Forecast Content and Content Horizons for some Important Macroeconomic Time Series

John W. Galbraith
Department of Economics
McGill University
855 Sherbrooke St. West
Montreal, Quebec, Canada
H3A 2T7

Greg Tkacz
Monetary and Financial Analysis
Bank of Canada
234 Wellington St.
Ottawa, Ontario, Canada
K1A 0G9

Abstract

For quantities that are approximately stationary, the information content of statistical forecasts tends to decline as the forecast horizon increases, and there exists a maximum horizon beyond which forecasts cannot provide discernibly more information about the variable than is present in the unconditional mean (the *content horizon).* The pattern of decay of forecast content (or skill) with increasing horizon is well known for many types of meteorological forecasts; by contrast, little generally-accepted information about these patterns or content horizons is available for economic variables. In this paper we attempt to develop more information of this type by estimating content horizons for a variety of macroeconomic quantities; more generally, we characterize the pattern of decay of forecast content as we project farther into the future. We find a wide variety of results for the different macroeconomic quantities, with models for some quantities providing useful content several years into the future, for other quantities providing negligible content beyond one or two months or quarters.

Key words: diffusion index, dynamic factor model, forecast content, forecast skill, macroeconomic forecast

JEL Classification numbers: C53, E17

* We thank Omer Acikgoz, Annie Guérard-Langlois and Victor Couture for valuable research assistance, and the Bank of Canada and CIRANO (*Centre Interuniversitaire de recherche en analyse des organisations)* for support of this research. Frédéric Demers and Pierre St. Amant, and the Co-editor and anonymous referees, provided very helpful comments. The views of Tkacz are exclusively his own and should not be attributed to the Bank of Canada.

1. Introduction

Macroeconomic forecasts are made at a variety of horizons, typically from one month (or quarter) to a number of years into the future. The information content of these forecasts in general falls with increasing horizon, and beyond a certain point, forecasts of an approximately-stationary quantity of interest convey negligibly more information about future values than does the unconditional mean of the series. This point, called the 'content horizon' by Galbraith (2003), varies substantially across data series.

The patterns of decay of the content of econometric forecasts with increasing horizon have not been widely studied. Some investigation has been made of the maximum horizon at which forecasts have positive content (in this sense or by related measures such as forecast *skill*, used in the meteorological literature: see for example Murphy and Winkler 1987) for measures of national output such as Gross Domestic Product (GDP) and to a lesser degree for consumer-price inflation; see Öller (1985), Galbraith (2003), Brisson et al. (2003). Galbraith and Kisinbay (2005) examine content horizons for financial volatility forecasts. Such information on the limitations of forecasts is useful in interpreting forecasts at longer horizons, and for forecasters can serve as a criterion for evaluation of models and development of technique. As well, it may suggest bounds on the forecast horizons at which models will be useful, as is typically the case in meteorological forecasting (for example, daily temperature forecasts are usually issued to the public at horizons not exceeding seven days; meteorologists work with horizons up to approximately ten days for this variable). Nonetheless, there is very little information of this type for macroeconomic variables other than GDP and inflation.[1]

The present study attempts to extend knowledge of the forecast content horizon, and more generally of the pattern of decay of forecast content, to a broader set of such variables. In addition to consumer price inflation and real GDP growth, we examine (in both U.S. and Canadian data) short-term and long-term interest rates, growth in real personal disposable income, the federal government surplus/deficit as a proportion of GDP, the current account balance as a proportion of GDP, and the balance on goods and services trade; for the US, we also consider the growth rate of the monthly industrial production index. In each case, we estimate forecast content for simple univariate autoregressive forecasts, as well as for multivariate forecasts produced by diffusion index (dynamic factor) models. We consider variables for which stationarity is a reasonable local approximation; variables in which there is a discernible trend are transformed to an approximately stationary form.

[1] Granger (1996) also makes this point, noting that 'In some sciences there seems to be a horizon beyond which useful forecasting is not possible; in weather forecasting there seems to be [a horizon of] four or five days, for example. It is unclear if economics has similar boundaries as variables differ in their forecastability...'.

It is important to emphasize that this study deals with statistical forecasts of particular transformations (such as the level, change, growth rate, or real growth rate), of economic variables. More than one transformation may be approximately stationary; for example, quarterly and annual real GDP growth rates may each be approximately stationary, but the annual growth rate, being an aggregate of the quarterly rates, will be smoother and will have a different time series representation. Each such transformation is a separate quantity and will in general require a separate analysis of the horizons at which forecasting models will have value. We select a variety of such variables for this study, but numerous other quantities are of interest as well and may be analyzed using these methods.

Finally, we note that this study treats statistical forecasts of deviations of a quantity from its unconditional mean. For most quantities, the existence of a fixed unconditional mean is a local approximation, not a literal truth; however the attempt to use theoretical reasoning to predict very long term structural changes in the mean of a variable, for example the effects of demographic or environmental changes on typical GDP growth, is outside the scope of our study of statistical forecasting methods. Such predictions could in principle be subject to evaluation, but the very long spans of data that would be necessary are typically not available.

Section 2 of the paper describes the techniques that we will use both for forecasting and to estimate forecast content at different horizons, given a particular forecasting technique. Section 3 describes the US and Canadian data and the implementation of the different forecasting techniques on each of the time series of interest, and estimates the content horizon for each. The final section offers some concluding remarks.

2. Forecasting methods and the content horizon

We begin by defining forecast content and the content horizon, which are our primary objects of interest in this study. This discussion follows Galbraith (2003). The concept of forecast content is similar to that of *forecast skill* used in evaluating meteorological forecasts, although the reference forecast used in meteorological evaluations differs to some degree among authors; see for example Murphy and Winkler (1987). Meteorologists routinely track the skill of their forecasts over time; see for example Ghelli (2004) on the forecast skill of temperature forecasts from the European Centre for Medium Range Weather Forecasting, or the U.S. National Oceanic and Atmospheric Administration (NOAA) on the skill of various temperature, precipitation, and other forecasts.[2]

2.1 Forecast content and content horizon

We treat approximately-stationary transformations of each of the variables of interest. Of course, stationarity is a local approximation in economic time series; for

[2]Various skill measures are given on the NOAA web site, http://www.noaa.gov .

the transformations used in this study, it appears to be a reasonable approximation over a period of several years. Similarly, climate change may make stationarity only a local approximation in meteorology, but it remains a good approximation for short-range forecasting purposes. We observe $\{y_t\}_{t=1}^{T}$, a sequence of T observations on the covariance stationary process y, and we forecast the value $y_{T+s},\ s>0$, using these observations; label the forecast $\tilde{y}_{T+s|T}$, so that the mean squared forecast error is $E(\tilde{y}_{T+s|T}-y_{T+s})^2$. The sample mean $\overline{y}=T^{-1}\sum_{t=1}^{T}y_t$ also provides a possible forecast of any future value, but one which of course does not exploit any conditioning information.

