Università Commerciale
Luigi Bocconi

Lecture 1: Introduction and review of key concepts: Loss functions and decision theory; forecast evaluation

Prof. Massimo Guidolin

Advanced Financial Econometrics III

Winter/Spring 2020

Overview

- Generalities on forecasting
- Mean squared, mean absolute, mean percentage forecast accuracy measures
- The difference between statistical and economic loss functions
- The recovery problem
- Limitations of quadratic loss functions and location-dependent losses
- Hints to forecast evaluation issues

Generalities

There are two key approaches to forecasting over time: structural vs. time series

- Prediction or forecasting indicates an attempt to determine the values that a series is likely to take in the future
  - Forecasts might also usefully be made in a cross-sectional environment but this is more rarely seen
- Determining the forecasting accuracy of a model is an important test of its adequacy
- Some econometricians would go as far as to suggest that statistical adequacy of a model in terms of whether it contains insignificant parameters, provides a poor fit etc.
- Two approaches to forecasting:

  1 Econometric (structural) forecasting -- relates one or more dependent variable to one or more explanatory variables
  - Such models often work well in the long run, when a relationship between variables arises from no-arbitrage or market efficiency

Generalities

Forecasts can be point and interval forecasts, in- and out-of-sample

② **Time series forecasting** -- predicting the future of a series given its previous values and/or previous values of an error term

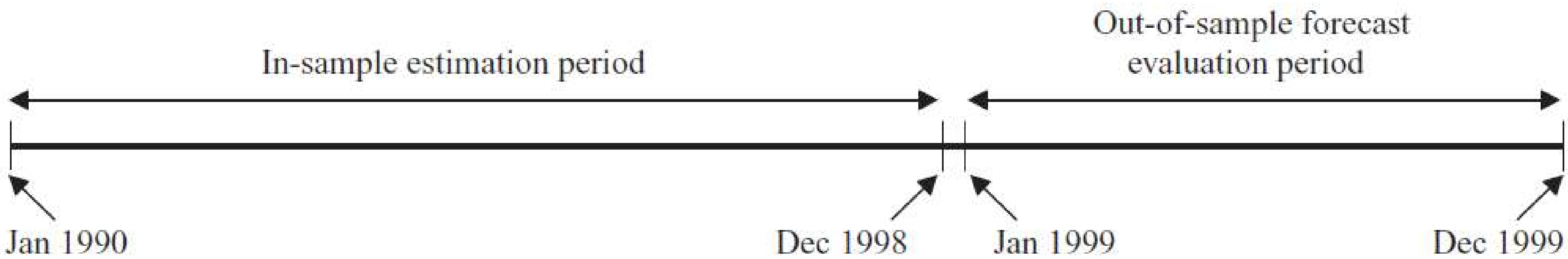
- The distinction between the two types is somewhat blurred -- for example, it is not clear where VAR models fit

- **Point forecasts** predict a single value for the variable, while **interval forecasts** provide a range of values in which the future value of the variable is expected to lie with a given level of confidence
- **In-sample forecasts** are those generated for the same set of data that was used to estimate the model's parameters
  - One would expect the 'forecasts' of a model to be relatively good in-sample, for this reason
  - Therefore, a sensible approach to model evaluation through an examination of forecast accuracy is not to use all of the observations in estimating the model parameters…
  - … but rather to hold some observations back (holdout sample)

Generalities

Forecasts can be in- and out-of-sample

- True (possibly) genuine out-of-sample (OOS) forecasts are those generated for data not used in estimation
  - This includes pseudo OOS forecasts, when estimates are not contaminated by data used in assessing the forecasts but some elements of model specification or selection may be based on the full sample
  - Typically, pseudo OOS exercises are recursive
  - OOS may be applied cross-sectionally, e.g., I estimate the ICAPM premia on N securities and see whether these price other, M securities
  - However, more commonly OOS exercises are time series in nature



  - E.g., use data from 1990M1 until 1998M12 to estimate the model parameters, and then the observations for 1999 would be forecasted from the estimated parameters

