Forecast Encompassing Tests and Probability Forecasts

No 774

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF WARWICK

Forecast Encompassing Tests and Probability Forecasts

Michael P. Clements*
Department of Economics,
University of Warwick

David I. Harvey
School of Economics
University of Nottingham

July 21, 2006

Abstract

We consider tests of forecast encompassing for probability forecasts, for both quadratic and logarithmic scoring rules. We propose test statistics for the null of forecast encompassing, present the limiting distributions of the test statistics, and investigate the impact of estimating the forecasting models' parameters on these distributions. The small-sample performance of the various statistics is investigated, both in terms of small numbers of forecasts and model estimation sample sizes. Two empirical applications show the usefulness of the tests for the evaluation of recession probability forecasts from logit models with different leading indicators as explanatory variables, and for evaluating survey-based probability forecasts.

JEL classification: C12, C15, C53.

Keywords: Probability forecasts, encompassing tests, recession probabilities.

*Corresponding author. Department of Economics, University of Warwick, Coventry, CV4 7AL, UK. Tel.: +44 (0)24 7652 2990. Email: m.p.clements@warwick.ac.uk. We are grateful to an Associate Editor and two referees of this journal for many useful comments and suggestions. Helpful comments were also received from seminar participants at the University of Kent at Canterbury and Carlos III, Madrid.

1 Introduction

There is an extensive literature in economic and management science attesting to the usefulness of forecast combination.[1] That is, a linear combination of two or more forecasts may often yield more accurate forecasts than using a single forecast. The forecast combination literature typically assesses the out-of-sample accuracy of combinations whose weights have been determined in-sample. Tests of forecast encompassing assess whether *ex post* a linear combination of forecasts results in a statistically significant reduction in the mean squared forecast error (MSFE) relative to using a particular forecast, and can be viewed as an indicator of when combinations might be useful *ex ante*, although Chong and Hendry (1986) originally presented such tests as model-specification tests for large-scale models where more conventional tests may be infeasible. Forecast encompassing is formally equivalent to 'conditional efficiency' due to Nelson (1972) and Granger and Newbold (1973), whereby a forecast is conditionally efficient if the variance of the forecast error from a combination of that forecast and a rival forecast is not significantly less than that of the original forecast alone.

Tests of forecast encompassing and the closely-related tests of equal forecast accuracy of Diebold and Mariano (1995) are routinely calculated in forecast comparison exercises in applied work. For the most part they are applied to forecasts of variables that, in principle at least, can take any value on the real line. Yet as noted by Diebold and Lopez (1996) in their review of forecast evaluation and combination, forecasts of economic and financial variables often take the form of probabilities, e.g., forecast probabilities of recession. These authors suggest that the significance of differences in accuracy between different probability forecasts can be tested using the Diebold-Mariano tests, and Granger and Pesaran (2000) mention combining probability forecasts. Kamstra and Kennedy (1998) propose a method of combining probability forecasts using logit regression. This is presented as a computationally simple way of combining forecasts, and no optimality properties for the derived combination are claimed. In other spheres, the combination of probability assessments is commonplace, although the emphasis tends to be different from ours.[2]

In this paper we apply forecast encompassing to probability forecasts. Because of the nature of the forecasts, our analysis will differ from the standard application of encompassing tests to point forecasts in a number of ways. One is the nature of the loss function. Tests of encompassing applied

[1] See *inter alia* Diebold and Lopez (1996) and Newbold and Harvey (2002) for recent surveys, and Clemen (1989) for an annotated bibliography.

[2] The literature on the combination of experts' subjective probability distributions (see e.g., Genest and Zidek (1986) and Clemen and Winkler (1999)) looks at ways of aggregating individual assessments such that the aggregate possesses desirable properties, rather than focusing on accuracy. See also Dawid (1986) and Winkler (1996) on probability forecasting and evaluation from a meteorological prespective, as well as a discussion of earlier contributions.

to point forecasts invariably use mean squared error (MSE). The same measure is commonly used to evaluate probability forecasts, in the guise of the quadratic probability score (QPS), but at least as commonplace is the logarithmic probability score (LPS). We consider forecast encompassing tests based on both. Secondly, our analysis differs from the standard approach because of the nature of the models that are used to generate forecasts. We will consider the effects of parameter estimation uncertainty on tests of forecast encompassing when the forecasting models are logit models estimated by nonlinear least squares, following West (1996), West and McCracken (1998), and West (2001). Thirdly, we wish to evaluate forecasts which are not necessarily based on formal models, or are based on well-defined information sets but do not make optimal use of that information.

A number of related tests of forecast encompassing appear in the literature for point forecasts. Contrary to what is often assumed, we show that these tests can give different outcomes even in the standard case (i.e., for point forecasts). With regard to the second and third points in the previous paragraph, we show that the asymptotic distributions of the tests are differentially affected by parameter estimation uncertainty, extending West and McCracken (1998), and West (2001), and that in particular, the outcomes may depend on whether or not the forecasts are the optimal predictors for a given information set.

Our analysis assumes linear combinations of forecasts, as in the literature for encompassing tests for standard point forecasts. Linear combination of probability forecasts is commonly used in the literature on combining experts' subjective probability distributions (see e.g., Genest and Zidek (1986) for a critique and annotated bibliography), where it is referred to as the 'linear opinion pool'. Linear combination (equal-weighted) of individuals' probability forecasts is also used to construct the Survey of Professional Forecasters' anxious index described in section 5.2.

The plan of the paper is as follows. In section 2 we discuss the two most commonly used loss functions for probability forecasts: QPS and LPS. One of the standard ways of evaluating (or 'scoring') probability forecasts (QPS) motivates the use of the standard forecast encompassing test framework, and we show that within the standard framework the different approaches to testing for forecast encompassing are not necessarily equivalent. In section 3 we consider a different way of testing for forecast encompassing, based on the alternative scoring method LPS. We show that the conditions for forecast encompassing are the same as under QPS, and moreover the different testing approaches behave in a similar fashion, in population. Section 4 presents the test statistics and corresponding null limiting distributions, for both QPS- and LPS-based encompassing testing methods. In addition, the impact on the tests of estimating the models' parameters is considered, extending some existing results on forecast encompassing tests and parameter estimation uncertainty, and illustrating the implications for probability forecasts. Monte Carlo simulation is employed to assess the finite sample properties of the different tests. Section 5 contains two empirical applications:

an evaluation of model-based forecasts, and an evaluation of survey-based forecasts. In keeping with the majority of the literature on forecast combination and encompassing, we do not consider the possibility of combining information at the modelling stage to produce a 'super model', as an alternative strategy for generating forecasts. In some cases this may be possible, as when the information sets on which the individual forecasts are based are observed, but is clearly impossible for survey-based forecasts (as in our second application). Section 6 concludes.

2 Probability forecasts and forecast encompassing tests based on QPS

In common with the literature on forecast combination, we consider linear combinations of forecasts. Denoting the two rival forecasts by f_1 and f_2, then the combined forecasts is $f_c = \alpha + \beta_1 f_1 + \beta_2 f_2$. Standard approaches to testing forecast encompassing (when the forecasts are point forecasts) are based on whether one forecast, say f_1, contains all the useful information in f_2, in the mean-squared-error sense. Forecast encompassing is said to hold when the combination of f_1 and f_2 does not have a MSE which is significantly smaller than that of f_1 alone. More generally then, given a loss function, forecast encompassing holds when setting $\beta_2 = 0$ does not result in a statistically significant increase in the loss function.

The two main ways of scoring probability forecasts are the quadratic and logarithmic scores. Given a probability forecast f, of the binary event $Y = 0$ or $Y = 1$, the Brier or quadratic probability score (QPS: Brier (1950)) is simply $(f - Y)^2$, corresponding to the usual notion of squared-error loss.[3] The logarithmic probability score (LPS: see Brier (1950) and Good (1952)) is defined as: $-Y \log(f) - (1 - Y)\log(1 - f)$.[4] For a sequence of probability forecasts and outcomes, $\{f_t, y_t\}$, $t = 1, \ldots, n$, these scores are calculated as:

$$QPS = \frac{1}{n}\sum_{t=1}^{n} 2\,(f_t - y_t)^2 \tag{1}$$

[3] Note that in contrast to the evaluation of point forecasts, the actual values are not observed, so that f is compared to a binary $0 - 1$ variable rather than the true probability. Probability forecasts can be evaluated by translating the forecasts into 'categorical' forecasts of the binary event, and then looking at the coincidence of predicted 1's and actual 1's, for example. In general this involves a loss of information, as a forecast of 0.99 is treated identically to a forecast probability of 0.51, if the usual 0.5 rule is used to translate forecast probabilities into event probabilities (see, e.g., Birchenhall, Jessen, Osborn and Simpson (1999)). Granger and Pesaran (2000) consider probability and categorical forecasts in a decison-based context, and show that only in specific circumstances will the two lead to the same actions.

[4] As written, both scores give possible values on $[0, +\infty)$.

and:

$$LPS = -\frac{1}{n}\sum_{t=1}^{n}\left[y_t \ln f_t + (1-y_t)\ln(1-f_t)\right] \tag{2}$$

As QPS is proportional to squared-error loss, the adoption of this measure leads to tests of forecast encompassing based on whether the combination of forecasts results in a statistically significant reduction in the expected squared error, as in the point forecast case. These two well-known measures of scoring probability forecasts have been used in economic applications by Diebold and Rudebusch (1989) and Anderson and Vahid (2001), *inter alia*. In the remainder of this section, we consider forecast encompassing and QPS, and in the following section LPS.

We consider three tests of forecast encompassing that have been applied to point forecasts. If we define the forecast errors as $e_{it} = y_t - f_{it}$, $i = 1, 2$, then these tests are based on the following three regressions:

FE(1)

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

FE(2)

$$y_t = \alpha' + \left(1 - \beta_2'\right) f_{1t} + \beta_2' f_{2t} + \varepsilon_t$$

or, equivalently:

$$e_{1t} = \alpha' + \beta_2' \left(e_{1t} - e_{2t}\right) + \varepsilon_t$$

FE(3)

$$y_t = \alpha'' + f_{1t} + \beta_2'' f_{2t} + \varepsilon_t$$

or, equivalently:

$$e_{1t} = \alpha'' + \beta_2'' f_{2t} + \varepsilon_t.$$

The null hypothesis that f_{1t} forecast encompasses f_{2t} is $\beta_2 = 0$ in FE(1), and similarly $\beta_2' = 0$ and $\beta_2'' = 0$ for FE(2) and FE(3) respectively. The alternative hypothesis is typically 1-sided, i.e., $\beta_2 > 0$ (or $\beta_2' > 0$, $\beta_2'' > 0$), precluding the possibility of a negative weight on f_{2t}. Regression FE(1) is the most general formulation, in that the coefficient on f_{1t} is left unrestricted under both the null and alternative. Fair and Shiller (1990) argue for the use of FE(1) over alternatives such as FE(2), which restricts $\beta_1 + \beta_2 = 1$ as part of the maintained hypothesis, so that only convex combinations are allowed. FE(3) tests $\beta_2'' = 0$ but with a maintained of $\beta_1 = 1$. FE(3) is the form of the original Chong and Hendry (1986) forecast encompassing test. Each of the above regressions allows the individual forecasts to be biased. Under the null $\beta_2 = 0$, f_{1t} is unbiased when $\alpha = (1 - \beta_1)E(f_{1t})$; this condition simplifies to $\alpha = 0$ when $\beta_1 = 1$, as assumed in FE(2) and FE(3) under the null. Recently, West and McCracken (1998) and West (2001) have considered FE(3) and FE(2) respectively, in the context of estimation uncertainty, and Harvey, Leybourne and Newbold (1998) have considered the effects of conditionally heteroscedastic forecast errors in FE(2).