The *forecast content* is defined as the proportionate reduction in a loss function, usually the mean squared error (MSE), available from the use of conditioning information in a forecasting model than the unconditional mean alone, and is specific to a forecast of a particular series, model type, and time aggregation. The MSE-loss forecast content at horizon s is described by the content function,

$$C(s)=1-\frac{MSE_{\tilde{y}(s)}}{MSE_{\overline{y}(s)}},\qquad s=1,\ldots,S, \tag{2.1.1}$$

where $\mathrm{MSE}_{\tilde{y}(s)}$ is the expected squared error of the model-based s-step forecast, and $\mathrm{MSE}_{\overline{y}(s)}$ is the corresponding expected squared error of the unconditional mean as a forecast of the process. Of course, forecast content can also be defined relative to another loss function. We use the MSE as a representative loss function, in part because analytical results on forecast content for AR forecasting models are available in that case, but also report empirical results below for the linex loss function $b[e^{a\epsilon}-a\epsilon-1]$, where ϵ is the forecast error and $a,\ b$ $(b>0)$ are parameters; the linex loss is approximately linear for negative (positive) losses and exponential for positive (negative) losses where $a>0$ $(a<0)$.[3] Having defined the forecast content, the forecast content horizon is defined as the horizon s_0 beyond which $C(s)\le 0$; we will also refer to the $\delta-$ level content horizon, beyond which the forecast content is less than δ (we use $\delta=0.05, 0.10$ in Tables 1 and 2 below.)

Although analytical expressions are available for computing (2.1.1) given the parameters of a pure autoregressive process (Galbraith 2003),[4] and although we will report these results, the present paper is also concerned with general multivariate forecasting methods, for which no analytical solutions exist. We must therefore

[3]The parameter b will cancel from numerator and denominator if linex loss is substituted for MSE in (2.1.1) and so plays no role in results below.

[4]For a stationary AR(p) process with error distribution $\varepsilon_t\sim IN(0,\sigma_\varepsilon^2)$, we have

$$C(s)=1-\frac{B_{[1,1]}}{\nu}+o(T^{-1}), \tag{2.1.2}$$

also use non-parametric estimates of the forecast content function, which can be computed using sample estimates of the MSE or linex losses.

2.2 Forecasting methods

Forecast content is a measure of the value of conditioning information for a forecast at a given horizon. Different forecast methods exploit conditioning information differently, and so may produce different content and content horizons. Our interest here is not in the relative merits of different forecasting methods, but in estimating the forecast content and the maximum horizon at which conditioning information can be valuable. We therefore need to investigate a variety of different methods in order to estimate the attainable content horizon, given existing forecasting technology and data constraints, for a particular time series. We consider univariate and multivariate methods, and examine the behavior of the forecast content functions as we vary the forecasting model. We use y_t to represent an observation on any one of the processes of interest discussed in Section 3 below.

(i) Autoregressive models

The AR model is a univariate linear model. In spite of its simplicity, it is often competitive with much more elaborate forecasting models. We will report results for each process using these benchmark forecasts. The AR(p) model is of the form

$$\alpha(L)(y_t - \mu) = \epsilon_t \quad \text{or } y_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j y_{t-j} + \epsilon_t, \tag{2.2.1}$$

where $\alpha(L)$ is a $p-$th order polynomial in the lag operator, ϵ_t is an iid disturbance, α_0 is a constant and α_j is the autoregressive parameter at lag j. Where the forecast horizon exceeds 1, forecasts are computed by the standard iterated method rather than by direct projection on the available lags.

---

where $B_{[1,1]}$ is the top-left element of the matrix

$$B = \sum_{j=0}^{s-1} A^j M A'^j + T^{-1} \sum_{j=0}^{s-1} \sum_{k=0}^{s-1} A^j M A'^k \cdot \mathrm{tr}[(A^{s-j-1}\Gamma)'(\Gamma^{-1} A^{s-k-1})], \tag{2.1.3}$$

$$\nu = \sum_{i=0}^{\infty} a_i^2 \left(1 - 2T^{-1} \sum_{h=s}^{T+s-1} \rho(h)\right) + T^{-1} \left(\sum_{i=0}^{\infty} a_i\right)^2, \tag{2.1.4}$$

and where M is the $(p+1) \times (p+1)$ matrix with 1 in the upper-left corner and zeroes elsewhere, $\Gamma = E(Y_t Y_t')$, $\rho(h)$ is the autocorrelation function at lag h, and A is the $(p + 1) \times (p + 1)$ matrix such that $Y_t = AY_{t-1} + \epsilon_t$, with $\epsilon_t = (\varepsilon_t, 0, \ldots 0)'$.

As we have noted, the forecast content of an autoregressive process may be characterized analytically, taking into account the uncertainty associated with parameter estimation. We report below both this analytical result, which applies to a true AR model of known order estimated on the full sample, and therefore tends to produce relatively high estimates of forecast content, as well as the 'empirical' measure based on a sequence of pseudo-out-of-sample forecasts.

The availability of these analytical results makes the pure autoregression an important benchmark to include, despite the fact that we might in general expect the best multivariate model to produce lower MSE and higher content horizon than the AR. As well, the empirical performance of the AR model class is remarkably competitive on economic data; see for example the large-scale comparison in Stock and Watson (1999a), who find the AR the best-performing forecasting model overall in a set of linear and non-linear univariate models, and Brisson et al. (2003), who find that general multivariate models such as dynamic factor (diffusion index) models tend not to produce appreciable increases in the content horizon for their series of interest (although they do find reductions in MSE at points inside the content horizon).

(ii) Diffusion index (dynamic factor) models

This class of models will be our general device for incorporating multivariate information. The diffusion index model is a dynamic factor model, exploited by Stock and Watson (2002a) for forecasting in (e.g.) the form:

$$y_t = \beta' F_{t-1} + \varepsilon_{t-1}, \tag{2.2.2}$$

where F_{t-1} is the set of factors obtained by exploiting a large set of conditioning variables available for forecasting, represented by the $N-$ dimensional time series $\{X_{t-1}\}_{t=1}^{T}$, modelled using the factor structure

$$X_{t-1} = \Lambda_{t-1} F_{t-1} + e_{t-1}. \tag{2.2.3}$$

Factor extraction begins by standardizing the matrix X to mean zero and variance one (X may contain lags of the series). The factors F are estimated using the q eigenvectors associated with the q greatest eigenvalues of this matrix, as (up to scale) $\hat{F} = X\hat{\Lambda}$, where $\hat{\Lambda}$ is the matrix of eigenvectors associated with the q largest eigenvalues. The estimated moment matrix is scaled by the number of series N, i.e. to $N^{-1}(X'X)$, for consistent estimation of an assumed true number of factors; an information criterion is known to estimate this population number of factors consistently. See Stock and Watson (2002a,b) for a complete exposition of the model.

