Bank of China (Hong Kong) Ltd

Enhancement of the Liquidity Risk Framework Project

Behavioral Modeling Methodology Report

October 2012

Table of Contents

1. Executive Summary

1.1. Document Objective

The Hong Kong Monetary Authority ("HKMA") issued the final version of Supervisory Policy Manual ("SPM") module LM-2 "Sound Systems and Controls for Liquidity Risk Management" ("LM-2") in April 2011.

LM-2 mainly follows the "Principles for Sound Liquidity Risk Management and Supervision" ("Sound Principles") issued by the Basel Committee on Banking Supervision ("BCBS") in August 2008 and also incorporates some requirements from the paper "Basel III: International framework for liquidity risk measurement, standards and monitoring" ("Basel III") issued by the BCBS in December 2010 .

To enhance the liquidity risk management framework in order to comply with the LM-2 requirements, Bank of China (Hong Kong) Limited ("BOCHK" or "the Bank") initiated the "Enhancement of the liquidity risk management framework Project" ("the LM-2 Project"). The overall goals of the LM-2 Project are to enhance the existing liquidity risk management framework to comply with the LM-2 requirements by Aug 2012 and to meet the Bank's internal management requirements, especially in the following areas:

- Cash flow projection

  Develop and implement cash flow projection methodologies, by taking into account estimated cash flows due to both contractual terms and customer behaviors.

- Stress testing

  Define liquidity stress scenarios and embed their impacts into the cash flow projection and thereby propose early warning indicators ("EWI") in the contingent funding plan ("CFP").

- Liquidity costs/premiums allocation

  Establish the liquidity costs and premiums allocation policy and incorporate the policy into the Bank's existing fund transfer pricing ("FTP") mechanism.

This report, which constitutes the deliverable "1.2 Behavioral Modeling Methodology Report" stipulated in the Statement of Work ("SOW"), is set to introduce the theoretical and technical aspects, including the rationale and assumptions behind, of the behavioral models that are commonly implemented in practice, and discuss the advantages and disadvantages of those models.

The models and approaches discussed in this report focus on behavioral models applicable for liquidity risk management purpose, and were not designed to alignment with application in other areas such as interest rate risk management.

1.2. Scope of implementation

The implementation scope covers BOCHK (solo) and two subsidiaries of the Bank, namely, Nayang Commercial Bank (“NCB”) and Chiyu Banking Corporation (“CYB”). Generally, the methodology, approaches and assumptions in this document universally apply to all the three entities in scope, unless clarified specifically.

1.3. Structure of the Report

The following part of this report is divided into two sections. Section 2 is a general introduction to the customer behavior models with the definition, approaches, and limitations of those models. Section 3 elaborates the detailed methodology, data requirement of the behavior models that are commonly implemented in practice by product type, i.e., savings deposits (including current accounts), term deposits, mortgages, and commitments. In this section, we also analyzed the advantages and disadvantages of those models and provide data requirement for implementing the models.

2. Introduction to Behavior Models

2.1. Behavior Model Definition

Customer behavior modeling is one of the important methods to help the management in the banks to forecast the future positions on balance sheet and off-balance-sheet. It is one of the components of the liquidity risk management. The purpose of building behavior models is to predict whether customers will exercise the options embedded in the financial products on or off balance sheet, which primarily include the following:

- Runoff and sedimentation: the customers of savings and current accounts can withdraw their money any time as they want thus causing the deposit run-off. But still, some of the deposits will become “sticky” and stay with the banks for a long time, which is the so-called core deposit.
- Roll over and early withdrawal: customers could early withdraw the term deposits before they mature or choose to continuing to deposit (roll over) when they reach maturity.
- Prepayment: the customers of loans, in particular the mortgages, probably will advance the repayment by partially or fully repaying their debt ahead of the contract schedule prior to the maturity.
- Contingent withdrawal: contingent withdrawal is usually associated with off-balance-sheet products, such as the commitment, letter of credit, overdraft and etc. The customers could make unexpected withdrawal as long as the authorized limit is not reached causing cash outflows for the banks.

By analyzing the expected exercise of the embedded options, inputs to cash flow forecast are forecasted. For example, in the term deposits, the behavior models usually will analyze the roll over rate and (early) withdrawal rate which will be used to adjust the contractual maturity of the deposits in cash flow projection to reflect the effect of the embedded options that the customers have. In addition, some models will go further to explore the effect of factors, such as market fluctuation, characteristics of customers, status of macro economy and etc on the cash flow characteristics of the financial products and build relationships between the behavior of customers (whether to exercise the options) and the driving factors based on the historical data. Details on the observation windows of historical data used in behavioral modeling can be referred to the Appendix IX of the Cash Flow Model Implementation Report.

