Macroeconomic Forecasting

Gianluca Cubadda
Università di Roma "Tor Vergata"

15th May 2018

Recursive forecasts

Let y_t denote the series to be forecast and $y_{t+h|t} = \mathrm{E}(y_{t+h|t}|I_t)$ denote the out-of sample forecasts of y_{t+h} based on I_t, the information available up to time to t. Out-of-sample forecasts are typically computed using one of two methods:

- Recursive (*expanding window*) forecasts: An initial sample using data from $t = 1, ..., T$ is used to estimate the models, and h-step ahead out-of sample forecasts are produced. The sample is increased by one, the models are re-estimated, and h-step ahead forecasts are produced

$$[1, 2, ..., T] \rightarrow T + h$$
$$[1, 2, ..., T + 1] \rightarrow T + h + 1$$
$$...$$
$$[1, 2, ..., T + M] \rightarrow T + h + M$$

Rolling forecasts

- Rolling (*moving window*) forecasts. An initial sample using data from $t = 1, ..., T$ is used to estimate the models, and to form h-step ahead out-of-sample forecasts. Then the window is moved ahead one time period, the models are re-estimated using data from $t = 2, ..., T + 1$ and h-step ahead out-of-sample forecasts are produced. T is the window width

$$[1, 2, ..., T] \rightarrow T + h$$
$$[2, ..., T + 1] \rightarrow T + h + 1$$
$$...$$
$$[M + 1, ..., T + M] \rightarrow T + h + M$$

Forecast Evaluation Statistics

Define the forecast error as $e_{t+h|t} = y_{t+h} - y_{t+h|t}$. Common forecast evaluation statistics based on M h-step ahead forecasts are

- Mean Square Forecast Error: $MSFE = \frac{1}{M} \sum_{t=T}^{T+M-1} e_{t+h|t}^2$
- Mean Absolute Forecast Error: $MAFE = \frac{1}{M} \sum_{t=T}^{T+M-1} |e_{t+h|t}|$
- Mean Absolute Percentage Forecast Error:

  $MAPFE = \frac{1}{M} \sum_{t=T}^{T+M-1} \frac{|e_{t+h|t}|}{|y_{t+h}|}$

Forecast Evaluation Statistics

Remarks:

- For $h > 1$ the forecast errors $\{e_{t+h|t}; t = T, ..., T + M - 1\}$ are serially correlated and follow an MA$(h-1)$ process.
- A model which produces small values of the forecast evaluation statistics is judged to be a good model.
- The forecast evaluation statistics are random variables and a formal statistical procedure should be used to determine if the difference among different models are "small".

Diebold-Mariano Test for Equal Predictive Accuracy (EPA)

- Let $\{e^{(1)}_{t+h|t}\}$ and $\{e^{(2)}_{t+h|t}\}$ be the h-step forecast errors associated with two competing models, e.g. an AR(p) model vs. a VAR(p) model.
- The accuracy of each forecast is measured by a particular loss function $L(e^{(i)}_{t+h|t})$ for $i = 1, 2$. Popular choices are:

$$L(e^{(i)}_{t+h|t}) = e^2_{t+h|t}; \text{ squared error loss}$$
$$L(e^{(i)}_{t+h|t}) = |e_{t+h|t}|; \text{ absolute error loss}$$

- To determine if one model predicts better than another we may compare the set of hypotheses

$$H_0 : E[L(e^{(1)}_{t+h|t})] = E[L(e^{(2)}_{t+h|t})]$$
$$H_1 : E[L(e^{(1)}_{t+h|t})] \neq E[L(e^{(2)}_{t+h|t})]$$

Diebold-Mariano Test for EPA

- The null hypothesis of equal predictive accuracy (EPA) is then

$$H_0 : E[d_{t+h|t}] = 0$$

where $d_{t+h|t} = L(e^{(1)}_{t+h|t}) - L(e^{(2)}_{t+h|t})$ is defined as the loss differential.

- The Diebold-Mariano test statistic is

$$DM = \overline{d} \Big/ [LRV(d_{t+h|t})/M]^{1/2}$$

where

$$\overline{d} = \frac{1}{M} \sum_{t=T}^{T+M-1} d_{t+h|t}$$

$$LRV(d_{t+h|t}) = \widehat{\gamma}_0 + 2 \sum_{k=1}^{h-1} \widehat{\gamma}_k, \ \gamma_k = \mathrm{Cov}(d_{t+h+k|t+t}, d_{t+h|t})$$

Diebold-Mariano Test for EPA

- Under the null of EPA, DM is asymptotically distributed as $N(0,1)$. One sided test can also be performed.

Remarks:

- The long-run variance is used in the test statistic because the loss differentials $\{d_{t+h|t}\}$ are serially correlated for $h > 1$.
- When the competing models are nested and an expanding window is used, the limit distribution under H_0 is no more $N(0,1)$. The reason is that, as T grows, the denominator of DM goes to 0.
- However, when T remains finite and M grows, parameter estimates do not reach their probability limits and the DM test remains valid even for nested models. This is the case when a rolling window is used.