This model class is an attractive representative of the class of multivariate models for several reasons. In addition to the fact that it tends to produce results comparable with the best theory-based regression models for forecasting, while allowing for automatic implementation in cases where no strong theory-based forecasting model is available, it is also useful that the diffusion index information may be complementary with autoregressive information, in that a model with both elements may gain relative to either class of model alone; see Stock and Watson (1999b, 2002a/b), Brisson et al. (2003). We also consider such combined diffusion+AR forecasts in the set examined below.

The diffusion forecasts are produced and evaluated in a sequence of pseudo-out-of-sample results. A large multivariate data set is required to implement these forecasts; see section 3.1.

(iii) Forecast combination

In addition to these methods, we explore fixed-weight combinations of forecasts made from diffusion index (or diffusion index with AR components) and pure autoregressive models; see Li and Tkacz (2004) for a general treatment of possible forecast combination methods. These forecasts are indicated as 'combined' in the figures, and often have the highest forecast content. The present implementation uses equally weighted forecasts, implicitly assigning equal variance to each component forecast. These weights may differ substantially from those of the optimal forecast where the two inputs forecasts have substantially different associated forecast variance, but equal weighting has generally been found to perform remarkably well and robustly; see Li and Tkacz (2004) and Stock and Watson (2004).

3. Data and forecast results

3.1 Data

Professional economic forecasters produce forecasts of a wide range of variables, although some variables are of general interest and are forecast by almost all public- and private-sector forecasters. Similarly, academic forecasters have produced many studies of different quantities, although some such as output growth and inflation are particularly widely studied. Although many of these studies use fairly long horizons, the content of forecasts at longer horizons, in the sense described above, remains an open question for most of these variables.

In this study we forecast each of the variables mentioned in the introduction, but use a much larger set of variables in the dynamic factor (diffusion index) models used for forecasting. An appendix available from the authors indicates the data available, at quarterly, monthly, and higher (weekly or bi-weekly) frequencies; the data sets begin in the first quarter or month of 1969 and end in the first quarter, or third or fourth month, of 2004. For forecasting at the quarterly frequency, monthly (and higher-frequency exchange rate data) are aggregated to the quarterly level. The numbers of series available for factor forecasts are 91 (Canadian monthly data), 229 (Canadian quarterly data), 143 (US monthly data) and 203(US quarterly data); there are 423 sample points in the monthly data sets, and 141 in the quarterly sets.

The order of integration of each of these series was assessed, and for approximately stationary series, no transformations were performed on the data. For non-stationary series we computed either monthly, quarterly or annual growth rates, which typically correspond to the values of interest to users of forecasts of integrated time series. Note also that the computation of, for example, a monthly series of annual values from monthly data introduces a moving average process of order 11 in the transformed series, because contiguous monthly measures pertain to annual periods overlapping by 11 months. This moving-average effect can readily be modelled, directly or by autoregressive (or other) approximation.

The series, and the transformations applied to each for use as a forecast target, are the following:

- 1. Output growth: $y_{1,t} = 100(\ln(Y_t) - \ln(Y_{t-\ell}))$, where Y_t is quarterly real GDP at time t, and $\ell = 1$ and 4 quarters, yielding quarterly and annual growth rates; for the US only, we also forecast $y_{1b,t}$ equal to monthly real industrial production, with $\ell = 1$, yielding monthly growth rates.
- 2. Inflation: $y_{2,t} = 100(\ln(P_t) - \ln(P_{t-\ell}))$, where P_t is the level of consumer prices at time t, and $\ell = 1$ and 12 months, for monthly and annual inflation respectively.
- 3. Current account balance as a proportion of GDP, $y_{3,t}$: no further transformation.
- 4. Balance on goods and services trade as a proportion of GDP, $y_{4,t}$: no further transformation.
- 5. Short-term ex-post real interest rates: $y_{5,t} = r_t^s - y_{2,t}$, where r_t^s is an annualized short term interest rate.
- 6. Long-term ex-post real interest rates: $y_{6,t} = r_t^l - y_{2,t}$, where r_t^l is an annualized long term interest rate.
- 7. Growth in real personal disposable income: $y_{7,t} = 100(\ln(Y_t) - \ln(Y_{t-\ell}))$, where Y_t is monthly (US) or quarterly (Canada) real personal disposable income at time t, and $\ell = 1$.
- 8. Consolidated government surplus (deficit) as a percentage of GDP, $y_{8,t}$: no further transformation.

As discussed in Tkacz (2001), the time-series properties of short and long-term rates are sufficiently different as to warrant separate analysis. Furthermore, short-term rates are closely related to the policies followed by central bankers, while long-term rates are largely driven by non-policy factors, such as inflation expectations and market liquidity; this difference suggests possible differences in predictability, in the sense of Diebold and Kilian (2001), or forecast content.

The inflation and real GDP series are the only ones for which we take two transformations, and the comparison of these (see for example Figures 1a vs 1b and

1e vs. 1f for inflation) underlines the importance of the transformation for forecast skill, content, or predictability. When we take the difference at lag 12 for inflation, we produce a series with an MA(11) structure relative to the lag 1 difference; that is, the annually-differenced log price series is a different time series, and has a different time series structure, than the monthly difference in log price. The annual inflation rate will typically show more autocorrelation, more predictability, and more forecast content or skill. This is an illustration of the general feature that forecast content, skill, content horizon, predictability, etc. are features of a particular transformation of a variable, and that different transformations may show very different results. The choice depends on which quantity is of interest, and different choices are legitimate as long as weak stationarity is an adequate approximation (so that the unconditional mean has a well-defined meaning) over the period under consideration.

3.2 Implementation details

Implementation of these methods involves a number of parameter choices, including t_0, the initial sample for estimation of models prior to pseudo-out-of-sample forecast comparison; p, the number of AR terms in AR or Diffusion+AR models; k, the number of factors for diffusion index forecasts. The aim is to optimize performance of each method in order to obtain a good estimate of the maximum obtainable horizon conditional on available sample size. We allow p to be no greater than 4 for quarterly data, 8 for monthly, and k to be no greater than 4. The lag length selected as optimal is retained throughout the pseudo-out-of-sample forecast period.

We take the initial sample for each empirical method as $t_0 = T/2$; as t_0 is larger, initial parameter estimates (and therefore forecast quality) is higher in the earlier pseudo-out-of-sample forecasts, but the sample of these forecasts is smaller. The results are not greatly sensitive to moderate variations in this initial sample size. The number of AR terms is chosen by the Schwarz information criterion, using an upper bound, and is used not only in the empirical methods but in defining the number of terms for the analytical AR computation. The number of principal components is chosen to optimize at an arbitrarily-chosen fixed horizon, and following results in Stock and Watson (2002b), is limited to an upper bound of four on the number of components.

