**I) Title: Model Uncertainty**

**II) Contents:** Encompassing, combination for forecast, comment by Hendry and Clements

**III) Documentation:**

- Diebold, Francis X. (1998). *Elements of Forecasting*, Chapter 12, SW College Publishing.
- Hendry, David F., and Clements Michael P. (2001). Economic Forecasting: Some Lessons From Recent Research. *European Central Bank Working Paper Series*, No. 82.
- Jordan, Thomas J., and Savioz, Marcel. (2003). Does it Make Sense to Combine Forecasts from VAR Models? An Empirical Analysis with Inflation Forecasts for Switzerland. *Quarterly Bulletin*, 4.
- Jungmittag, Andre. (1993). The Combination of Forecasts: An Overview with Applications. *Jahrbücher für Nationalökonomie und Statistik*. Vol 212/1-2

**Working Title: Model Uncertainty**

**Table of contents:**



## Model uncertainty

When a central bank has to make a (monetary policy) decision based on forecasts, it finds itself often with several different forecasts, obtained from different forecasting methods. Before making the main decision, it therefore has to decide on which forecast to retain. So far we have seen several statistical criteria one can use to evaluate forecasts and we have found that one should choose the forecast which performs best with respect to a loss function.

Sometimes, however, one might not want to rely on only one particular forecast, but rather on a combination of multiple forecasts. It seems likely that one forecast contains data that was not used in another and in this case it is of particular interest to combine the forecasts. An additional advantage of a combined forecast is that it is likely to be more robust than any of the single forecasts alone, since a combined forecast does not depend too much on the model specification of a single forecast and is therefore less affected by underlying structural changes in the economy.

## Forecast encompassing

First, we will want to test if one forecast incorporates, i.e. encompasses all the relevant information from different forecasts, because there will be no gain in combining forecasts in this case.

Consider two forecasts a and b and the regression

(1) $y_{t+h} = \beta_a y^{(a)}_{t+h,t} + \beta_b y^{(b)}_{t+h,t} + \varepsilon_{t+h,t}$

where $y^{a}_{t+h,t}$ and $y^{b}_{t+h,t}$ represent the forecasts made by methods a and b respectively. When we run the regression, we must make sure to take into account that $\varepsilon_{t+h,t}$ may be serially correlated for $h>1$. If $\beta_a = 1$ and $\beta_b = 0$ we can conclude that forecast a encompasses forecast b. Vice versa, we would say that forecast b encompasses forecast a if $\beta_a = 0$ and $\beta_b = 1$. For any other values there is no encompassing and there might be a potential gain in combining the two forecasts.

## Forecast combination

Similarly to the evaluation of a single forecast we say that a good combination of forecasts minimizes some loss function. In practice one chooses to minimize the variance of the combined forecast. Consider again two forecasts with

$E\left[y_{t+h} - E\left[y^{(i)}_{t+h} \middle| y_t\right]\right] = E\left[y_{t+h} - y^{(i)}_{t+h,t}\right] = E\left[e^{(i)}_{t+h,t}\right] = 0$, $E\left[\left(e^{(i)}_{t+h,t}\right)^2\right] = \sigma_i^2$ for $i = 1,2$

and

$E\left[e^{(1)}_{t+h,t} e^{(2)}_{t+h,t}\right] = \sigma_{12}$.

In order to obtain a combined forecast that outperforms each single forecast we consider the linear combination

(2) $y^{(c)}_{t+h,t} = (1-\omega) y^{(1)}_{t+h,t} + \omega y^{(2)}_{t+h,t}$

where $0 < \omega < 1$ is the weighting factor. Since $\omega + (1-\omega) = 1$ and both forecasts are assumed to be unbiased, $y^{c}_{t+h,t}$ will be unbiased as well. Its variance is given by

(3) $\sigma_c^2 = (1-\omega)^2 \sigma_1^2 + \omega^2 \sigma_2^2 + 2\omega(1-\omega)\sigma_{12}$

which we seek to minimize.

Solving the FOC

(4) $\dfrac{\partial \sigma_c^2}{\partial \omega} = -2\sigma_1^2 + 2\omega\sigma_1^2 + 2\omega\sigma_2^2 + 2\sigma_{12} - 4\omega\sigma_{12} = 0$

for optimal $\omega$ yields

(5) $\omega^* = \dfrac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$

Plugging this optimal weight in (3) gives the minimal variance

(6) $\sigma^2_{c,\min} = \dfrac{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} = \dfrac{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho}$

where $\rho$ is the correlation coefficient between $y^{(1)}_{t+h,t}$ and $y^{(2)}_{t+h,t}$.

It can be shown that $\sigma^2_{c,\min} \le \min\left(\sigma_1^2, \sigma_2^2\right)$ except for some particular cases ($\sigma_1^2 = 0$, or $\sigma_1^2 = \sigma_2^2$ and $\rho = 1$, or $\rho = \dfrac{\sigma_1}{\sigma_2}$) which we will ignore.