Although there is little to suggest the use of one formulation over another in the literature (except when there are unit roots: see Ericsson (1993, p. 650)), it is straightforward to show that $\beta_2 = 0$ in FE(1) does not imply either $\beta_2' = 0$ (in FE(2)) or $\beta_2'' = 0$ (in FE(3)). Recall that $\beta_2 = 0$ is the condition for forecast encompassing. Thus, f_1 may forecast encompass f_2 but tests based on FE(2) and FE(3) will be invalid. Without making any specific assumptions about the process $\{y_t, f_{1t}, f_{2t}\}$, from the population values of the least squares estimators of β_2, β_2' and β_2'' we have:

$$
\begin{aligned}
\text{FE(1)} &: \quad \beta_2 = 0, \quad \text{if } V(f_{1t})C(y_t, f_{2t}) - C(f_{1t}, f_{2t})C(y_t, f_{1t}) = 0 \\
\text{FE(2)} &: \quad \beta_2' = 0, \quad \text{if } V(f_{1t}) - C(y_t, f_{1t}) + C(y_t, f_{2t}) - C(f_{1t}, f_{2t}) = 0 \\
\text{FE(3)} &: \quad \beta_2'' = 0, \quad \text{if } C(y_t, f_{2t}) - C(f_{1t}, f_{2t}) = 0.
\end{aligned} \tag{3}
$$

It follows that provided f_{2t} is uncorrelated with f_{1t} and y_t, then both $\beta_2 = 0$ and $\beta_2'' = 0$, so that f_1 encompasses f_2, and FE(3) will also indicate encompassing. But the same is not true of FE(2) in general: for $\beta_2' = 0$ the additional condition that $V(f_{1t}) - C(y_t, f_{1t}) = 0$ has to be satisfied. When f_{2t} is correlated with f_{1t} and y_t, it is possible that $\beta_2 = 0$ without either $\beta_2' = 0$ or $\beta_2'' = 0$. In which case tests based on both FE(2) and FE(3) are invalid as tests of encompassing. As we define encompassing by $\beta_2 = 0$, it would seem natural to base tests on $\beta_2 = 0$ in FE(1). There are two reasons to also consider FE(2) and FE(3). We might expect tests based on these regressions to be more powerful when the population values of the parameters satisfy $\beta_1 + \beta_2 = 1$ and $\beta_1 = 1$ respectively. Secondly, as these tests are commonplace in the literature, we wish to see whether they are likely to be misleading in practice. In order to assess the practical usefulness of FE(2) and FE(3) we check whether β_2' and β_2'' are zero for typical data generation processes for which $\beta_2 = 0$.

2.1 Data generation process

Below we describe the probability-forecast data generation process used to investigate the three forecast encompassing tests. We begin by deriving the population values of β_2, β_2' and β_2'' for a range of cases: when the forecasts are correlated, biased, and when there is model mis-specification; this allows us to ascertain circumstances under which forecast encompassing holds. The same DGP will later be used to investigate the small-sample properties of the tests of forecast encompassing, including the effects of parameter estimation uncertainty. The forecasts can be viewed as the forecasts from estimated logit models, but the formulation is general enough to support other interpretations.

The DGP is given by:

$$\begin{aligned} y_t &= 1\left(\frac{\exp(\delta_0+\delta_1 X_{1t}+\delta_2 X_{2t}+\delta_3 Z_t)}{1+\exp(\delta_0+\delta_1 X_{1t}+\delta_2 X_{2t}+\delta_3 Z_t)} > \upsilon_t\right) \\ f_{1t} &= \frac{\exp(\theta_{01}+\theta_{11}X_{1t})}{1+\exp(\theta_{01}+\theta_{11}X_{1t})} \\ f_{2t} &= \frac{\exp(\theta_{02}+\theta_{12}X_{2t})}{1+\exp(\theta_{02}+\theta_{12}X_{2t})} \end{aligned} \tag{4}$$

where:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \\ Z_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ \mu_Z \end{bmatrix}, \begin{bmatrix} 1 & \rho_{X_1,X_2} & \rho_{X_1,Z} \\ \rho_{X_1,X_2} & 1 & \rho_{X_2,Z} \\ \rho_{X_1,Z} & \rho_{X_2,Z} & 1 \end{bmatrix}\right).$$

Here, y_t is a binary variable that depends on three explanatory variables, X_{1t}, X_{2t} and Z_t, when $\delta_i \neq 0$, for $i = 1, 2, 3$. $y_t = 1$ when $\frac{\exp(\delta_0+\delta_1 X_{1t}+\delta_2 X_{2t}+\delta_3 Z_t)}{1+\exp(\delta_0+\delta_1 X_{1t}+\delta_2 X_{2t}+\delta_3 Z_t)}$ exceeds υ_t, a uniform random variable on the unit interval, that is independent of X_{1t}, X_{2t} and Z_t, and otherwise $y_t = 0$. The two rival forecasts depend on X_{1t} and X_{2t}, respectively. The form of the dependence restricts $f_{it} \in (0, 1)$, $i = 1, 2$, so that f_{1t} and f_{2t} can be interpreted as probabilities. The explanatory variables follow a multivariate normal, with variances normalized to unity and non-zero correlations, so that $C(f_{1t}, f_{2t}) \neq 0$ in general, allowing the rival forecasts to be correlated.[5]

A necessary condition for f_1 to encompass f_2 is that $\delta_2 = 0$ (and $\delta_1 \neq 0$) otherwise y_t depends on X_{2t} directly. If in addition we set $\delta_3 = 0$, then the model underlying f_1 is correctly-specified in the sense that it contains all the variables that affect y_t. Conversely, setting $\delta_3 \neq 0$ allows us to consider the effects of model mis-specification on the tests of encompassing.

2.2 Forecast encompassing based on QPS

Although conceptually straightforward, the form of the DGP does not allow us to obtain the probability limits of β_2, β_2' and β_2'' analytically. However, expressions for the moments involving $\{y, f_1, f_2\}$ for this DGP can be obtained, and subsequently evaluated by numerical integration for given values of the underlying parameters (δ_i, $i = 1, 2, 3$; μ_Z; ρ_{X_1,X_2}; $\rho_{X_1,Z}$; $\rho_{X_2,Z}$), with the resulting values substituted into (3) to obtain β_2, β_2' and β_2''. In Appendix 1 we report expressions for these

[5] Here and throughout the paper, we denote variance by $V(.)$ and covariance by $C(.,.)$.

moments for the case $\delta_0 = 0$, $\delta_1 = 1$. Note that in order to evaluate the moments numerically, we first need the population values of θ_{ij}—these cannot be obtained analytically from the underlying parameters. We assume the interpretation that the forecasts are obtained from estimated logit models, with the θ_{ij} parameters representing the population coefficients from logit regressions of y_t on a constant and X_{1t}, and a constant and X_{2t}, for f_{1t} and f_{2t} respectively. We therefore simulate the population θ_{ij} values by taking the average of logit parameter estimates from $10,000$ replications of the DGP with $n = 10,000$. Note that when f_{2t} is independent of f_{1t} and y_t, θ_{12} converges in probability to zero, and therefore f_{2t} approaches a constant in the limit. This case is of little interest in practice, and we focus exclusively on cases where $\rho_{X_1,X_2} > 0$.

Table 1 (Panel A) gives the population values of the parameters in the forecast encompassing regressions for a range of values of the underlying parameters. The table displays results for $\delta_3 = 0$ and $\delta_3 \neq 0$, allowing for both correctly specified and mis-specified forecasts, and results are reported for both biased and unbiased forecasts. Unbiased forecasts result when $\mu_Z = 0$, and the θ_{ij} determining f_{it} are the population parameters associated with logit regressions of y_t on a constant and X_{it}. We obtain biased forecasts when $\mu_Z = 1$ (with $\delta_3 \neq 0$) and the forecasts are given by:

$$
\begin{aligned}
f_{1t} &= \frac{\exp(\theta_{11} X_{1t})}{1 + \exp(\theta_{11} X_{1t})} \\
f_{2t} &= \frac{\exp(\theta_{12} X_{2t})}{1 + \exp(\theta_{12} X_{2t})}
\end{aligned}
$$

with the θ_{ij} being the population parameters associated with logit regressions of y_t on X_{1t} and X_{2t} respectively, without a constant term. Results are reported for representative combinations of the design parameter settings $\rho_{X_1,X_2} = \{0.2, 0.5, 0.8\}$, $\rho_{X_1,Z} = \{0, \pm 0.5\}$, $\rho_{X_2,Z} = \{0, 0.5\}$.[6]

The results for the model-based forecasts using QPS can be summarised as follows (assuming throughout that $\delta_1 \neq 0$, $\delta_2 = 0$ and $\rho_{X_1,X_2} > 0$):

(a) when the model giving rise to f_1 is correctly specified (because $\delta_3 = 0$), f_1 encompasses f_2 regardless of the correlation between X_1 and X_2 (and therefore f_1 and f_2),

(b) when the model generating f_1 is mis-specified (e.g., $\delta_3 \neq 0$), encompassing fails when (i) X_1 is correlated with the omitted variable Z, or (ii) X_2 is correlated with the omitted variable Z,

(c) model mis-specification is not sufficient for forecast encompassing to fail: f_1 encompasses f_2 provided Z is uncorrelated with both X_1 and X_2,

(d) the encompassing results for the three testing approaches FE(1), FE(2) and FE(3) coincide,

[6] Note that the parameter combination $\rho_{X_1,X_2} = 0.5$, $\rho_{X_1,Z} = -0.5$, $\rho_{X_2,Z} = -0.5$ yields a singular covariance matrix Ω and is therefore omitted.

i.e., when $\beta_2 = 0$, it is also the case that $\beta_2' = \beta_2'' = 0$.

Results (a)-(c) are intuitive except perhaps for (b) (i), i.e., the lack of encompassing when X_1 is correlated with Z. For example, encompassing fails under (b) (ii) because X_2 acts as a proxy for the omitted variable, but under (b) (i), X_2 is not correlated with Z, and it may seem surprising that the omission of Z from the model for f_1 causes a failure of encompassing in general (i.e., unless X_1 and X_2 are uncorrelated). This finding is not specific to the probability forecasting case. In Appendix 2 we illustrate the failure of encompassing in the analogous case for the standard point forecast setup, for which analytical results can be derived.

In both the unbiased and biased model-based forecast cases, the f_{it} are optimal predictors given the forecasting model and the information set (an intercept and X_{it}, or just X_{it}). By construction, the f_{it} satisfy $E\left[(y_t - f_{it})\,\mathbf{X}_{it}\right] = \mathbf{0}$ in the first case, where $\mathbf{X}_{it} = [1, X_{it}]'$, and $E\left[(y_t - f_{it})\,X_{it}\right] = 0$ in the second. The table also records the population values of the encompassing regression parameters when instead the forecasts are given by $f_{1t} = \exp(X_{1t})\,(1 + \exp(X_{1t}))^{-1}$ and $f_{2t} = \exp(X_{2t})\,(1 + \exp(X_{2t}))^{-1}$, for which $E\left[(y_t - f_{it})\,\mathbf{X}_{it}\right] \neq \mathbf{0}$. We refer to these as 'non-model'-based forecasts. They allow for an inefficient use of information in the sense that the model's forecast errors and explanatory variables are correlated. The expressions for the moments involving $\{y, f_1, f_2\}$ for the non-modelled forecasts can be obtained from Appendix 1 by setting $\theta_{01} = \theta_{02} = 0$ and $\theta_{11} = \theta_{12} = 1$ in the general expressions given there. These moments are evaluated numerically for particular paramater values, and substitution into the expressions for the FE(1) – FE(3) population parameters gives the values reported in the table.

In contrast to the results for the model-based forecasts, the three testing approaches FE(1) – FE(3) do not coincide for the non-model-based forecasts. While encompassing holds (i.e., $\beta_2 = 0$) in the same circumstances as for the model-based forecasts, the FE(2) and FE(3) regressions imply a failure of encompassing unless $\delta_3 = 0$, in which case the non-model-based forecasts are actually correctly specified, coinciding with the optimal model-based forecasts. When mis-specification occurs, however, i.e., when $\delta_3 = 1$, β_2' and β_2'' are non-zero even though $\beta_2 = 0$. In these instances, both $\beta_1 \neq 1$, and $\beta_1 + \beta_2 \neq 1$ (see table), that is, the restrictions underlying FE(2) and FE(3) (relative to FE(1)) are violated. The non-model-based case therefore highlights a potential weakness in the application of tests based on FE(2) and FE(3), while FE(1) remains the more general approach.

