f1 - 2, 5, …

f2 - 4, 7, …

f_avg - (2+4)/2, (5+7)/2, …

Economics 144

Economic Forecasting

Auto-arima may tend to overshoot - ets may be all the way down - some bias on upside
some bias on downside - to avoid bias, why not just combine the forecast
- Based on the graph, u may not need combination and one of these will be fine
- offset biases that individual forecast may have

Lecture 12

Evaluating and Combining Forecasts

Part I (Theory)

Dr. Randall R. Rojas

Today’s Class

- Introduction
- Evaluating a Single Forecast
- Testing Properties of Optimal Forecasts
- Assessing Optimality w.r.t. an Information Set
- Comparing Forecast Accuracy
- Statistical Comparison of Forecast Accuracy
- Forecast Encompassing and Forecast Combination
- R Example

Introduction

- Q: Can competing forecasts be combined to produce a composite forecast superior to all original forecasts?
- A: Yes!

→ Main reasons for combining forecasts:

- Many models or forecasts that have very similar predictive accuracy makes it difficult to identify a single best forecast.
- Diversification gains.

Portfolio - each one of these forecasts has its individual biases - combine them - reduced volatility/risk - getting comparable return
Uncertainty - bias upward/bias downward - reduce errors.
Strong parallel between portfolio theory optimization

Introduction

- **When** should we combine forecasts?
  - Individual forecasts are misspecified don't necessarily know - how do I know it's mis-specified
  - Unstable forecasting environment (e.g., past track record unreliable) highly dynamic environment - more uncertainty - errors can be very large.

    Smooth it out - smoothing out fluctuation - combining them all - smoother result.
  - Short track record; use 1-over-N weights?

y = Data

- **What** should we combine?
  - Forecasts using different information sets Don't need to limit myself to that info set - could use many others
  - Forecasts based on different modeling approaches (e.g., linear vs. nonlinear)
  - Surveys, econometric model forecasts

pulling data from different sources all together. - kalman filters - censors fusion

Introduction 3 of 6

Advantages of combining forecasts

- Dimensionality Reduction:
  - Combination reduces the information in a vector of forecasts to a single summary measure using a set of combination weights

Optimal weight - objective - best explain something - explain relation in y

allocation of weights to forecasts need to be rebalanced. - allow for weights to capture most recent dynamic/vary in time

- Optimal combination chooses weights to minimize the expected loss of the combined forecast

Real time modeling - Kalman filters - take actions fast & right. - dynamic fashion - quickly forecast/refine that forecast.

  - More accurate forecasts tend to get larger weights.
  - Combination weights also reflect correlations across forecasts.
  - Estimation error is important to combination weights

Introduction

Successful Applications of Forecast Combinations

All these methods are applicable

- Gross National Product
- Currency market volatility and exchange rates
- Inflation, interest rates, money supply
- Stock returns
- Meteorological data
- City populations
- Sports (e.g., outcomes of football games)
- Wilderness area use
- Political risks
- Estimation of GDP

Introduction 5 of 6

Two Types of Forecast Combinations

1. Data underlying the forecasts are not observed:

forecast are just based on what observations we have

not observed - high degree of uncertainty - forecast can take on anything

   – Treat individual forecasts like any other conditioning information (data) and estimate the best possible mapping from the forecasts to the outcome.

Really high type of uncertainty environment

Introduction

Two Types of Forecast Combinations

2. Data underlying the model forecasts are observed: `model combination’

not the best way to characterize it - track history of data, good handle on dynamics - evidence on model - model can be built based on track record and combine them

Using a middle step of first constructing forecasts limits the flexibility of the final forecasting model. Why not directly map the underlying data to the forecasts?

- Estimation error plays a key role in the risk of any given method. Model combination yields a risk function which, through parsimonious use of the data, could result in an attractive risk function.

This is what we’re dealing with

- Combined forecast can be viewed simply as a different estimator of the final model.

