Lecture 4- Forecast Evaluation and Combination

Pedro Valls EESP-FGV & CEQEF-FGV

July 6, 2024

FGV EESP

CENTRO DE ESTUDOS QUANTITATIVOS EM ECONOMIA E FINANÇAS

Outline

- Introduction
- Unbiasedness and efficiency
- Test for Unbiasedness
- Test for Weak Efficiency
- Test for Strong Efficiency
- Evaluation of fixed event forecast
- Tests of Predictive Accuracy
- Forecast Comparison Tests
- The Combination od Forecasts
- Forecast Encompassing
- Evaluation and Combination of density forecasts

Introduction

- We will try to answer the following questions:
  - (i) How "good" , in some sense, is a particular set of forecasts?
  - (ii) Is one set of forecasts better than another one?
  - (iii) Is it possible to get a better forecast as a combination of various forecasts for the same variable?
- For "good" we should have how to test. - Forecast Evaluation ([Clements and Hendry, 1998] and [Clements and Hendry, 1999])
- For (ii) we introduce some basic statistics to assess it or some given criterion, for example, *MSFE* - Forecast Comparison ([Clark and McCracken, 2013])
- How to combine the forecasts and why the pooled forecasts is expected to perform better - Forecast Combination ([Timmermann, 2006])
- For a rigorous and exhaustive treatment of these topics [Elliott and Timmermann, 2013].

Unbiasedness and efficiency

- A good prediction should have the unbiasedness and efficiency properties.
  - Unbiasedness
    - the optimal forecast under MSFE loss is $E_t(y_{t+h}) = \widehat{y}_{t+h|t}$
    - $E(e_t) = 0$ where $e_t = y_{t+h} - \widehat{y}_{t+h|t}$
  - Efficienct
    - the optimal $h-$step ahead forecast error should be at most correlated of order $h-1$ and uncorrelated with available information at the time the forecast is made.

Test for unbiasedness I

- Let us consider the regression

$$y_{i+h} = \alpha + \beta \widehat{y}_{i+h|i} + \varepsilon_{i+h} \quad \text{for} \quad i = T, \cdots, T+H-h \quad \text{and} \quad h < H \tag{1}$$

- where h is the forecast horizon and $T+1, \cdots, T+H$ is the evaluation sample.
- The forecasts $\widehat{y}_{i+h|i}$ are recursively updated in periods $i = T, \cdots, T+H-h$
- Note that

$$E_i(y_{i+h}) = E_i(\alpha + \beta \widehat{y}_{i+h|i} + \varepsilon_{i+h}) = \alpha + \beta E_i(\widehat{y}_{i+h|i}) = \widehat{y}_{i+h|i}$$

- if

$$\alpha = (1-\beta) E(\widehat{y}_{i+h|i}) \tag{2}$$

Test for unbiasedness II

- But when $\alpha = 0$ and $\beta = 1$, $\widehat{y}_{i+h|i}$ is also an unbiased forecast for y_{i+h}.
- The sufficient condition $\alpha = 0$ and $\beta = 1$ can be tested by a "robust" $F-$test where ε_{i+h} the autocorrelated of order $h-1$ is taken into account in the derivation of the HACSE.
- The necesssary condition (2) is equivalent to $\tau = 0$ in:

$$e_{i+h} = y_{i+h} - \widehat{y}_{i+h|i} = \tau + \varepsilon_{i+h}$$

- which can be tested with a robust version of the $t-$test.
- The sufficient condition also implies that the forecast and the forecast error are uncorrelated.

Test for unbiasedness III

- Expression (1) can be rewritten as

$$
\begin{aligned}
y_{i+h} &= \alpha + \beta \widehat{y}_{i+h|i} + \varepsilon_{i+h} \\
y_{i+h} - \widehat{y}_{i+h|i} &= \alpha + (\beta - 1)\widehat{y}_{i+h|i} + \varepsilon_{i+h} \\
e_{i+h} &= \alpha + (\beta - 1)\widehat{y}_{i+h|i} + \varepsilon_{i+h} \qquad (3)
\end{aligned}
$$

- so that

$$
E(\widehat{y}_{i+h|i} e_{i+h}) = \alpha E(\widehat{y}_{i+h|i}) + (\beta - 1) E(\widehat{y}^2_{i+h|i}) + \underbrace{E(\widehat{y}_{i+h|i} \varepsilon_{i+h})}_{=0} = 0 \qquad (4)
$$

- so the sufficient condition also guarantees that the forecasts cannot be used to "reduce" the forecast error
- It is also a condition that relates to efficiency on the forecasts.