3 Forecast encompassing and the logarithmic scoring rule (LPS)

The LPS loss function (2) for the combined forecast is:

$$LPS = -\frac{1}{n}\sum\left[y_t \ln(\alpha + \beta_1 f_{1t} + \beta_2 f_{2t}) + (1-y_t)\ln(1-\alpha-\beta_1 f_{1t} - \beta_2 f_{2t})\right]$$

Unconstrained minimisation with respect to $\{\alpha, \beta_1, \beta_2\}$ gives the first order population conditions:

$$\begin{aligned}
E\left[\frac{y_t}{\alpha+\beta_1 f_{1t}+\beta_2 f_{2t}} - \frac{1-y_t}{1-\alpha-\beta_1 f_{1t}-\beta_2 f_{2t}}\right] &= 0 \\
E\left[\left(\frac{y_t}{\alpha+\beta_1 f_{1t}+\beta_2 f_{2t}} - \frac{1-y_t}{1-\alpha-\beta_1 f_{1t}-\beta_2 f_{2t}}\right) f_{1t}\right] &= 0 \\
E\left[\left(\frac{y_t}{\alpha+\beta_1 f_{1t}+\beta_2 f_{2t}} - \frac{1-y_t}{1-\alpha-\beta_1 f_{1t}-\beta_2 f_{2t}}\right) f_{2t}\right] &= 0
\end{aligned}$$

If f_{2t} is independent of y_t and f_{1t}, and encompassing held (i.e., $\beta_2 = 0$), some simplification is possible. For example, the third condition reduces to the first, and becomes redundant, so for forecast encompassing to hold, we require a unique solution to the following two equations, for admissible values of α and β_1:

$$E\left[\frac{y_t}{\alpha+\beta_1 f_{1t}} - \frac{1-y_t}{1-\alpha-\beta_1 f_{1t}}\right] = 0 \tag{5}$$

$$E\left[\left(\frac{y_t}{\alpha+\beta_1 f_{1t}} - \frac{1-y_t}{1-\alpha-\beta_1 f_{1t}}\right) f_{1t}\right] = 0 \tag{6}$$

The simplest approach to investigating the tests of forecast encompassing for LPS is by simulation. For given values of the underlying parameters ($\delta_i, i = 1,2,3; \mu_Z; \rho_{X_1,X_2}; \rho_{X_1,Z}; \rho_{X_2,Z}$), the population values of the θ_{ij} are taken to be the values that we determined by simulation when investigating the population QPS combination weights. These values are then used in the simulations proper to construct f_{1t} and f_{2t} as model-based forecasts. Bias is introduced as discussed in section 2, namely, by setting $\mu_Z \neq 0$ and omitting the intercepts from the logit model forecasts. For the non-model-based forecasts, the pre-simulations are unnecessary, as we simply set $f_{1t} = \exp(X_{1t})(1+\exp(X_{1t}))^{-1}$ and $f_{2t} = \exp(X_{2t})(1+\exp(X_{2t}))^{-1}$, as for QPS.

Table 1 (Panel B) records the population values of the combination weights for LPS for the same parameter settings as for QPS, so that the results for QPS and LPS are directly comparable. As in the case of QPS, the FE(2) test imposes $\beta_1 = (1-\beta_2)$, and the FE(3) test sets $\beta_1 = 1$. It is apparent that the results match: when forecast encompassing holds for LPS it also holds for QPS, and the three different LPS tests match in the case of the model-based forecasts, but not otherwise.

The equivalence between the values of $\boldsymbol{\beta}$ $(= [\alpha, \beta_1, \beta_2]')$ that minimize LPS and QPS in population is perhaps unsurprising as the two alternative loss functions can be viewed as two different estimators of the combination weights. Given that y_t is binary, the combined probability $P(y_t = 1) = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t}$ implies that

$$E(y_t |\, f_{1t}, f_{2t}) = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t}$$

with associated regression model:

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

OLS estimation is then the same as QPS. Alternatively, the likelihood route is based on maximizing (the log of) $L = \prod_{y_t=1} P(y_t = 1)^{y_t} \prod_{y_t=0} P(y_t = 0)^{1-y_t}$, which corresponds to LPS. In the next section we consider tests based on LPS and QPS, as the different properties of the loss functions (LPS penalises large errors relatively more heavily) means that a combined forecast may result in a statistically significant reduction in loss for one loss function (forecast encompassing rejected) but not for the other (forecast encompassing not rejected).

4 Forecast encompassing test statistics for QPS and LPS

In this section we present forecast encompassing test statistics for use with FE(1), FE(2) and FE(3) for both QPS and LPS.

4.1 QPS tests

We follow Harvey *et al.* (1998) and West (2001) in considering tests of QPS forecast encompassing based on the Diebold and Mariano (1995) (DM)-type approach. Following these papers, we assume that the forecasts are generated from a non-nested structure. When the forecasts are from models which are nested, an analysis along the lines of Clark and McCracken (2001) is required (see also the review article by West (2006)). Extending the Harvey *et al.* (1998) analysis of tests based on FE(2) to allow for biased forecasts, by working with deviations from means, it is straightforward to define DM tests for both FE(2) and FE(3) as:

$$\text{DM} = \frac{n\overline{d}}{\sqrt{\sum_{\tau=-(h-1)}^{h-1} \sum_{t=|\tau|+1}^{n} (d_t - \overline{d})(d_{t-|\tau|} - \overline{d})}} \tag{7}$$

$$\text{MDM} = n^{-1/2}[n + 1 - 2h + n^{-1}h(h-1)]^{1/2} DM$$

where

$$d_t = (e_{1t} - \overline{e}_1)[(e_{1t} - \overline{e}_1) - (e_{2t} - \overline{e}_2)] \qquad \text{for FE(2)}$$
$$d_t = (e_{1t} - \overline{e}_1)(f_{2t} - \overline{f}_2) \qquad \text{for FE(3)}$$

and $\overline{d} = n^{-1} \sum_{t=1}^{n} d_t$. Harvey *et al.* (1998) provide the rationale for the modification to improve the small-sample performance of the DM test, that results in the MDM statistic. We also define DM and MDM tests for FE(1) by making use of the Frisch-Waugh theorem. The population parameter β_2 in FE(1) is identical to β_2 in:

$$\eta_{1t} = \beta_2 \eta_{2t} + \nu_t$$

where η_{1t} and η_{2t} are the errors from the regression of y_t and f_{2t}, respectively, on a constant and f_{1t}. The null of $\beta_2 = 0$ therefore holds when $E(\eta_{1t}\eta_{2t}) = 0$, allowing use of the DM and MDM tests with:

$$d_t = \hat{\eta}_{1t}\hat{\eta}_{2t} \qquad \text{for FE(1).}$$

Abstracting temporarily from the forecasts being based on models, whereby the estimation of the models' parameters imparts an additional source of uncertainty, then the asymptotic normality of DM and MDM follows from standard results. In practice, we compare the MDM tests to a Student t distribution with $n-1$ degrees of freedom. Alternatively, one could consider a regression-based t-statistic of the null that $\beta_2 = 0$ in FE(1) (and similarly for FE(2) and FE(3)), calculated using an appropriate heteroskedasticity and autocorrelation consistent variance estimator. Such a test also has a limiting standard normal distribution under the null. Preliminary findings indicated that tests based on the DM approach (especially MDM) generally outperformed regression-based t-statistics. The DM approach can also be readily extended to consider the effects of parameter estimation: see section 4.3.

A final point is that although it may appear that the form of the denominator in (7) is inappropriate for FE(1) and FE(3) when $\{y_t, f_{1t}, f_{2t}\}$ is autocorrelated, allowing for dependence only up to order $h-1$ in d_t is in fact appropriate for all three approaches for 'well-specified' forecasts. In the case of FE(2), d_t depends only on the forecast errors, with dependence structure that can reasonably be assumed to be no more than an MA($h-1$) for a horizon of h. Thus the form of the denominator in (7) is readily justified for FE(2). For FE(1) and FE(3) we arrive at the same conclusion, although the argument is a little more involved. First consider FE(3) when $h = 1$. Regardless of any autocorrelation in f_{1t} and y_t, for well conceived forecasts we expect e_{1t} to be free of autocorrelation. As d_t is the product of $(e_{1t} - \overline{e}_1)$ and a potentially autocorrelated process $(f_{2t} - \overline{f}_2)$, it follows that d_t should also be serially uncorrelated. For FE(1) a similar argument establishes that d_t is serially uncorrelated for $h = 1$: $\hat{\eta}_{1t}$ is the residual from a regression of y_t on a constant and f_{1t}, and will therefore behave like a (bias-corrected) 1-step forecast error, so that

d_t will be free of autocorrelation regardless of the autocorrelation structure of $\hat{\eta}_{2t}$. Turning now to h-step-ahead forecasts, the same arguments go through, except now we would expect $(e_{1t} - \bar{e}_1)$ and $\hat{\eta}_{1t}$ to be autocorrelated to a maximum order of MA$(h-1)$. Thus for both FE(1) and FE(3) d_t should be autocorrelated no more than order $(h-1)$.

The non-model-based forecasts do not give rise to h-step forecast errors which exhibit correlation of no more than $h-1$. For forecasts such as these additional covariance terms need to be included in the long-run variance estimator. Various kernel functions for weighting the covariance terms could be used, although the simplest approach is to continue with the uniform, but with the Newey-West lag truncation $L = trun\left(4(n/100)^{\frac{2}{9}}\right)$ replacing $h-1$ in the denominator. The advantage of maintaining uniform weighting is that the MDM small-sample modification is derived assuming the use of this variance estimator in the DM statistic. Because $L < h-1$ is possible for small estimation sample sizes, we suggest using $\max(L, h-1)$.

4.2 LPS tests

LPS-based tests of forecast encompassing are based on the MLE of the likelihood:

$$\ln L = \sum \ln L_t = \sum \left[y_t \ln f_t + (1-y_t)\ln(1-f_t)\right] \tag{8}$$

noting that this likelihood corresponds to $-n^{-1}$ times LPS. For the test based on the FE(1) approach, $f_t = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t}$, with obvious modifications for tests based of FE(2) and FE(3). For the DGP given in section 2.1 it is straightforward to show that (8) is the true likelihood so that the consistency of the MLE $\hat{\beta}$ for β follows under standard conditions. We use the OPG estimator of the covariance matrix of $\hat{\beta}$, denoted $\hat{V}^G = (g'g)^{-1}\Big|_{\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2}$, where g is the $n \times 3$ gradient matrix with typical row $g_t = \frac{\partial \ln L_t}{\partial \alpha}, \frac{\partial \ln L_t}{\partial \beta_1}, \frac{\partial \ln L_t}{\partial \beta_2}$, and let $\hat{g}_t$ (and $\hat{g}$) denote this quantity evaluated at $\hat{\beta}$. For LPS, the 'sandwich estimator' is identical to the OPG estimator, because the hessian $H = \frac{\partial^2 \ln L}{\partial\beta\partial\beta'}$ is identical to $-g'g$. Therefore the sandwich estimator $\hat{V}^S = \hat{V}^H \hat{g}'\hat{g}\hat{V}^H = \hat{V}^G$ where $\hat{V}^H = (-H)^{-1}\Big|_{\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2}$.

When the $\{y_t, f_{1t}, f_{2t}\}$ are autocorrelated, we need to use an autocorrelation-consistent estimator of the covariance of $\hat{\beta}$. Unlike for the (M)DM test denominator for QPS-forecast encompassing, we need to allow for general autocorrelation in excess of order $h-1$, as the arguments that permit us to restrict the dependence to $h-1$ in that context no longer hold. Put another way, we need an autocorrelation correction even for $h=1$ when the data and forecasts display serial correlation. The autocorrelation-consistent variance estimator we employ is given by:

$$\hat{V}_E = \hat{V}^G \hat{\Gamma} \hat{V}^G$$

where:

$$\hat{\Gamma} = \sum_{\substack{i=1 \\ |i-j|\leq L}}^{n} \sum_{j=1}^{n} w\left(_{i-j}\right) \hat{g}_i' \hat{g}_j$$

with $w\left(_{i-j}\right) = 1 - \frac{|i-j|}{1+L}$ and the Newey-West lag truncation $L = trun\left(4(n/100)^{\frac{2}{9}}\right)$ rather than $L = h-1$ (provided $L > h-1$). Under the null, $t = \hat{\beta}_2 / \sqrt{\left(\hat{V}_E\right)_{3,3}}$ has a standard normal distribution in the limit. In practice, we compare the test statistic to a Student t with $n-3$ degrees of freedom ($n-2$ for the FE(2) and FE(3) versions).