3.3 Content results

The results of the estimation of MSE forecast content are presented in Figures 1 and 2, for each of 21 (10 Canadian and 11 US) quantities, and are summarized in the corresponding Tables 1 and 2; in addition the tables contain linex-loss forecast content with parameter a equal to 1 and -1. The graphical results display estimates of the general pattern of decay of forecast content with horizon for the various methods. In the tables the point of interest is the horizon achievable by the best method. We are somewhat conservative in reporting these horizons; in particular, where a method produces very low or negative content for several periods, followed by positive content within the range of sampling error, we interpret these fluctuations as noise and do not record the later periods of estimated positive content. Several such instances arise with the dynamic factor models. We report both the analytical result in principle achievable from a pure AR process with MSE loss (computed via equations 2.1.2–2.1.4 above) as well as the best result obtained using the estimated losses from pseudo-out-of-sample sequences of empirical forecast observations, with content evaluated via equation 2.1.1 or its linex analogue. The maximum horizons considered for the two data frequencies are 36 months and 16 quarters respectively; horizons exceeding these values are reported as 36+ or $16+$.

Confidence bands are not given in Figures 1 and 2, but an indication of the precision of the estimates is given by the standard error in a parametric bootstrap of the forecast content for the results computed from the analytical expressions (2.1.2–2.1.4). Since these analytical results come from a parametric (AR) process, white noise residuals can be generated conditional on this model. Bootstrap samples can be generated by re-sampling from these residuals or by drawing from a parametric distribution; in the present case we draw from the normal distribution on which the analytical results are based. Properties of bootstrap sampling in autoregressive processes were established by Freedman (1984) and Bose (1988); see also the recent exposition in Horowitz (2001, p. 3188 ff.) For a given estimated AR representation, a set of bootstrap sequences is generated from the simulated residuals. On each of these bootstrap samples the forecast content is estimated at each horizon using (2.1.2–2.1.4). Confidence intervals and standard errors at each horizon h are then obtained from the set of bootstrap values of forecast content at each horizon. For monthly series where substantial content remains at the maximum horizon (all three Canadian series, US 12-month real interest rates and inflation), the standard errors are in the region of 0.01 to 0.03 at horizon 1, rising to 0.15-0.20 at the maximum horizon; in the remaining monthly cases where content is near zero at the longest horizons, standard errors are similar at horizon 1 but decline to well under 0.01 at long horizons, reflecting the fact that variation around zero in the AR parameters has little effect on long-horizon content, so that uncertainty attributable to parameter estimation is small. The same patterns hold in quarterly data, but with larger standard errors reflecting the smaller sample sizes; typically 0.04-0.08 at horizon 1, rising to around 0.20 where content remains substantial at the maximum lag (all US series except real GDP growth, Canadian goods and services trade balance) or declining to around 0.01 or less where content is near zero (remaining quarterly series). Note that the bootstrap confidence bands are slightly asymmetrical around the estimate.

Table 1[5][6]

Forecast content horizons C_δ, monthly data,
by analytical and best empirical methods

	analytical AR MSE		empirical MSE		empirical linex, $a = 1$		empirical linex, $a = -1$	
δ :	10%	5%	10%	5%	10%	5%	10%	5%
quantity forecast:								
Canada								
1 π_1	6	9	36^+	36^+	36^+	36^+	36^+	36^+
2 π_{12}	36^+	36^+	36^+	36^+	36^+	36^+	36^+	36^+
3 r_3	36^+	36^+	36^+	36^+	36^+	36^+	36^+	36^+
4 r_{12}	36^+	36^+	36^+	36^+	36^+	36^+	36^+	36^+
US								
1 π_1	3	7	36^+	36^+	36^+	36^+	36^+	36^+
2 π_{12}	36^+	36^+	36^+	36^+	36^+	36^+	36^+	36^+
3 rpdi	0	0	0	0	0	1	0	1
4 ip	1	1	0	1	1	1	1	1
5 r_3	9	13	4	6	5	6	6	7
6 r_{12}	36^+	36^+	21	25	23	25	26	31

The figures indicate a number of general points, as well as results specific to the series of interest. In particular, we note that:

. The best empirical method differs across time series. Where diffusion and AR methods are similar in performance, the equally-weighted combination tends to produce improvements.

[5]Variables are presented in the same order as in the figures below. Variable definitions are: π_1, 1-month inflation; π_{12}, 12-month inflation; r_3, three-month ex post real interest rate; r_{12}, twelve-month ex post real interest rate; rpdi, percentage change in real personal disposable income; ip, percentage change in the industrial production index.

[6]In both Tables 1 and 2, 5% and 10% refer to 5% and 10% forecast content.

. The analytical results, representing potentially achievable results from pure persistence forecasts on the given sample size, tend to be similar to the best empirical outcome, although this is not invariably the case. Note that the empirical results use a sequence of results on sample sizes smaller than the full sample, whereas the theoretical AR results take into account parameter uncertainty at the level pertaining to the full sample of data. This feature would tend to produce higher content in the theoretical AR results; however the empirical results based on diffusion index or combined forecasts use additional information.

. The results on $\delta-$level content tend not to differ radically across empirical methods (the results from each individual empirical method are not recorded in the tables, but are visible in the figures).

. Data series such as monthly growth in industrial production and quarterly growth in real GDP are difficult to forecast beyond the shortest horizon. Annual real GDP growth, measured quarterly, shows a longer horizon (approximately four quarters), in part reflecting the moving average effect noted above; however there is also some indication of forecast content reappearing at longer horizons in the three-year range (see in particular Figures 2b, 2h).

. Inflation and real interest rate series are relatively persistent and have correspondingly high content horizons. Note that we treat the raw series as approximately stationary; studies in which these series are treated as I(1), so that differences are required for approximate stationarity, would show very short content horizons on the *differenced* series.[7] Monthly inflation is much less persistent than the annual (12-month difference) inflation, and forecast content is correspondingly lower at all horizons. The maximum horizon at which measurable content remains is, however, also at least 36 months for the monthly inflation series. U.S. short-term (3-month) rates show substantially lower content horizons than Canadian rates or U.S. longer-term rates.

. Moderately high content horizons also tend to be observed in the trade balance and government surplus/ deficit (relative to GDP) measures.

. Differences in results between empirical MSE and linex loss functions are generally minor.

[7]Recall however that we treat real interest rates, which are much less persistent than nominal. Stationarity of the real rate is also compatible with non-stationarity of the corresponding nominal interest rate, as for example where the nominal rate and price level are I(1) but co-integrated with co-integrating vector (1, -1).

