**, 57–69.

Newbold, P., and Harvey, D. I. (2002). Forecasting combination and encompassing. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*, pp. 268–283: Oxford: Blackwells.

Patton, A. J., and Timmermann, A. (2003). Properties of optimal forecasts. Cepr discussion papers 4037, C.E.P.R. Discussion Papers.

Patton, A. J., and Timmermann, A. (2007). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association*, **102**, 1172–1184.

Pesaran, M. H., and Weale, M. (2006). Survey expectations. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 715–776: Elsevier, Horth-Holland.

Reifschneider, D. L., and Tulip, P. (2007). Gauging the uncertainty of the economic outlook from historical forecasting errors. Finance and economics discussion series, Board of Governors of the Federal Reserve System (U.S.).

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, **48**, 1–48.

Stock, J. H., and Watson, M. W. (1999). A comparison of linear and nonlinear models for forecasting macroeconomic time series. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting*, pp. 1–44. Oxford: Oxford University Press.

Stock, J. H., and Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, **41**, 788–829.

Stock, J. H., and Watson, M. W. (2011). Dynamic factor models. In Clements, M. P., and Hendry, D. F. (eds.), *Oxford Handbook of Economic Forecasting*, pp. 35–60. Oxford: Oxford University Press. Chapter 2.

Theil, H. (1958). *Economic Forecasts and Policy.* Amsterdam: North-Holland Publishing Company. (2nd edition 1961).

Timmermann, A. (2006). Forecast combinations. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 135–196: Elsevier, Horth-Holland.

Timmermann, A., and Granger, C. W. J. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, *20*(1), 15–27.

Varian, H. R. (1975). A Bayesian approach to real estate assessment. In Fienberg, S. E., and Zellner, A. (eds.), *Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage*, pp. 195–208. Amsterdam: North Holland.

Wallis, K. F. (2005). Combining Density and Interval forecasts: A Modest Proposal. *Oxford Bulletin of Economics and Statistics*, **67(s1)**, 983–994.

Whittle, P. (1963). *Prediction and Regulation by Linear Least-Square Methods.* Princeton: D. Van Nostrand.