4.3 Parameter estimation uncertainty

When forecasts are derived from models with estimated parameters, as will typically be the case in practice, it has recently been shown that the asymptotic distributions of tests of predictive accuracy may be affected: see West (1996), West and McCracken (1998) and West (2001). West (2001) establishes that forecast encompassing tests of the sort we describe as FE(2) will be affected by the additional source of uncertainty from estimating the models' parameters, and West and McCracken (1998) show that the same is true for the FE(3) tests. West and McCracken (1998) and West (2001) show by Monte Carlo the influence of estimation uncertainty and illustrate the properties of tests which make an allowance for estimation uncertainty. Both studies consider linear models estimated by OLS, and assume a squared-error loss function (corresponding to QPS). In this section we extend their results to cover the FE(1) tests, and forecasts obtained from estimated logit models, maintaining the assumption of a QPS loss function. We show that, in contrast to tests based on FE(2) and FE(3), tests based on FE(1) are unaffected by estimating parameters in linear models, and by simulation that the same is true of logit/probit models estimated by maximum likelihood (ML).

Beginning with the linear model setup in West (2001), we suppose forecasts are obtained from models with regressors X_{1t} and X_{2t}, assumed here to be single variables for convenience:[7]

$$\begin{aligned} y_t &= X_{1t}\theta_1 + e_{1t} \\ y_t &= X_{2t}\theta_2 + e_{2t} \end{aligned} \tag{9}$$

When the parameters are known, and when the forecast errors are assumed to have zero mean, the FE(2) DM test that the first model encompasses the second utilises $d_t = e_{1t}(e_{1t} - e_{2t})$. Under the

[7] If X_{1t} and X_{2t} are vectors, they may have variables in common, but we require that neither is a subset of the other. For the nested case, see Clark and McCracken (2001).

null that $E(d_t)=0$, the sample average $\widetilde{d}=n^{-1}\sum_{t=R+1}^{R+n} d_t$ scaled as $\sqrt{n}\widetilde{d}/\sqrt{S}$ is asymptotically $N(0,1)$, where $S=V(d_t)$, and S is estimated by the sample variance of d_t.[8] When instead the forecasts are generated from models with estimated parameters, so $\overline{d}=n^{-1}\sum_{t=R+1}^{R+n}\hat{e}_{1t}(\hat{e}_{1t}-\hat{e}_{2t})$, with $\hat{e}_{it}=X_{it}\left(\theta_i-\hat{\theta}_i\right)+e_{it}$, West and McCracken (1998) show that:

$$\sqrt{n}\left[\overline{d}-E(d_t)\right]\overset{a}{\sim}N(0,\Omega) \tag{10}$$

where $\Omega=S+\pi DV_\theta D'$.[9] Here, V_θ is the asymptotic covariance matrix of θ, where $\theta=(\theta_1,\theta_2)'$, $D=E\left(\frac{\partial d_t}{\partial\theta'}\right)$, and $\pi=n/R$, as $n,R\to\infty$, $\pi<\infty$. We can show that estimation uncertainty affects tests of predictive accuracy of probability forecasts from logit (or probit) models estimated by ML in the same *general* way that it affects OLS estimation of linear models: that is, (10) holds for estimated logit forecasts. This follows by checking that the assumptions in West (1996) that give rise to (10) are also satisfied by ML estimation of logit models. That the assumptions are satisfied follows immediately from West (1996, Assumption 2). In the nonlinear regression equation, $y_t=g(X_t,\theta^*)+e_t$, we need to be able to write $\hat{\theta}-\theta^*=B(t)H(t)$, where $H(t)$ is a sample average of the population orthogonality conditions, and $B(t)$ weights these conditions. We then require that $B(t)\overset{a.s}{\to}B$ and $H(t)=t^{-1}\sum_{s=1}^{t}h_s(\theta^*)$ with $E[h_s(\theta^*)]=0$. For the logit model, $g(X_t,\theta^*)=\Lambda(X_t,\theta^*)=\exp\left(\theta^{*\prime}X_t\right)\left(1+\exp\left(\theta^{*\prime}X_t\right)\right)^{-1}$ is a smooth function of θ (as required). The model is estimated by ML, so that, as noted by West, $H(t)$ is the score and $B(t)$ the inverse of the Hessian evaluated at a point θ^0, between θ^* and $\hat{\theta}$. For the logit model:

$$h_s=\left[y_s-\exp\left(\theta^{*\prime}X_s\right)\left(1+\exp\left(\theta^{*\prime}X_s\right)\right)^{-1}\right]X_s$$

and

$$B(t)=\left[-t^{-1}\sum_{s=1}^{t}\left(\exp\left(\theta^{0\prime}X_s\right)\left(1+\exp\left(\theta^{0\prime}X_s\right)\right)^2\right)X_sX_s'\right]^{-1}.$$

In general (that is, for linear models estimated by OLS, and logit/probit models estimated by ML), for the three tests of forecast encompassing we have the following expressions for d_t:

[8] We follow the convention in West (1996) and subsequent papers of R denoting the estimation sample, and $R+1$ to $R+n$ the set of (1-step) forecasts. We consider the 'fixed' forecasting scheme, whereby model estimates are obtained once only on the first R observations and are subsequently held fixed for all the forecasts.

[9] For 'rolling' and 'recursive' schemes (a fixed window moved through the data, and an expanding window) there are additional terms in the expression for Ω, but these are all zero when $D=0$.

$$
\begin{aligned}
\text{FE(1) } d_t &= \left\{[y_t - E(y_t)] - \frac{C(y_t, f_{1t})}{V(f_{1t})}[f_{1t} - E(f_{1t})]\right\}\left\{[f_{2t} - E(f_{2t})] - \frac{C(f_{1t}, f_{2t})}{V(f_{1t})}[f_{1t} - E(f_{1t})]\right\} \\
\text{FE(2) } d_t &= \{[y_t - E(y_t)] - [f_{1t} - E(f_{1t})]\}\{[f_{2t} - E(f_{2t})] - [f_{1t} - E(f_{1t})]\} \qquad (11) \\
\text{FE(3) } d_t &= \{[y_t - E(y_t)] - [f_{1t} - E(f_{1t})]\}[f_{2t} - E(f_{2t})]
\end{aligned}
$$

The expressions for d_t for FE(2) and FE(3) follow immediately, while FE(1) is discussed below. Although our interest is in the logit models, the linear model set up (9) turns out to be informative. The results for FE(2) and FE(3) are in the literature we have cited. For the DM test based on the FE(2) regression, we have $D = [-2E(e_{1t}X_{1t}) + E(e_{2t}X_{1t}), E(e_{1t}X_{2t})] = [E(e_{2t}X_{1t}), E(e_{1t}X_{2t})]$, using $E(e_{1t}X_{1t}) = 0$. Because $D \neq (0, 0)$, $\Omega \neq S$ and parameter estimation affects the asymptotic distribution of the forecast encompassing test. Similarly for the DM test based on FE(3): in population terms, $d_t = [e_{1t} - E(e_{1t})][f_{2t} - E(f_{2t})]$, so that $D = [-\theta_2 C(X_{1t}, X_{2t}), C(e_{1t}, X_{2t})] \neq (0, 0)$ in general.[10]

The effects of estimation uncertainty on tests based on FE(1) have not been considered to date. Recall that our DM type test for FE(1) is based on testing $E(d_t) = 0$, where $d_t = \eta_{1t}\eta_{2t}$, with η_{1t} and η_{2t} denoting the errors from the regressions

$$
\begin{aligned}
y_t &= \gamma_1 + \gamma_2 f_{1t} + \eta_{1t} \\
f_{2t} &= \phi_1 + \phi_2 f_{1t} + \eta_{2t}. \qquad (12)
\end{aligned}
$$

Using the population values of the regression parameters we obtain:

$$
\begin{aligned}
\eta_{1t} &= y_t - E(y_t) - \frac{C(y_t, f_{1t})}{V(f_{1t})}[f_{1t} - E(f_{1t})] \\
\eta_{2t} &= f_{2t} - E(f_{2t}) - \frac{C(f_{1t}, f_{2t})}{V(f_{1t})}[f_{1t} - E(f_{1t})] \qquad (13)
\end{aligned}
$$

For forecasts from linear models, e.g., $f_{it} = X_{it}\theta_i$, so that $E(f_{it}) = E(X_{it})\theta_i$, the errors η_{1t} and η_{2t} can be written as functions of moments of the data:

$$
\eta_{1t} = [y_t - E(y_t)] - \frac{C(y_t, X_{1t})}{V(X_{1t})}[X_{1t} - E(X_{1t})]
$$

[10] When f_{2t} is independent of f_{1t} and y_t, then it follows immediately from (11) that $D = E\left(\frac{\partial d_t}{\partial \theta'}\right) = 0$ for FE(1) – FE(3), so that the limiting distributions of the tests are unaffected by estimation uncertainty. However, as noted earlier, this case is of little interest in practice, with $f_{2t} = X_{2t}\hat{\theta}_2$ approaching a constant as the estimation sample R gets large.

$$\eta_{2t} = \theta_2 \left\{ [X_{2t} - E(X_{2t})] - \frac{C(X_{1t}, X_{2t})}{V(X_{1t})} [X_{1t} - E(X_{1t})] \right\}$$

so that we obtain d_t as:

$$d_t = \theta_2 \left\{ [y_t - E(y_t)] - \frac{C(y_t, X_{1t})}{V(X_{1t})} [X_{1t} - E(X_{1t})] \right\} \left\{ [X_{2t} - E(X_{2t})] - \frac{C(X_{1t}, X_{2t})}{V(X_{1t})} [X_{1t} - E(X_{1t})] \right\}. \tag{14}$$

Clearly, $\partial d_t / \partial \theta_1 = 0$, and:

$$E\left(\frac{\partial d_t}{\partial \theta_2}\right) = C(y_t, X_{2t}) - \frac{C(X_{1t}, X_{2t}) C(y_t, X_{1t})}{V(X_{1t})} \tag{15}$$

The null hypothesis, $\beta_2 = 0$, implies that $V(f_{1t})C(y_t, f_{2t}) - C(f_{1t}, f_{2t})C(y_t, f_{1t}) = 0$. Substituting for f_{1t} and f_{2t} in this expression, and dividing both sides by $\theta_1^2 \theta_2$ (noting that $\theta_1 \neq 0$, $\theta_2 \neq 0$), we find that the right-hand-side of (15) equals zero. So $D = (0, 0)$ under the null, and estimation uncertainty is irrelevant asymptotically. This contrasts with the findings using the DM variants of the FE(2) and FE(3) tests.

Consider now logit model forecasts, $f_{it} = \frac{\exp(\theta_{0i} + \theta_{1i} X_{it})}{1 + \exp(\theta_{0i} + \theta_{1i} X_{it})}$, where the parameter vector is now $\theta = (\theta_{01}, \theta_{11}, \theta_{02}, \theta_{12})$. In order to determine the effects of estimation uncertainty, we need $D = E\left(\frac{\partial d_t}{\partial \theta'}\right)$ for each of the three cases FE(1), FE(2) and FE(3). For FE(1) we have:

$$\frac{\partial d_t}{\partial \theta_{j1}} = [f_{2t} - E(f_{2t})] \frac{\partial \{[y_t - E(y_t)] - [f_{1t} - E(f_{1t})]\}}{\partial \theta_{j1}}, \qquad j = 0, 1$$

$$\frac{\partial d_t}{\partial \theta_{j2}} = \{[y_t - E(y_t)] - [f_{1t} - E(f_{1t})]\} \frac{\partial [f_{2t} - E(f_{2t})]}{\partial \theta_{j2}}, \qquad j = 0, 1.$$

and similar expressions can be obtained for FE(2) and FE(3). Unlike in the linear case (where $f_{it} = X_{it}\theta_i$), f_{it} is a non-linear function of θ_i, and calculating D appears to be analytically intractable, and so we proceed by Monte Carlo. It also appears that little progress can be made when instead the loss function is given by LPS, so that the impact of parameter estimation on tests of forecast encompassing for logit-model forecasts under LPS is also investigated by simulation.