Table 2[8]

Forecast content horizons C_δ, quarterly data, by analytical and best empirical methods

	analytical AR MSE		empirical MSE		empirical linex, $a = 1$		empirical linex, $a = -1$	
δ :	10%	5%	10%	5%	10%	5%	10%	5%
quantity forecast:								
Canada								
1 rgdp:q	1	1	1	1	1	2	1	2
2 rgdp:a	3	4	3	4	4	6	4	6
3 ca	6	9	14	14	14	16^+	14	16^+
4 gs	16^+	16^+	15	16^+	16^+	16^+	16^+	16^+
5 rpdi	0	3	11	16^+	16^+	16^+	16^+	16^+
6 s/d	6	8	5	7	7	8	7	8
US								
1 rgdp:q	0	1	0	1	1	2	1	2
2 rgdp:a	3	4	3	3	3	4	3	4
3 gs	16^+	16^+	15	16^+	16^+	16^+	16^+	16^+
4 ca	16^+	16^+	16^+	16^+	16^+	16^+	15	16^+
5 s/d	16^+	16^+	16^+	16^+	16^+	16^+	16^+	16^+

High forecast content and long content horizons tend to arise in persistent series, that is, series for which deviations from the mean are long-lasting. The apparent persistence of these deviations may however be exaggerated by a changing mean. This is related to the well-known phenomenon by which structural change may lead to apparent long memory (slow hyperbolic decay in the autocorrelation

[8]Variables are presented in the same order as in the figures below. Variable definitions are: rgdp, percentage change in real gross domestic product, :q for quarterly, :a for annual; ca, current account balance as a proportion of GDP; gs, goods and services trade balance as a proportion of GDP; rpdi, percentage change in real personal disposable income; s/d, surplus or deficit as a proportion of GDP.

function) of a time series; see for example Teverosky and Taqqu (1997), Granger and Hyung (2004). To the extent that important structural changes are present in the time series examined here, then, the content horizons computed will tend to be generous. This problem may be particularly likely to have afflicted our inflation results; if inflation has in fact shifted to a different mean over the sample period, then these horizons may exaggerate the true content horizons. Again, however, we note that in spite of this potential phenomenon, it is interesting that a number of the estimated horizons, particularly for one-quarter output growth measures, appear to be quite short.

4. Concluding remarks

Estimates of the content or skill of forecasts, and of the horizon at which these become approximately zero, are specific to particular forecasting methods, although this horizon may turn out to be similar for different methods. The present study offers some estimates of the pattern of decay of forecast content and of these maximum horizons, but it is of course always possible that another method or further data may provide better results than existing techniques: for this reason no such study can indicate the maximum that is achievable. These results should be seen instead as estimates of the content and content horizon for some standard, generally-successful methods, which provide reasonable benchmarks for future studies in which forecasters attempt to extend the horizons, or examine new methods in comparison with the properties of existing techniques.

Although these results cannot be conclusive, they may also be of some value in guiding forecast reporting at longer horizons. Forecasts at horizons beyond the content horizon provide no information relative to unconditional properties of the series (that is, have no content or skill); while we cannot locate this horizon exactly, we can certainly find horizons beyond which content is very low. Reporting practices for forecasts beyond these horizons, if they are to be made, might well involve an acknowledgment of the limitations of information content of the forecasts and a reference to the unconditional properties of the series. Of course, forecasts beyond this point which are equal to the unconditional mean, with constant confidence intervals, implicitly convey the same point. We nonetheless suggest that approximate knowledge of the forecast content horizon for various quantities of interest is a useful form of knowledge for forecasters to accumulate.

For a number of important macroeconomic variables, the content horizon is quite substantial; in other cases, it is low. The differences across series are large (from zero to over 36 months, for example, in our monthly data), whereas differences in reported maximum horizons tend to be less substantial. This information may be useful in guiding the horizons chosen not only for forecast research studies, but in routine forecast reporting to economic agents acting on forecasts.

References

Bose, A. (1988) Edgeworth correction by bootstrap in autoregressions. *Annals of Statistics* 16, 1709-1722.

Brisson, M., B. Campbell and J.W. Galbraith (2003) Forecasting some Low- Predictability Time Series Using Diffusion Indices. *Journal of Forecasting* 22, 515-531.

Diebold, F.X. and L. Kilian (2001) Measuring predictability: theory and macroeconomic applications. *Journal of Applied Econometrics* 16, 657-669.

Freedman, D.A. (1984) On bootstrapping two-stage least-squares estimates in stationary linear models. *Annals of Statistics* 12, 827-842.

Galbraith, J.W. (2003) Content horizons for univariate time series forecasts. *International Journal of Forecasting* 19, 43-55.

Galbraith, J.W. and T. Kışınbay (2005) Content horizons for conditional variance forecasts. *International Journal of Forecasting* 21, 249-260.

Ghelli, A. (2004). Forecast verification: probabilistic aspects. From *Use and Interpretation of ECMWF products,* European Centre for Medium Range Weather Forecasting, Meteorological Training Course Lecture Programme.

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. and N. Hyung (2004) Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance* 11, 399-421.

Horowitz, J.L. (2001) The Bootstrap. In J.J. Heckman and E.E. Leamer, eds., *Handbook of Econometrics* Vol. 5, Elsevier.

Li, F. and G. Tkacz (2004) Combining forecasts with nonparametric kernel regressions. *Studies in Nonlinear Dynamics and Econometrics* 8, article 2.

Murphy, A.H. and R.L. Winkler (1987) A general framework for forecast verification. *Monthly Weather Review* 115, 1330-1338.

Öller, L.-E. (1985). How far can changes in business activity be forecast? *International Journal of Forecasting* 1, 135-141.

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

Stock, J.H. and M.W. Watson (1999b) Forecasting inflation. *Journal of Monetary Economics* 44, 293–335.

Stock, J.H. and M.W. Watson (2002a) Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20, 147-162.

Stock, J.H. and M.W. Watson (2002b) Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167-1179.

Stock, J.H. and M.W. Watson (2004) Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* 23,405-430.

Teverosky, V. and M. Taqqu (1997) Testing for long range dependence in the presence of shifting means or a slowly-declining trend, using a variance-type estimator. *Journal of Time Series Analysis* 18, 279-304.

Tkacz, G. (2001) Estimating the fractional order of integration of interest rates using a wavelet OLS estimator. *Studies in Nonlinear Dynamics and Econometrics* 5, 19-32.