Test for unbiasedness IV

- Also when $\alpha = 0$ and $\beta = 1$ we have

$$Var(y_{i+h}) = Var(\widehat{y}_{i+h|i}) + Var(\varepsilon_{i+h})$$

- implying that the "volatility" of the variable is larger than the "volatility" of the optimal forecast
- by (3) the "volatility" of the forecast error is larger that the "volatility" of the optimal forecast.
- The coefficient of determination, R^2, from (1) can also be used as an indicator of the forecast quality, with good forecasts associated with high R^2

Test for Weak Efficiency

- e_{t+h} is correlated across time, at most order $h-1$ which implies that no lagged information beyond $h-1$ can explain the forecast errors.
- This property can be tested by fitting a $MA(h-1)$ to the $h-$step ahead forecast error and testing that the residuals are white noise.

Test for Strong Efficiency

- A test for a good forecast, strong efficiency is that $\gamma = 0$ in the following regression:
$$e_{i+h} = \gamma' \mathbf{z}_i + \varepsilon_{i+h} \tag{5}$$
- where $\mathbf{z}_i$ is a vector of variables explaining the forecast error is correlated across time, at most order $h-1$ which implies that no lagged information beyond $h-1$ can explain the forecast errors.
- This property can be tested by fitting a $MA(h-1)$ to the $h-$step ahead forecast error and testing that the residuals are white noise.

Evaluation of fixed event forecast I

- We can consider the forecast $\{\widehat{y}_{\tau|\tau-h}\}$ for $h = 1, 2, \ldots$, i.e. forecasts for a fixed target value y_τ made at different time periods that become closer and closer to τ.
- $\{\widehat{y}_{\tau|\tau-h}\}$ are known as fixed event forecasts .
- For an $AR(1)$ it is

$$\widehat{y}_{\tau|\tau-h} = \rho^h y_{\tau-h} \tag{6}$$

- Let the forecast error be decomposed as

$$e_{\tau|\tau-h} = y_\tau - \widehat{y}_{\tau|\tau-h} = \upsilon_{\tau|\tau-h+1} + \upsilon_{\tau|\tau-h+2} + \cdots + \upsilon_{\tau|\tau} \tag{7}$$

- where

$$\upsilon_{\tau|J} = \widehat{y}_{\tau|J} - \widehat{y}_{\tau|J-1} \quad \text{and} \quad \widehat{y}_{\tau|\tau} = y_\tau \quad \text{for} \quad J = \tau - h + 1, \cdots, \tau \tag{8}$$

Evaluation of fixed event forecast II

- and for the $AR(1)$ we have

$$
\begin{aligned}
v_{\tau|\tau} &= \widehat{y}_{\tau|\tau} - \widehat{y}_{\tau|\tau-1} = y_\tau - \rho y_{\tau-1} = \varepsilon_\tau \\
v_{\tau|\tau-1} &= \widehat{y}_{\tau|\tau-1} - \widehat{y}_{\tau|\tau-2} = \\
&= \rho y_{\tau-1} - \rho^2 y_{\tau-2} = \rho(y_{\tau-1} - \rho y_{\tau-2}) = \rho\varepsilon_{\tau-1} \\
v_{\tau|\tau-2} &= \widehat{y}_{\tau|\tau-2} - \widehat{y}_{\tau|\tau-3} = \\
&= \rho^2 y_{\tau-2} - \rho^3 y_{\tau-3} = \rho^2(y_{\tau-2} - \rho y_{\tau-3}) = \rho^2\varepsilon_{\tau-2} \\
&\vdots \\
v_{\tau|\tau-h+1} &= \widehat{y}_{\tau|\tau-h+1} - \widehat{y}_{\tau|\tau-h+2} = \rho^{h-1} y_{\tau-h+1} - \rho^{h-2} y_{\tau-h+2} \\
&= \rho^{h-1}(y_{\tau-h+1} - \rho y_{\tau-h+2}) = \rho^{h-1}\varepsilon_{\tau-h+1} \\
v_{\tau|J} &= \rho^{\tau-J}\varepsilon_J
\end{aligned}
$$