The Model Confidence Set (MCS)

- The DM test is suited to compare either two competing models or a given model vs. a benchmark model (e.g., several multivariate models vs. a univariate model). However, a forecaster may need to compare a large variety of models.
- The Model Confidence Set (MCS, Hansen *et al.*, 2011) selects a set of models that contains the best-performing model with a probability that is no less than $1 - \alpha$, with α being the size of the test.
- The MCS does not necessarily select a single model; instead the number of models in the superior set will depend on how informative are the data.

The MCS

- Define a set M_0 that contains the set of models under evaluation indexed by $i = 0, 1, ..., m_0$.
- Define the loss differential between models (i, j) as

$$d_{t+h|t}^{(i,j)} = L(e_{t+h|t}^{(i)}) - L(e_{t+h|t}^{(j)})$$

- The set of superior models is defined as

$$M^* = \left\{ i \in M_0 : \mathrm{E}(d_{t+h|t}^{(i,j)}) \leq 0, \text{ for any } j \in M_0 \right\}$$

  The MCS uses a sequential testing procedure to determine M^*
- The set of hypotheses to be compared is

$$H_{0,M} : \mathrm{E}(d_{t+h|t}^{(i,j)}) = 0 \text{ for any } i, j \in M \subset M_0$$

$$H_{1,M} : \mathrm{E}(d_{t+h|t}^{(i,j)}) > 0 \text{ for some } i, j \in M \subset M_0$$

The MCS

- When the test rejects the null hypothesis, at least one model in the set M is considered inferior and the model that contributes the most to the rejection of the null is eliminated from the set M.
- This procedure is repeated until the null is accepted and the remaining models in M equal $\widehat{M}^*_{1-\alpha}$ i.e. the $(1-\alpha)\%$ model superior set.
- Several test statistics can be used for the sequential testing of the null hypothesis, see Hansen *et al.* (2011) for details.
- Since the distributions of the test statistics depend on unknown parameters, a bootstrap procedure is used to estimate the distribution.

Factor Models

- The exact factor model is defined as

$$X_t = \Lambda f_t + \epsilon_t$$

  where X_t is a N-vector of stationary time series, f_t is a q-vector of unobserved common factors, Λ is $N \times q$ loading matrix, and ϵ_t is a N-vector of idiosyncratic (possibly autocorrelated) errors such that: (i) $\mathrm{E}(f'_{tt}\epsilon_{t-j}) = 0$; (ii) $\mathrm{E}(\epsilon'_t\epsilon_{t-j})$ is a diagonal matrix $\forall j$.
- When assumption (ii) is relaxed, the model above is defined as the approximate factor model. Additional conditions are needed to ensure that the cross-correlation among elements of ϵ_t is mild.

Factor Models

- When both N and the sample size T diverge, the factors f_t can be consistently estimated by the first q principal components of X_t. In general, the speed of convergence for this estimator is $\min\left\{\sqrt{N}, \sqrt{T}\right\}$. When $\sqrt{T}/N \to 0$, then the factor estimation error is asymptotically irrelevant (Bai and Ng, 2006).
- The diffusion index approach by Stock and Watson (2002a, 2002b) is

$$\delta_h(L) y_{t+h} = \beta_h' f_t + \varepsilon_{t+h}$$

where y_t is a stationary scalar time series, $\delta_h(L)$ is a polynomial of order p in the lag operator L, and ε_{t+h} is an innovation w.r.t the past of $[y_t, f_t']'$, and $h \geq 1$.

- Under some technical conditions, the model above can be estimated by OLS having estimated the factors f_t with the PC's of X_t. Bai and Ng (2002) offer some information criteria for the choice of q.

Other approaches

- De Mol *et al.* (2008) consider ridge regression as a forecasting method for high-dimensional time series. They show that ridge regression provides consistent forecasts as both N and T diverge. Empirically, they show that ridge regression perform equally well as PCR.
- C. and Guardabascio (2012) consider methods for forecasting macroeconomic time series in a "medium N" framework where Their interest is motivated by a body of empirical research suggesting that popular data-rich prediction methods perform best when N ranges from 20 to 40.They resort to PLS and PCR to consistently estimate a stable dynamic regression model with many predictors as T only diverges. They show that PLS compare well to other popular models in macroeconomic forecasting.
- Macroeconomic and financial forecasting is a very active research area, and new methods are continuously proposed.

References

- C., G., & Guardabascio, B. (2012). A medium-N approach to macroeconomic forecasting. *Economic Modelling*, 29, 1099–1105.
- De Mol, C., Giannone, D., Reichlin, L. (2008). Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components? *Journal of Econometrics* 146, 30–66.
- Hansen, P. R., Lunde, A., & Nason, J. (2011). The model confidence set. *Econometrica*, 79, 453–497.
- Stock, J.H., Watson, M.W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.
- Stock, J.H., Watson, M.W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economics Statistics* 20, 147–162.