Figure 1a
Forecast content functions, Canadian monthly data:
1-month inflation

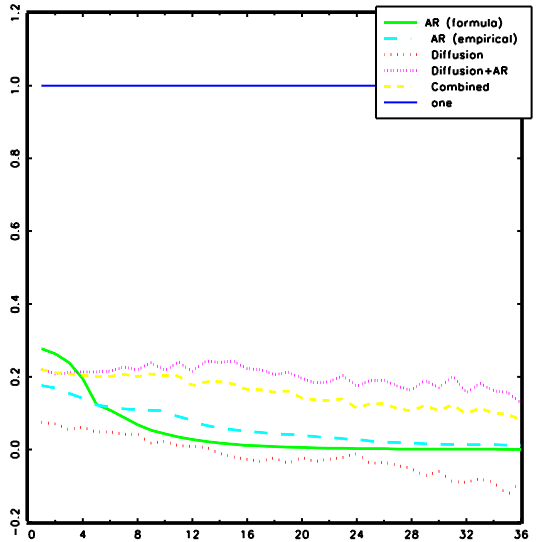

Figure 1b
Forecast content functions, Canadian monthly data:
12-month inflation

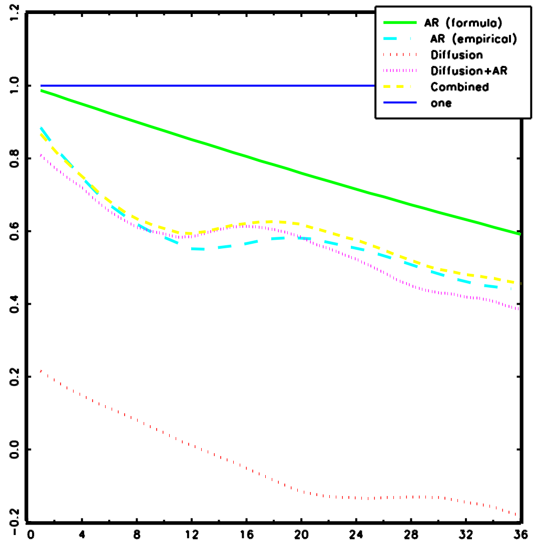

Figure 1c
Forecast content functions, Canadian monthly data:
3-month ex post real interest rate

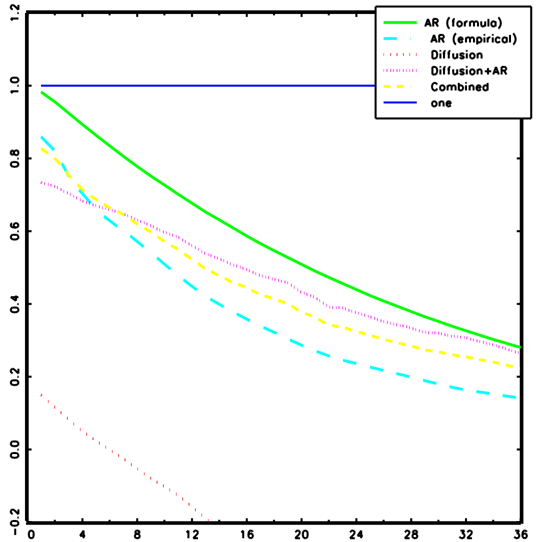

Figure 1d
Forecast content functions, Canadian monthly data:
12-month ex post real interest rate

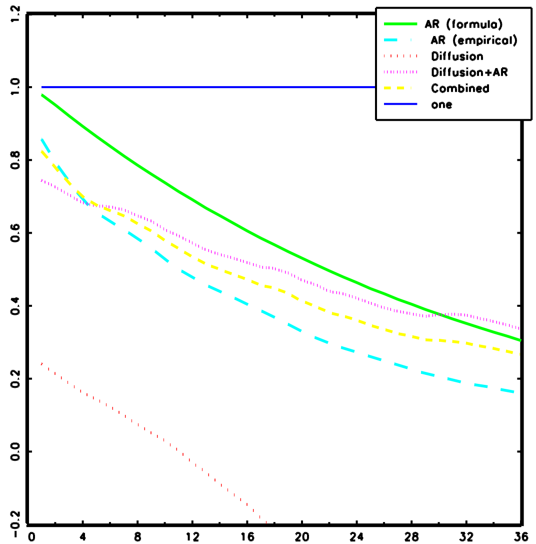

Figure 1e
Forecast content functions, US monthly data:
1-month inflation

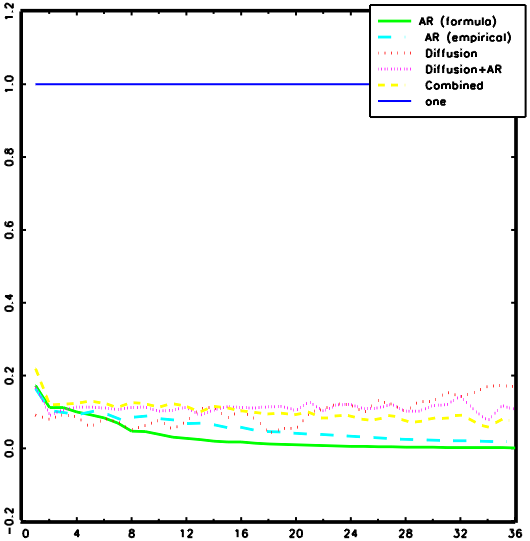

Figure 1f
Forecast content functions, US monthly data:
12-month inflation

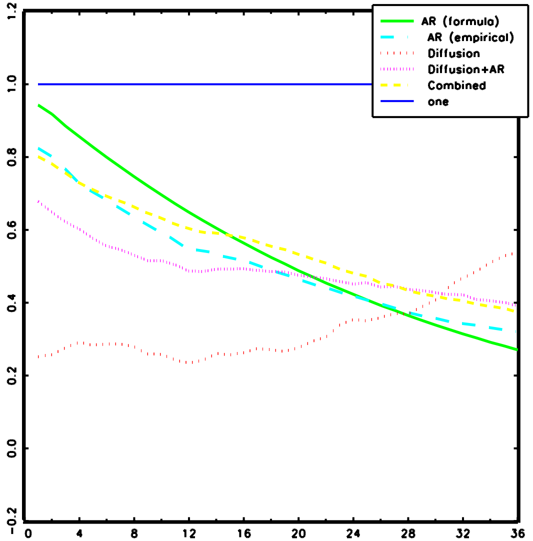

Figure 1g
Forecast content functions, US monthly data:
real personal disposable income growth

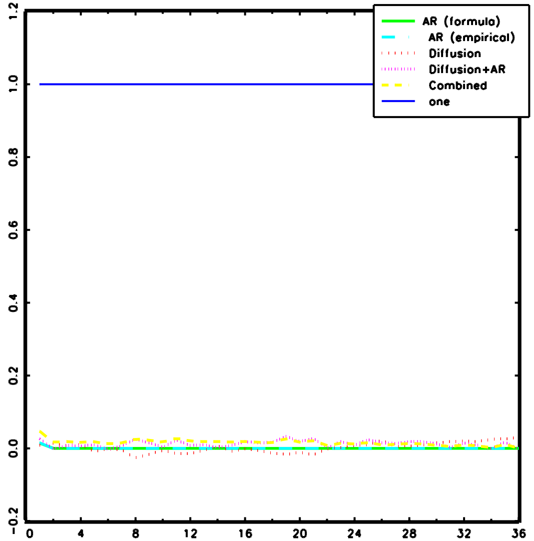

Figure 1h
Forecast content functions, US monthly data:
industrial production growth

