ĐẠI HỌC KINH TẾ TÀI CHÍNH

FINAL COURSE REPORT

Title: Outstanding Home Loan Balance Analytics

Group: 08
Class: A01E
Lecturer supervisor: MSC. Ngô Thuận Dủ

Ho Chi Minh city, 30th October 2025

Đoàn Nguyên Trí

Group members

- Lê Trương Minh Quân - 25210786
- Trần Minh Quân - 25210187
- Nguyễn Hồng Phúc - 225053466
- Nguyễn Minh Nhật - 225210135
- Đoàn Nguyên Trí - 225210468

Đoàn Nguyên Trí

Reason for choosing the topic

- The housing and credit markets have become highly volatile after the pandemic.
- Forecasting home loan balances is essential for financial planning and credit risk management.
- Traditional models capture trends but fail to reflect nonlinear patterns in real-world data.
- Combining SARIMA and XGBoost enhances accuracy by merging statistical and machine learning strengths.

Research objectives

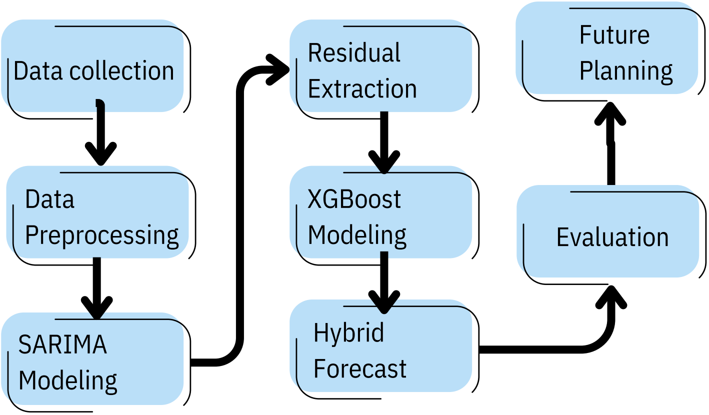
- To develop a SARIMA–XGBoost hybrid model for more accurate forecasting of home loan balances.

Analytical tool:

- R on Google colab

Đoàn Nguyên Trí

Implementation process



Pipeline chart

Đoàn Nguyên Trí

Checking and summarizing missing data

```
sum(is.na(DL))
```

```
13
```

```
percentage_na = (sum(is.na(DL)) / length(DL)) * 100
percentage_na
```

```
13.5416666666667
```

```
miss_index <- which(is.na(DL))
miss_index
```

```
5 · 11 · 16 · 26 · 36 · 43 · 51 · 59 · 66 · 72 · 80 · 87 · 93
```

```
data= read_xlsx('/content/mis.xlsx')
data
```