Generalities

Forecasts can be one- vs. multi-step ahead, rolling vs. expanding

- A one-step-ahead forecast is a forecast generated for the next observation only, whereas multi-step-ahead forecasts are those generated for 2, 3, . . . , H steps
  - When H is large enough, given a sample of T observation, a researcher would then lose T – H observations, because the last H-step ahead forecasts cannot be compared to any data
- A way around the problem is a **recursive** implementation
- A recursive **expanding** model is one where the initial estimation date is fixed, but additional obs. are added one at a time
- A **rolling window** is one where the length of the in-sample period to estimate is fixed

Objective: to produce	*Data used to estimate model parameters*	
1-, 2-, 3-step-ahead forecasts for:	*Rolling window*	*Recursive window*
1999M1, M2, M3	1990M1–1998M12	1990M1–1998M12
1999M2, M3, M4	1990M2–1999M1	1990M1–1999M1
1999M3, M4, M5	1990M3–1999M2	1990M1–1999M2
1999M4, M5, M6	1990M4–1999M3	1990M1–1999M3
1999M5, M6, M7	1990M5–1999M4	1990M1–1999M4
1999M6, M7, M8	1990M6–1999M5	1990M1–1999M5
1999M7, M8, M9	1990M7–1999M6	1990M1–1999M6
1999M8, M9, M10	1990M8–1999M7	1990M1–1999M7
1999M9, M10, M11	1990M9–1999M8	1990M1–1999M8
1999M10, M11, M12	1990M10–1999M9	1990M1–1999M9

Generalities

Optimal forecasts often coincide with conditional expectations

- The start date and end date successively increase by one observation

- How do we construct forecasts? There is no specific rule: as we shall see, it depends on loss functions and econometric models how "optimal forecasts" may be obtained
- One key idea in the literature: conditional expectations!

$$E[Y_{t+h}|\Omega_t] = E[Y_{t+h}|I_t] = E_t[Y_{t+h}] \text{ (when content of } \Omega_t/I_t \text{ is clear)}$$

  - Imporant note: this is true only for specific loss functions!
- Often optimal forecasts are compared to benchmarks, e.g., the random walk and uncond. mean (historical sample average)

Naive forecasting methods

(1) Assume no change so that the forecast, f, of the value of y, s steps into the future is the current value of y

$$E(y_{t+s}|\Omega_t) = y_t$$

Such a forecast would be optimal if y_t followed a random walk process.

(2) In the absence of a full model, forecasts can be generated using the long-term average of the series. Forecasts using the unconditional mean would be more useful than 'no change' forecasts for any series that is 'mean-reverting' (i.e. stationary).

Generalities

The mean squared forecast error (MSE) and mean absolute error (MAE) are the two key measures of predictive accuracy

- In practice, forecasts would usually be produced for the whole OOS period and then compared with the actual values
  - The differences between them, forecast errors, are then aggregated
  - The forecast error for observation *i* is defined as the difference between actual value for observation *i* and the forecast made for it
- Because the forecast error will be positive (negative) if the forecast was too low (high), it is not possible to simply sum the forecast errors, since the positive and negative errors will cancel out
- Before the forecast errors are aggregated, they are usually squared or the absolute value taken, which renders them all positive

$$MSE = \frac{1}{T-(T_1-1)}\sum_{t=T_1}^{T}(y_{t+s}-f_{t,s})^2 \quad MAE = \frac{1}{T-(T_1-1)}\sum_{t=T_1}^{T}|y_{t+s}-f_{t,s}|$$

- Little can be gleaned from considering MSE or MAE because they are unbounded from above

Generalities

Theil's U-statistic assess relative predictive accuracy

- Normally, MSE or MAE from one model would be compared with those of other models for the same data and forecast period
- Mean absolute percentage error (MAPE) is a relative measure:

$$MAPE = \frac{100}{T-(T_1-1)}\sum_{t=T_1}^{T}\left|\frac{y_{t+s}-f_{t,s}}{y_{t+s}}\right|$$

