Default dependence in Thailand's housing and automobile loan portfolios: A copula approach

Supanut Wanchai

November 2023

1 Introduction

According to Bank of Thailand (Banking Sector Quarterly Brief (Q3 2023))[2], the banking system's loans contracted 0.9% YoY, and Non-performing loan (NPL) increased to 2.7%. In the same time, Thailand's household debt remained at 90.7% which is very high. This makes a concern if Thai economy slows down, NPL will increase.

According to KKP Bank's analyst meeting Q3 2023[1], I saw an increasing trend in NPL in its loan portfolio. If NPL increases, the bank's profit will decrease. Moreover, housing and automobile loans are top 2 largest proportion in its loan portfolio. They account for more than half of its portfolio. This inspired me to study default dependence of housing and automobile loans in Thailand which will help the bank to manage credit risk.

In this paper, I present the numerical results that show what copula works best in modeling dependence between housing and automobile loan defaults in the case of Thailand. Also, I want to promote copulas as alternative models to a linear model that is usually used.

2 Literature review

Traditional credit risk models employed the linear correlation as a measure of default dependence. A major assumption of the models is that the return from the loans are normally distributed. However, there are studies showed that asset returns are not normally distributed. Thus, copulas were adopt in credit risk analyses. According to Crook and Moreira (2011)[4] which was claimed as the first empirical study of copulas for consumer loans, copulas, such as Galambos, could be used to model the dependence between some segments of a large UK bank's credit card portfolio.

A study by Chan and Kroese (2010)[3] found that Gaussian copula did not well performed detecting loan default dependence during period of high financial distress. Fenech, Vosgha, and Shafik (2015)[5] employed Archimedean copulas to data of post global financial crisis period. They found that Archimedean copulas is better approach to model loan default correlations than Gaussian approaches.

In this study, I conducted empirical study of 3 Archimedean copulas, Clayton, Frank and Gumbel copulas, and a normal copula to housing and automobile loan of Thailand's banking system. Also, I performed model selection by information criteria.

3 Methodology

3.1 Data

I retrieved the data set "FI_NP_003_S2 Gross NPLs Outstanding Classified by Business Sector_ISIC Rev.4.0 1/" from Bank of Thailand website. The data set contains quarterly NPL outstanding and %NPL from Quarter1 2013 to Quarter2 2023. I selected only %NPL of housing and automobile loans. Hence, there are 2 variables and 42 observations to be modeled.

3.2 Jarque-Bera (JB) test

To check if the marginal distribution of each variable, I employ JB test. The null hypothesis is that the sample data follows normal distribution. I used 5% critical value so if the p-value is less than 5%, I reject the null hypothesis and conclude that the data is not normally distributed.

3.3 Pseudo-observation

Since input of copula must be in the range from 0 to 1, I compute pseudo-observation for data that is not fit normal distribution using the following definition.

Given n realizations $x_i = (x_{i1}, ..., x_{id})^T, i \in \{1, ..., n\}$ of a random vector X, the pseudo-observation are defined via $u_{ij} = r_{ij}/(n+1)$ for $i \in \{1, ..., n\}$ and $j \in \{1, ..., d\}$, where r_{ij} denotes the rank of x_{ij} among all $x_{kj}, k \in \{1, ..., n\}$

3.4 Copula

I fitted 4 bi-variate copulas using copula package in R. The first copula is normal copula. The other three copulas are Clayton, Frank and Gumbel copulas, which are Achimedean copulas.

A bivariate copula C is a function mapping the unit square [0, 1] × [0, 1] to the unit interval [0, 1], which is nondecreasing and right-continuous, and satisfies:

1. $\lim_{u\to 0} C(u, v) = 0, \quad \lim_{v\to 0} C(u, v) = 0$
2. $\lim_{u\to 1} C(u, v) = v, \quad \lim_{v\to 1} C(u, v) = u$
3. $C(u_1, v_1) - C(u_1, v_2) - C(u_2, v_1) + C(u_2, v_2) \geq 0, \quad \forall u_1 \geq u_2, \ v_1 \geq v_2$ (Supermodular)

Based on Sklar's theorem, for all $(x_1, x_2) \in R^2$ and joint distribution function $F_X \in R_2(F_1, F_2)$, there exists a unique copula C such that $F_X(x_1, x_2) = C(F_1(x_1), F_2(x_2))$. Thus, a copula could be interpreted as joint distribution function. In addition, each copula has different features. They suit for different dependence structure as per Table 1.

Copula	Dependence structure
Normal	Symmetric dependence without tail dependence
Clayton	Left (lower) tail dependence
Frank	Symmetric dependence without tail dependence
Gumbel	Right (upper) tail dependence

Table 1: Candidate copulas and their respective features

3.4.1 Normal copula

$C_\alpha(u_1, u_2) = \frac{1}{2\pi\sqrt{1-\alpha^2}} \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} exp(\frac{-(v_1^2 - 2\alpha v_1 v_2 + v_2^2)}{2(1-\alpha^2)}) dv_1 dv_2, \ \alpha \in (-1, 1)$
where Φ is the distribution function of $N(0, 1)$

3.4.2 Clayton copula

$C_\alpha(u_1, u_2) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}}, \ \alpha > 0$

3.4.3 Frank copula

$C_\alpha(u_1, u_2) = -\frac{1}{\alpha} ln(1 + \frac{(e^{-\alpha u_1} - 1)(e^{-\alpha u_2} - 1)}{e^{-\alpha} - 1}), \ \alpha \neq 0$

3.4.4 Gumbel copula

$C_\alpha(u_1, u_2) = e^{-((-ln u_1)^\alpha + (-ln u_2)^\alpha)^{\frac{1}{\alpha}}}, \ \alpha \in [1\infty)$

3.5 Model selection

Finally, I used Akaike information criterion (AIC) and Bayesian information criterion (BIC) to select the best fit copula. Generally, the best model has the lowest values of AIC and BIC.