An interesting case is when $\rho = 0$ and (4) becomes

(7) $\omega^* = \dfrac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$

Formula (7) implies that the forecast with the smaller error variance receivers a greater weight. Example: If $\sigma_1^2$ is getting smaller, forecast one becomes more accurate and $\omega^*$ decreases which corresponds to putting less weight on forecast two. In fact, if we rewrite (2) as

(8) $y_{t+h,t}^{(c)} = y_{t+h,t}^{(1)} + \omega\left(y_{t+h,t}^{(2)} - y_{t+h,t}^{(1)}\right)$

one can interpret the combined forecast as forecast one corrected by the weighted difference between forecast two and one. If $\omega^*$ is getting smaller, the correction part becomes less important.

**<u>Combination methods</u>**

In practice we do not know the true forecast error variances and we therefore have to come up with an estimate of the combining weights.
The simplest way to determine the combining weights is to form an arithmetical average of all forecasts. For N forecasts we would have:

(9) $y_{t+h,t}^{(c)} = \dfrac{1}{N} y_{t+h,t}^{(1)} + \ldots + \dfrac{1}{N} y_{t+h,t}^{(N)}$

This method completely ignores any information we have about the forecast errors and it seems logical that taking into account the sample variances of the observed forecast errors improves the accuracy of the combined forecast. If we assume that
$corr(y_{t+h,t}^{(i)}, y_{t+h,t}^{(j)}) = \rho_{ij} = 0$ for $i, j = 1,\ldots N$ with $i \neq j$ we can use

(10) $\hat{\omega}_N^{(i)} = \dfrac{\left(\hat{\sigma}_i^2\right)^{-1}}{\sum_{j=1}^{N}\left(\hat{\sigma}_j^2\right)^{-1}}$

where $\hat{\omega}_N^{(i)}$ is the estimated weight for forecast i and $\hat{\sigma}_i^2$ is its sample forecast error variance.
We could also compute sample covariances between the observed forecast errors and use this information in a formula similar to (4), which is a rather tedious task for $N > 2$.
However, it is useful to note that when we are combining different forecasts, we are actually forming a portfolio of forecasts and seek to "diversify" the forecast errors of the combined forecast. A standard result from finance is that the optimal weighting factors (the betas) for the included assets depend on their variances and covariances. The beta has a regression interpretation and it is normally computed by regressing the assets on the market index. (STIMMT DAS?)
Similarly to the finance intuition we consider in the context of forecast combination the regression of the observed values on the forecasts we wish to combine:

(11) $y_{t+h} = \beta_0 + \beta_1 y_{t+h,t}^{(1)} + \ldots + \beta_N y_{t+h,t}^{(N)} + \varepsilon_{t+h,t}$

The OLS regression procedure seeks to minimize

(12) $\min\limits_{\beta_0,\ldots,\beta_N}\left\{\sum_{t=1}^{T-h}\left[y_{t+h}-\left(\beta_0+\beta_1 y_{t+h,t}^{(1)}+\ldots+\beta_N y_{t+h,t}^{(N)}\right)\right]^2\right\}=\min\limits_{\beta_0,\ldots,\beta_N}\left\{\sum_{t=1}^{T-h}\left(\hat{\varepsilon}_{t+h}\right)^2\right\}=\min\limits_{\beta_0,\ldots,\beta_N}\left\{\sum_{t=1}^{T-h}\left(e_{t+h}^{(c)}\right)^2\right\}$

which can be interpreted as minimizing the sample variance of the combined forecast error $e_{t+h}^{(c)}$, which is exactly what we want.

There are many variations and extensions to the general concept of **the regression-based forecast combination method** outlined by equation (11).

**Constrained weights and shrinkage of combining weights**

The **Shrinkage Principle** states that imposing restrictions on forecasting models often improves forecast performance. The term shrinkage comes form the idea of imposing restrictions on forecasting models to "shrink" or "coax" estimates in a certain direction. Note that the arithmetical average can be seen as a particular case of constrained weights. We now also have a simple way to find the combining weights similar to (10) but this time we don't need to assume $\rho_{ij}=0$ for computational ease. We run a restricted OLS regression on (11) with the restrictions that $\beta_0=0$ and $\sum_{i=1}^{N}\beta_i=1$.

It might as well be of interest to the forecaster to only restrict $\beta_0=0$. As a practical matter it is desirable to try out different sets of restrictions and analyze if the results turn out very different.

**Serial correlation**

Of course there is the usual caveat about the error terms in regression (11), namely that the forecast errors of the combined forecast may be serially correlated for $h>1$. Therefore we should regress (11) while allowing for $\varepsilon_{t+h} \sim MA(h-1)$ because forecast errors of an optimal forecast will necessarily exhibit this pattern.

To be on the safe side and admit that the forecasts which are to be combined may not entirely capture the dynamics in y, we could even allow for serially correlated disturbances or lagged dependent variables, thus letting $\varepsilon_{t+h} \sim ARMA(p,q)$.

**Time varying combining weights**

The performance of different forecasting methods may vary over time, so it is only natural that we want to weigh the forecasts in function of their relative accuracy over time. One way to allow for trend-like changing weights is to regress

(13) $y_{t+h}=\left(\beta_0^0+\beta_0^1 t\right)+\left(\beta_1^0+\beta_1^1 t\right)y_{t+h,t}^{(1)}+\ldots+\left(\beta_N^0+\beta_N^1 t\right)y_{t+h,t}^{(N)}+\varepsilon_{t+h,t}$

where $t$ represents the time index. The size and the significance of the estimates for $\beta_i^1$ tell us something about the relative importance of the time trend.

**Nonlinear combining regressions**

If we suspect the behaviour of $y_t$ not to be perfectly linear, we might want to include quadratic terms in (11) in order to capture deviations from linearity.

For N=2 we run the regression

(14) $y_{t+h}=\beta_0+\beta_1 y_{t+h,t}^{(1)}+\beta_2 y_{t+h,t}^{(2)}+\beta_{11}\left(y_{t+h,t}^{(1)}\right)^2+\beta_{22}\left(y_{t+h,t}^{(2)}\right)^2+\beta_{12}\left(y_{t+h,t}^{(1)}y_{t+h,t}^{(2)}\right)+\varepsilon_{t+h,t}$

Examining the significance of the estimates for $\beta_{11}, \beta_{12}$ and $\beta_{22}$ enables us to infer whether the nonlinearity assumption can and should be maintained.

**A comment by Hendry and Clements**

A further argument in favour of forecast combination is given by the comment on current forecasting methods which has been put forward by Hendry and Clements (HC hereinafter).

HC take a closer look at the reasons and the proposed explanations for the often observed poor forecast accuracy. They find that structural breaks are a major source for forecast failure. In this context one could argue in favour of using a combined forecast since it is likely to be less affected by a structural break due to it being based on a combination different forecasting methods.

HC's critique is directly related to the commonly adapted optimality theory whose basic assumptions it is useful to recall:
- The model is a good representation of the economy and
- the structure of the economy will remain relatively unchanged.

The optimality theory and those two assumptions are unable to provide sufficient explanations for the commonly observed forecast failures. One often makes use of the vague terms of mis-specified models, poor methods, inaccurate data, incorrect estimation, and data-based model selection as a possible explanation for forecast failure without proving it.

HC suggest two alternative assumptions which are less stringent than those of the optimality theory:
- Models are simplified representations which are incorrect in many ways and
- economies both evolve and suddenly shift.

Those assumptions allow for more reasonable and founded explanations of forecast errors of which HC find the main sources to be:
- shifts in the coefficients of deterministic/stochastic terms
- mis-specification of deterministic/stochastic terms
- mis-estimation of the coefficients of deterministic/stochastic terms
- mis-measurement of the data
- changes in the variances of the errors
- errors cumulating over the forecast horizon.

Theoretical analysis, Monte Carlo simulations and empirical evidence suggest that shifts in the coefficients of deterministic terms have most the harmful influence on forecast accuracy, leading to systematic forecast failure.

According to HC, a fundamental problem that arises in the presence of structural breaks is that causal variables (variables that actually determine the outcome) cannot be proven to dominate non-causal variables in forecasting models. After a structural break, the best causal model may forecast less accurately than a model without causal variables and this is the case despite the causal model performed best before the shift happened. Its incapability to adapt to the new situation will result in it making biased forecasts after the shift and a non-causal model which adapts immediately to new situations is therefore able to outperform it in the long run.

HC propose that the usage of intercept corrections (IC) could be one possible solution to deterministic shifts. IC consists in adjusting the constant of an equation, based on realized equation errors immediately prior to the forecast origin. ICs are typically used if one has recently observed forecast failure and suspects that a deterministic shift has occurred, which one wishes to offset. Note that by their very nature ICs are non-causal remedies to deterministic shifts.

Further points put forward by HC:
- Role of causal models: Bad forecast performance cannot be used to disqualify causal models because EqCM, VAR and VAR in differences do equally bad.
- Model selection: Model selection strategies appear to have little impact on forecast performance.
- Results of forecasting competitions: Conclusions are such (simple models do best, accuracy measure matters, pooling helps, evaluation horizon matters) because economies are not reducible to stationarity by differencing and some models are relatively robust to deterministic shifts.
- Simplicity in forecasting: Simple models perform well due to their adaptability to shifts in intercepts and trends.
- Evaluating forecasts: Choice of forecasts should depend on their purpose, rather than just a statistical criterion.

HC conclude: "Thus, in practice, economic forecasts end up being a mixture of science – based on econometric systems that embody consolidated economic knowledge and have been carefully evaluated – and art, namely judgements about perturbations from recent unexpected events."