2.2. Basic Approach

Generally speaking, behavior models are more suitable to retail customers. The reason is that retail customers are more diversified and are driven by certain common factors. In contrast, the corporate customers are highly concentrated and thus the behavior of a single customer, whose motives exhibit quite arbitrary features, may dominate the fluctuation of the whole portfolio. Hence it is difficult to find the common factors and the performance of the behavior models may not be satisfactory.

Broadly, there are basically two approaches in market practice:

- Statistic analysis approach: this approach analyses the historical data of the balance of the products and utilize some statistics (such as the percentile) to predict the balance in future. Statistic analysis approach requires comparatively less data and the calculation of the descriptive statistics is relatively simple. Nevertheless, it is based on the assumption that "history will reoccur" and does not explore the driving factors behind, hence if there is a significant change to the economic environment or characteristics of the customers, it may not be appropriate.

- Statistic modeling approach: this approach predicts the future behavior by linking it to variables such as macro economic factors, customer specific information, trading account information and etc with regression models. Hence, this modeling approach is more complicated and data requirement are higher. However, it can find the dominant driving factors to the customer behaviors inherent in the products and forecast the future trend.

We presented and elaborated the following different models from the above approaches in this report, including:

- Volatility Analysis;
- Time series regression Analysis;
- Replicating Portfolio;
- Roll-over and Early Withdrawal Analysis;
- OTS Prepayment Model;
- Linear Regression Model;
- Logistic Regression Model;
- Static Utilization Ratio Analysis;

- Life Cycle Utilization Rate Analysis;
- Peaks and Valleys Analysis and etc.

The detailed methodology of the above models is discussed and the pros and cons, as well as the data requirement are illustrated in Section 3.

2.3. Model Limitation

In implementing behavior models for liquidity risk and balance sheet management, the following limitations inherent in behavior models should be noted:

- Behavior models are based on historical data, implying that history will happen in future. This is a shortcoming for all models. As a result, stress testing is good supplement.
- Some behavior models take into account the growth rate using dynamic data. But some use static data by fixing the account or customer. This feature should be understood in implementing behavior models.
- The performance of behavior models depends on data quality and sufficiency. Sufficient and high quality data is a pre-condition for sound behavior modeling.
- To apply the results of customer behavior models to liquidity risk management, other management measures, for instance the financial plans and customer relationships, need to be integrated to forecast the whole picture of the risk.

2.4. Model Granularity

Product	Products Granularity	Customer Segment Granularity	Currency Granularity
Non-defined Maturity Deposit	▶ Saving Account Deposits ▶ Current Account Deposits ▶ Call Deposits and 1-Day Fixed deposits ▶ Other Saving Deposits	▶ Personal [1] ▶ Corporate [2]	▶ HKD ▶ RMB ▶ USD
	▶ Vostro Account	▶ All	
Fixed deposit	▶ Fixed deposits at different tenor (E.g. 1M, 3M, 6M, 9M, 1Y, 2Y)	▶ Personal ▶ Corporate	

[1] Personal customer segment includes personal customer, personal wealth management customers, SSE

[2] Corporate customer segment includes SME, Corporates, Banks and Non-Bank FI

Mortgage	▶ Private and Commercial Property Mortgages ▶ Government Housing Scheme Mortgages	▶ n/a	▶ HKD
Commitment Off-balance sheet	▶ Undrawn Lending Commitment	▶ Personal ▶ Corporate	▶ HKD
	▶ Trade-related Commitment	▶ Corporate	

3. Behavior Model Methodology

3.1. Saving Deposits

Saving deposit is one of the most important funding sources for the Bank. However, as the saving deposits can be withdrawn by the customers at any time without notice to the Bank, the Bank faces challenges in its liquidity management, duration matching and internal transfer pricing ("FTP"). To deal with the problem, behavior models can be used to analyze the expected cash flows resulting from saving deposits so that the results can be applied in the above mentioned areas.