- Another criterion which is popular is Theil's U-statistic (1966), where $fb_{t,s}$ is the forecast obtained from a benchmark model (e.g., a simple model such as a naive or random walk)

$$U = \frac{\sqrt{\sum_{t=T_1}^{T}\left(\frac{y_{t+s}-f_{t,s}}{y_{t+s}}\right)^2}}{\sqrt{\sum_{t=T_1}^{T}\left(\frac{y_{t+s}-fb_{t,s}}{y_{t+s}}\right)^2}}$$

- A U-statistic of one implies that the model under consideration and the benchmark model are equally (in)accurate
- A value of less than one implies that the model is superior to the benchmark, and vice versa for U > 1
- However, if $fb_{t,s}$ is the same as y_{t+s} , U will be infinite

Statistical vs. Economic Loss Functions

Often statistical measures of predictive accuracy poorly correlate with economic measures, such as trading profits

- Many econometric forecasting studies evaluate the models' success using statistical loss functions
- It is not necessarily the case that models classed as accurate because they have small MSFE are useful in practical situations
  - We have many examples that the accuracy of forecasts according to traditional statistical criteria may give little guide to the potential profitability of those forecasts in a market trading strategy
  - Models that perform poorly on statistical grounds may still yield a profit if used for trading, and vice versa
- Models that can accurately forecast the sign, or can predict turning points have been found to be more profitable, i.e., to perform best in terms of economic loss functions (Leitch and Tanner, 1991, AER)
  - Two possible indicators of the ability of a model to predict direction changes irrespective of their magnitude are those suggested by Pesaran and Timmerman (1992, JBES) and by Refenes (1995, JFor)

Economic Loss Functions

The mean percentage proportions of sign predictions is a measure that often associates with economic measures

- The relevant formulae are, respectively:

$$\% \text{ correct sign predictions} = \frac{1}{T-(T_1-1)} \sum_{t=T_1}^{T} z_{t+s} \quad \text{where} \quad \begin{array}{ll} z_{t+s} = 1 & \text{if } (y_{t+s} f_{t,s}) > 0 \\ z_{t+s} = 0 & \textit{otherwise} \end{array}$$

$$\% \text{ correct direction change predictions} = \frac{1}{T-(T_1-1)} \sum_{t=T_1}^{T} z_{t+s}$$

$$\text{where} \quad \begin{array}{ll} z_{t+s} = 1 & \text{if } (y_{t+s} - y_t)(f_{t,s} - y_t) > 0 \\ z_{t+s} = 0 & \textit{otherwise} \end{array}$$

- In each case, the criteria give the proportion of correctly predicted signs and directional changes for some lead time *s*, respectively
- There exist forecasters who may be operating independently so that forecasts are produced and there are several potential users

- A decision maker will typically have a payoff or utility function U(*x*, α), which depends upon some uncertain variable or vector *x* which will be realized and observed at a future time T

Economic Loss Functions: The Set-Up

- The variable or vector α is a decision variable which must be chosen out of a set A at some earlier time $t < T$
  - For instance, $x \in X \subset R^1$ and $\alpha \in A \subset R^1$ after the forecast is learned, but before x is realized
- The decision maker can base her choice of α upon a current scalar forecast (a "point forecast") x_F of the variable x, and make the choice $\alpha(x_F) \equiv \text{argmax}_{\alpha \in A} U(x_F, \alpha)$
- Given the realized value x_R, the decision maker's ex post utility $U(x_R, \alpha(x_F))$ can be compared with the maximum possible utility they could have attained, namely $U(x_R, \alpha(x_R))$
- This shortfall can be averaged over a number of such situations, to obtain the decision maker's average loss in terms of foregone utility
- $U(x, \alpha(x))$ does not necessarily correspond to a standard statistical metrics, such as RMSFE
  - E.g., if one forecasting method has a lower (or zero) bias but higher average squared error than a second one, clients with different goals or preferences may disagree on which of the two techniques is "best"

Economic Loss Functions: The Set-Up

- A decision problem consists of the following components:

$$\begin{array}{|ll|} \hline \text{uncertain variable} & x \in \mathcal{X}, \\ \text{choice variable and choice set} & \alpha \in \mathcal{A}, \\ \text{objective function} & U(\cdot,\cdot) : \mathcal{X} \times \mathcal{A} \to \mathbf{R}^1 \\ \hline \end{array}$$