3.6 Linear regression

Due to the nature of default, people tend to default automobile loan before housing loan. Thus, I will fit a linear regression model with housing loan NPL as a dependent variable and automobile loan NPL as a predictor. Then, I will compare it with the best copula, again using AIC and BIC.

4 Preliminary results and discussions

I tested if the variables follow normal distribution by Jarque-Bera test. According to figure 1: Histogram, I expected that housing is normally distributed but automobile is not.

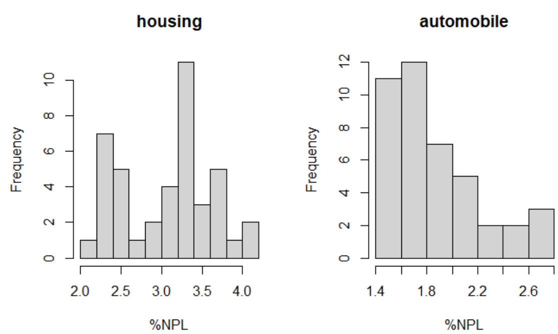

Figure 1: Histogram

Variable	Jarque-Bera statistic	p-value
housing	2.5858	0.2745
automobile	6.6873	0.03531

Table 2: Jarque-Bera test

Table 2 showed that housing is normally distributed because p-value is larger than 5%. On the other hand, automobile is not normally distributed. As a result, I computed cumulative probability of housing loan default by normal distribution function with it mean (3.077) and standard deviation (0.548). And, I transformed automobile loan default to range 0 to 1 by using Pseudo-observation.

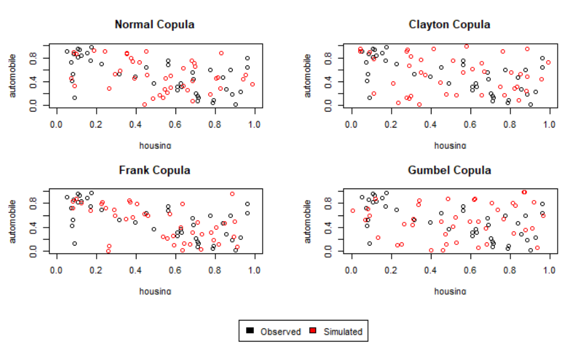

Figure 2: Copula simulation

Family	α	AIC	BIC
Normal	-0.5542	-10.45036	-8.712686
Clayton	-0.1861	-2.914093	-1.176423
Frank	-4.552	-16.40813	-14.67046
Gumbel	1	2.000001	3.73767

Table 3: Copula coefficient and information criteria

The simulation of copulas can be found in figure 2. It can be observed that Frank and Normal copula fit better than the others. From table 3, Frank copula had both the lowest AIC and BIC so I concluded that it is the best model to represent dependence between Thailand's housing and automobile loans.

Coefficients	Estimate	p-value
Intercept	4.8344	1.07e-15
Automobile	-0.9444	2.85e-05
AIC	55.1099	
BIC	60.314	

Table 4: Linear regression result

According to table 4, all coefficients were significant. However, Linear regression's AIC and BIC were much greater than those of all the copulas implemented in this study. Thus, copulas are good alternatives to the traditional linear regression models.

5 Conclusion

Again, I concluded that Frank copula is the best model to represent dependence between Thailand's housing and automobile loans. However, this study showed that Archimedean copulas is not always performed better than the Gaussian copula. In this experiment, it can be observed than normal copula had lower AIC and BIC than Clayton and Gumbel copulas. Finally, the simulation could be implied the similar result as AIC and BIC that Frank copula performed best. Moreover, the result supported that copulas are great alternatives to linear model since the linear regression had much higher AIC and BIC.

One of my motivation is to help the bank mange credit risk but I experimented on country level data. Thus, data is a limitation to achieve insight to help the bank. To achieve this goal, customer level data of the bank is required so that I will have more accurate models. Then, the approriate policy could be produced.

Moreover, there are a few areas that I would like to explore. First, I wanted to implement Goodness-of-fit test. This is an alternative to AIC and BIC that help in model selection process. Second, it would be better to fit a distribution to variables that is not normally distributed than to compute pseudo-observation. Finally, there are additional copulas such as Student t, Galambos, that could be tested.

References

[1] Analyst meeting 3q 2023. https://kkp.listedcompany.com/misc/PRESN/20231025-kkp-am3q2023.pdf. Accessed: 2023-11-24.

[2] Banking sector quarterly brief (q3 2023). https://www.bot.or.th/en/news-and-media/news/news-20231120.html. Accessed: 2023-11-24.

[3] Joshua C.C. Chan and Dirk P. Kroese. Efficient estimation of large portfolio loss probabilities in t-copula models. *European Journal of Operational Research*, 205(2):361–367, 2010.

[4] Jonathan Crook and Fernando Moreira. Checking for asymmetric default dependence in a credit card portfolio: A copula approach. *Journal of Empirical Finance*, 18(4):728–742, 2011.

[5] Jean Pierre Fenech, Hamed Vosgha, and Salwa Shafik. Loan default correlation using an archimedean copula approach: A case for recalibration. *Economic Modelling*, 47:340–354, 2015.