Problem Set 2

Due Wednesday January 22, 4pm

Data Exercises

(1) An excel data file realgdpgrowth.xlsx is posted on canvas. It contains quarterly observations from 1947q2 to 2024q3 for most categories of U.S. GDP, calculated as quarterly percentage changes at annual rates in real GDP and its component categories. To answer the following questions, it should be sufficient to use the "import data wizard" and the mean/std/quantile commands (e.g. `quantile(gdp, [0.2 0.7])`).

(a) The series *pce_nondurables* is personal consumption expenditures on nondurable goods. What was the percentage change (annual rate) in 2024q3?

(b) Assuming growth rates are independent across quarters, so using no other information other than from the basic summary commands (such as `mean/std/quantile/etc.`), what is our point forecast for future values of *pce_nondurables*?

$Z_{\alpha/2}$,

(c) Using the normal rule, give an 90% forecast interval.

(d) Using empirical percentiles, give an 90% forecast interval.

(2) Repeat question 1 for the series *pce_durables* (personal consumption expenditures on durable goods).

(a) What was the percentage change (annual rate) in 2024q3?

(b) What is our point forecast for future values of *pce_durables*?

(c) Using the normal rule, give an 90% forecast interval.

(d) Using empirical percentiles, give an 90% forecast interval.

(e) What are the important differences between your answers to questions 1 and 2. Why are there such large differences for these two different components of personal consumption?

Theoretical Questions

(3) Suppose you are an intern at a company. For a meeting, you are responsible to bring photocopies of the agenda (a one-page sheet, cost is $0.05 each) for everyone attending

the meeting. You are told that it is important that everyone has a copy. You do not know exactly how many people will be at the meeting so you are uncertain exactly how many copies to print. You need to forecast the number of people attending. Describe the loss problem. Is the loss function symmetric?

(4) Imagine a similar situation, but instead you are told to bring copies of Hamilton's *Time Series Analysis*, one per person, which you need to purchase and the cost is $150 per copy. You are told that it would be okay if participants shared copies. Describe the loss problem. Is the loss function similar to question (3), or how is it different?

(5) *In this exercise you will verify that the median is the best point forecast under absolute loss.* Suppose that Y is a random variable with the cdf[1] $F(y) = y^2$ and pdf[2] $f(y) = 2y$ both on the interval $[0, 1]$.

   (a) Calculate the expectation of $|Y - c|$,

$$\mathbb{E}|Y - c| \equiv \int_0^1 |y - c| f(y) dy,$$

   where c is a constant.

   (b) Find c which minimizes $\mathbb{E}|Y - c|$.

   (c) Find the median of F.

   (d) What is the optimal (best) point forecast for Y under absolute loss? How is it related to your answers to (b) and (c)?

(6) Suppose that X is distributed uniformly on $[0, 1]$, i.e., $Prob(X \leq c) = c$ for $c \in [0, 1]$.

   (a) Find the point optimal forecast of X given the loss function

$$L(e) = \begin{cases} 3e & e \geq 0 \\ -e & e < 0 \end{cases}, \quad e = X - \hat{X}.$$

   (b) Compare $L(e)$ and your result in (a) with the absolute loss and the point optimal forecast under the absolute loss. Discuss.

---

[1] Cumulative distribution function

[2] Probability density function

(5) *In this exercise you will verify that the median is the best point forecast under absolute loss.* Suppose that Y is a random variable with the cdf[1] $F(y) = y^2$ and pdf[2] $f(y) = 2y$ both on the interval $[0, 1]$.

prob$(Y \leq 1) = 1^2 = 1$,

0.5,

(a) Calculate the expectation of $|Y - c|$,

$$\mathbb{E}|Y - c| \equiv \int_0^1 |y - c| f(y) dy,$$

where c is a constant.

(b) Find c which minimizes $\mathbb{E}|Y - c|$.

(c) Find the median of F.

(d) What is the optimal (best) point forecast for Y under absolute loss? How is it related to your answers to (b) and (c)?

$F(y) = P(Y \leq y)$, $y \in [0,1]$.

1, (a) $E|Y-c| \equiv \int_0^1 |y-c| \cdot f(y) \cdot dy = \int_0^c (c-y) \cdot f(y) \cdot dy + \int_c^1 (y-c) \cdot f(y) \cdot dy$,

$= \int_0^c (c-y) \cdot 2y \cdot dy + \int_c^1 (y-c) \cdot 2y \cdot dy = \int_0^c (2cy - 2y^2) \cdot dy + \int_c^1 (2y^2 - 2yc) \cdot dy$

$= \left[cy^2 - \frac{2}{3}y^3\right]_0^c + \left[\frac{2}{3}y^3 - cy^2\right]_c^1 = \left[(c^3 - \frac{2}{3}c^3) - 0\right] + \left[(\frac{2}{3} - c) - (\frac{2}{3}c^3 - c^3)\right]$