- For point forecasts, the decision maker's optimal action function $\alpha(\cdot)$ is given by:
$$\alpha(x_F) \equiv \arg\max_{\alpha \in \mathcal{A}} U(x_F, \alpha) \quad \text{all } x_F \in \mathcal{X}$$
- Assume that max $U(\cdot, \cdot)$ has interior solutions $\alpha(x_F)$, and also that it satisfies the following conditions on its second and cross-partial derivatives, which ensure that $\alpha(x_F)$ is unique and increasing in x_F:
$$U_{\alpha\alpha}(x, \alpha) < 0, \quad U_{x\alpha}(x, \alpha) > 0 \quad \text{all } x \in \mathcal{X}, \text{ all } \alpha \in \mathcal{A}$$
- The "loss" arising from a forecast x_F, when x is realized to be x_R, is the loss in utility or profit due to the imperfect prediction
  - The amount by which utility falls short of what it would have been if the decision maker had instead exactly foreseen the realized value x_R

Economic Loss Functions: The Set-Up

The "loss" arising from a forecast x_F, when x is realized to be x_R, is the loss in utility or profit due to the imperfect prediction

- We define the point-forecast/point-realization loss function induced by the decision problem as

$$L(x_R, x_F) \equiv U\big(x_R, \alpha(x_R)\big) - U\big(x_R, \alpha(x_F)\big) \quad \text{all } x_R, x_F \in \mathcal{X}$$

  - In defining the loss, realized utility is compared to what would have been if the forecast had been the realized value, and not with what utility would be if the realization had been the forecast, $U(x_F, \alpha(x_F))$
- Therefore there is no reason why $L(x_R, x_F)$ should necessarily be symmetric in x_R and x_F
- Under our assumptions, the loss function $L(x_R, x_F)$ satisfies the following properties:

$$L(x_R, x_F) \geqslant 0, \quad L(x_R, x_F)|_{x_R = x_F} = 0,$$

$$L(x_R, x_F) \text{ is increasing in } x_F \text{ for all } x_F > x_R,$$

$$L(x_R, x_F) \text{ is decreasing in } x_F \text{ for all } x_F < x_R.$$

Economic Loss Functions: The Set-Up

- These properties generalize in interesting ways to **density forecasting**, see Granger and Machina (2006)
  - Whereas a point forecast x_F conveys information on the general "location" of x, it conveys no information as to x's potential variability
  - Forecasters who seek to formally communicate their own extent of uncertainty, or alternatively, who seek to communicate their knowledge of the stochastic mechanism that generates x, would report a distribution forecast FF, e.g., the CFD over the interval X
  - A decision maker receiving a distribution forecast, and who seeks to maximize expected utility or expected profits, would have an optimal action function $\alpha(\cdot)$ and loss functions defined by:

$$\alpha(F_F) \equiv \arg\max_{\alpha \in \mathcal{A}} \int U(x, \alpha)\,\mathrm{d}F_F(x) \quad \text{all } F_F(\cdot) \text{ over } \mathcal{X}$$

$$L(x_R, F_F) \equiv U\big(x_R, \alpha(x_R)\big) - U\big(x_R, \alpha(F_F)\big) \quad \text{all } x \in \mathcal{X}, \text{ all } F_F(\cdot) \text{ over } \mathcal{X}$$

- For point forecasts, the optimal action function $\alpha(\cdot)$ satisfies the first-order conditions: $\boxed{U_\alpha\big(x, \alpha(x)\big) \underset{x}{\equiv} 0}$

The Recovery Problem

The standard, rationality properties of loss functions impose restrictions on the structure of the underlying loss functions