The most commonly used approach for saving deposits is to estimate the core ratio and non-core ratio of saving deposits. In the approach, the saving deposits are divided into two portions: the core portion and non-core portion. The core portion by definition is categorized as long-term and stable funds, while the non-core portion is categorized as short-term funding source. For example, the Royal Bank of Scotland ("RBS") performs volatility analysis on saving deposits to estimate the core ratio defined as the ratio of core saving deposits to the total saving deposits and then amortize the core deposits (core ratio is 95%) over five years where the amortized cash flows are used in the cash flow projection and FTP management.

Commonly used behavioral models for saving deposits and current accounts include:

- Volatility Analysis
- Time series regression Analysis
- Replicating Portfolio.

In the following sections, we will discuss the methodology of the above three models in details.

3.1.1. *Volatility Analysis*

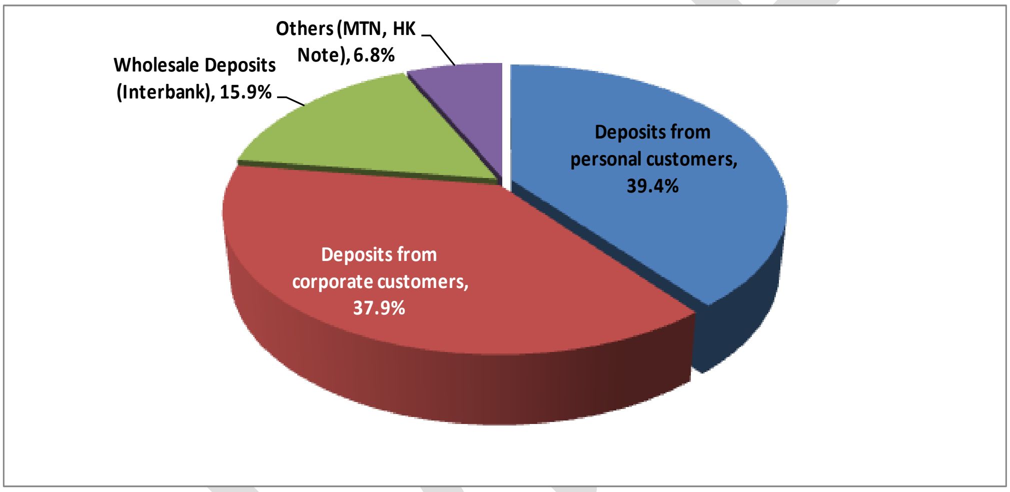
3.1.1.1. Model Methodology

Volatility analysis is widely used by banks in the estimation of core ratio of saving deposits. Core deposits are categorized as long-term and stable funds, while non-core deposits are categorized as short-term funding source. The core deposits will be allocated the longer time bucket such as equal or greater than 2 years while the non-core deposits will be amortized over a shorter period based on the volatility analysis results.

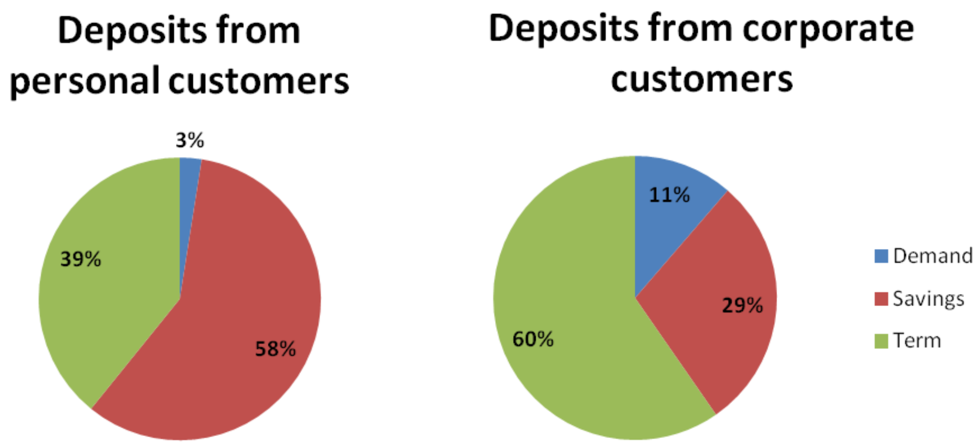
Volatility analysis is usually performed at portfolio level. Volatility is generally larger at account level than at portfolio level as the fluctuations of individual accounts may offset each other when aggregated. Liquidity risk management is conducted at portfolio level which means that the focus is on the balance of the whole portfolio rather than individual accounts. Therefore, volatility analysis is usually performed on saving and demand deposits at portfolio level. For data requirement, historical balance and basic information (region, product and customer segment) is needed for the analysis. Most of the Banks, when applying this methodology, would fix the number of accounts and consider only the negative change in balance to filter out the growth trend of the deposit balance, which could be applied as the proxy for the run-off rate of the portfolio.