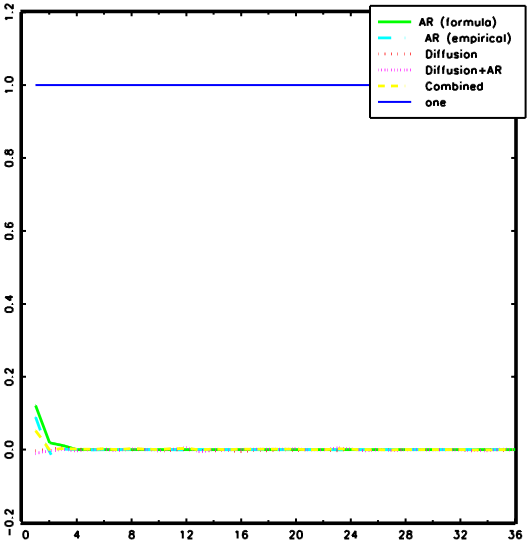

Figure 1i
Forecast content functions, US monthly data:
3-month ex post real interest rate

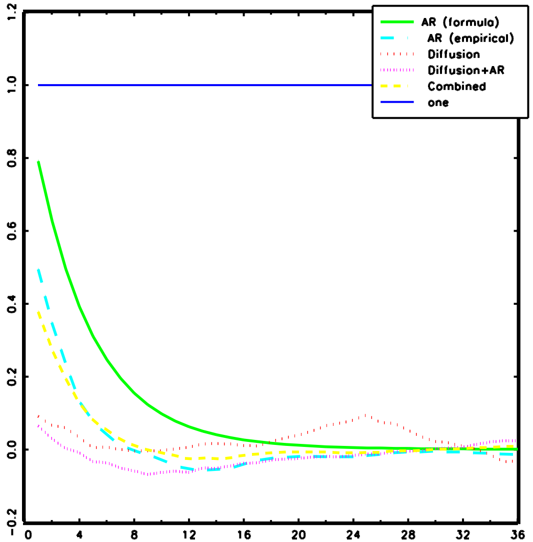

Figure 1j
Forecast content functions, US monthly data:
12-month ex post real interest rate

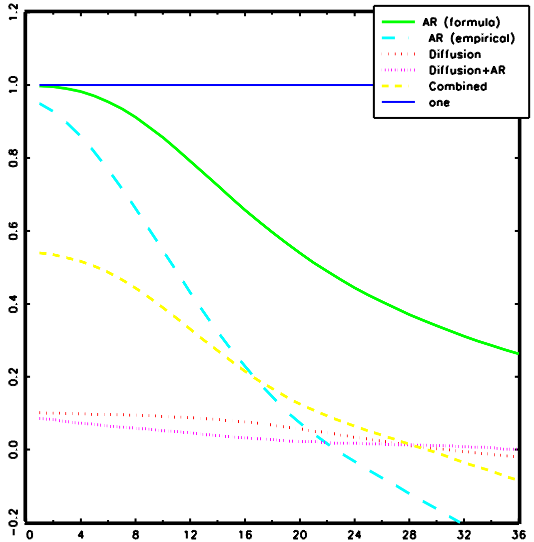

Figure 2a
Forecast content functions, Canadian quarterly data:
quarterly real GDP growth

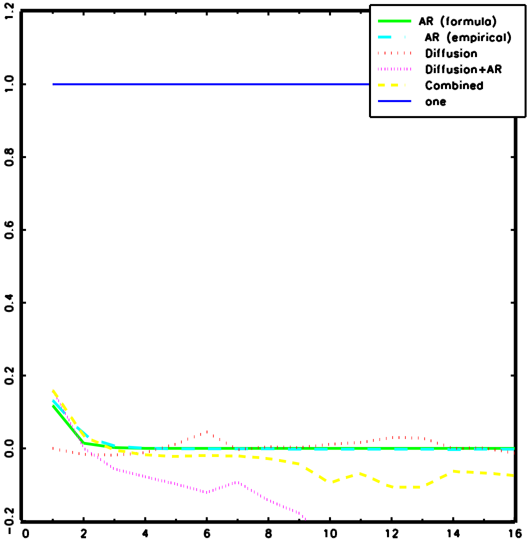

Figure 2b
Forecast content functions, Canadian quarterly data:
annual real GDP growth

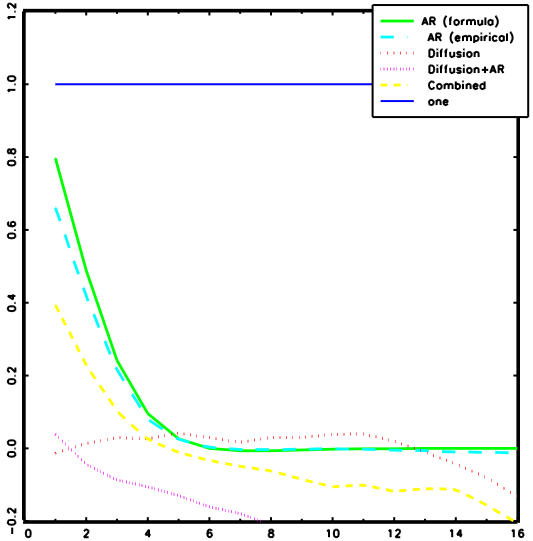

Figure 2c
Forecast content functions, Canadian quarterly data:
current account balance as proportion of GDP

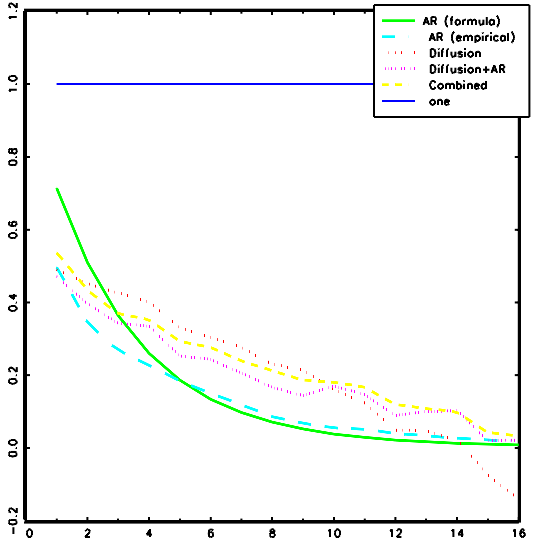

Figure 2d
Forecast content functions, Canadian quarterly data:
goods and services balance as proportion of GDP

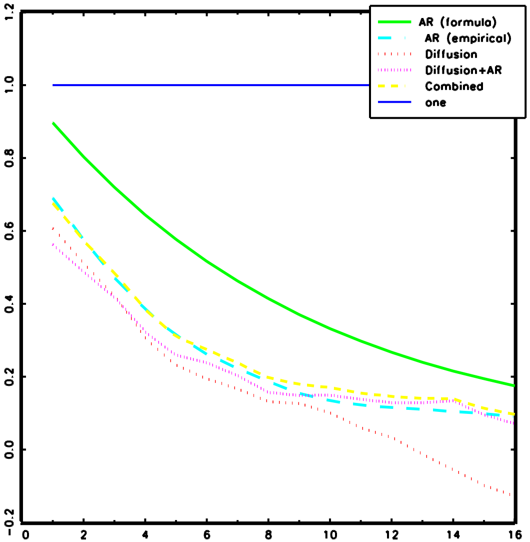

Figure 2e
Forecast content functions, Canadian quarterly data:
real personal disposable income