4.4 A Monte Carlo of the small-sample test performance

As part of the overall assessment of the small-sample performance of the tests of forecast encompassing based on QPS and LPS, we design experiments that show up the effects of estimation uncertainty, as well as the effects of the size of the sample of available forecasts. Data is generated from the DGP (equation (4)) for $t = 1, \ldots, R, R+1, \ldots, R+n$, and the models are estimated on observations 1 through R, and forecasts are generated of $R+1$ to $R+n$. We use $n = \{25, 50, 100, 200, 500\}$, and values of R such that for each n, $n/R = \{2, 1, 0.5, 0.25, 0.125, 0.0625, 0\}$, where the limiting value of zero corresponds to the case of no parameter estimation uncertainty (using the simulated population θ_{ij} values as described in section 2.2), and $n/R = 2$ indicates twice as many forecasts as in-sample observations. All simulations were computed using 10,000 Monte Carlo replications.

The parameter values used to simulate data in the Monte Carlo match those underlying the estimates reported in Table 1. Note that the forecasting models match the logit models of the first empirical application in section 5, and so the results of the Monte Carlo will serve as a guide to the properties of the tests in that application, as well as showing up the effects of parameter estimation uncertainty. The encompassing tests are implemented as described in sections 4.1 and 4.2. The LPS tests we report include an autocorrelation-correction. This is not necessary here as we consider iid data and $h = 1$, but we do so as it is prudent in practical applications to apply such a correction.

Table 2 presents the Monte Carlo estimates of size for 1-sided 5% level QPS-MDM and LPS-t tests, for the three test approaches (FE(1), FE(2) and FE(3)), using the DGP given by (4) with $\delta_3 = 0$ and $\rho_{X_1,X_2} = 0.5$. The results for the probability forecast setup match the analytical findings for the linear-model case, in that the QPS and LPS tests based on FE(1) are immune to parameter estimation effects, while the tests based on FE(2) and FE(3) are not. For FE(1), the empirical sizes depend on n, and the LPS-t test is clearly over-sized for the smaller values of n, but for a given n are largely invariant to changes in n/R (that is, across columns). For FE(2) and FE(3) the empirical sizes approach the nominal 5% as n gets large *and* $n/R \to 0$, but are over-sized for larger n/R. The actual sizes are approximately 14% for FE(2) when $n = 2R$, for large n, and approximately $8\% - 10\%$ for FE(3). The limiting case of no estimation uncertainty is recorded in the final column. For both the QPS and LPS tests, tests based on FE(2) display under-sizing in small samples, while those based on FE(3) exhibit over-sizing. Qualitatively similar results were obtained for the mis-specified case (the DGP given by (4) with $\delta_3 = 1$, $\mu_Z = 0$, $\rho_{X_1,Z} = \rho_{X_2,Z} = 0$) and for the biased foreast case (the DGP given by (4) with $\delta_3 = 1$, $\mu_Z = 1$, $\rho_{X_1,Z} = \rho_{X_2,Z} = 0$).[11]

Table 3 reports empirical sizes for the same DGPs but using non-model-based forecasts. Here, the additional autocorrelation correction is applied to the QPS-MDM tests, as described in section 4.1. The FE(1) tests are correctly sized in large samples for both QPS and LPS, and are somewhat over-sized for small n. When the non-model-based forecasts are correctly specified, i.e., when $\delta_3 = 0$, the FE(2) and FE(3) tests are also correctly sized for large n, as would be expected given the population results of Table 1. However, for large n when $\delta_3 = 1$, the FE(2) tests are over-sized and the FE(3) tests are under-sized. This is again consistent with our earlier results, since, in

[11] These results are available from the authors on request.

population, $\beta_2' > 0$ and $\beta_2'' < 0$ for non-model-based forecasts when $\delta_3 = 1$, and the size results reported pertain to 1-sided tests. Size distortions also persist for the FE(2) and FE(3) tests in small samples.

Finally, Table 4 presents estimated powers of the tests for the case $\delta_3 = 0$, $\rho_{X_1,X_2} = 0.5$, for two alternative hypothesis DGPs, given by $\delta_1 = 0.5$, $\delta_2 = 0.5$, and $\delta_1 = 0$, $\delta_2 = 1$. We abstract from the issue of parameter estimation uncertainty for the purposes of this power comparison. The power of all the tests is increasing in n, and higher for the $\delta_1 = 0$, $\delta_2 = 1$ DGP than the $\delta_1 = 0.5$, $\delta_2 = 0.5$ DGP, as would be expected. For the model-based forecasts, the FE(1) based tests have generally the higher power compared to FE(2) and FE(3), under both QPS and LPS. The power advantages of FE(1) over both FE(2) and FE(3) are often substantial, while on the relatively few occasions where FE(1) is outperformed, the power losses are slight. For the non-model-based forecasts, FE(3) is always considerably inferior in terms of power compared to the other two approaches. Neither FE(1) nor FE(2) dominates the other, and although FE(2) has greater power than FE(1) in many cases, the differences in power are not great.

Overall, tests based on FE(1) appear to have the most desirable finite sample properties. They are robust to the effects of parameter estimation uncertainty, have good finite sample size and power relative to approaches based on FE(2) and FE(3), and are more reliable when the forecasts are non-model-based. We therefore recommend that the QPS-MDM and LPS-t tests based on the FE(1) approach should be used in practical application.

5 Empirical illustrations

Our first empirical application of the QPS and LPS probability forecast encompassing tests is to the evaluation of rival model-based forecast probabilities of recession. The second assesses the accuracy of individual survey respondents' probability forecasts relative to the 'consensus' or average probability forecasts of the phenomena of interest.

5.1 Evaluating model-based US recession forecast probabilities

The forecast encompassing tests are used to assess the relative information value of two leading indicators for post War US recessions. Forecasts of the probabilities of recession are generated from logit models that use, respectively, interest rate spreads, and some transformation of the oil price. Recessionary periods are those defined by the NBER's Business Cycle Dating Committee[12]. The motivation for focusing on spreads and oil prices is as follows. The usefulness of financial variables

[12] See `http://www.nber.org/cycles.html`.

for predicting US recessions has been established by Estrella and Mishkin (1998), Anderson and Vahid (2001) and Hamilton and Kim (2000), *inter alia*. The slope of the yield curve tends to dominate other financial variables as a predictor of recessions, in that the yield curve is often the single best choice, and the incremental benefit of including other financial variables is often small. Chauvet and Potter (2002) allow for structural breaks in probit models that use the yield curve to predict recessions, but we use simple constant-parameter models as our aim is simply to illustrate the use of the forecast encompassing tests. Recent research also suggests a role for oil prices as an indicator for output growth (e.g., Hamilton (1983), Mork (1989), Hooker (1996), Hamilton (1996), Carruth, Hooker and Oswald (1998) and Hamilton (2000)), and by association, for recessions, although there is some debate over the form of the relationship. We follow Lee, Ni and Ratti (1995) and Hamilton (1996) and use the *net increase* in oil prices over the previous year, and downweight increases in oil prices at times of relatively high volatility.

We use quarterly spread and oil price data for the period 1960:1–1999:4.[13] We divide the period into an in-sample estimation period (1960:1–1979:4) and a forecast period (1980:1–1999:4). Preliminary work suggested that the logit models of the recession indicator on the spread (and an intercept) appeared to work best if the spread is lagged three or four periods, and we chose a three period lag. For the oil price model, the recession indicator was regressed on the net increase lagged two periods. The in-sample fits of the two logit regression models are summarized in Table 5. The superior individual performance of the spread is apparent from the QPS and LPS, although from the p-values of the pseudo-R^2 statistics for both models it is evident that both leading indicators have significant in-sample predictive power for recessions.[14] It is therefore of interest to assess whether the oil price-logit model forecast probabilities contain useful information in terms of delivering a statistically-significant reduction in QPS and/or LPS when used in combination with the yield curve predicted probabilities.

Table 6 reports results using the two recommended forecast encompassing tests—the FE(1) based QPS-MDM test, and the FE(1) based autocorrelation-consistent LPS-t test (given that the

[13] The spread data is the 10-year treasury constant maturity rate less the 3-month treasury bill at secondary market rate. These are popular choices of long and short rates. Quarterly data is obtained from the monthly data by averaging. The data were taken from the FRED website `http://research.stlouisfed.org/fred2`. The oil price data are available from James Hamilton's web page `http://weber.ucsd.edu/~jhamilto`.

[14] Various R^2 measures have been proposed in the literature for discrete choice models. We follow Estrella and Mishkin (1998) and report a measure defined as:

$$R^2 = 1 - \left(\frac{ll_u}{ll_r}\right)^{-\left(\frac{2}{n} ll_r\right)}$$

where ll_u and ll_r are the unrestricted maximised value of log likelihood, and the value imposing the restriction that the slope coefficients are zero. The p-value is of the standard likelihood ratio test compared to a χ^2 distribution with one degree of freedom, as there is a single slope parameter.

data is autocorrelated). The forecast probabilities underlying the results are 1-step ahead forecasts for 1980:1–1999:4 ($n = 80$), generated using the coefficient estimates for 1960:1–1979:4. The tests are conducted using a 1-sided alternative, and the corresponding p-values are reported in the table. The null that the oil price-model forecasts encompass the spread-model forecasts is clearly rejected. This is consistent with the superior in-sample performance of the spread-model probability forecasts. This holds regardless of whether accuracy is assessed by QPS or LPS; for LPS the null is rejected at the 1% level. The more interesting hypothesis is whether the spread-model forecasts encompass the oil-price model forecasts. The null of encompassing in this direction is not rejected at the 5% level for either QPS or LPS. The p-value for the LPS-t test is 0.060, but we have already established that it is likely to be over-sized in small samples. Consequently, we arrive at the conclusion that encompassing holds in one direction only, with the spread-model forecasts encompassing the oil-price-model forecasts.

5.2 Evaluation of survey-based expectations

The tests of forecast encompassing are also illustated using the probabilities of output decline recorded by the respondents to the Survey of Professional Forecasters (SPF).[15] Amongst other things, respondents report the probabilities they attach to declines in real output in the current quarter and in each of the following four quarters. We focus on the next quarter and the four-quarter ahead forecasts, which correspond to $h = 2$ and $h = 5$ step-ahead forecasts. The individual forecasts will be compared to an average or 'consensus' forecast of the individual assessments. For $h = 2$, this average forecast corresponds to the 'anxious index' reported by the SPF. The question we address is whether the average indices contain all the useful information in the individuals' probability assessments: do the average indices encompass the individual forecasts? We calculate tests in a pairwise fashion for the anxiety indices against each of the individuals' forecasts in turn. Comparing each individual to the average of those who respond has the advantage of allowing the tests to employ all the occasions on which a particular individual responded.

The binary series of whether output declines in quarter t, $\{y_t\}$, is calculated from the series of quarterly real-time data sets for real output maintained by the Federal Reserve Bank of Philadelphia (see Croushore and Stark (2001)). Consider a respondent to the 1995 first-quarter SPF survey who reports forecast probabilities of a decline in 95:Q2 relative to 95:Q1 (our $h = 2$ forecast), and of a decline in 96:Q1 relative to 95:Q4 (a $h = 5$ forecast). We determine whether there was a decline

[15] The SPF is a quarterly survey of macroeconomic forecasters of the US economy that began in 1968 as the ASA-NBER survey, administered by the American Statistical Association and the National Bureau of Economic Research, and since June 1990 has been run by the Philadelphia Fed, under its current name. Detailed information on the survey is available at the URL `http://www.phil.frb.org/econ/spf`.

in 95:Q2 relative to 95:Q1 based on the values of real output in these two quarters in the August 1995 real-time data set (which contains the advance release figures for 95:Q2), and we use the May 1996 real-time data set to evaluate whether there was a decline in 96:Q1. In this way, declines are calculated based on the first-available figures. This seems preferable to using later vintage series which will typically contain revisions and definitional changes that were largely unpredictable at the time (see, e.g., Koenig, Dolmas and Piger (2003), and Faust, Rogers and Wright (2005, 2003)).