Concentration analysis should also be considered before performing volatility analysis. The accuracy of forecast and granularity of behavioral model in liquidity risk management relies on the concentration level of the whole portfolio.

We performed analysis of the Bank's funding mix during the project phase to determine the concentration of the Bank's funding. As shown in the below figure, more than 77% of the funding needs for BOCHK come from the deposit base, of which personal and corporate make up 39.4% and 37.9% respectively. The wholesale unsecured funding (interbank) deposits make up about 16% of the funding source. The rest of the funding comes in the form of Medium Term Notes (MTN) and HK notes, which sums to approximately, 6.8%.



As shown in below figure, personal deposits accounted for 39.4% of the deposits base of which savings account make up the lion share with 58.3% followed by 39.2% of term deposits and the rest in demand deposits. For corporate deposit which is 37.9% of the total deposits base, 59.7% is term deposits followed by 29.0% and 11.3% of savings and demand deposits, respectively.



It can be concluded that the concentration of the Bank’s saving and demand deposit portfolio can be generally defined as deposits from personal and corporate customers.

As sometimes a small proportion of the customers take up a large proportion of saving deposits, volatility analysis on account level and portfolio level should be carried on at the same time by segment such as personal, SSE, SME and corporate. Details of the behavioral model granularity can be referred to the cash flow model implementation report Appendix IX. The segment is not only based on their product natures but also their features of liquidity. As such, it usually can be done from both quantitative and qualitative perspectives. For quantitative analysis, products with the similar volatility trend in rough volatility analysis can be categorized into the same portfolio. For qualitative analysis, the characteristics of the products are analysed based on the product analysis results. Characteristics of the Bank’s products are studied and analyzed in details to determine the model granularity. Details of the product analysis can be referred to the Appendix II of the cash flow implementation report. It is worthy to point out that this segment approach also applies to other behavior models in the subsequent sections.

For a simple illustration, suppose we used the sample the daily historical data of the total balance for a portfolio of saving accounts of BOCHK constructed based on a pre-determined granularity setting from 1 Jan 2006 to 1 Jan 2008. The composition accounts are fixed, which means that no new accounts come into the portfolio. The data are shown as following:

Date	Aggregate Balance (million)
1/1/2006	1,000,000
1/2/2006	850,000
1/3/2006	900,000
1/4/2006	875,000
1/5/2006	1,020,000
1/6/2006	950,000

1/7/2006	950,000
…	…
12/31/2006	975,000
1/1/2007	1,100,000
…	…
1/1/2008	830,000

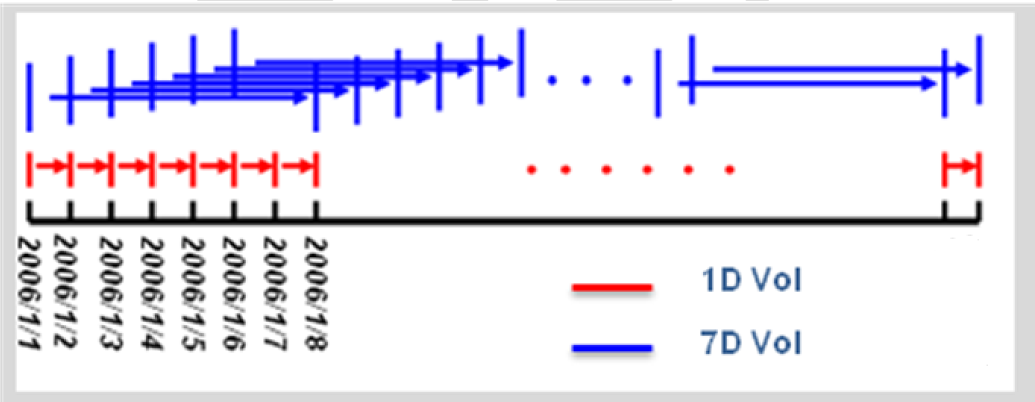
Standing at 1 Jan 2006, we can calculate the 1-day percentage change (which is generally referred to as “1-day volatility”) using the total balances as of 1 Jan 2006 and 2 Jan 2006:

(850,000-1000,000)/1000,000=-15%

The 7-day volatility, 1-month volatility, 3-month volatility and volatilities of longer tenors can be calculated with the balances as of 7 Jan 2010, 1 Feb 2010, 1 Apr 2010, and the corresponding date analogously. Then we have one observation of the volatility curve:

Observation	1-Day	7-Day	1-Month	3-Month	6-Month	1-Year
1	-15%	-5%	-30%	-32%	-50%	10%

Moving the standing date from 1 Jan 2006 forward till 1 Jan 2007 and following the calculation at 1 Jan 2006 shown above, we will get 365 observations for each volatility tenor.