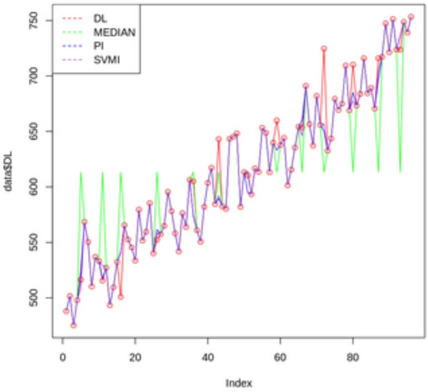
```
print(DL)
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2017	488.15	501.66	475.30	497.93	NA	568.52	550.47	510.29	536.97	533.11	NA	527.22
2018	493.42	509.55	532.45	NA	565.55	552.60	545.45	533.59	579.56	551.68	559.55	585.33
2019	540.21	NA	557.49	564.92	595.51	578.22	558.14	541.98	576.30	563.91	606.32	NA
2020	561.02	550.66	582.08	603.61	617.22	584.54	NA	582.62	580.28	643.44	645.24	648.23
2021	582.09	613.31	NA	593.50	616.64	613.60	653.26	648.77	613.34	639.88	NA	637.65
2022	644.12	601.43	615.54	635.51	654.13	NA	691.10	656.60	637.28	682.03	655.87	NA
2023	632.61	643.68	679.35	669.49	674.95	709.38	669.10	NA	673.27	683.81	715.89	684.49
2024	689.12	670.63	NA	717.14	747.37	721.54	751.26	723.81	NA	748.64	739.20	753.31

Nguyễn Minh Nhật

Comparing missing methods

Method	MAE	RMSE	MAPE
PI	3,371	11,74	0,529
SVMI	3,224	11,239	0,511
MEDIAN	9,548	29,673	1,556



Choose PI, SVMI

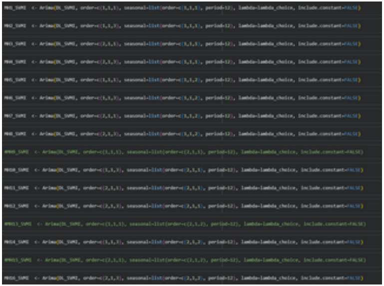
Nguyễn Minh Nhật

SVMI Method

Seasonal Data:
P=1,2; Q=1,2; D=1

Non-seasonal Data:
p=1,2; q=1,3; d=1

16 SARIMA Models

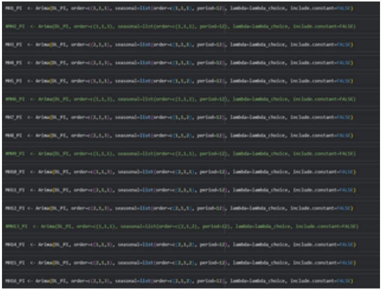


PI Method

Seasonal Data:
P=1,2; Q=1,2; D=1

Non-seasonal Data:
p=1,2; q=1,3; d=1

16 SARIMA Models



Nguyễn Hồng Phúc

SVMI Method

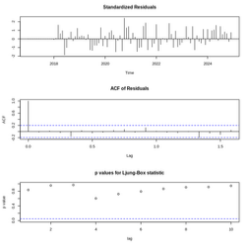
Model	AIC	BIC
MH1_SVMI(1,1,1)(1,1,1)	931.6148	943.7090
MH5_SVMI(1,1,1)(1,1,2)	931.5828	946.0959

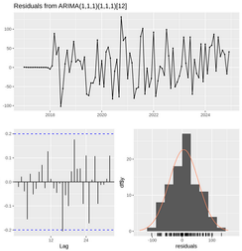
PI Method

Model	AIC	BIC
MH1_PI(1,1,1)(1,1,1)	827.9442	840.0384
MH8_PI (2,1,3)(1,1,2)	826.5767	848.3463

Nguyễn Minh Nhật

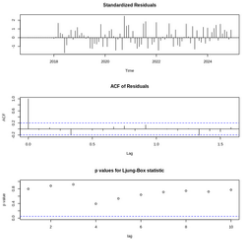
Compare models

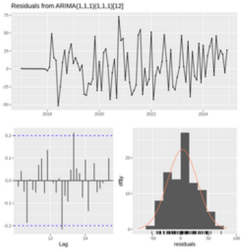





```
    Box-Ljung test

data:  MH1_SVMI$residuals
X-squared = 16.802, df = 20, p-value = 0.6658
    Ljung-Box test

data:  Residuals from ARIMA(1,1,1)(1,1,1)[12]
Q* = 12.95, df = 15, p-value = 0.6062

Model df: 4.   Total lags used: 19
```

→ MH1_SVMI (1,1,1)(1,1,1)



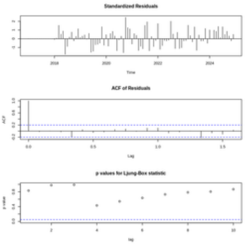


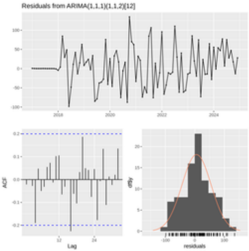