Table 7 records p-values of the FE(1) forecast encompassing tests of the null that the individual forecast encompasses the average, and of encompassing in the reverse direction. As we have no knowledge of how the forecasts are formed, the test statistics for corrected for general autocorrelation, as described earlier. Results are presented for some individuals who made at least sixty 2- and 5-step forecasts. Note that the historical time periods for which we have observations will vary from individual to individual, and that as the averages were constructed from all available forecasts at each time point and horizon, the number of constituent forecasts differ over time period, horizon and individual.

For $h = 2$, for all but the first individual, we reject the null that the individuals' forecasts encompass the consensus for both QPS and LPS (at a 10% significance level). Tests that the average encompasses the individual forecasts reject in only one instance for both QPS and LPS (the same individual). We conclude that for $h = 2$ the equal-weighted linear combinations of the individuals' forecasts (such as the 'anxiety index') convey all the useful information in the individual forecasts with one exception. For $h = 5$ the QPS and LPS tests indicate that encompassing holds in both directions: the respondents encompass the average and are encompassed by the average. This is consistent with the individuals' forecasts and the average forecast containing similar information.

The last rows of each panel of table 7 report the results of encompassing tests of the average SPF forecasts against the logit-model forecasts using the spread as a leading indicator, as in the previous section. To ensure a fair comparison with the survey predictions, the logit-model forecasts are generated recursively using at each point in time the vintage of data that was available to the respondents to the corresponding survey. For $h = 2$, both the QPS and LPS tests indicate that encompassing is rejected in both directions: both the survey and model-based forecasts contain useful additional information not contained in the other. For $h = 5$ there is clear evidence that the model-based forecasts dominate: they encompass, but are not encompassed by, the survey-based forecasts. This is consistent with Estrella and Mishkin (1998) who find that the yield curve has good predictive power at more than two quarters ahead.

6 Conclusions

We have investigated the use of tests of forecast encompassing for evaluating probability forecasts. Probability forecasts are commonly evaluated using both quadratic and logarithmic scoring rules (QPS and LPS), and we have considered tests of forecast encompassing based on both. The use of QPS leads to encompassing tests based on OLS estimation of the encompassing regression, as in the standard point forecast evaluation literature. For LPS we use maximum likelihood. We derive tests which have limiting standard normal distributions under the null of encompassing for both loss functions.

In the literature on forecast encompassing for standard (i.e., not probability) point forecasts, tests are often based on one of a number of regression approaches which differ in terms of the restrictions imposed in the maintained model. We show that the choice of which of these approaches to use may matter for at least two reasons. Firstly, we are able to establish analytically that, in the standard setup, the limiting distribution of the test based on the more general regression (that advocated by Fair and Shiller (1990)) is unaffected by parameter estimation uncertainty, although this is not the case for tests based on the two more restricted regressions which have been the focus of the recent literature (see, e.g., West and McCracken (1998) and West (2001) on the more restricted regressions FE(3) and FE(2)). Analytical results for the (non-linear) probability forecasting case are difficult to obtain, and instead we rely on a Monte Carlo study to show that the same is true for probability forecasts for tests based on both QPS and LPS.

Secondly, tests based on the more restricted forms of the encompassing regression (as frequently used in the point forecast context) may reject encompassing when encompassing holds. This will occur when the implicit assumptions underpinning the more restricted regressions do not hold. This may occur for forecasts which do not make an efficient use of the information on which they are based, as may be the case for some non-model-based forecasts.

The tests are illustrated with two empirical applications. The first is typical of the literature on assessing the usefulness of various leading indicators for predicting the probabilities of post War US recessions. Logit models are used to derive predicted recession probabilities using two leading indicators of recession—the yield curve and the oil price. We conclude that encompassing holds in one direction only, with forecasts from the model with the yield curve encompassing predictions from the oil price model. The second empirical illustration compared individual SPF respondents' probability assessments of declines in output against the 'consensus', to see whether the implicit equal-weighting of the consensus could be improved upon. Comparing the consensus and yield-curve model forecasts, we also found that there was useful incremental information in the latter at longer horizons.

Appendix 1. Moments of forecasts for probability forecast DGP

In this Appendix we derive the moments of $\{y_t, f_{1t}, f_{2t}\}$ for the data generating process given by equation (4) in section 2.1. The expressions for the moments can then be evaluated using numerical methods to derive population values of the parameters in the QPS encompassing regressions.

Firstly, consider the unbiased forecast case where $\mu_Z = 0$. Recall that the forecasts are given by:

$$f_{1t} = \frac{\exp(\theta_{01} + \theta_{11}X_{1t})}{1+\exp(\theta_{01} + \theta_{11}X_{1t})}, \quad f_{2t} = \frac{\exp(\theta_{02} + \theta_{12}X_{2t})}{1+\exp(\theta_{02} + \theta_{12}X_{2t})}$$

where θ_{ij} can be interpreted as the population parameters associated with logit regressions of y_t on a constant and X_{1t}, and a constant and X_{2t}, respectively. Then, for $i = 1, 2$:

$$\begin{aligned}
E(f_i) &= \int_{-\infty}^{\infty} \frac{\exp(\theta_{0i} + \theta_{1i}X_i)}{1+\exp(\theta_{0i} + \theta_{1i}X_i)} f(X_i) dX_i \\
&= \int_{-\infty}^{\infty} \frac{\exp(\theta_{0i} + \theta_{1i}X_i)}{1+\exp(\theta_{0i} + \theta_{1i}X_i)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}X_i^2\right) dX_i \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\exp(\theta_{0i} + \theta_{1i}X_i - \frac{1}{2}X_i^2)}{1+\exp(\theta_{0i} + \theta_{1i}X_i)} dX_i
\end{aligned}$$

$$\begin{aligned}
E(f_i^2) &= \int_{-\infty}^{\infty} \left[\frac{\exp(\theta_{0i} + \theta_{1i}X_i)}{1+\exp(\theta_{0i} + \theta_{1i}X_i)}\right]^2 f(X_i) dX_i \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\exp(2\theta_{0i} + 2\theta_{1i}X_i - \frac{1}{2}X_i^2)}{[1+\exp(\theta_{0i} + \theta_{1i}X_i)]^2} dX_i
\end{aligned}$$

$$\begin{aligned}
E(f_1 f_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(\theta_{01} + \theta_{11}X_1)}{1+\exp(\theta_{01} + \theta_{11}X_1)} \frac{\exp(\theta_{02} + \theta_{12}X_2)}{1+\exp(\theta_{02} + \theta_{12}X_2)} f(X_1, X_2) \partial X_1 \partial X_2 \\
&= \frac{1}{2\pi\sqrt{|\Omega_{2\times 2}|}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(\theta_{01} + \theta_{02} + \theta_{11}X_1 + \theta_{12}X_2)}{[1+\exp(\theta_{01} + \theta_{11}X_1)][1+\exp(\theta_{02} + \theta_{12}X_2)]} \\
&\quad \times \exp\left(-\frac{1}{2}x_{2\times 1}'\Omega_{2\times 2}^{-1}x_{2\times 1}\right) \partial X_1 \partial X_2
\end{aligned}$$

where:

$$x_{2\times 1} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \qquad \Omega_{2\times 2} = \begin{bmatrix} 1 & \rho_{X_1,X_2} \\ \rho_{X_1,X_2} & 1 \end{bmatrix}$$

so that, for example, when $\rho_{X_1,X_2} = 0.5$:

$$E(f_1 f_2) = \frac{1}{2\pi\sqrt{0.75}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(\theta_{01} + \theta_{02} + \theta_{11}X_1 + \theta_{12}X_2 - \frac{2}{3}X_1^2 - \frac{2}{3}X_2^2 + \frac{2}{3}X_1X_2)}{[1+\exp(\theta_{01}+\theta_{11}X_1)]\,[1+\exp(\theta_{02}+\theta_{12}X_2)]} \partial X_1 \partial X_2$$

In the specification for y_t, consider the potentially encompassing situation where $\delta_2 = 0$. Also, let $\delta_0 = 0$ and $\delta_1 = 1$. Then:

$$\begin{aligned} E(y_t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{1} 1\left(\frac{\exp(X_1+\delta_3 Z)}{1+\exp(X_1+\delta_3 Z)} > v\right) f(X_1,X_2,Z)\partial v \partial X_1 \partial X_2 \partial Z \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\frac{\exp(X_1+\delta_3 Z)}{1+\exp(X_1+\delta_3 Z)}} 1.\partial v f(X_1,X_2,Z)\partial X_1 \partial X_2 \partial Z \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(X_1+\delta_3 Z)}{1+\exp(X_1+\delta_3 Z)} f(X_1,X_2,Z)\partial X_1 \partial X_2 \partial Z \\ &= \frac{1}{(2\pi)^{3/2}\sqrt{|\Omega|}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(X_1+\delta_3 Z)}{1+\exp(X_1+\delta_3 Z)} \exp\left(-\frac{1}{2}x'\Omega^{-1}x\right) \partial X_1 \partial X_2 \partial Z \end{aligned}$$

where:

$$x = \begin{bmatrix} X_1 \\ X_2 \\ Z \end{bmatrix}, \qquad \Omega = \begin{bmatrix} 1 & \rho_{X_1,X_2} & \rho_{X_1,Z} \\ \rho_{X_1,X_2} & 1 & \rho_{X_2,Z} \\ \rho_{X_1,Z} & \rho_{X_2,Z} & 1 \end{bmatrix}$$

In addition, we obtain for $E(y_t f_{1t})$:

$$\begin{aligned} & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{1} 1\left(\frac{\exp(X_1+\delta_3 Z)}{1+\exp(X_1+\delta_3 Z)} > v\right) \frac{\exp(\theta_{01}+\theta_{11}X_1)}{1+\exp(\theta_{01}+\theta_{11}X_1)} f(X_1,X_2,Z)\partial v \partial X_1 \partial X_2 \partial Z \\ = & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(X_1+\delta_3 Z)}{1+\exp(X_1+\delta_3 Z)} \frac{\exp(\theta_{01}+\theta_{11}X_1)}{1+\exp(\theta_{01}+\theta_{11}X_1)} f(X_1,X_2,Z)\partial X_1 \partial X_2 \partial Z \\ = & \frac{1}{(2\pi)^{3/2}\sqrt{|\Omega|}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(\theta_{01}+(1+\theta_{11})X_1+\delta_3 Z)}{[1+\exp(X_1+\delta_3 Z)][1+\exp(\theta_{01}+\theta_{11}X_1)]} \exp\left(-\frac{1}{2}x'\Omega^{-1}x\right) \partial X_1 \partial X_2 \partial Z \end{aligned}$$

and for $E(y_t f_{2t})$:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{1} 1\left(\frac{\exp(X_1+\delta_3 Z)}{1+\exp(X_1+\delta_3 Z)}>v\right)\frac{\exp(\theta_{02}+\theta_{12}X_2)}{1+\exp(\theta_{02}+\theta_{12}X_2)}f(X_1,X_2,Z)\partial v\partial X_1\partial X_2\partial Z$$

$$= \frac{1}{(2\pi)^{3/2}\sqrt{|\Omega|}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp(\theta_{02}+X_1+\theta_{12}X_2+\delta_3 Z)}{[1+\exp(X_1+\delta_3 Z)][1+\exp(\theta_{02}+\theta_{12}X_2)]}\exp\left(-\frac{1}{2}x'\Omega^{-1}x\right)\partial X_1\partial X_2\partial Z$$