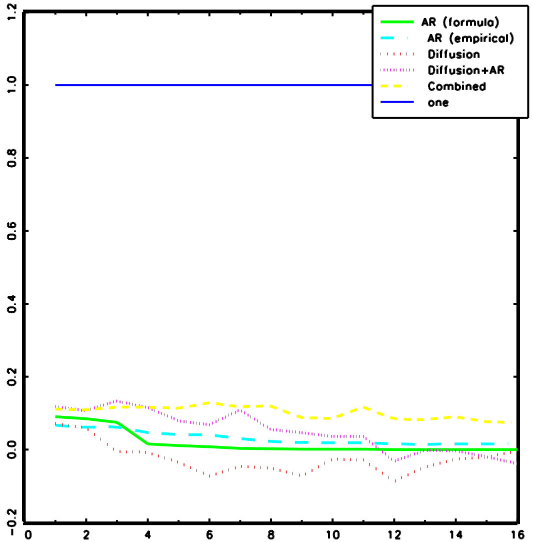

Figure 2f
Forecast content functions, Canadian quarterly data:
government surplus or deficit as proportion of GDP

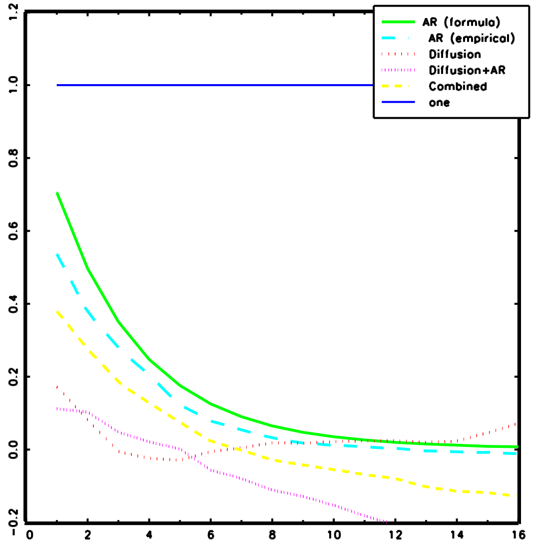

Figure 2g
Forecast content functions, US quarterly data:
quarterly real GDP growth

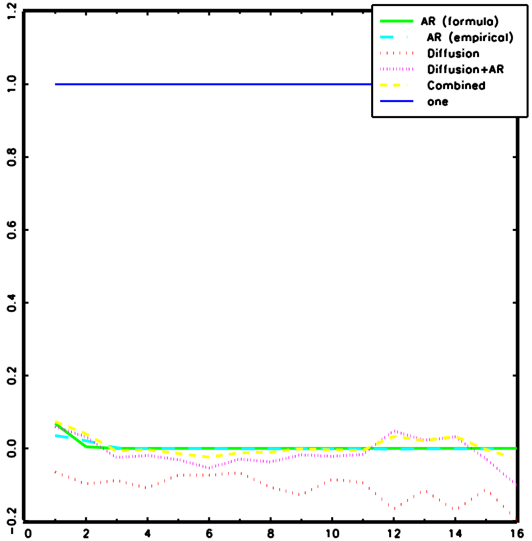

Figure 2h
Forecast content functions, US quarterly data:
annual real GDP growth

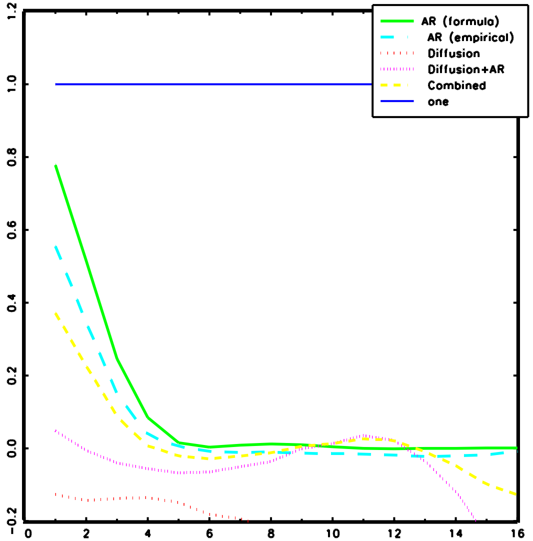

Figure 2i
Forecast content functions, US quarterly data:
goods and services balance as proportion of GDP

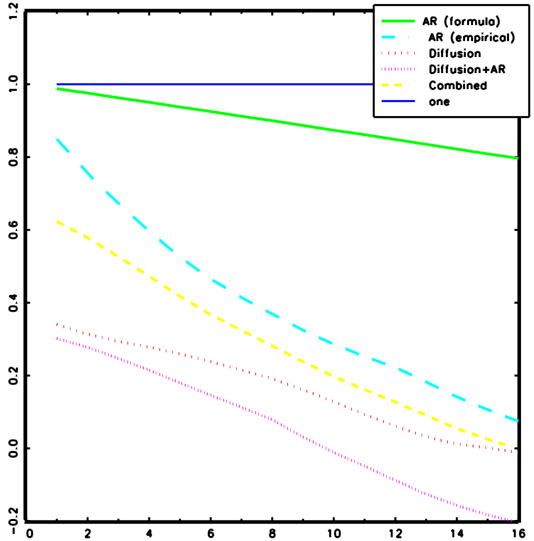

Figure 2j
Forecast content functions, US quarterly data:
current account balance as proportion of GDP

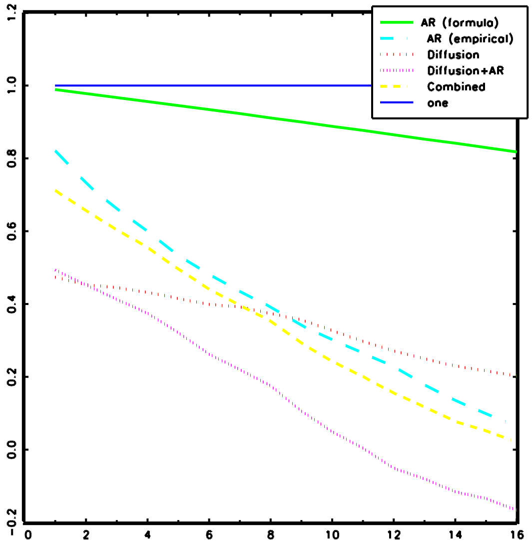

Figure 2k
Forecast content functions, US quarterly data:
government surplus or deficit as proportion of GDP