24.04.2022

Model.

Cross Sectional. — Indep. of time

$y = f(x_1, x_2, \ldots)$

Longitudinal — Dep. on time

$x_t = f(x_{t-1}, x_{t-2}, \ldots e_{t-1} \ldots)$

Panel. — Combination of Cross sectional & Longitudinal.

Model.

Validation — 84 months

2015 April. – 2022 Mar. } observed data points (known)

70% of 84 = 60 months → Training

24 months for Testing.

Training & Testing.

Training	Testing
70%	30%.
80%.	20%.
65%.	35%.
60%.	40%.

Fit a Arima. model on 60 months Data.

1 ... 61 62 ... 84

1 — x_1
2 — x_2
3 — x_3
4 — x_4 Known
5 — ~~x_5~~
6 — x_6
7 — x_7

8 — x_8
9 — x_9
10 — x_{10}

Testing

$x_t = 0.3 + 0.5\, x_{t-1}$ Training

$x_1 = 0.3 + 0.5\, x_0$

$x_2 = 0.3 + 0.5\, x_1$

$x_8 = 0.3 + 0.5\, x_7$

validation

Bootstraping

validation

Machine Learning

1) Recurrent validation

2) K. fold validation

3) LOOCV [Leave one out Cross validation]

Data analysis

K-Fold

Fold

K=5

K=10

n = 100

Fold 1 — 20

2 — 20

3 — 20

4 — 20

5 — 20

Error — Training Step 1 → Fold 1 + Fold 2 + Fold 3 + Fold 4 — 80%.

Testing → Fold 5.

Error — Step 2 — Fold 2 + Fold 3 + Fold 4 + Fold 5 = 80%

Testing → Fold 1.

Error — Step 3 — F1 + F3 + F4 + F5 — 25 Tol.

F2

Error — Step 4 — F1 + F2 + F4 + F5

F3

Step 5 — F1 + F2 + F3 + F5.

F4

5. Fold

x = 123.

123/5. 24.6

Fold 1	24
2	24
3	24
4	24

5. 27

y = observed value.

$\hat{y}$ = Model.

Error = $(y - \hat{y})$.

$\sum (\text{Error})^2 = \sum (y - \hat{y})^2$.

Known y	x_1	x_2	$\hat{y}$	$(y-\hat{y})^2$
2	1	2	✓	—
4	3	3	✓	—
6	5	9	✓	—
8	7	15	✓	—
10	9	20	✓	—

$\hat{y}$ = Const + $b_1 x_1 + b_2 x_2$

$= 2 + 0.5 \times x_1 + 0.3 x_2$

$(y-\hat{y})^2$

$\Sigma (y-\hat{y})^2$

$$MSE = \frac{\Sigma (y-\hat{y})^2}{n}$$

RMSE = Root mean sum of square error

$= \sqrt{MSE}$

MSE & RMSE for Training & Testing.

Over fitting, under fitting

under fitting, fitting, over fitting

LOOCV

Goodness fit.

Cross sectional.

R^2, $\uparrow$ Adj R^2?

Model 1 . .

Badness fit

AIC, BIC, SIC.

Akaike information criteria.

AIC in less.

Models.

LOO CV. Leave one out cross validation

~~RMSE vs MSE~~.

60 months → ROR of 60 months.

ARoR.

60 months → 5 yr → AROR.

$\bar{x} = \,..$

1.

2

3

4

5

$\bar{x}_1 =$

$\bar{x}_2 =$

$\bar{x}_3 =$

$\bar{x}_4 =$

$\bar{x}_5$

$\bar{x} =$ based on 5 obs.

LOOCV

$$\bar{x} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4 + \bar{x}_5}{5}.$$

Average of average error to minimise..

ANOVA. [vv Imp]

Analysis of variance.

Avg. Prediction Error. = Avg. Reducible error + Avg. Irreducible error.

Measurement error → Control.

Inherent. → Not in control.

Obj of Data Analysis: Data Sample

✓ Anova:

One way Anova

1) Between

mpe

	M	T	W	T2	F
Machine 1	P_1	P_2	P_3	P_4	P_5
Machine 2	P_6	P_7	P_8	P_9	P_{10}
Machine 3	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}

2) within

Contrast

Mixed index.

Production = f(M/c)

Two way Anova

machine \ Oper.	01	02	03
M1			
M2	Production		
M3			

Production = f(M/c) → f(op)

✓ Production = f(M/c, op, M/c × op)

Two way Anova.

84 Months

7 years.

Sector — Banking

36 Bank.

21) Public

11) Private

↳ 4 Public

↳ 4 Private

	y_1	y_2	..	y_7
Bank 1	0	0	.	0
Bank 2	0	0	.	
Bank 3	0	0		0
Bank 4.	0	0	0	0

If no of Banks > 2 ⇒ ANOVA.
if the obs. follow normal dist.

non-parametric
Kruskal Wallis Test

no of Banks = 2 ⇒ t-Test ( )
⇒ Mann Whitney U-Test

obj: All Bank are equally efficient
if not rank the bank.
using ANOVA.

Apply Anova

$H_0: \mu_{B_1} = \mu_{B_2} = \mu_{B_3} = \mu_{B_4}$

H_1: At least two banks are not equally efficient

If H_0 is accepted ... then
Statistically mathematically there is a difference of their sample mean.

Ho is rejected. then we can easily find their rank using paired - t - Test. (Tukey).

Source.	Sum of Square.	DF.	Mean Sum of Square.	F.	p-value.
Between					< 0.05.
within.					
Total.					

✓ Ho is rejected

To rank the banks or app's

Post Hoc. Test

Gr 1.	Gr. 2	Gr 3

$\bar{x}_1 < \bar{x}_2 < \bar{x}_3$

Gr - 3
- Rank 1.

Gr - 2
- Rank 2

Gr. 1 - Rank 3.

Post Hoc

↓

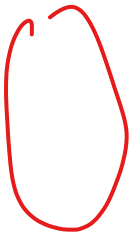
Inter homogeneous comparison.

Gr 1 - Gr 2.

Gr 1 - Gr 3

Gr 2 = Gr 3

Hd' . .



6x1.

6x2

6x3.

Dimension

~ 2 inch. .