```
    Box-Ljung test

data:  MH1_PI$residuals
X-squared = 21.91, df = 20, p-value = 0.3454
    Ljung-Box test

data:  Residuals from ARIMA(1,1,1)(1,1,1)[12]
Q* = 16.316, df = 15, p-value = 0.3614

Model df: 4.   Total lags used: 19
```

→ MH1_PI (1,1,1)(1,1,1)

Nguyễn Minh Nhật

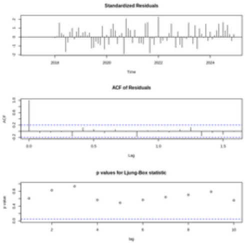
Compare models

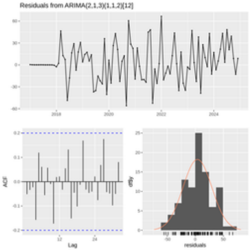





```
    Box-Ljung test

data:  MH5_SVMI$residuals
X-squared = 20.612, df = 20, p-value = 0.4202
     Ljung-Box test

data:  Residuals from ARIMA(1,1,1)(1,1,2)[12]
Q* = 16.274, df = 14, p-value = 0.2969

Model df: 5.   Total lags used: 19
```

→ MH5_SVMI (1,1,1)(1,1,2)





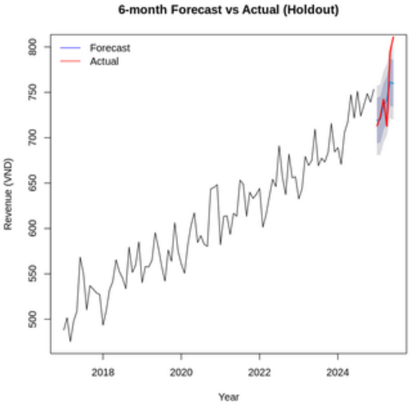

```
    Box-Ljung test

data:  MH8_PI$residuals
X-squared = 19.114, df = 20, p-value = 0.5144
```

```
    Ljung-Box test

data:  Residuals from ARIMA(2,1,3)(1,1,2)[12]
Q* = 15.57, df = 11, p-value = 0.1579