The positive values of observations of each volatility tenor are filtered and the negative percentage changes are ordered separately, and percentiles can be calculated as shown in the following table:

Percentile	1-Day	7-Day	1-Month	3-Month	6-Month	1-Year
…	…	…	…	…	…	…
50%	-0.02%	-0.08%	-3.88%	-12.72%	-28.91%	-22.34%
…	…	…	…	…	…	…
90%	-15.76%	-40.03%	-41.49%	-39.20%	-48.56%	-47.29%
95%	-30.70%	-53.47%	-55.09%	-48.91%	-60.80%	-52.36%
99%	-52.08%	-66.03%	-71.82%	-69.43%	-76.78%	-72.61%

The results above can be interpreted as the distribution of run-off rate of the bank's deposit portfolio. In some Banks, the minimum percentage change (or run-off rate) within 1-Year at 95th percentile are used as the benchmark for the determination of the composition of core and non-core portion of the portfolio. In the above example, it is noted that the run-off rate at 1-Year 95th percentile is 53.36%. This implies that at 95% confidence level, at least 46.64% (100% - 53.36%) of the deposit portfolio will stay with the bank for one year, which is the defined as the core portion of deposit in the above example.

Pros and Cons

Volatility analysis is straight forward, easy to implement, easy to validate and has a relatively less data requirement, which makes it a widely used methodology in the liquidity risk management. Instead of using complex time series models that require extensive period of data to filter out portfolio growth trend, volatility analysis can be easily applied to filter growth trend based on the historical analysis of the change in balance.

However, the use of volatility analysis would results in the following model deficiencies:

- Analysis dimensions: volatility analysis does not analyze the factors that affect the customer behaviors. That is to say, we can only observe the fluctuations in the balance of saving deposits, while the reasons behind them are not explored.
- The way that volatility analysis used to filter out growth trend might be less sophisticated than other time series model such as the time series regression analysis (to be discussed in details in the following sections). time series regression could be applied to decompose growth trend and volatility in a more efficient way given the sufficiency of data available.

3.1.1.2. Data Requirement

The data requirement as demonstrated in the above section for the volatility analysis is low, as only the balances of the portfolio under analysis are required without consideration of the factors that affect the balance fluctuation. However, if we want to run the analysis by segment, we still need some other info regarding the deposits, such as currency, customer segment and etc. To get the aggregate balance of a portfolio of saving deposits, the daily balance of each account in the portfolio is required and the aggregate balance of the portfolio is the sum of the balances of each account in the portfolio. The required daily data for each account are:

Data	Details
Observation Date	Date of Deposit Balance.

Account Number	The account number of the observed deposit account. Unique identifier of a saving account.
Branch No.	The Branch No. indicates the location of deposit.
Deposit Balance	The deposit balance of the account as of the Observation Date.
Currency	Currency of the deposit.
Customer ID	Identify the customer segment.

3.1.2. *Time Series Regression Analysis*

3.1.2.1. Model Methodology

Similar to Volatility Analysis, time series model is applied to the total balance of the saving deposits as one of popular models for dynamic cash flow projection to estimate the future change in deposit balance. The autoregressive integrated moving average (ARIMA) model which analyzes and forcasts equally spaced univariate time series data is adopted. It predicts a value in a response time series as a linear combination of its own past values and past errors.

The general form for ARMA(p,q) model is:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

Where Φ is the parameter of the autoregressive part of the model, θ is the parameter of the moving average part of the model and a is the white noise.

In this case, the Zt is the time series of the monthly balance of the deposit. By fitting the model with the SAS procedure PROC ARIMA, the parameters of Φ and θ will be solved. Therefore Zt can be predicted by using the estimated parameters and Zt-1 which is known at time t.