- and

$$
e_{\tau|\tau-h} = \rho^{h-1}\varepsilon_{\tau-h+1} + \rho^{h-2}\varepsilon_{\tau-h+2} + \cdots + \varepsilon_\tau \tag{9}
$$

Evaluation of fixed event forecast III

- Unbiasedness requires that $E(e_{\tau|\tau-h}) = 0 \;\; \forall \tau - h$
- For weak efficiency

$$E(e_{\tau|\tau-h}|\upsilon_{\tau|\tau-h}, \cdots, \upsilon_{\tau|1}) = 0 \;\; \forall \tau - h \tag{10}$$

- the forecast error at time $\tau - h$ is uncorrelated with all previous forecast revisions up to time $\tau - h$
- Condition (10) is equivalent to

$$E(\upsilon_{\tau|\tau-h}|\upsilon_{\tau|\tau-h-1}, \cdots, \upsilon_{\tau|1}) = 0 \;\; \forall \tau - h \tag{11}$$

- the forecast revision at time $\tau - h$ is independent of all previous revisions up to $\tau - h - 1$.
- It also imply that

$$\widehat{y}_{\tau|J} - \widehat{y}_{\tau|J-1} = \rho^{\tau-J}\varepsilon_J \tag{12}$$

- i.e., the evolution of the fixed event forecasts should follow a random walk or the forecast revision should be white noise.

Tests of Predictive Accuracy I

- Suppose that y_t has a $MA(\infty)$ representation:

$$y_t = \psi(L)\varepsilon_t \tag{13}$$

- the $h-$step ahead minimum MSFE predictor is given by:

$$\widehat{y}_{T+h|T} = \sum_{J=h}^{\infty} \psi_J \varepsilon_{T+h-J} \tag{14}$$

- with associated forecast error

$$\begin{aligned} e_{T+h} &= y_{T+h} - \widehat{y}_{T+h|i} \\ &= \sum_{J=0}^{\infty} \psi_J \varepsilon_{T+h-J} - \sum_{J=h}^{\infty} \psi_J \varepsilon_{T+h-J} \\ &= \sum_{J=0}^{h-1} \psi_J \varepsilon_{T+h-J} \end{aligned} \tag{15}$$

Tests of Predictive Accuracy II

- with $\psi_0 = 1$
- Group the errors in forecasting $(y_{T+1}, \cdots, y_{T+h})$ conditional on period T in

$$\mathbf{e}_h = \psi \varepsilon_h \tag{16}$$

- where

$$\mathbf{e}_h = (e_{T+1}, \cdots, e_{T+h})' \tag{17}$$

$$\varepsilon_h = (\varepsilon_{T+1}, \cdots, \varepsilon_{T+h})' \tag{18}$$

$$\psi = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 & 0 \\ \psi_1 & 1 & \dots & \dots & 0 & 0 \\ \psi_2 & \psi_1 & \ddots & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & 1 & 0 \\ & & & & \psi_1 & 1 \end{pmatrix} \tag{19}$$

Tests of Predictive Accuracy III

- Define

$$\Phi_h = E(\mathbf{e}_h \mathbf{e}_h') = \psi E(\varepsilon_h \varepsilon_h')\psi' = \sigma_\varepsilon^2 \psi\psi' \tag{20}$$

- assuming that the model that was estimated using the sample $1, \cdots, T$ remains valid over the forecast period $T+1, \cdots, T+h$ then:

$$Q = \mathbf{e}_h' \Phi_h^{-1} \mathbf{e}_h \sim \aleph^2(h) \tag{21}$$

- where

$$\Phi_h^{-1} = \sigma_\varepsilon^{-2} (\psi^{-1})' \psi^{-1} \tag{22}$$

- Rewriting (13) into an AR representation

$$\varphi(L) y_t = \varepsilon_t \quad \text{where} \quad \varphi(L) = \psi^{-1}(L) \tag{23}$$