Evaluating a Single Forecast 1 of 2

- Recall the Wold representation:

Sanity check: feed residuals acf/pacf - if residuals look like white noice - consider forecast combination

$$y_t = \mu + \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \cdots$$
$$\varepsilon_t \sim WN(0, \sigma^2)$$

Take the residuals and see if we can model them

model for residuals - MA model

maybe we should combine with other models

- The *h*-step-ahead linear LS forecast:

$$y_{t+h,t} = \mu + b_h\varepsilon_t + b_{h+1}\varepsilon_{t-1} + \cdots$$

- The *h*-step-ahead forecast error

$$e_{t+h,t} = y_{t+h} - y_{t+h,t} = \varepsilon_{t+h} + b_1\varepsilon_{t+h-1} + \cdots + b_{h-1}\varepsilon_{t+1}$$

- Variance

$$\sigma_h^2 = \sigma^2\left(1 + \sum_{i=1}^{h-1} b_i^2\right)$$

Evaluating a Single Forecast 2 of 2

- Properties of Optimal Forecasts:
  - Optimal forecasts have *1*-step-ahead errors that are white noise.

    Variances can only get larger
  - Optimal forecasts have *h*-step-ahead errors that are at most MA(*h-1*).

    more distant future, greater uncertainty, represeted by variances - explain this non-decreasing as u move to the future
  - Optimal forecasts have *h*-step-ahead variances that are nondecreasing in *h* and that converge to the unconditional variance of the process.

Assessing Optimality w.r.t. an Information Set 1 of 3

- Unforecastability Principle: Optimal forecast errors should be unforecastable given the information available at the time of the forecast.

- Consider a regression of the form:

h-step erros

$$e_{t+h,t} = \alpha_0 + \sum_{i=1}^{k-1} \alpha_i x_{it} + u_t$$

Right hand side - some variable x that allows me to explain the error

where the hypothesis is that all α's are 0. Optimality therefore implies that $(\alpha_0, \alpha_1)=(0,0)$.

if alpha that are statistically significant are 0, optimality seems to hold - can't find any other data that can explain the forecast errors - whatever on right hand side, so long as they're statistically significant, you are done

- The regression of interest is therefore:

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

Actual value - see if they can be explained

Assessing Optimality w.r.t. an Information Set 2 of 3

- Mincer-Zarnowitz Regression:

$$y_{t+h} = \beta_0 + \beta_1 y_{t+h,t} + u_t$$

- In this case, optimality implies that $(\beta_0, \beta_1)=(0,1)$. Therefore, $y_{t+h} = y_{t+h,t} + u_t$

approaches the same - testing the same hypothesis - how will u test for the errors if they can't be forecasted

$$\rightarrow \; e_{t+h,t} = \alpha_0 + \alpha_1 y_{t+h,t} + u_t$$

i.e., $(\alpha_0, \alpha_1)=(0,0)$ when $(\beta_0, \beta_1)=(0,1)$

auto-arima it to test - if there's no forecastibility, output will have non of the parameters. - (0,0,0),(0,0,0) - no trend, seasonality - cycles - throw them into auto-arima

- The two approaches are the same!

Assessing Optimality w.r.t. an Information Set 3 of 3

Irrelevance Proposition

In a world with no model misspecification, infinite data samples (no estimation error) and complete access to the information sets underlying the individual forecasts, there is no need for forecast combination.

not realistic - based on this, we should just do a foreacast combineation
- combination may not get better result than individual forecast

forecast bad - worsen my individual forecast

Comparing Forecast Accuracy

- Popular Accuracy Measures
- Let $e_{t+h,\,t} = y_{t+h} - y_{t+h,t}$ and $p_{t+h,\,t} = (y_{t+h} - y_{t+h,t})/y_{t+h}$

accuracy(arima) - gives u all errors matrix in one go

- Mean Error: $\mathrm{ME} = \frac{1}{T}\sum_{t=1}^{T} e_{t+h,t}$
- Error Variance: $\mathrm{EV} = \frac{1}{T}\sum_{t=1}^{T} (e_{t+h,t} - \mathrm{ME})^2$
- Mean Squared Error: $\mathrm{MSE} = \frac{1}{T}\sum_{t=1}^{T} e_{t+h,t}^2$

Comparing Forecast Accuracy

- Mean Squared Percent Error: $\mathrm{MSPE} = \frac{1}{T}\sum_{t=1}^{T} p_{t+h,t}^2$

- Root Mean Squared Error: $\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_{t+h,t}^2}$

- Root Mean Squared Percent Error:

$$\mathrm{RMSPE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} p_{t+h,t}^2}$$