Model df: 8.   Total lags used: 19
```

→ MH8_PI (2,1,3)(1,1,2)

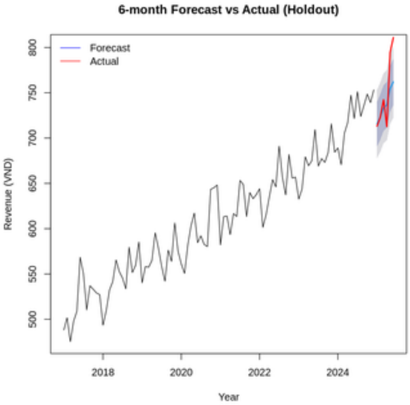
Nguyễn Hồng Phúc

Compare SARIMA forecasting SVMI models



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2025	719.1929	694.0636	744.1776	680.7000	757.3473
Feb 2025	720.1515	694.6621	745.4922	681.1063	758.8488
Mar 2025	734.4071	708.7777	759.8891	695.1482	773.3210
Apr 2025	741.9919	716.1900	767.6459	702.4690	781.1688
May 2025	761.2071	735.3024	786.9664	721.5282	800.5460
Jun 2025	759.8553	733.7305	785.8321	719.8386	799.5257

MH1_SVMI

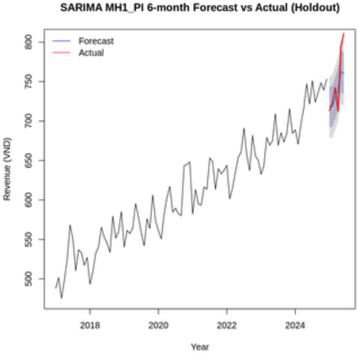


	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2025	715.6463	690.9719	740.1805	677.8511	753.1134
Feb 2025	723.4060	698.2602	748.4077	684.8882	761.5867
Mar 2025	733.2605	707.9481	758.4288	694.4880	771.6960
Apr 2025	736.6418	711.1464	761.9918	697.5888	775.3546
May 2025	754.6338	729.0436	780.0811	715.4367	793.4963
Jun 2025	762.0549	736.3102	787.6563	722.6215	801.1529

MH5_SVMI

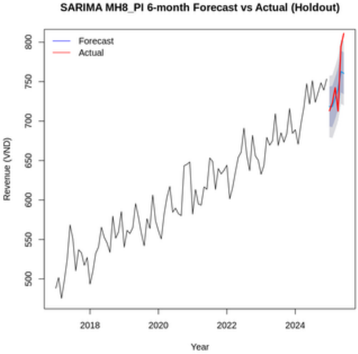
Nguyễn Hồng Phúc

Compare SARIMA forecasting PI models



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2025	718.4615	692.9181	743.9429	679.3702	757.4077
Feb 2025	718.4872	692.4430	744.4670	678.6290	758.1946
Mar 2025	729.2402	703.0318	755.3843	689.1310	769.1990
Apr 2025	740.8438	714.4957	767.1279	700.5211	781.0168
May 2025	762.7809	736.3212	789.1780	722.2880	803.1272
Jun 2025	760.6103	733.9813	787.1757	719.8581	801.2137

MH1_PI



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2025	727.2865	702.8681	751.6489	689.9184	764.5237
Feb 2025	735.2674	710.6280	759.8503	697.5612	772.8417
Mar 2025	744.8428	720.1157	769.5138	707.0026	782.5518
Apr 2025	746.9785	722.1915	771.7094	709.0465	784.7791
May 2025	761.8305	736.9962	786.6095	723.8266	799.7051
Jun 2025	764.1643	739.1618	789.1110	725.9030	802.2950

MH8_PI

Nguyễn Hồng Phúc

Compare Accuracy SARIMA

MAE	RMSE	MAPE
21.807316	28.002643	2.825773

MH1_SVMI

MAE	RMSE	MAPE
20.824139	27.874747	2.678033

MH5_SVMI

MAE	RMSE	MAPE
22.244434	27.607365	2.887656

MH1_PI

MAE	RMSE	MAPE
23.780373	28.303641	3.118151

MH8_PI

Đoàn Nguyên Trí

Compare models with XGBoots

	Month	SARIMA	Residual	Hybrid
1	2025-01-01	719.1929	21.71065	740.9036
2	2025-02-01	720.1515	38.98716	759.1386
3	2025-03-01	734.4071	23.98032	758.3874
4	2025-04-01	741.9919	-68.07921	673.9127
5	2025-05-01	761.2071	16.83514	778.0422
6	2025-06-01	759.8553	-37.52914	722.3262

MH1_SVMI

	Month	SARIMA	Residual	Hybrid
1	2025-01-01	715.6463	-10.984674	704.6616
2	2025-02-01	723.4060	40.631748	764.0377
3	2025-03-01	733.2605	-3.613548	729.6469
4	2025-04-01	736.6418	15.269107	751.9109
5	2025-05-01	754.6338	-31.765160	722.8687
6	2025-06-01	762.0549	25.578274	787.6332

MH5_SVMI

	Month	SARIMA	Residual	Hybrid
1	2025-01-01	718.4615	29.563969	748.0254
2	2025-02-01	718.4872	18.647486	737.1347
3	2025-03-01	729.2402	-5.819032	723.4212
4	2025-04-01	740.8438	41.536366	782.3801
5	2025-05-01	762.7809	30.992546	793.7735
6	2025-06-01	760.6103	38.280087	798.8904

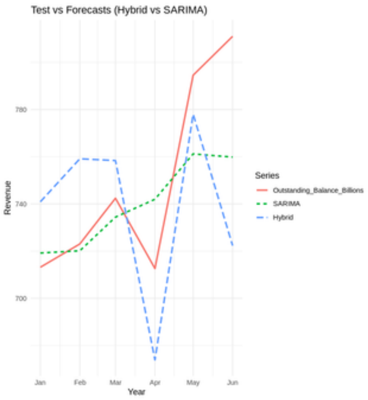
MH1_PI

	Month	SARIMA	Residual	Hybrid
1	2025-01-01	727.2865	-0.2424077	727.0441
2	2025-02-01	735.2674	13.7868004	749.0542
3	2025-03-01	744.8428	-19.2334175	725.6093
4	2025-04-01	746.9785	-21.2168427	725.7617
5	2025-05-01	761.8305	-17.5402946	744.2902
6	2025-06-01	764.1643	12.9422321	777.1066

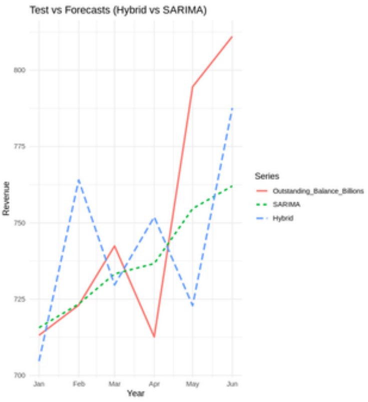
MH8_PI

Đoàn Nguyên Trí

Forecasting results (Hybrid SVMI models)



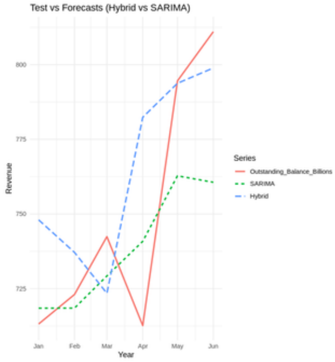
MH1_SVMI



MH5_SVMI

L.T.Minh Quân

Forecasting results (Hybrid PI models)



MH1_PI

MH8_PI

L.T.Minh Quân

Compare Accuracy Hybrid

MAE	RMSE	MAPE
33.200000	40.189311	4.322374

MH1_SVMI: shows the lowest errors (MAE, RMSE, MAPE) and strong short-term accuracy.

MAE	RMSE	MAPE
38.280754	40.106397	5.104445

MH5_SVMI: has slightly higher errors but captures the overall trend and seasonal movement more accurately.

MAE	RMSE	MAPE
53.466953	63.114203	7.377818

MH1_PI: performs less effectively with larger deviations.

MAE	RMSE	MAPE
44.765517	48.011387	6.091414

MH8_PI: deviation is smaller than MH1_PI but error values are large

→Although MH1_SVMI has lower errors, MH5_SVMI fits the real trend better selected for SARIMA–XGBoost hybrid model.

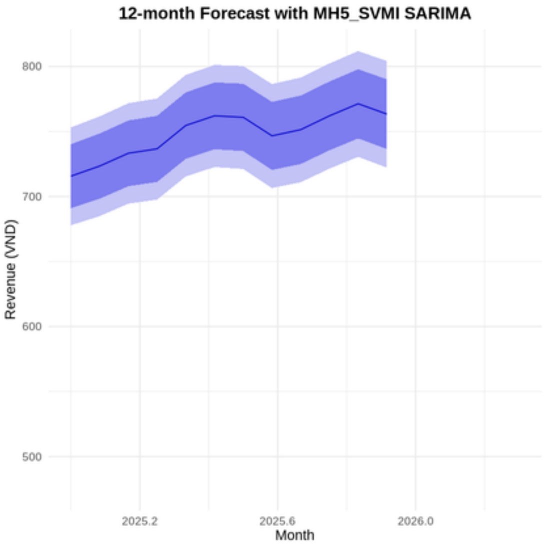
Trần Minh Quân

Forecasting for the next 12 months

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2025	715.6463	690.9719	740.1805	677.8511	753.1134
Feb 2025	723.4060	698.2602	748.4077	684.8882	761.5867
Mar 2025	733.2605	707.9481	758.4288	694.4880	771.6960
Apr 2025	736.6418	711.1464	761.9918	697.5888	775.3546
May 2025	754.6338	729.0436	780.0811	715.4367	793.4963
Jun 2025	762.0549	736.3102	787.6563	722.6215	801.1529
Jul 2025	760.9032	734.9566	786.7039	721.1600	800.3052
Aug 2025	746.6418	720.4176	772.7142	706.4714	786.4569
Sep 2025	751.3810	724.9921	777.6173	710.9583	791.4464
Oct 2025	762.0180	735.5008	788.3832	721.3994	802.2809
Nov 2025	771.3321	744.6799	797.8326	730.5071	811.8020
Dec 2025	763.3637	736.4742	790.0971	722.1740	804.1881

L.T.Minh Quân

Forecasting for the next 12 months



L.T.Minh Quân

Conclusion

- This study successfully applied both statistical and hybrid approaches to forecast outstanding home loan balances.
- Among the tested SARIMA models, MH5_SVMI (1,1,1)(1,1,2) demonstrated the best trend-following performance, stable residuals, and reasonable accuracy.
- Although MH1_SVMI showed slightly lower errors, MH5_SVMI provided a better representation of real economic trends, making it the optimal baseline for hybrid modeling.
- The upcoming SARIMA–XGBoost integration aims to further improve predictive accuracy by combining the seasonal structure of SARIMA with the nonlinear learning power of XGBoost.

➔Overall, the research highlights that hybrid time series models can effectively enhance financial forecasting, supporting credit risk management and policy planning in the housing loan sector.

Trần Minh Quân

Strategic Recommendations

January–March 2025: Maintain current lending policies while closely monitoring inflation and interest rates to prevent early-year credit distortions.

April–June 2025: Gradually increase credit availability to meet rising demand, prioritizing low-risk borrowers and fixed-rate loan products.

July–November 2025: Strengthen liquidity reserves and tighten post-disbursement monitoring to manage potential repayment risks during this expansion phase.

December 2025: Conduct portfolio risk assessments and adjust loan pricing strategies to maintain long-term sustainability.

→ **Overall:** The strategy emphasizes a balance between credit expansion and risk control, ensuring financial stability and sustainable loan portfolio growth.

Trần Minh Quân

REFERENCES

https://www.appsilon.com/post/r-time-series-forecasting by Dario Darecic, July 2, 2024

https://www.r-bloggers.com/2021/03/time-series-forecasting-with-xgboost-and-feature-importance/ , posted on March 2, 2021 by Selcuk Disci in R bloggers

https://www.appsilon.com/post/imputation-in-r by Dario Darecic, January 10, 2023

https://www.rdocumentation.org/packages/astsa/versions/2.2/topics/sarima posted by rdocumentation.org

Little, R. J. A., & Rubin, D. B., Statistical Analysis with Missing Data, 3rd ed., Wiley, 2019.

A. Quarteroni, R. Sacco, and F. Saleri, Numerical Mathematics, 2nd ed., Springer, 2007.

T. Wu, J. Y. Chung, and J. S. Lee, “Constructing support vector machines with missing data,” WIREs Data Mining and Knowledge Discovery, vol. 8, no. 3, 2018.

M. C. P. de Souto, I. G. Costa, D. A. de Araujo, and D. M. B. Couto, “Impact of missing data imputation methods on gene expression clustering and classification,” BMC Bioinformatics, vol. 16, 2015.

E. T. Emmanuel, “A survey on missing data in machine learning,” Journal of Big Data, vol. 8, no. 1, 2021.

Trần Minh Quân

Thank You!

Special thanks to our lecturer for the continuous support and guidance, and to our classmates for your attention and collaboration.

Group 8 – Final Course Project, UEF Data Science.