Tests of Predictive Accuracy IV

- we also have

$$
\begin{aligned}
\varepsilon_h &= \varphi \varepsilon_h \qquad \varphi = \psi^{-1} \\
\Phi_h^{-1} &= \sigma_\varepsilon^{-2} \varphi' \varphi
\end{aligned} \tag{24}
$$

- we can rewrite the test statistic (21) as

$$
Q = \frac{\mathbf{e}'_h \varphi' \varphi \mathbf{e}_h}{\sigma_\varepsilon^2} = \frac{\varepsilon'_h \varepsilon_h}{\sigma_\varepsilon^2} = \frac{1}{\sigma_\varepsilon^2} \sum_{J=1}^{h} \varepsilon_{T+J}^2 \tag{25}
$$

- Then (25) is the sum of squares of the one-step ahead forecast error in forecasting y_{T+J} for $J = 1, \cdots, h$.
- Replacing σ_ε^2 by and consistent estimate $\widehat{\sigma}_\varepsilon^2$ we can rewrite (25) as

$$
Q = \frac{1}{\widehat{\sigma}_\varepsilon^2} \sum_{J=1}^{h} e_{T+J|T+J-1}^2 \sim F(h, T-p) \tag{26}
$$

- where p is the number of parameter used in the model

Forecast Comparison Tests

- The most common approach is to rank the forecast according to the associated loss function, typically the MSFE or MAFE.
- These comparisons are deterministic, it evaluated whether one MSFE is larger than the others but not whether their difference is statistically significant.
- The first test is due to [Granger and Newbold, 1986] also known as the Morgan-Granger-Newbold test and requires the forecast errors to be zero mean, normally distributed and uncorrelated.
- If $\mathbf{e}_1$ and $\mathbf{e}_2$ indicate the forecast errors from two competing models the test is based on the auxiliary variables:

$$u_{1,T+J} = e_{1,T+J} - e_{2,T+J} \tag{27}$$

$$u_{2,T+J} = e_{1,T+J} + e_{2,T+J} \tag{28}$$

- Note that

$$E(\mathbf{u}_1'\mathbf{u}_2) = MSFE_1 - MSFE_2 \tag{29}$$

- where $\mathbf{u}_j' = (u_{j,T+1}, \cdots, u_{j,T+H})$ and $j = 1, 2$.

Forecast Comparison Tests - Morgan-Granger-Newbold

- The hypothesis of interest is whether $\mathbf{u}_1$ and $\mathbf{u}_2$ are uncorrelated or not.
- The test statistics is

$$\frac{r}{\sqrt{(H-1)^{-1}(1-r^2)}} \sim t_{H-1} \tag{30}$$

- where
  - t_{H-1} is a Student t distribution with $H-1$
  - H is the length of the evaluation sample
- $r = \frac{\sum_{i=1}^{H} u_{1,T+i} u_{2,T+i}}{\sqrt{\sum_{i=1}^{H} u_{1,T+i}^2 \sum_{i=1}^{H} u_{2,T+i}^2}}$

Forecast Comparison Tests - Diebold & Mariano I

- The second teste is due to [Diebold and Mariano, 1995] with test statistic given by

$$DM = H^{1/2} \frac{\sum_{j=1}^{H} d_j / H}{\sigma_d} = H^{1/2} \frac{\overline{d}}{\sigma_d} \tag{31}$$

- where

$$d_j = g(e_{1j}) - g(e_{2j}) \tag{32}$$

- and g is a loss function of interest, e.g. the quadratic loss $g(e) = e^2$ or absolute loss $g(e) = |e|$, e_1 and e_2 are the errors from two competing forecasts and σ_d^2 is the variance of $\overline{d}$.