$= \frac{1}{3}c^3 + \left[\frac{2}{3} - c + \frac{1}{3}c^3\right] = \frac{2}{3}c^3 - c + \frac{2}{3}$

(b), we do F.O.C, and check S.O.C,

F.O.C, $\frac{\partial E|Y-c|}{\partial c} = (\frac{2}{3}c^3 - c + \frac{2}{3})' = (2c^2 - 1) = 0$, $c^* = \sqrt{\frac{1}{2}} = \frac{\sqrt{2}}{2}$

S.O.C. $\frac{\partial^2 E|Y-c|}{\partial c^2} = (2c^2 - 1)' = 4c \geq 0$ since $c \in [0,1]$.

Thus, the optimal c^* we solve is the minimizer

(c), Let c be the median of F. Hence, c satisfied $F(c) = 0.5$,

$F(c) = Pr(Y \leq c) = c^2 = 0.5 \Rightarrow c = \sqrt{0.5} = \frac{\sqrt{2}}{2}$.

(d), Under absolute loss function, the optimal point forecast is the point c that minimize $E|Y-c|$, which is equal to c^* we solved in part (b), $c = \frac{\sqrt{2}}{2}$, that c is the median of Y solved in (c). Thus, we know the median is the optimal point forecast under absolute loss.

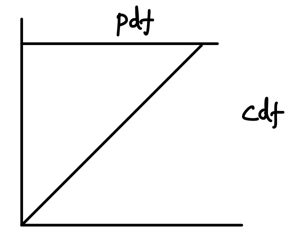
(6) Suppose that X is distributed uniformly on $[0,1]$, i.e., $Prob(X \le c) = c$ for $c \in [0,1]$.

(a) Find the point optimal forecast of X given the loss function

$$L(e) = \begin{cases} 3e & e \ge 0 \\ -e & e < 0 \end{cases}, \quad e = X - \hat{X}. \qquad L(X-\hat{X}) = \begin{cases} 3(X-\hat{X}) & X-\hat{X} \ge 0 \\ -(X-\hat{X}) & X-\hat{X} < 0 \end{cases}$$

$X < \hat{X}$

(b) Compare $L(e)$ and your result in (a) with the absolute loss and the point optimal forecast under the absolute loss. Discuss.



$F(x) = x \ , \ f(x) = 1$

(a). $\hat{X}$ is the forecast point. we want to find $\hat{X}^*$ under $L(e)$.

That $\hat{X}^*$ will minimizing $E[L(e)]$

$$E[L(e)] = \int_0^1 L(e)\cdot f(x)\cdot dx = \overset{X-\hat{X}<0\ (e<0)}{\int_0^{\hat{X}} -(X-\hat{X})\cdot f(X-\hat{X})\cdot dx} + \overset{X-\hat{X}>0\ (e>0)}{\int_{\hat{X}}^1 3(X-\hat{X})\cdot f(X-\hat{X})\cdot dx}$$

$$= \int_0^{\hat{X}} -(X-\hat{X})\cdot dx + \int_{\hat{X}}^1 3(X-\hat{X})\cdot dx = \left[-\left(\frac{X^2}{2} - \hat{X}\cdot X\right)\right]_0^{\hat{X}} + \left[3\left(\frac{X^2}{2} - \hat{X}\cdot X\right)\right]_{\hat{X}}^1$$

$$= \left[-\frac{\hat{X}^2}{2} + \hat{X}^2 - 0\right] + \left[\left(\frac{3}{2} - 3\hat{X}\right) - \left(\frac{3}{2}\hat{X}^2 - 3\hat{X}^2\right)\right] = \frac{1}{2}\cdot\hat{X}^2 + \frac{3}{2} - 3\hat{X} + \frac{3}{2}\hat{X}^2$$

$= 2\hat{X}^2 - 3\hat{X} + \frac{3}{2}$. Then we use F.O.C to find optimal $\hat{X}$.

$\frac{\partial E[L(e)]}{\partial \hat{X}} = 4\hat{X} - 3 = 0 \Rightarrow \hat{X}^* = \frac{3}{4}$ S.O.C $= 4 > 0$, So $\hat{X}^*$ is the minimizer,

The optimal point forecast is $\frac{3}{4}$. $L(X) = \begin{cases} 3(X-\frac{3}{4}) & X \ge \frac{3}{4} \\ -(X-\frac{3}{4}) & X \le \frac{3}{4} \end{cases}$

(b). Under the absolute loss function, the optimal point forecast is equal to the median of the distribution, which is $\hat{X} = \frac{1}{2}$.

Comparison: The optimal point forecast under $L(e)$ is higher than under absolute loss function. that's because the loss function is not symmetric. The $L(e)$ punish a lot more for underestimation than overestimation. To minimize expected loss, we prefer to have a higher point forecast.