- In practice, loss functions are typically not derived from an underlying decision problem, but rather are postulated exogenously
- Because decision-based loss functions inherit certain necessary properties, it is worth asking precisely when a given loss function (or functional form) can or cannot be viewed as being derived from an underlying decision problem
  - When they can, it is then worth asking about the restrictions this loss function implies about the underlying utility function or constraints
  - When they cannot, one wonders of the rationality of decision-makers
- Granger and Machina (2006, JoE) demonstrated that for an arbitrary loss function $L(\cdot, \cdot)$, the class of objective functions that generate $L(\cdot, \cdot)$ has the following specification:

$$U(x, \alpha) \equiv f(x) - L\big(x, g(\alpha)\big)$$

for some function $f(\cdot): X \rightarrow R^1$ and monotonic function $g(\cdot): A \rightarrow X$

The Recovery Problem

The relationship between decision makers' loss functions and their underlying decision problems is tight but far from unique

- This theorem states that an objective function U(x,α) and choice space A are consistent with the loss $L(x_R, x_F)$ if and only if they can be obtained from the function $-L(x_R, x_F)$ by one or both of the two types of transformations (called "inessential")
- These consist of either transforming the action monotonically or of adding some quantity f(x) that does not depend on the optimal action

- Therefore **the relationship between decision makers' loss functions and their underlying decision problems is tight but far from unique**
- The most frequently used loss function in statistics is unquestio-nably the squared-error form:
$$L_{\mathrm{Sq}}(x_R, x_F) \equiv k \cdot (x_R - x_F)^2, \quad k > 0$$
which satisfies standard properties
- It must be that for an arbitrary squared-error function with $k > 0$, an objective function U(·, ·): $X \times A \rightarrow R^1$ with strictly monotonic optimal action α(·) will generate $L_{Sq}(\cdot, \cdot)$ as its loss iff:

The Recovery Problem: the Odd Case of Ms. Quadratic

The utility functions recoverable from standard quadratic loss functions are rather unusual and have odd properties

$$U(x, \alpha) \equiv f(x) - k \cdot \left(x - g(\alpha)\right)^2$$

for some function f (·): $X \rightarrow R^1$ and monotonic function g(·): $A \rightarrow X$

- Utility functions of this form are not particularly standard!
- One property is that changes in the level of the choice variable α do not affect the curvature (i.e. the second and higher order derivatives) of U(x, α) with respect to x, but only lead to uniform changes in the level and slope with respect to x
  - For any pair of values $\alpha_1, \alpha_2 \in A$, the difference $U(x, \alpha_1) - U(x, \alpha_2)$ is an affine function of x:

$$U(x, \alpha_1) - U(x, \alpha_2) \equiv -k \cdot [g(\alpha_1)^2 - g(\alpha_2)^2] + 2 \cdot k \cdot [g(\alpha_1) - g(\alpha_2)] \cdot x$$

- Another disturbing property is revealed by the canonical form of the utility function that leads to a squared loss:

$$\hat{U}(x, x_F) \equiv f(x) - L_{\mathrm{Sq}}(x, x_F) \equiv f(x) - k \cdot (x - x_F)^2$$