Forecast Comparison Tests - Diebold & Mariano II

- This variance is estimated as

$$\widehat{\sigma}_d^2 = \left(\gamma_0 + 2\sum_{i=1}^{h-1}\gamma_j\right) \quad \text{with} \quad \gamma_k = \frac{1}{H}\sum_{t=k+1}^{H}(d_t - \overline{d})(d_{t-k} - \overline{d}) \tag{33}$$

- where h is the forecast horizon
- The null hypothesis is

$$H_0 : E(d) = 0 \tag{34}$$

- and under the null DM has an asymptotic standard normal distribution.
- A modified version of the DM statistic was proposed by [Harvey et al., 1998] and is given by

$$HLN = \left(\frac{H + 1 - 2h + H^{-1}h(h-1)}{HH}\right) DM \tag{35}$$

Forecast Comparison Tests - Diebold & Mariano III

- and it is distributed as a Student t with $H-1$ degrees of freedom.
- Then the models are nested the distribution of the test statistic becomes non-standard and its a function of Brownian Motions
- A solution is to use rolling rather than recursive estimation, see [Giacomini and White, 2006]

Forecast Comparison Tests - The model confidence set (MCS) I

- The model confidence set (MCS) is a forecasting model selection technique developed by [Hansen et al., 2011]
- It consists of an algorithm that ranks a set of predictions from a set of models. M^* contains the set of best type selected from an initial set of models.
- The M^0 is the set that contains the best models defined from a predictive quality criteria.
- The set that contains the best models is defined by $M^* = \{i \in M^0 :: E(d_\tau^{i,j} \leq 0 \text{ for all } j \in M^0\}$

Forecast Comparison Tests - The model confidence set (MCS) II

- Let $M^{\dagger}$ be the complementary set, i.e. $M^{\dagger} = \{i \in M^0 :: E(d_{\tau}^{i,j} > 0$ for all $j \in M^0\}$ in which $g(e_{\tau}^{i})$ is some loss function and

$$d_{\tau}^{i,j} = g(e_{\tau}^{i}) - g(e_{\tau}^{j}) \tag{36}$$

$$e_{\tau}^{i} = \tilde{y}_{t+\tau+h}^{t+\tau-i} - y_{t+\tau+h} \tag{37}$$

- MCS selects models using an equivalency test, δ_M, and an elimination rule, ρ_M.
- The equivalence rule is applied to the set $M = M^0$.
- If the equivalence rule is rejected at a selected confidence level, then there is, with high probability, a group of bad models in terms of predictive power that must be eliminated from the set of good models.
- In this case an elimination rule, ρ_M is used to remove models with low predict power form the set of good models.

Forecast Comparison Tests - The model confidence set (MCS) III

- Having done this, we use the equivalence rule again.
- The procedure is repeated until the equivalency predictive hypothesis in the analyzed set, δ_M is not rejected.
- The set of models of the last step $(\widetilde{M}_F)$ is selected and must contains the best models to a certain level of significance.
- The null hypothesis of equivalence test is given by:

$$H_M^0 : E(d_\tau^{i,j}) = 0 \text{ for all } i, j \in M \tag{38}$$

- where $M \subset M^0$.
- The alternative hypothesis is given by:

$$H_M^1 : E(d_\tau^{i,j}) \neq 0 \text{ for all } i, j \in M \tag{39}$$

Forecast Comparison Tests - The model confidence set (MCS) IV

- An important point to emphasize is that there may be better models out of the initial set of models "candidates" M^0.
- The goal is to rank a particular set of models to obtain M^*
- The null hypothesis can be tested from the following statistics

$$T_{R,M} = \max_{i,j \in M} |t_{ij}| \tag{40}$$

- where $t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{Var(\bar{d}_{ij})}}$ for some i and j
- The test statistic given by (40) has a non-stantdard statistical distribution but can be simulated using bootstrap techniques.

Forecast Comparison Tests - The model confidence set (MCS) V

- The elimination rule is given by

$$\rho_M = \arg\max_{i \in \in M} \left\{ \sup_{j \in M} (t_{ij}) \right\} \tag{41}$$

- The MCS algorithm has the following steps:

  (i) Initializes the procedure by setting the initial set of model to be analyzed $M = M^0$;

  (ii) Tests H_M^0 using δ_M and a significance level α;

  (iii) If H_M^0 is not rejected the procedure ends and the final set is $\widehat{M}^*_{1-\alpha} = M$, otherwise we use the elimination rule ρ_M to delete an object from M set and back to step (i).