A special case exists where $\delta_3 = 0$ so that Z_t does enter the DGP. In this case we have:

$$\begin{aligned} E(y_t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp(X_1)}{1+\exp(X_1)}f(X_1,X_2)\partial X_1\partial X_2 \\ &= \frac{1}{2\pi\sqrt{|\Omega_{2\times 2}|}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp(X_1)}{1+\exp(X_1)}\exp\left(-\frac{1}{2}x'_{2\times 1}\Omega^{-1}_{2\times 2}x_{2\times 1}\right)\partial X_1\partial X_2 \end{aligned}$$

$$E(y_t f_{1t}) = \frac{1}{2\pi\sqrt{|\Omega_{2\times 2}|}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp(\theta_{01}+(1+\theta_{11})X_1)}{[1+\exp(X_1)][1+\exp(\theta_{01}+\theta_{11}X_1)]}\exp\left(-\frac{1}{2}x'_{2\times 1}\Omega^{-1}_{2\times 2}x_{2\times 1}\right)\partial X_1\partial X_2$$

$$E(y_t f_{2t}) = \frac{1}{2\pi\sqrt{|\Omega_{2\times 2}|}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp(\theta_{02}+X_1+\theta_{12}X_2)}{[1+\exp(X_1)][1+\exp(\theta_{02}+\theta_{12}X_2)]}\exp\left(-\frac{1}{2}x'_{2\times 1}\Omega^{-1}_{2\times 2}x_{2\times 1}\right)\partial X_1\partial X_2$$

Consider now the biased forecast case where $\mu_Z = 1$. The forecasts are given by:

$$f_{1t} = \frac{\exp(\theta_{11}X_{1t})}{1+\exp(\theta_{11}X_{1t})}, \quad f_{2t} = \frac{\exp(\theta_{12}X_{2t})}{1+\exp(\theta_{12}X_{2t})}$$

with the θ_{ij} being the population parameters associated with logit regressions of y_t on X_{1t} and X_{2t} respectively, without a constant term. We have, for $i = 1, 2$:

$$\begin{aligned} E(f_i) &= \int_{-\infty}^{\infty}\frac{\exp(\theta_{1i}X_i)}{1+\exp(\theta_{1i}X_i)}f(X_i)dX_i \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\exp(\theta_{1i}X_1-\frac{1}{2}X_i^2)}{1+\exp(\theta_{1i}X_i)}dX_i \end{aligned}$$

$$\begin{aligned} E(f_i^2) &= \int_{-\infty}^{\infty}\left[\frac{\exp(\theta_{1i}X_i)}{1+\exp(\theta_{1i}X_i)}\right]^2 f(X_i)dX_i \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\exp(2\theta_{1i}X_i-\frac{1}{2}X_i^2)}{[1+\exp(\theta_{1i}X_i)]^2}dX_i \end{aligned}$$

$$
\begin{aligned}
E(f_1 f_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(\theta_{11}X_1)}{1+\exp(\theta_{11}X_1)}\frac{\exp(\theta_{12}X_2)}{1+\exp(\theta_{12}X_2)} f(X_1,X_2)\partial X_1 \partial X_2 \\
&= \frac{1}{2\pi\sqrt{|\Omega_{2\times 2}|}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(\theta_{11}X_1+\theta_{12}X_2)}{[1+\exp(\theta_{11}X_1)][1+\exp(\theta_{12}X_2)]}\exp\left(-\frac{1}{2}x'_{2\times 1}\Omega^{-1}_{2\times 2}x_{2\times 1}\right)\partial X_1\partial X_2
\end{aligned}
$$

$$
\begin{aligned}
E(y_t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{1} 1\left(\frac{\exp(X_1+\delta_3 Z)}{1+\exp(X_1+\delta_3 Z)} > v\right) f(X_1,X_2,Z)\partial v\partial X_1\partial X_2\partial Z \\
&= \frac{1}{(2\pi)^{3/2}\sqrt{|\Omega|}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(X_1+\delta_3 Z)}{1+\exp(X_1+\delta_3 Z)}\exp\left(-\frac{1}{2}(x-\mu)'\Omega^{-1}(x-\mu)\right)\partial X_1\partial X_2\partial Z.
\end{aligned}
$$

For $E(y_t f_{1t})$ and $E(y_t f_{2t})$, respectively:

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{1} 1\left(\frac{\exp(X_1+\delta_3 Z)}{1+\exp(X_1+\delta_3 Z)} > v\right)\frac{\exp(\theta_{11}X_1)}{1+\exp(\theta_{11}X_1)} f(X_1,X_2,Z)\partial v\partial X_1\partial X_2\partial Z \\
= &\frac{1}{(2\pi)^{3/2}\sqrt{|\Omega|}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp((1+\theta_{11})X_1+\delta_3 Z)}{[1+\exp(X_1+\delta_3 Z)][1+\exp(\theta_{11}X_1)]}\exp\left(-\frac{1}{2}(x-\mu)'\Omega^{-1}(x-\mu)\right)\partial X_1\partial X_2\partial Z
\end{aligned}
$$

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{1} 1\left(\frac{\exp(X_1+\delta_3 Z)}{1+\exp(X_1+\delta_3 Z)} > v\right)\frac{\exp(\theta_{12}X_2)}{1+\exp(\theta_{12}X_2)} f(X_1,X_2,Z)\partial v\partial X_1\partial X_2\partial Z \\
= &\frac{1}{(2\pi)^{3/2}\sqrt{|\Omega|}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(X_1+\theta_{12}X_2+\delta_3 Z)}{[1+\exp(X_1+\delta_3 Z)][1+\exp(\theta_{12}X_2)]}\exp\left(-\frac{1}{2}(x-\mu)'\Omega^{-1}(x-\mu)\right)\partial X_1\partial X_2\partial Z
\end{aligned}
$$

where $\mu = [0\ 0\ 1]'$.

Appendix 2. Lack of encompassing when X_1 and Z correlated

We consider a linear version of our setup, whereby the DGP is given by:

$$
y_t = \delta_0 + \delta_1 X_{1t} + \delta_2 X_{2t} + \delta_3 Z_t + \varepsilon_t,
$$

with $\delta_2 = 0$, to allow the possibility that f_1 encompasses f_2. X_{1t}, X_{2t} and Z_t are generated as in equation (4), but with $\rho_{X_2,Z} = 0$, so X_2 is uncorrelated with the omitted variable, Z, and

$\rho_{X_1,X_2} \neq 0$. Consider forecasts

$$f_{1t} = \hat{\theta}_{01} + \hat{\theta}_1 X_{1t}$$
$$f_{2t} = \hat{\theta}_{02} + \hat{\theta}_2 X_{2t},$$

obtained from the linear regressions $y_t = \theta_{01} + \theta_1 X_{1t} + u_{1t}$ and $y_t = \theta_{02} + \theta_2 X_{2t} + u_{2t}$, respectively. In the limit the forecasts will be given by:

$$\begin{aligned} f_{1t} \xrightarrow{p} \frac{C(y_t, X_{1t})}{V(X_{1t})} X_{1t} &= \left[\delta_1 + \delta_3 \frac{C(X_{1t}, Z_t)}{V(X_{1t})}\right] X_{1t} \\ f_{2t} \xrightarrow{p} \frac{C(y_t, X_{2t})}{V(X_{2t})} X_{2t} &= \left[\delta_1 \frac{C(X_{1t}, X_{2t})}{V(X_{2t})}\right] X_{2t}. \end{aligned}$$

After some algebra, we find that the population value of β_2 in the encompassing regression is:

$$\beta_2 = \frac{-\delta_3 C(X_{1t}, Z_t) V(X_{2t})}{\delta_1 \left[V(X_{1t}) V(X_{2t}) - C(X_{1t}, X_{2t})^2\right]} = \frac{-\delta_3 \lambda_1}{\delta_1}$$

where λ_1 is a population coefficient from the regression:

$$Z_t = \lambda_0 + \lambda_1 X_{1t} + \lambda_2 X_{2t} + \eta_t.$$

For $\beta_2 = 0$ we require that $\lambda_1 = 0$, so that X_1 and Z are uncorrelated, or that the model with X_1 is correctly specified, i.e., $\delta_3 = 0$.

References

Anderson, H. M., and Vahid, F. (2001). Predicting the probability of a recession with nonlinear autoregressive leading indicator models. *Macroeconomic Dynamics*, **5**, 482–505.

Birchenhall, C. R., Jessen, H., Osborn, D., and Simpson, P. (1999). Predicting U.S. business cycle regimes. *Journal of Business & Economic statistics*, **17, No. 3, (July)**, 313–323.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **75**, 1–3.

Carruth, A. A., Hooker, M. A., and Oswald, A. J. (1998). Unemployment equilibria and input prices: Theory and evidence from the United States. *Review of Economics and Statistics*, **80**, 621–628.

Chauvet, C., and Potter, S. (2002). Predicting a recession: evidence from the yield curve in the presence of structural breaks. *Economic Letters*, **77**, 245–253.

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.

Clark, T. E., and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, **105**, 85–110.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, **5**, 559–583. Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar.

Clemen, R. T., and Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, **19**, 187–203.

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

Dawid, A. P. (1986). Probability forecasting. In Kotz, S., Johnson, N. L., and Read, C. B. (eds.), *Encyclopedia of Statistical Sciences, vol. 7*, pp. 210–218: John Wiley and Sons.

Diebold, F. X., and Lopez, J. A. (1996). Forecast evaluation and combination. In Maddala, G. S., and Rao, C. R. (eds.), *Handbook of Statistics*, Vol. 14, pp. 241–268: Amsterdam: North–Holland.

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 253–263. Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar.

Diebold, F. X., and Rudebusch, G. D. (1989). Scoring the leading indicators. *Journal of Business*, **62**, 369–391.

Ericsson, N. R. (1993). Comment on 'On the limitations of comparing mean squared forecast errors', by M.P. Clements and D.F. Hendry. *Journal of Forecasting*, **12**, 644–651.

Estrella, A., and Mishkin, F. S. (1998). Predicting US recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, **80**, 45–61.

Fair, R. C., and Shiller, R. J. (1990). Comparing information in forecasts from econometric models. *American Economic Review*, **80**, 39–50.

Faust, J., Rogers, J. H., and Wright, J. H. (2003). Exchange rate forecasting: The errors we've really made. *Journal of International Economic Review*, **60**, 35–39.

Faust, J., Rogers, J. H., and Wright, J. H. (2005). News and noise in G-7 GDP announcements. *Journal of Money, Credit and Banking*, **37 (3)**, 403–417.

Genest, C., and Zidek, J. V. (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, **1**, 114–148.

Good, I. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B*, **14 (No. 1)**, 107–114.

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 Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, **19**, 537–560.

Hamilton, J. D. (1983). Oil and the Macroeconomy since World War II. *Journal of Political Economy*, **91**, 228–248.

Hamilton, J. D. (1996). This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics*, **38**, 215–220.

Hamilton, J. D. (2000). What is an Oil Shock?. *Journal of Econometrics*, **113**, 363–398.

Hamilton, J. D., and Kim, D. H. (2000). A re-examination of the predictability of economic activity using the yield spread. *NBER Working Papers*, **7954**.

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

Hooker, M. A. (1996). Whatever happened to the oil price-macroeconomy relationship?. *Journal of Monetary Economics*, **38**, 195–213.

Kamstra, M., and Kennedy, P. (1998). Combining qualitative forecasts using logit. *International Journal of Forecasting*, **14**, 83–93.

Koenig, E. F., Dolmas, S., and Piger, J. (2003). The use and abuse of real-time data in economic forecasting. *The Review of Economics and Statistics*, **85(3)**, 618–628.

Lee, K., Ni, S., and Ratti, A. (1995). Oil shocks and the macroeconomy: The role of price variability. *Energy Journal*, **16**, 39–56.

Mork, K. A. (1989). Oil and the Macroeconomy when prices go up and down: An extension of Hamilton's results. *Journal of Political Economy*, **97**, 740–744.

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.

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.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, **64**, 1067–1084.

West, K. D. (2001). Tests for forecast encompassing when forecasts depend on estimated regression parameters. *Journal of Business and Economic Statistics*, **19**, 29–33.

West, K. D. (2006). Forecasting evaluation. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 99–134: Elsevier, Horth-Holland.

West, K. D., and McCracken, M. W. (1998). Regression-based tests of predictive ability. *International Economic Review*, **39**, 817–840.

Winkler, R. L. (1996). Scoring rules and the evaluation of probabilities (with discussion). *Test*, **5 (No. 1)**, 1–60.