Comparing Forecast Accuracy 3 of 3

- Mean Squared Error: $\mathrm{MSE} = \mathrm{EV} + \mathrm{ME}^2$

Bias variance tradeoff - when modelling, u risk overfitting - cost u on the test error - sweet spot - turn us around - bias/variance tradeoff - reduce bias but then increase variance - used a lot in the backend of algorithm - cross-validation - optimize variable parameters - tradeoff - Reduce MSE as a whole - settled at low enough value

- Mean Absolute Error: $\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{T} |e_{t+h,t}|$

- Mean Absolute Percent Error:

$$\mathrm{MAPE} = \frac{1}{T}\sum_{t=1}^{T} |p_{t+h,t}|$$

Statistical Comparison of Forecast Accuracy

- All accuracy measures are sample estimates of population accuracy.
- Using Hypothesis Testing, we can test the Equal Accuracy Hypothesis: $E(L(e^a_{t+h,t})) = E(L(e^b_{t+h,t}))$ against the alternative that one or the other is better.

- Equivalently, we can test the hypothesis that:

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

Theoretical value - 0 - close enough to 0 - there is no difference in between them

- If d_t is covariance stationary, then:

covaraicne stationary - error close enough - differences oscillate around 0

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

follows a normal centered at 0 with the variance f

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

Forecast Encompassing and Forecast Combination 1 of 3

- Can competing forecasts be combined to produce a composite forecast superior to all original forecasts?→Yes
- Forecast Encompassing Tests: Determine whether one forecast incorporates all the relevant information in competing forecasts.
- Forecast Combination: There are two* main methods for combining forecasts:
  - Variance-Covariance done with rrors/everything else
  - Regression ($y_{t+h} = \beta_0 + \beta_1 y^a_{t+h,t} + \beta_2 y^b_{t+h,t} + \varepsilon_{t+h,t}$)

*Technically, Variance-Covariance is a special case of Regression.

Forecast Encompassing and Forecast Combination

- To find the optimal weight (ω*), we can minimize the variance of the combined forecast error:

$$\rightarrow \omega^* = \underbrace{\frac{\sigma_{bb}^2 - \sigma_{ab}}{\sigma_{bb}^2 + \sigma_{aa}^2 - 2\sigma_{ab}}}_{\text{Optimal combining weight}}$$

- You can replace σ_{ij}^2 with $\hat{\sigma}_{ij}^2 = \frac{1}{T}\sum_{t=1}^{T} e_{t+h,t}^{i} e_{t+h,t}^{j}$

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

assign weights based on varcovar matrix
- assign based on optimal weights

Forecast have degree of sensitity of dynamics

→ We are forming a portfolio of forecasts!

step forward beyond equal-weighted average

Forecast Encompassing and Forecast Combination 3 of 3

- Time-Varying Combining Weights:

What would be the proper profile of weights - have a linear equation to model the growth of these weights

$$y_{t+h} = (\beta_0^0 + \beta_0^1 \mathrm{TIME}) + (\beta_a^0 + \beta_a^1 \mathrm{TIME}) y_{t+h,t}^a + (\beta_b^0 + \beta_b^1 \mathrm{TIME}) y_{t+h,t}^b + \varepsilon_{t+h,t}$$

Set time to 0 - OLS result

values from model a

- Serial Correlation:

$$y_{t+h} = \beta_0 + \beta_a y_{t+h,t}^a + \beta_b y_{t+h,t}^b + \varepsilon_{t+h,t}$$

$$\varepsilon_{t+h,t} \sim \mathrm{ARMA}(p,q)$$

Weights can vary with time

- Nonlinear Combining Regressions:

$$y_{t+h} = \beta_0 + \beta_a y_{t+h,t}^a + \beta_b y_{t+h,t}^b + \beta_{aa} (y_{t+h,t}^a)^2 + \beta_{bb} (y_{t+h,t}^b)^2 + \beta_{ab} y_{t+h,t}^a y_{t+h,t}^b + \varepsilon_{t+h,t}$$

Vary quadratically - get out of control quickly - require a lot more work - whole multitude of different time-valuing wiegint option

Take into account all these dynamics

For Next Class

- Readings about today's class:
  Chapter 9[a], 12[b]

- Review Exercises / Problems:
  Chapter 12[b]: 1, 3, 5, 9, 11

- Readings for next class:
  Chapter 13[b], 10[a]