Pros and Cons

One of the most obvious advantages of time series analysis is that it predicts the future balances only using the input of previous balances which is easy in implementation. Also, the impact of seasonality can be analyzed by this method if necessary.

The primary shortcomings of the time series analysis are summarized as below:

- Since it only uses the input of previous balances, it fails to explain the change in balance with other macro economic factors.

- It is a results of a complex mathematical procedures and requires more understanding on the model assumption and underlying computation

3.1.2.2. Data Requirement

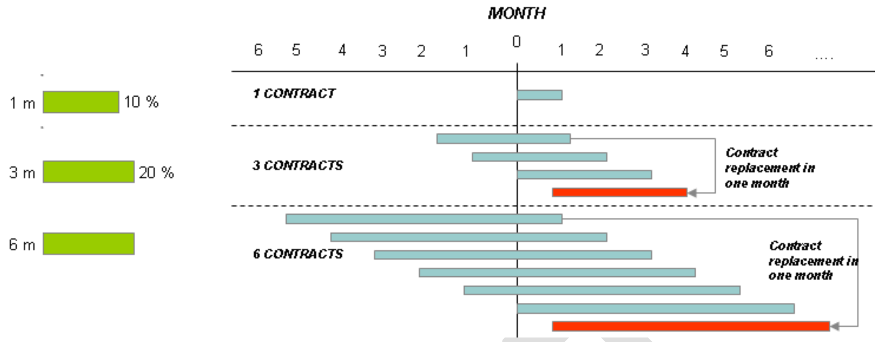
The data types required in performing the time series regression analysis is similar to the volatility analysis. However, the length of historical data required by time series regression analysis should cover at least one economic cycle. As time series regression analysis is also performed by segment as the volatility analysis does, hence other info regarding the deposits, such as currency, customer segment and etc are also required. To summarize, the required data for each account are:

Data	Details
Observation Date	Date of deposit balance.
Account Number	The account number of the observed deposit account.
Branch No.	The Branch No. indicates the location of deposit.
Deposit Balance	The deposit balance of the account as of the Observation Date.
Currency	Currency of the deposit.
Customer ID	Identify the customer segment.

3.1.3.1. Replicating PortfolioModel Methodology

Replicating portfolio is to use a portfolio of financial instruments with clear maturities to replicate the performance of saving deposits. This model aims to transform savings deposits into a portfolio of simple plain vanilla instruments, which have similar characteristics. The cash flows of this portfolio replicate the cash flows of saving deposits as closely as possible. We can estimate and forecast the interest rate, economic value and future cash flows of the saving deposit by utilizing the replicating portfolio.

The following graph below is an example showing how replicating portfolio strategy works. For ease of illustration, we suppose the balance of the savings deposits is stable over time. The deposits of the saving accounts can be invested in different buckets, i.e., one-month, three-month, and six-month instruments. Each of these buckets is then divided into monthly maturities. For instance, the six-month bucket consists of six different contracts with monthly maturities. As time goes by, every month one contract per bucket matures and is replaced by a corresponding new contract. Thus, the coupons of the constituents determine the yield of the replicating portfolio. The yield of the replicating portfolio in this example is given by current one-month, three-month, and six-month rates and by a set of historical rates. Adding up all maturities in all buckets, the replicating portfolio here consists of 10 contracts (1+3+9), and the yield of the replicating portfolio is an average of these 10 rates. The weights are determined such that this yield is as close as possible to the rate paid on the saving accounts plus a constant margin over time.



Picture： Replicating Portfolio Strategy

Actually, the assumption of stable balance of saving deposits is far too realistic, hence the assumption should be relaxed and in implementing the replicating portfolio model, the first step is to form a balance trend function. Typically trend structures are linear, quadratic, exponential or a combination of these. For the linear trend:

$$V_t = \beta_0 + \beta_1 \cdot \Delta_t + \sum_i \kappa_i \cdot \left(r_{i,t} - \bar{r}_i\right) + \delta \cdot \left(cr_t - \overline{cr}\right) + \varepsilon_t$$

where,

$i \in \{1,\ldots, I\}$: Maturity of buckets in month

V_t : Total volume at time t

$r_{i,t}$: Interest rate with maturity i at time t

$\bar{r}_i$: Average interest rate with maturity i over estimation period

cr_t : Customer's rate at time t

$c\bar{r}$: Average customer's rate over estimation period

Δ_t : Time in months between time 0 and t

$\beta_i, \kappa_i, \delta$: Parameters to be estimated

ε_t : