Forecasting Economic and Financial Time Series Week 6: Evaluating and Combining Forecasts

Ron Smith
MSc/PGCE Option, EMS Birkbeck

Spring 2019, Revised

Evaluating Forecasts

- Good forecasts enable good decisions, we therefore need to know how to evaluate forecasts.
- What constitutes a good forecast
  - Does the forecast add value
  - Does the evaluation suggest ways that we might improve the forecasting method.
- Hence we seek to analyse the track record of forecasts, $y^f_{t+h,t}$, their corresponding realisations, y_{t+h} and errors $e_{t+h,h} = y_{t+h} - y^f_{t+h,t}$.
- Any evaluation is conditional on loss function and information set of forecaster, so always a joint test.
- May be able to "reverse engineer" the loss function from the errors.

Some Statistical Accuracy Measures

For forecast error, $e_{t+h,h} = y_{t+h} - y^f_{t+h,t}$ or percent error $p_{t+h,h} = (y_{t+h} - y^f_{t+h,t})/y_{t+h}$

- **Mean error** $ME = \sum_{t=1}^{T} e_{t+h,t}/T$ measures bias,
- **Error Variance** $EV = \sum_{t=1}^{T}(e_{t+h,t} - ME)^2/T$
- **Mean squared error** combines them $MSE = \sum_{t=1}^{T} e^2_{t+h,t}/T = ME^2 + EV$:
- **Root mean squared forecast error** $RMSFE = \sqrt{\sum_{t=1}^{T} e^2_{t+h,t}/T}$ preserves units
- **Mean absolute error** $MAE = \sum_{t=1}^{T} |e_{t+h,t}|/T$

Questions

1. **Decision environment and loss function. Why are you doing it?**
2. **Forecast object.** What are we forecasting? Assume continuous variable for now
3. **Forecast Statement.** Will assume a point forecast for now, but might be probability, interval or density forecast.
4. **Forecast Horizon.** How far into the future do you wish to forecast? This can have implications for what is the best modelling strategy for the series of object of interest.
5. **Information set.** What information is available at the time of forecasting, is just past history or are other data available that are related to the series of interest - do they need forecasting?
6. **Method and complexity.**

Economic and Statistical Evaluation

- The forecast is going to be used to guide a decision so we need to understand the decision environment and the consequences of error to choose an appropriate measure of forecast accuracy.
- We will consider both economic and statistical (e.g. root mean square forecast error RMSFE) measures of forecast accuracy.
- The economic evaluation of our forecasts typically will be specific to the forecasting problem.
- A common economic measure is the profits earned from the actions taken as a result of the forecast.
- Often the ranking of forecasting methods by profit is very different from ranking by RMSFE.

Decision theory Involves

- A parameter space $\mathfrak{B}$ that reflects the possible states of nature relative to the unknown parameter vector $\boldsymbol{\beta}$. (Future inflation)..
- A set of possible actions $\mathfrak{A}$ and particular actions $\mathbf{a}$. (Changing interest rates)
- A loss function (negative of profit or utility) $L(\boldsymbol{\beta}, \mathbf{a})$ defined for all $(\boldsymbol{\beta}, \mathbf{a})\epsilon(\mathfrak{B}, \mathfrak{A})$. (Missing inflation target)
- A set of observations $\mathbf{y}$ which depend on the true $\boldsymbol{\beta}$**.** (Basis for the forecast, including past inflation)
- A decision rule $\mathbf{a} = \delta(\mathbf{y})$. (Raise interest rates if forecast is above target)
- with consequent loss $L(\boldsymbol{\beta}, \delta(\mathbf{y}))$ (Send a letter to the Chancellor).
- and risk or expected loss of $\rho(\boldsymbol{\beta}, \mathbf{a}) = \int_y L(\boldsymbol{\beta}, \delta(\mathbf{y}))f(\mathbf{y} \mid \boldsymbol{\beta})\mathbf{dy}$

Portfolio Choice: actions: invest in equities or bonds?

- From available data you estimate a model that forecasts return on equities $\hat{R}^E$ and return on bonds $\hat{R}^B$
- Decision rule: invest in equities if $\hat{R}^E - \hat{R}^B > K$. Choice of K?
- Let the indicator function $\mathbf{1}(\hat{R}^E - \hat{R}^B - K) = 1$ if $(\hat{R}^E - \hat{R}^B - K) > 0$, 0 otherwise
- Profit $\Pi(\boldsymbol{\beta}, \delta(\mathbf{y})) = \mathbf{1}(\hat{R}^E - \hat{R}^B - K)R^E + (1 - \mathbf{1}(\hat{R}^E - \hat{R}^B - K))R^B$
- Choose forecasting model and decision rule to maximise profit (averaged over possible states of world R^E and R^B).
- Say $K = 0$, $\hat{R}^B = 3\%$, $\hat{R}^E = 2\%$, so you invest in bonds, but $R^E = -20\%$, in RMSE terms you forecast was terrible, in profit terms it was fine: it got you out of equities.
- Diversified decision rules: proportion equities function of $\hat{R}^E - \hat{R}^B$.

Newsvendor model

- You buy q newspapers for c each and sell them for p each, demand is a random variable with cumulative distribution function $F(.)$.
- Expected profit is

$$E(\Pi) = E(p(\min[q, D]) - cq$$

- Optimal number of papers to order is

$$q = F^{-1}\left(\frac{p-c}{p}\right)$$

Wikipedia has a derivation. But intuitively the larger the relative profit lost from not being able to sell the larger the number you order.

- Suppose $p = 8$, and $c = 6$, $(p-c)/p = 2/8 = 0.25$.
- If D has a uniform distribution from $\underline{D} = 50$ to $\overline{D} = 70$, then $q = \underline{D} + (\overline{D} - \underline{D}) \times 0.25 = 55$.
- If $D\tilde{}N(60, 5)$; $q = \mu + \sigma Z^{-1}(0.25) = 60 + 5 \times (-0.33) = 58.35$

Data for evaluation

- Overfitting is a problem for models that are chosen on the basis of their performance in the sample they are estimated on.
- They tend to be over-parameterised: fitting to peculiarities of the estimation sample.
- Real ex ante forecasts can be evaluated ex post when the data are realised. They are not available when you are first constructing a forecasting model.
- Short samples of real forecasts are often a problem.
- What is actual when there are revisions? First estimate? Final estimate?
- May be difficult to distinguish action from forecast. Actions such as the choice of time allowed to go to the airport or the newsvendors choice of number of papers to buy are both forecasts and actions.

Cross-validation/back-testing/Pseudo ex ante forecasts

- Estimate the model from T_0 up to T_1, forecast $T_1 + h, .h = 1, 2, ..., H$ using information up to T_1
- Re-estimate the model up to $T_2 = T_1 + 1$, forecast $T_2 + h, .h = 1, 2, ..., H$ using information up to T_2
- Go up to $T - H$, with data up to present T.
- Choice of T_0 : expanding or rolling window of what size?
- Information on variables at T_1 now is different from it was at T_1 because of revisions to the data. Need "real time data", which is often not available,

Informal methods of evaluation

- Line graphs of forecasts against actual over time, plot $y^f_{t+h,t}$ and y_{t+h}
- Scatter diagrams of $y^f_{t+h,t}$ against y_{t+h}
- Scatter diagrams of $y^f_{t+h,t} - y_t$ against $y_{t+h} - y_t$ useful for non stationary variables, e.g. exchange rates
- Plot forecasts errors, $y^f_{t+h,t} - y_{t+h}$ against time or histograms of forecast errors
- Plot loss $L(y^f_{t+h,t}, y_{t+h})$ against time or histograms of losses

Types of Loss Function

- **Quadratic Loss**: $L(e) = e^2$ Mathematically convenient, symmetric around the origin, increases at an increasing rate on each side, so large errors are penalised more than small ones. Evaluation uses root mean squared forecast error, RMSFE. Best forecast is the conditional mean (expected value). RMSFE not invariant to transformations, e.g. best model for forecasting level may not be best model for first differences.
- **Absolute Loss:** $L(e) = |e|$ Symmetric with loss increasing linearly with the error. Forecast evaluation uses mean absolute error. Best forecast is the conditional median.
- Direction -of -change, DoC, forecast loss function, e.g. go long on the asset going up in price, short on the asset going down. So a DoC forecast takes one of two values - up or down (include no change in one). Therefore the loss function might be:

$$\begin{aligned} L(y,\widehat{y}) &= 0 \quad if \quad sign\Delta y = sign\Delta\widehat{y} \\ L(y,\widehat{y}) &= 1 \quad if \quad sign\Delta y \neq sign\Delta\widehat{y} \end{aligned}$$

Linear- Exponential (Linex), Varian (1974) and Zellner (1986)

- **Non symmetric** One parameter version of Linex loss function for forecast error $e_{T+1} = y^f_{T+1,T} - y_{T+1}$ is

$$L(e_{T+1}) = \exp\left(\alpha e_{T+1}\right) - \alpha e_{T+1} - 1$$

  where $\alpha \neq 0$ controls the degree of asymmetry.
- $L(e_{T+1}) = 0$, when $e_{T+1} = 0$
- Limit $\alpha \to 0$ is quadratic.
- Underpredicting more costly when $\alpha > 0$. E&T p22.

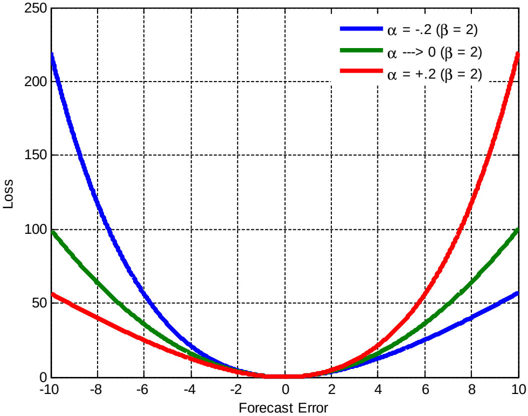
Linex

- More generally over a sequence of forecasts made at $t = 1, 2, ..., T$, where loss is scaled by $\beta > 0$:

$$L_{linex} = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\beta}{\alpha^2}\left[\exp\left(\alpha e_{t+h,t}\right) - \alpha e_{t+h,t} - 1\right]\right\}$$

  where α and β are constants with $\beta > 0$.
- The parameter α determines the degree of asymmetry.
  - For $\alpha > 0$, losses are approximately exponential for $e_{t+h,t} > 0$, and approximately linear for $e_{t+h,t} < 0$. If the error $e_{t+h,t}$ is defined as realized value of $y_{t+h} - \widehat{y}_{t+h|t}$ (as above), then $\alpha > 0$ corresponds to a case where under-predictions are most costly than over-predictions.
  - Conversely, with $\alpha < 0$ the function is exponential to the left of the origin, and linear to the right.
  - As $\alpha \to 0$ (and $\beta = 2$), the loss function simplifies to the standard quadratic case.

Some Accuracy Measures: Linex



Optimal forecasts with MSE and Linex, E&T p48

- Assume $y_t = \mu + u_t$, $u_t \sim IN(0, \sigma^2)$, with sample $t = 1, 2, ..., T$, with estimates $\overline{y}_T = \sum y_t / T$, $.s_T^2 = \sum (y_t - \overline{y})^2 / (T - 1)$
- Under MSE loss forecast

$$y^{fm}_{T+1,T} = \overline{y}_T$$

- Under Linex loss $L(e) = \exp\left(\alpha e_{T+1}\right) - \alpha e_{T+1} - 1$ forecast

$$y^{fl}_{T+1,T} = \overline{y}_T + \frac{\alpha}{2} s_T^2$$

Evaluation with Linex Loss

- Evaluating optimality of forecasts under Linex loss is more difficult since it depends on estimates of both mean and variance. The error is

$$
\begin{aligned}
e_{T+1} &= y^{fl}_{T+1,T} - y_{T+1} = \overline{y}_T - y_{T+1} + \frac{\alpha}{2} s^2_T \\
E(e) &= \frac{\alpha}{2}\sigma^2
\end{aligned}
$$

  For large samples.
- So using the average error we could infer the loss function from the errors and estimate

$$\widehat{\alpha} = 2\overline{e}/s^2$$

- Here we assume the forecasts are optimal and back out the loss function parameter from the bias.

Optimal Forecasts

- "Claims of forecast optimality are typically limited to restrictive classes of models and so can be very weak - optimality often does not extend too far in that there may be better forecast models available given the same data" E&T p39.
- Least Squares, $\hat{y} = X\widehat{\beta}$ gives the best (minimum variance), linear, unbiased, predictor, BLUP, conditional on the given X.
- But the loss function may not be quadratic (mean square error, minimum variance) and there may be better non-linear, biased predictors, or ones using a different X matrix.
- Often a trade-off between bias and variance.

Mean Square Error (MSE) Loss

- Consider an estimator $\hat{\beta}$
- MSE Loss is variance plus bias squared

$$L(\beta, \hat{\beta}) = E(\hat{\beta} - \beta)^2 = E(\hat{\beta} - E(\hat{\beta}))^2 + E(E(\hat{\beta}) - \beta)^2$$

- Biased estimator/forecast may have smaller MSE if it has reduced variance
- Similarly for a forecast where loss $L(Y_{T+h}, \hat{Y}_{T+h,T})$ is MSE

$$\begin{aligned} & E(\hat{Y}_{T+h,T} - Y_{T+h})^2 \\ = \; & E(\hat{Y}_{T+h,T} - E(\hat{Y}_{T+h,T}))^2 + E(E(\hat{Y}_{T+h,T}) - Y_{T+h})^2 \end{aligned}$$

- In many cases loss function is not symmetric so unbiased forecasts are not necessarily optimal.

Evaluating a single forecast

Four properties of forecasts of stationary variables under MSE loss that can in principle be checked are:

1. Whether forecasts are unbiased
2. Whether forecasts have one-step ahead errors that are white noise
3. Whether forecasts have $h-$step-ahead errors that are at most $MA(h-1)$
4. Whether forecasts have h-step-ahead errors with variances that are nondecreasing in h and that converge to the unconditional variance of the process

Testing if forecast is unbiased 1

- The test that the forecast error has zero mean depends on the assumptions we make.
- If $e_{t+h,t}$ is $IN(0, \sigma^2)$ (might be expected for 1-step-ahead errors), then use a t-test: regress the forecast errors on a constant and use $t-$stat to test mean is zero.
- If the error are non-Gaussian but remain (iid), then the t-test is still applicable in large samples.
- If the forecast errors are dependent, then more sophisticated procedures are required (e.g. Newey-West Heteroskedasticity and Autocorrelation, HAC, robust standard errors).
- Serial correlation in 1-step-ahead errors may indicate a sub-optimal forecast.
- Multi-step-ahead forecast errors will be serially correlated, even if the forecasts are optimal, because of forecast period overlap.
- When regressing forecast errors on an intercept make sure any serial correlation is appropriately modeled.

Testing if forecast is unbiased 2

- Suppose we have a long sequence of forecasts, $y^f_{t+h,t}$, e.g. the society of professional forecasters, or forward rates,

$$y_{t+h} = y^f_{t+h,t} + u_{t+h} \tag{1}$$

where u_{t+h} is uncorrelated with $y^f_{t+h,t}$, we can test this by a regression of actual on forecast *Mincer-Zarnowitz regression*

$$y_{t+h} = \alpha + \beta y^f_{t+h,t} + u_{t+h} \tag{2}$$

then unbiasedness implies that $\alpha = 0$, $\beta = 1$. We can do a test of the joint hypothesis using suitable standard errors.

- If y_{t+h} is $I(1)$ e.g. an exchange rate, there is a danger of a spurious regression. To avoid this we can subtract y_t from both sides of (1), making both sides $I(0)$

$$y_{t+h} - y_t = \alpha + \beta(y^f_{t+h,t} - y_t) + u_{t+h}$$

Testing if forecast is efficient

- Efficiency says that if we add any information available at time t, vector x_t

$$y_{t+h} = \alpha + \beta y^f_{t+h,t} + \gamma' x_t + u_{t+h}$$

  then $\gamma = 0$. All information should be embodied in the forecast. Candidates for x_τ include y_t, and forecast errors, known at t.
- A reasonable starting point for regressions involving h-step-ahead forecast errors is $MA(h-1)$ disturbances, which we'd expect if the forecast were MSE efficient.
- The errors should be uncorrelated with the forecast, in

$$e_{t+h,t} = \alpha_0 + \alpha_1 y_{t+h,t} + u_t \tag{3}$$

  we should have $(\alpha_0, \alpha_1) = (0, 0)$, Note that if you start with the *Mincer-Zarnowitz* form (2) and the subtract $y_{t+h,t}$ from both sides then we obtain (3) with $(\alpha_0, \alpha_1) = (0, 0)$ when $(\beta_0, \beta_1) = (0, 1)$. Therefore the two approaches are identical

Market Forecasts for interest rates

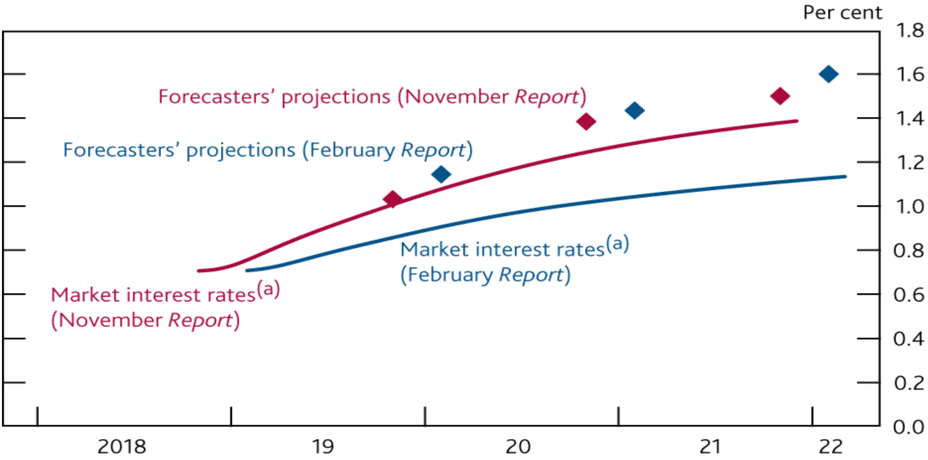
- Often market forward or futures rates are used as forecasts. The Bank of England uses the implied forecasts of future Bank Rate from the yeild curve to forecast interest rates in its published forecasts. February 2019 Inflation Report.
- Conditioning path for Bank Rate implied by forward market interest rates. The data are 15 working day averages of one day forward rates to 30 January 2019 and 24 October 2018 respectively. The curve is based on overnight index swap rates.
- February figure for 2019 Q1 is an average of realised overnight rates to 30 January 2019, and forward rates thereafter.

Market interest rate forecasts

Per cent

	2019				2020				2021				2022
	Q1[b]	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
February	0.7	0.7	0.8	0.9	0.9	0.9	1.0	1.0	1.0	1.1	1.1	1.1	1.1
November	0.8	0.9	0.9	1.0	1.1	1.1	1.2	1.2	1.3	1.3	1.4	1.4	

Expected interest rates: markets have changed, forecasters have not



Exchange rates 1

- Consider the log spot exchange rate, s_t, and the forward rate at time t, for t+1, $f_{t+1,t}$. Under rational expectations (and risk neutrality) the forward rate should be an unbiased predictor of the spot

$$E(s_{t+1} \mid I_t)) = f_{t+1,t}$$

which from the property of expectations implies

$$s_{t+1} = f_{t+1,t} + \varepsilon_{t+1} \tag{4}$$

where ε_{t+1} is white noise. Subtracting s_t from both sides

$$(s_{t+1} - s_t) = (f_{t+1,t} - s_t) + \varepsilon_{t+1} \tag{5}$$

the percentage change in the exchange rate should equal the percent forward premium.

- We can test unbiasedness by removing the restrictions from either (4) giving the levels version

$$s_{t+1} = \alpha^1 + \beta^l f_{t+1,t} + \varepsilon^l_{t+1} \tag{6}$$

or from (5) giving the change version

$$(s_{t+1} - s_t) = \alpha^c + \beta^c (f_{t+1,t} - s_t) + \varepsilon^c_{t+1} \tag{7}$$

We can then test $\alpha^i = 0, \beta^i = 1$.

- Although (4) and (5) are identical (6) and (7) are not. In fact while $\beta^l \approx 1$ ($f_{t+1,t} \approx s_t$ and $s_{t+1} = s_t + \varepsilon^s_t$ since it is a random walk), β^c is often negative. The forward premium does not predict the change. The hypotheses tested in (6) and (7) are different.

Diebold-Mariano test 1

- Once we have decided on a loss function, it is often of interest to know whether one forecast is more accurate than another.
- In hypothesis testing term, we might want to test the equal accuracy hypothesis

$$E(L(e^a_{t+h,t})) = E(L(e^b_{t+h,t}))$$

  against the alternative hypothesis that one or the other is better
- Equivalently, we might want to test the hypothesis that the expected loss differential is 0

$$E(d_t) = E(L(e^a_{t+h,t})) - E(L(e^b_{t+h,t})) = 0.$$

Diebold-Mariano test 2

- The hypothesis concerns population expected loss; we test it using sample average loss, $\overline{d}$.
- If d_t is stationary, then for T large

$$\sqrt{T}(\overline{d} - \mu) \sim N(0, f)$$

where

$$\overline{d} = \frac{1}{T}\sum_{t=1}^{T}(L(e^a_{t+h,t}) - L(e^b_{t+h,t}))$$

f is the variance of the sample mean loss differential and μ is the population mean loss differential.

- Small sample properties may not be good.

Diebold-Mariano test 3

- This implies that in large samples, under the null hypothesis of a zero population mean loss differential, the standardised sample mean loss differential has a standard normal distribution

$$B = \frac{\bar{d}}{\sqrt{\frac{\widehat{f}}{T}}} \sim N(0,1)$$

where $\widehat{f}$ is a consistent estimator of f

- In practice, using

$$\widehat{f} = \sum_{\tau=-M}^{M} \widehat{\gamma}_d(\tau)$$

where $M = T^{1/3}$ and $\widehat{\gamma}_d(\tau)$ denotes the sample autocovariance of the loss differential at displacement τ, provides an adequate estimator in many cases

Diebold-Mariano test 4

- Note the B statistic is just a t-statistic for the hypothesis of a zero population mean loss differential, adjusted to reflect the fact that the loss differential is not necessarily white noise
- We can compute it by regressing the loss differential series on an intercept, taking care to correct the equation for serial correlation
- The procedure used here amounts to a "nonparametric" way of doing it, because instead of assuming a particular model for the serial correlation, we use the sample autocorrelations of the loss differential directly
- The non-parametric serial correlation correction, involves selecting a truncation lag, M
- A parametric alternative is to regress the loss differential on an intercept, allowing for $ARMA(p, q)$ disturbances, using information criteria to select p and q

Forecast Encompassing

- In forecast accuracy comparison, a natural question is which forecast is best?
- Often forecasts are not very distinct, but even when one forecast is deemed to be significantly better than another the question arises as to whether competing forecasts may be fruitfully combined to produce a composite forecast superior to all the original forecasts
- Thus forecast combination, although obviously related to forecast accuracy comparison, is logically distinct and of independent interest
- Practical conclusion: combining forecasts improves accuracy and simple average seems difficult to beat.

Forecast Encompassing

- We use **forecast encompassing** tests to determine whether one forecast incorporates (or encompasses) all the relevant information in competing forecasts.
- If one forecast incorporates all the relevant information, nothing can be gained by combining forecasts
- Take the case of two forecasts, $y^a_{t+h,t}$ and $y^b_{t+h,t}$ - consider the regression

$$y_{t+h} = \alpha + \beta_a y^a_{t+h,t} + \beta_b y^b_{t+h,t} + \varepsilon_{t+h,t}$$

- If $(\beta_a, \beta_b) = (1, 0)$ we would conclude that model *a* forecast-encompasses model *b*, and if $(\beta_a, \beta_b) = (0, 1)$ we say model *b* forecast-encompasses model *a*
- For other values of (β_a, β_b), neither model encompasses the other, and both forecasts contains useful information about y_{t+h}
- In covariance stationary environments, encompassing hypothesis can be tested using standard methods. If neither forecast encompasses the other then combination is potentially desirable.

Forecast Combination

- In general forecast combination problems seeks an aggregator that reduces the information in a potentially high-dimensional vector of forecasts, $\widehat{\mathbf{y}}_{t+h,t}$, to a lower dimensional summary measure, $C(\widehat{\mathbf{y}}_{t+h,t}; \boldsymbol{\omega}_c)$, where $\boldsymbol{\omega}_c$ are the weights or parameters associated with the combination
- Many different types of loss function may be relevant depending on the problem, but here we simplify matters by following standard practice and assuming that the loss function only depends on the forecast error from the combination
- Simplifying still further, we adopt an objective function underlying the problem to be mean squared error (MSE) loss

$$L(y_{t+h}, \widehat{y}_{t+h,t}) = \theta(y_{t+h} - \widehat{y}_{t+h,t})^2, \qquad \theta > 0$$

- Under this type of quadratic loss it is straightforward to illustrate the gains from different forecast combination schemes

Forecast Combination

- Suppose we have two unbiased forecasts from which we form a composite as
$$y^c_{t+h,t} = \omega y^a_{t+h,t} + (1-\omega) y^b_{t+h,t}$$
because the weights sum to unity, the composite forecast will necessarily be unbiased (linear combinations of zero-mean variables - with weights that sum to one - give zero mean composite variables which is an unbiased forecast)
- Moreover, the combined forecast errors will satisfy the same relation as the combined forecast, with zero mean:
$$e^c_{t+h,t} = \omega e^a_{t+h,t} + (1-\omega) e^b_{t+h,t}$$
and variance
$$\sigma^2_c = \omega^2 \sigma^2_{aa} + (1-\omega)^2 \sigma^2_{bb} + 2\omega(1-\omega)\sigma^2_{ab}$$
where σ^2_{aa} and σ^2_{bb} are the forecast error variances and σ^2_{ab} is their covariance

Forecast Combination

- As we have assumed unbiased (zero mean) forecasts (single and combined), our quadratic loss function, where we wish to minimise $E[(e^c_{t+h,t})^2] = E(y_{t+h} - \widehat{y}^c_{t+h,t})^2$, is the variance of the combined forecast error
- The optimal combining weights are found, like when defining weights for a portfolio combining two risky assets, by minimising the variance of the combined forecast error with respect to ω,
- $$\sigma^2_c = \omega^2\sigma^2_{aa} + (1 + \omega^2 - 2\omega)\sigma^2_{bb} + 2\omega\sigma^2_{ab} - 2\omega^2\sigma^2_{ab}$$
  $$\frac{\partial\sigma^2_c}{\partial\omega} = 2\omega\sigma^2_{aa} + 2\omega\sigma^2_{bb} - 2\sigma^2_{bb} + 2\sigma^2_{ab} - 4\omega\sigma^2_{ab} = 0$$

Forecast Combination

This yields

$$\omega^* = \frac{\sigma^2_{bb} - \sigma^2_{ab}}{\sigma^2_{aa} + \sigma^2_{bb} - 2\sigma^2_{ab}}$$

- The optimal combining weight is a simple function of the variance and covariances of the underlying forecast errors weight
- So in

$$y^c_{t+h,t} = \omega y^a_{t+h,t} + (1-\omega) y^b_{t+h,t}$$

A greater weight is assigned to the models producing more precise forecasts(lower forecast error variance). The bigger σ^2_{bb} the greater weight put on $y^a_{t+h,t}$.

- But future variances may not be the same as past variances.

Forecast Combination

- Consider now the regression method of forecast combination. The form of the forecast-encompassing regressions immediately suggests combining forecasts by simply regressing realisations on forecasts
- The optimal variance-covariance combining weights have a or moving window regression interpretation as the coefficients of a linear projection of y_{t+h} onto forecasts, subject to two constraints
  - the weights sum to unity
  - the intercept is excluded
- In practice we cannot do population projection, so we simply run the regression on the available data
- Moreover it is usually preferable *not* to force the weights to add to unity, or to exclude an intercept
- Inclusion of an intercept, for example, facilitates bias correction and allows biased forecasts to be combined.

Forecast Combination

- Typically therefore we run the regression

$$y_{t+h} = \beta_0 + \beta_1 y^a_{t+h,t} + \beta_2 y^b_{t+h,t} + \varepsilon_{t+h,t}$$

  Extension to the fully general case of more than two forecasts is immediate but may have short samples.
- Confidence intervals?
- Extensions to this method are straight forward
  - time-varying weights, using for example recursive or rolling window estimation
  - accounting for serial correlation when the forecast horizon is greater than one and when forecasts are not necessarily optimal with respect the information set
  - Shrinkage of combining weights toward equality: simple arithmetic averages i.e. equal weights - often performs better in out-of-sample forecasting, than optimal combinations, because of changing structures and estimation errors.

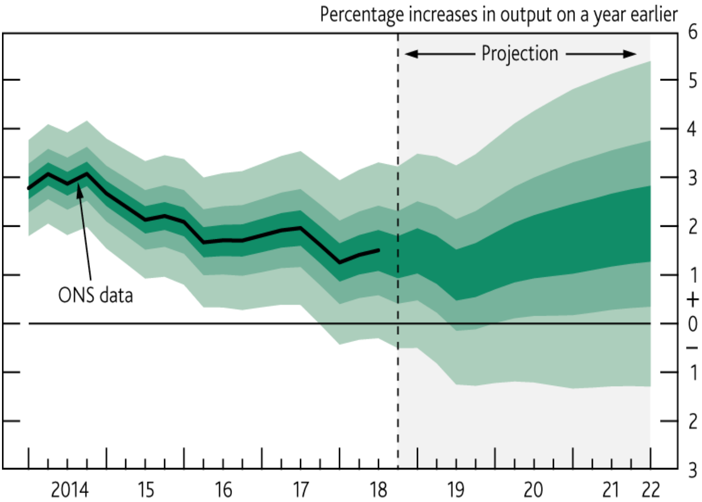
Alternatives to point forecasts

- Often need more than point forecasts, for instance when we do not know users' loss functions or we want to convey our uncertainty.
- They may care about the uncertainty so we may want to forecast the variance, assuming that it changes through time to get a **volatility forecast** need it to get optimal forecast with linex loss.
- We may use our estimate of the variance (and perhaps higher moments) together with an assumed distribution to give a parametric **interval forecast**, e.g. a 95% probability that it lies between $y^-_{t+h,t}$ and $y^+_{t+h,t}$.
- We may use our estimate of the probability distribution of the possible future values of the variable to give a **density forecast**

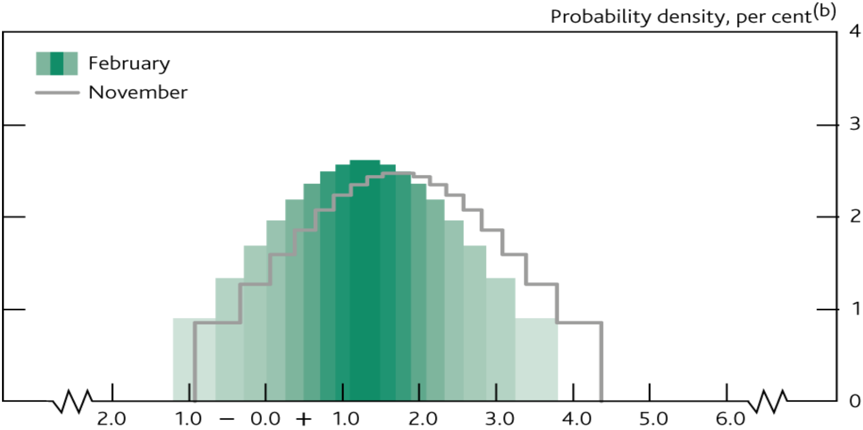
Density forecasts may be

- **parametric,** based on an assumed distribution, typically normal, but could be GARCH with t distribution.
- **bootstrapped**, drawn from the distribution of past residuals
- **judgemental**, like the Bank of England fan chart, which provides a sequence of density forecasts, structured around half normal distributions, which can represent the MPCs judgement about upside and downside risks..

Feb Inflation report GDP 90% projection

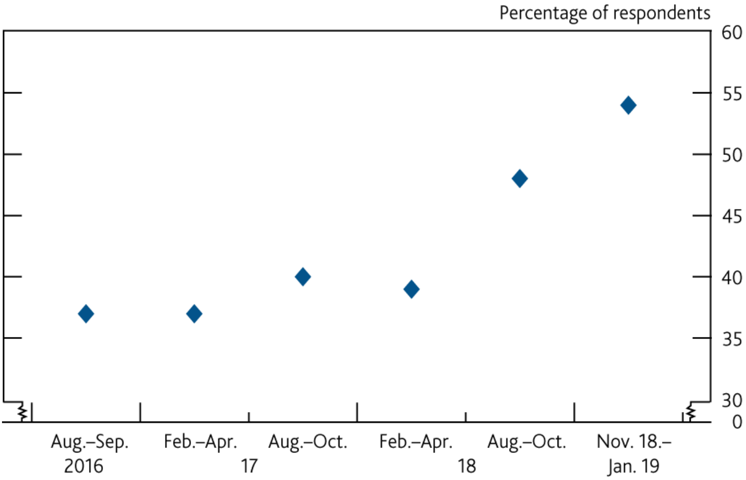


Feb 90% probabilities for GDP growth 2020Q1



Key judgement	Likely developments in 2019 Q1 to 2019 Q3 if judgements evolve as expected
1: global GDP growth weakens further and settles at close to its potential rate	• Quarterly euro-area GDP growth to average ¼%. • Quarterly US GDP growth to average ½%. • Indicators of activity consistent with four-quarter PPP-weighted emerging market economy growth of around 4¼%; within that, GDP growth in China to average around 6%. • The contribution of net trade to quarterly UK GDP growth to be close to zero, on average.
2: UK domestic demand growth is soft over much of 2019, due in part to elevated Brexit uncertainties, before picking up	• Business investment to fall by ½% per quarter, on average. • Quarterly real post-tax household income growth to average ¼%. • Quarterly consumption growth to average ¼%. • Mortgage spreads to widen a little. • Mortgage approvals for house purchase to average around 65,000 per month. • The UK house price index to increase by around ¼% per quarter, on average. • Housing investment to fall by ½% per quarter, on average.
3: potential supply continues to grow at subdued rates and excess demand emerges over the forecast	• Unemployment rate to average around 4%. • Participation rate to average around 63¾%. • Average weekly hours worked to remain around 32. • Cumulative growth in hourly labour productivity to be ¼% to ½%.
4: CPI inflation is supported by strengthening domestic inflation, although it falls slightly below the target temporarily due to lower energy prices	• Non-fuel import prices to rise by just over ¾% in the year to 2019 Q3. • Electricity and gas prices to contribute around ¼ percentage point to CPI inflation in 2019 Q2, as Ofgem's energy price cap is raised. • Commodity prices and sterling ERI to evolve in line with the conditioning assumptions set out in this *Report*. • Four-quarter growth in whole-economy AWE regular pay to average around 3¼%. • Four-quarter growth in whole-economy unit labour costs to average around 3¼%. • Four-quarter growth in whole-economy unit wage costs to average just over 3%; growth in private sector regular pay based unit wage costs to average around 3¼%. • Indicators of medium-term inflation expectations to continue to be broadly consistent with the 2% target.

% of firms reporting that Brexit is in their top three sources of uncertainty



Exercises

- Consider a real forecaster and explain their decision problem.
- Explain why multi-step ahead forecast errors will be serially corelated.
- Explain why (6) and (7) are not identical while (2) and (3) are identical.
- Comment on how the Bank of Englands interest rate forecasts were revised between the November and February inflation reports.
- Comment on how the Bank of Englands 2020Q1 GDP growth forecasts were revised between the November and February inflation reports.
- Explain how you would evaluate the 90% interval forecasts for GDP given in the inflation report.
- How would you evaluate and revise your beginning of term forecasts of the financial indicators?