Forecast Comparison Tests - The model confidence set (MCS) VI

- The authors suggest that the MCS have the following statistical properties:

  (i) $\lim_{n\to\infty}(M^* \subset \widehat{M}^*_{1-\alpha}) > 1-\alpha$ and;

  (ii) $\lim_{n\to\infty}(i^\dagger \in \widehat{M}^*_{1-\alpha}) = 0$ for all $i^\dagger \in M^\dagger$

Forecast Comparison Tests - Ranking the models: p-values for MCS

- The elimination rule, ρ_M defines a sequence of random sets, $M_0 = M_1 \supset M_2 \supset \cdots \supset M_{m_0}$, where $M_i = \{\rho_i, \cdots, \rho_{m_0}\}$ and m_0 is the number of elements in M_0, ρ_{m_0} is the first element to be eliminated, ρ_{m_1} is the second to be eliminated, and so on.
- In the end, only one model survives.
- Set up the p-value of the final model to one.
- It is stored in the p-values of the deleted models if they are larger than the p-value of the model previously eliminated.
- If the p-value of the current removal is stored. The p-values of MCS are important as make it easier for the analyst to assess a given set $\widehat{M}^*_{1-\alpha}$.

The Combination of Forecasts I

- When alternative models are available we could combine them, constructing a pooled forecast.
- Let us assume that two forecast $\widehat{y}_1$ and $\widehat{y}_2$ are available for the same target y
- The associated forecasts errors are e_1 and e_2
- We want to construct the linear combined forecast

$$\widehat{y}_c = \alpha \widehat{y}_1 + (1 - \alpha)\widehat{y}_2 \tag{42}$$

- the weights can be chosen in order to minimize the *MSFE* of $\widehat{y}_c$
- From (42) we have

$$\begin{aligned} e_c &= y - \widehat{y}_c \\ &= \alpha y - \alpha \widehat{y}_1 + (1 - \alpha)y - (1 - \alpha)\widehat{y}_2 \\ &= \alpha e_1 + (1 - \alpha)e_2 \end{aligned} \tag{43}$$

The Combination of Forecasts II

- so the variance of (43) $(MSFE)$ is given by

$$\sigma_{e_c}^2 = \alpha^2 \sigma_{e_1}^2 + (1-\alpha)^2 \sigma_{e_2}^2 + 2\alpha(1-\alpha)\rho\sigma_{e_1}\sigma_{e_2} \tag{44}$$

- where ρ is the correlation coefficient between e_1 and e_2.
- The optimal pooling weight, the minimizers of (44) are (see [Bates and Granger, 1969])

$$\alpha^* = \frac{\sigma_{e_2}^2 - \rho\sigma_{e_1}\sigma_{e_2}}{\sigma_{e_1}^2 + \sigma_{e_2}^2 - 2\rho\sigma_{e_1}\sigma_{e_2}} \tag{45}$$

- which yields

$$\sigma_{e_c,optimal}^2 = \frac{\sigma_{e_1}^2 \sigma_{e_2}^2 (1-\rho^2)}{\sigma_{e_1}^2 + \sigma_{e_2}^2 - 2\rho\sigma_{e_1}\sigma_{e_2}} \tag{46}$$

The Combination of Forecasts III

- and we can show that

$$
\begin{aligned}
\sigma^2_{e_c,optimal} - \sigma^2_{e_1} &= \frac{\sigma^2_{e_1}\sigma^2_{e_2}(1-\rho^2)}{\sigma^2_{e_1}+\sigma^2_{e_2}-2\rho\sigma_{e_1}\sigma_{e_2}} - \sigma^2_{e_1} \\
&= \frac{\sigma^2_{e_1}\sigma^2_{e_2}(1-\rho^2) - \sigma^2_{e_1}(\sigma^2_{e_1}+\sigma^2_{e_2}-2\rho\sigma_{e_1}\sigma_{e_2})}{\sigma^2_{e_1}+\sigma^2_{e_2}-2\rho\sigma_{e_1}\sigma_{e_2}} \\
&= \frac{\sigma^2_{e_1}(\sigma^2_{e_2}-\rho^2\sigma^2_{e_2}-\sigma^2_{e_1}-\sigma^2_{e_2}+2\rho\sigma_{e_1}\sigma_{e_2})}{(\sigma_{e_1}-\rho\sigma_{e_2})^2+\sigma^2_{e_2}(1-\rho^2)} \\
&= \frac{\sigma^2_{e_1}(-\rho^2\sigma^2_{e_2}-\sigma^2_{e_1}+2\rho\sigma_{e_1}\sigma_{e_2})}{(\sigma_{e_1}-\rho\sigma_{e_2})^2+\sigma^2_{e_2}(1-\rho^2)} \\
&= \frac{-\sigma^2_{e_1}(\sigma_{e_1}-\rho\sigma_{e_2})^2}{(\sigma_{e_1}-\rho\sigma_{e_2})^2+\sigma^2_{e_2}(1-\rho^2)} \leq 0 \\
\Longrightarrow \quad & \sigma^2_{e_c,optimal} \leq \sigma^2_{e_1} \qquad (47)
\end{aligned}
$$

The Combination of Forecasts IV

- where equality holds if $\rho = \frac{\sigma_{e_1}}{\sigma_{e_2}}$ which implies that $\widehat{y}_1$ or $\widehat{y}_2$ is optimal forecasts.
- If the forecast errors are uncorrelated $\rho = 0$ (45) reduces to

$$\alpha^* = \frac{\sigma^2_{e_2}}{\sigma^2_{e_1} + \sigma^2_{e_2}} \tag{48}$$

- which are commonly used weigths in empirical applications even with correlated error.
- In practice α is not known and must be estimated.
- An estimate of α is obtained by the regression, over the estimation sample,

$$y = \alpha \widehat{y}_1 + (1 - \alpha)\widehat{y}_2 + e \tag{49}$$

- or

$$e_2 = \alpha(\widehat{y}_1 - \widehat{y}_2) + e \tag{50}$$

Forecasts Encompassing

- Forecast Encompassing concerns whether the one-step forecast of one model can explain the forecast errors made by another (which is therefore not strongly efficient)
- Can use regression (50) and test for $\alpha = 0$.
- If $\alpha \neq 0$ the difference between $\widehat{y}_1$ and $\widehat{y}_2$ can partly explain e_2 and therefore the second model cannot forecast encompass the first one.
- A more direct test can be based on the regression

$$e_1 = \delta \widehat{y}_2 + u \tag{51}$$

- and it requires $\delta = 0$ for the second model not to forecast encompass the first one

Evaluation and Combination of density forecasts I

- Let the density forecast be $f_{T+h|T}$, given informarion up to T (and X_{T+h}) with horizon h and cumulative distribution function $F_{T+h|T}$.
- Let the true density of the target variable be $g_{T+h|T}$ and its CDF by $G_{T+h|T}$.
- For example in the case of the linear regression model under the assumption of normal errors, the optimal density forecast $(f_{T+h|T})$ is

$$y_{T+h} \sim N(\widehat{y}_{T+h}, V(e_{T+h}))$$

- where $\widehat{y}_{T+h} = X_{T+h}\widehat{\beta}_T$ and $V(e_{T+h})$ denotes the variance of the forecast error.
- The true density $(g_{T+h|T})$ is instead

$$y_{T+h} \sim N(X_{T+h}\beta, \sigma_{\varepsilon}^2)$$

Evaluation and Combination of density forecasts II

- Evaluation of point forecasts we just compare the forecast and actual values
- Evaluation of density forecasts we must compare the entire forecast and actual densities, or the corresponding CDF and the evaluation is more complex
- Introduce the Probability Integral Transformation (PIT) defined as

$$PIT_t(x) = F_{t+h|t}(x) \tag{52}$$

- for any x.
- It can be shown that if $F_{t+h|t} = G_{t+h|t}$ for all t, then the PIT_t's are independent $U[0,1]$ variables.
- For a formal assessment of probability calibration, let us consider the inverse normal transformation:

$$z_t = \Phi^{-1}(PIT_t) \tag{53}$$

Evaluation and Combination of density forecasts III

- where Φ is the CDF of a standard normal variable.
- If PIT_t is $\overset{iid}{\sim} U(0,1)$ then z_t is $\overset{iid}{\sim} U(0,1)$
- May want to directly compare two (or more) competing density forecasts.
- Introduce the logarithmic score defined as

$$\log S_j(x) = \log f_{j,t+h|t}(x) \tag{54}$$

- where j indicates the alternative densities.
- If one of the densities under comparison coincides with $g_{t+h|t}$ (the true density) then the expected value of the difference in the logarithmic scores coincides with the well known Kullback-Leibler Information Criterion (KLIC):

$$KLIC_{j,t} = E_g[\log g_{t+h|t}(x) - \log f_{j,t+h|t}(x)] = E[d_{j,t}(x)] \tag{55}$$

Evaluation and Combination of density forecasts IV

- Can interpret $d_{j,t}$ as a density forecast error.
- To compare two densities f_j and f_k we can use:

$$\Delta L_t = \log S_j(x) - \log S_k(x) \tag{56}$$

- To test if the two densities are different we can use:

$$\sqrt{T}\left(\frac{\sum \Delta L_t}{T} / std.dev\right) \to N(0,1) \tag{57}$$

- Starting from n forecast densities f_j, $j = 1, \cdots, n$ the combined density forecast is:

$$f_c = \sum_{j=1}^{n} \omega_j f_j \tag{58}$$

- where $\omega_j \geq 0$, $j = 1, \cdots, n$ and $\sum_{j=1}^{n} \omega_j = 1$.

Evaluation and Combination of density forecasts V

- The combined density f_c is therefore a finite mixture distribution
- Can use $\omega_j = 1/n$ for $j = 1, \cdots, n$
- The weights could be chosen optimally to maximize a certain objective function or minimize the KLIC with respect to the true unknown density

Example using simulated data

- The following data generating process will be used
  - Model 1 - linear regression $y_t = \alpha_1 + \alpha_2 x_t + \varepsilon_t$
  - Model 2 - model with dummy variables
    $y_t = \alpha_1 + \beta_1 D_t + \alpha_2 x_t + \beta_2 D_t x_t + \varepsilon_t$
  - Model 3 - dynamic model
    $y_t = \alpha_1 + \beta_1 D_t + \alpha_2 x_t + \beta_2 D_t x_t + \gamma_1^y y_{t-1} + \gamma_1^x x_{t-1} + \varepsilon_t$
- The following R commands are used to estimate this model by OLS.
  - Aula4_2022

frametitleReference

Bates, J. M. and Granger, C. W. J. (1969).
The combination of forecasts.
Operational Research Quarterly, 20(4):451–468.

Clark, T. E. and McCracken, M. W. (2013).
Advances in forecasts evaluation.
In Elliot, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume Volume 2 part B, pages 1107–1197. Elsevier.

Clements, M. P. and Hendry, D. F. (1998).
Forecasting Economic Time Series.
Cambridge University Press.

Clements, M. P. and Hendry, D. F. (1999).
Forecasting Non-Stationary Economic Time Series.
MIT Press.

Diebold, F. X. and Mariano, R. S. (1995).
Comparing predictive accuracy.
Journal of Business and Economic Statistics, 13:253–263.

Elliott, G. and Timmermann, A. (2013).
Handbook of Economic Forecasting, volume 2A and 2B.
North Holland.

Giacomini, R. and White, H. (2006).
Tests of conditional predictive ability.
Econometrica, 74:1545–1578.

Granger, C. W. J. and Newbold, P. (1986).
Forecasting Economic Time Series.
Academic Press, 2 edition.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011).
The model confidence set.
Econometrica, 2(79):453–497.

Harvey, D. S., Leybourne, S. J., and Newbold, P. (1998).
Tests for forecast encompassing.
Journal of Business and Economic Statistics, 16:254–259.

Timmermann, A. (2006).
Forecast combination.
In Elliot, G., Granger, C. W. J., and Timmermann, A., editors, *Handbook of Economic Forecast*, volume Volume 1, pages 135–196. Elsevier.