Table 1. Population values of FE(1), FE(2) and FE(3) optimal forecast combination parameters

δ_3	μ_Z	ρ_{X_1,X_2}	$\rho_{X_1,Z}$	$\rho_{X_2,Z}$	α	β_1	β_2	α'	β'_2	α''	β''_2
Panel A. QPS											
Unbiased model-based forecasts											
0	-	0.2	-	-	0.00	1.00	0.00	0.00	0.00	0.00	0.00
		0.5	-	-	0.00	1.00	0.00	0.00	0.00	0.00	0.00
		0.8	-	-	0.00	1.00	0.00	0.00	0.00	0.00	0.00
1	0	0.5	0.0	0.0	0.00	1.00	0.00	0.00	0.00	0.00	0.00
			0.0	0.5	−0.17	0.67	0.67	0.00	0.50	−0.25	0.50
			−0.5	0.0	−0.17	0.67	0.67	0.00	0.50	−0.25	0.50
			0.5	0.0	0.28	1.11	−0.66	0.00	−0.07	0.25	−0.50
			0.5	0.5	−0.11	0.89	0.33	0.00	0.14	−0.13	0.25
Biased model-based forecasts											
1	1	0.5	0.0	0.0	0.18	1.00	0.00	0.18	0.00	0.18	0.00
			0.0	0.5	0.01	0.67	0.67	0.18	0.50	−0.08	0.50
			−0.5	0.0	0.03	0.67	0.67	0.20	0.50	−0.05	0.50
			0.5	0.0	0.44	1.10	−0.66	0.16	−0.07	0.41	−0.50
			0.5	0.5	0.05	0.89	0.33	0.16	0.14	0.03	0.25
Non-model-based forecasts											
0	-	0.5	-	-	0.00	1.00	0.00	0.00	0.00	0.00	0.00
1	0	0.5	0.0	0.0	0.06	0.87	0.00	0.00	0.06	0.03	−0.06
1	1	0.5	0.0	0.0	0.28	0.78	0.00	0.18	0.11	0.23	−0.11

δ_3	μ_Z	ρ_{X_1,X_2}	$\rho_{X_1,Z}$	$\rho_{X_2,Z}$	α	β_1	β_2	α'	β'_2	α''	β''_2
Panel B. LPS											
Unbiased model-based forecasts											
0	-	0.2	-	-	0.00	1.00	0.00	0.00	0.00	0.00	0.00
		0.5	-	-	0.00	1.00	0.00	0.00	0.00	0.00	0.00
		0.8	-	-	0.00	1.00	0.00	0.00	0.00	0.00	0.00
1	0	0.5	0.0	0.0	0.00	1.00	0.00	0.00	0.00	0.00	0.00
			0.0	0.5	−0.17	0.67	0.67	0.00	0.50	−0.25	0.50
			−0.5	0.0	−0.17	0.67	0.67	0.00	0.50	−0.25	0.50
			0.5	0.0	0.27	1.11	−0.65	0.00	−0.05	0.25	−0.50
			0.5	0.5	−0.11	0.89	0.33	0.00	0.10	−0.13	0.25
Biased model-based forecasts											
1	1	0.5	0.0	0.0	0.17	1.00	0.00	0.17	0.01	0.17	0.00
			0.0	0.5	0.01	0.67	0.67	0.18	0.50	−0.08	0.51
			−0.5	0.0	0.03	0.67	0.67	0.20	0.50	−0.05	0.50
			0.5	0.0	0.43	1.11	−0.66	0.16	−0.20	0.41	−0.50
			0.5	0.5	0.04	0.89	0.34	0.17	0.13	0.03	0.26
Non-model-based forecasts											
0	-	0.5	-	-	0.00	1.00	0.00	0.00	0.00	0.00	0.00
1	0	0.5	0.0	0.0	0.06	0.87	0.00	0.00	0.07	0.03	−0.06
1	1	0.5	0.0	0.0	0.28	0.78	0.00	0.17	0.11	0.23	−0.10

Table 2. Empirical sizes of 1-sided nominal 0.05-level forecast encompassing tests with parameter estimation uncertainty: unbiased model-based forecasts, $\delta_3 = 0$, $\rho_{X_1,X_2} = 0.5$

Panel A. QPS-MDM

		n/R						
	n	2	1	0.5	0.25	0.125	0.0625	0
FE(1)	25	0.059	0.054	0.057	0.055	0.057	0.055	0.053
	50	0.057	0.053	0.049	0.053	0.054	0.053	0.054
	100	0.053	0.051	0.052	0.050	0.052	0.050	0.050
	200	0.054	0.051	0.050	0.052	0.047	0.048	0.050
	500	0.050	0.051	0.048	0.050	0.049	0.049	0.051
FE(2)	25	0.165	0.087	0.051	0.039	0.034	0.031	0.027
	50	0.145	0.082	0.056	0.045	0.037	0.034	0.030
	100	0.136	0.088	0.064	0.052	0.042	0.041	0.035
	200	0.138	0.087	0.071	0.057	0.044	0.043	0.039
	500	0.135	0.103	0.071	0.060	0.047	0.045	0.040
FE(3)	25	0.075	0.071	0.076	0.075	0.075	0.076	0.076
	50	0.075	0.075	0.069	0.070	0.072	0.070	0.073
	100	0.078	0.074	0.069	0.067	0.065	0.061	0.062
	200	0.087	0.077	0.068	0.060	0.059	0.059	0.059
	500	0.083	0.071	0.065	0.058	0.057	0.055	0.054

Panel B. LPS-t

		n/R						
	n	2	1	0.5	0.25	0.125	0.0625	0
FE(1)	25	0.107	0.101	0.111	0.110	0.112	0.118	0.114
	50	0.082	0.086	0.084	0.091	0.088	0.090	0.095
	100	0.072	0.071	0.074	0.066	0.074	0.073	0.075
	200	0.066	0.059	0.063	0.068	0.060	0.060	0.064
	500	0.055	0.058	0.056	0.052	0.056	0.053	0.055
FE(2)	25	0.078	0.060	0.049	0.046	0.040	0.040	0.036
	50	0.082	0.061	0.051	0.046	0.038	0.040	0.035
	100	0.098	0.077	0.063	0.053	0.045	0.040	0.038
	200	0.128	0.084	0.070	0.053	0.047	0.043	0.038
	500	0.139	0.106	0.072	0.056	0.047	0.041	0.039
FE(3)	25	0.128	0.124	0.136	0.135	0.139	0.135	0.137
	50	0.121	0.118	0.115	0.119	0.117	0.120	0.117
	100	0.117	0.114	0.111	0.097	0.103	0.097	0.104
	200	0.118	0.104	0.091	0.090	0.089	0.083	0.089
	500	0.105	0.087	0.084	0.070	0.075	0.073	0.070

Table 3. Empirical sizes of 1-sided nominal 0.05-level forecast encompassing tests: non-model-based forecasts, $\rho_{X_1,X_2} = 0.5$, $\rho_{X_1,Z} = \rho_{X_2,Z} = 0$

		$\delta_3 = 0$		$\delta_3 = 1$, $\mu_Z = 0$		$\delta_3 = 1$, $\mu_Z = 1$	
	n	QPS-MDM	LPS-t	QPS-MDM	LPS-t	QPS-MDM	LPS-t
FE(1)	25	0.077	0.105	0.075	0.095	0.075	0.095
	50	0.073	0.084	0.076	0.077	0.080	0.069
	100	0.064	0.067	0.062	0.059	0.069	0.065
	200	0.055	0.058	0.054	0.052	0.056	0.067
	500	0.053	0.051	0.055	0.053	0.054	0.056
FE(2)	25	0.059	0.043	0.072	0.059	0.095	0.059
	50	0.061	0.042	0.085	0.070	0.120	0.063
	100	0.057	0.039	0.094	0.084	0.143	0.077
	200	0.047	0.039	0.104	0.107	0.179	0.114
	500	0.047	0.040	0.156	0.176	0.329	0.253
FE(3)	25	0.095	0.127	0.074	0.103	0.057	0.099
	50	0.089	0.107	0.064	0.075	0.044	0.060
	100	0.077	0.097	0.042	0.052	0.028	0.055
	200	0.066	0.083	0.029	0.032	0.013	0.046
	500	0.057	0.066	0.015	0.013	0.004	0.019

Table 4. Estimated powers of 1-sided nominal 0.05-level forecast encompassing tests: $\delta_3 = 0$, $\rho_{X_1,X_2} = 0.5$

		Model-based forecasts, $n/R = 0$				Non-model-based forecasts			
		$\delta_1 = 0.5$, $\delta_2 = 0.5$		$\delta_1 = 0$, $\delta_2 = 1$		$\delta_1 = 0.5$, $\delta_2 = 0.5$		$\delta_1 = 0$, $\delta_2 = 1$	
	n	QPS-MDM	LPS-t	QPS-MDM	LPS-t	QPS-MDM	LPS-t	QPS-MDM	LPS-t
FE(1)	25	0.249	0.323	0.578	0.618	0.248	0.293	0.576	0.569
	50	0.404	0.429	0.852	0.847	0.401	0.386	0.850	0.777
	100	0.625	0.619	0.985	0.984	0.622	0.562	0.985	0.964
	200	0.875	0.855	1.000	1.000	0.874	0.820	1.000	1.000
	500	0.997	0.995	1.000	1.000	0.997	0.994	1.000	1.000
FE(2)	25	0.176	0.196	0.586	0.613	0.233	0.235	0.620	0.581
	50	0.299	0.293	0.855	0.848	0.420	0.378	0.903	0.851
	100	0.497	0.445	0.987	0.984	0.684	0.618	0.996	0.988
	200	0.770	0.701	1.000	1.000	0.926	0.888	1.000	1.000
	500	0.986	0.968	1.000	1.000	0.999	0.998	1.000	1.000
FE(3)	25	0.262	0.328	0.536	0.559	0.195	0.269	0.342	0.412
	50	0.385	0.407	0.789	0.767	0.268	0.286	0.518	0.453
	100	0.576	0.558	0.964	0.950	0.382	0.339	0.751	0.549
	200	0.823	0.770	0.999	0.999	0.584	0.417	0.941	0.666
	500	0.991	0.978	1.000	1.000	0.903	0.642	1.000	0.888

Table 5. In-sample fits of recession logit models

	QPS	LPS	R^2	Slope p-value
Spread	0.154	0.258	0.316	0.000
Oil price	0.233	0.394	0.070	0.019

Table 6. FE(1) forecast encompassing tests of model-based probability forecasts of recession

	H_0: Spread $\mathcal{E}$ Oil price	H_0: Oil price $\mathcal{E}$ Spread
QPS-MDM	0.111	0.026
LPS-t	0.060	0.004

Notes: The entries are p-values of the forecast encompassing null against a 1-sided alternative. $\mathcal{E}$ denotes "forecast encompasses".

Table 7. FE(1) forecast encompassing tests of SPF respondents' probability forecasts of recession

Panel A. 1-quarter forecasts ($h = 2$)					
		QPS-MDM	LPS-t	QPS-MDM	LPS-t
		H_0: Individual $\mathcal{E}$ Average		H_0: Average $\mathcal{E}$ Individual	
I.D.	n				
84	104	0.29	0.56	0.02	0.05
65	95	0.03	0.00	0.87	1.00
70	75	0.00	0.00	0.76	0.41
20	67	0.05	0.01	0.48	0.66
82	64	0.08	0.00	0.21	0.10
		H_0: Average $\mathcal{E}$ Spread		H_0: Spread $\mathcal{E}$ Average	
		0.04	0.00	0.08	0.00
Panel B. 4-quarter forecasts ($h = 5$)					
		QPS-MDM	LPS-t	QPS-MDM	LPS-t
		H_0: Individual $\mathcal{E}$ Average		H_0: Average $\mathcal{E}$ Individual	
I.D.	n				
84	100	0.63	0.33	0.98	0.96
65	90	0.12	0.08	0.97	0.98
70	70	1.00	0.98	0.50	0.51
20	63	0.71	0.73	0.30	0.36
82	63	0.72	0.74	0.36	0.34
		H_0: Average $\mathcal{E}$ Spread		H_0: Spread $\mathcal{E}$ Average	
		0.01	0.00	0.75	0.78

Notes: The entries are p-values of the forecast encompassing null against a 1-sided alternative. $\mathcal{E}$ denotes "forecast encompasses".