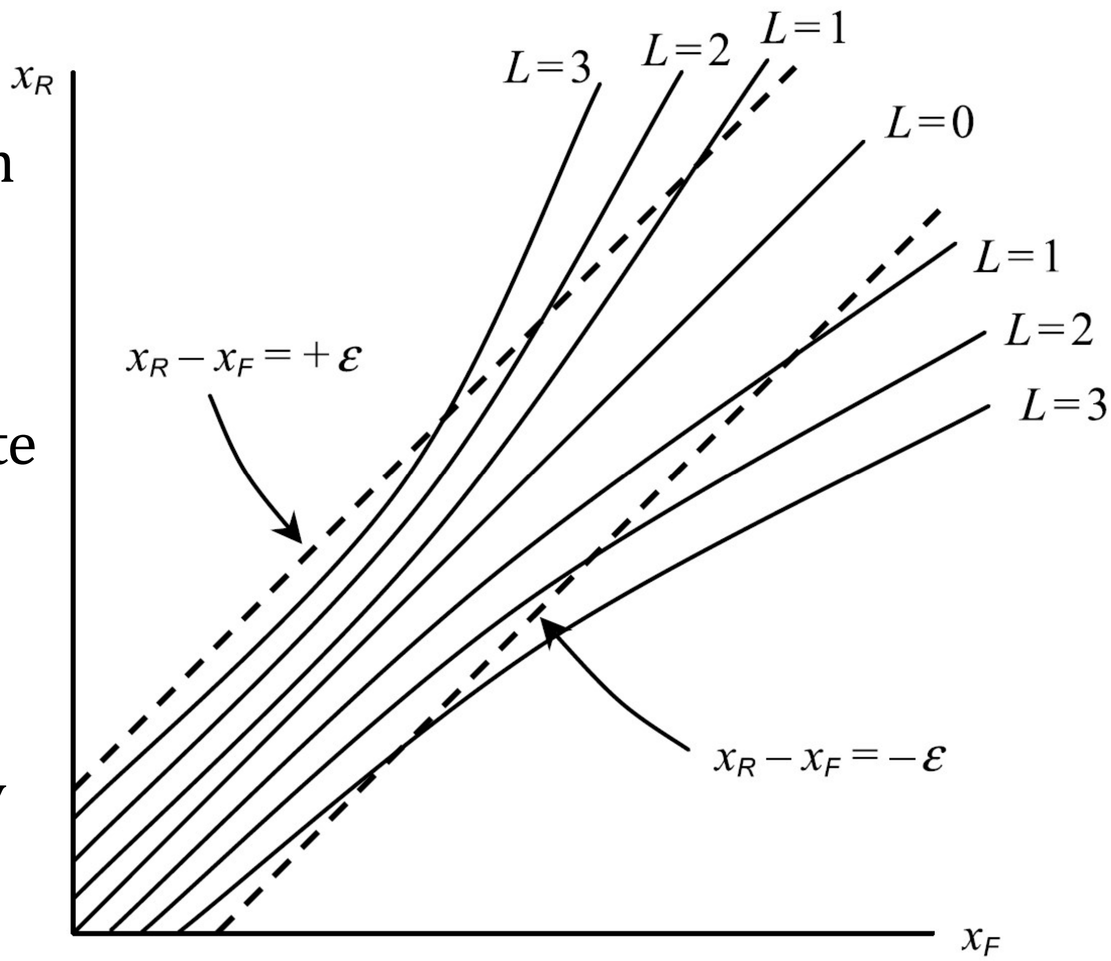
The Recovery Problem: the Odd Case of Ms. Quadratic

- E.g., this implies that when a firm faces a realized output price of x, its shortfall from optimal profits due to having planned for an output price of x_F only depends upon the difference between x and x_F (their square), and not upon how high or how low the two values are
- Thus, the profit shortfall from having underpredicted a realized output price of $10 by one dollar is the same as the profit shortfall from having underpredicted a realized price of $2 by one dollar
- This is clearly unrealistic in any decision problem which exhibits "wealth effects" or "location effects" in the uncertain variable
- E.g., as a firm could make money if the realized output price was $7, but would want to shut down if the realized output price was only $4 (in which case there would be no profit loss at all from having underpredicted the price by $1)

- One argument for the squared-error form is that if the forecast errors x_R - x_F are not too big then this functional form is the natural second-order approximation to any smooth loss function that exhibits the necessary properties stated earlier
- However, this argument is fallacious

The Recovery Problem: the Odd Case of Ms. Quadratic

- The figure shows the level curves of some smooth loss $L(x_R, x_F)$, along with the region where $|x_R - x_F| \leq$ some small ε, a constant-width band about the 45° line
- This region does not constitute a small neighborhood in R^2, even as $\varepsilon \rightarrow 0$
- The 2nd order approximation to $L(x_R, x_F)$ when x_R and x_F are both small and approximately equal to each other is not the same as the 2nd order approximation to $L(x_R, x_F)$ when x_R and x_F are both large



- Legitimate 2nd approx. can only be taken in over small neighborhoods of points in R^2, and not over bands (even narrow bands about the 45∘ line
- The "quadratic approximation" $L_{Sq}(x_R, x_F) \equiv k\,(x_R - x_F)^2$ over such bands is not justified by Taylor's theorem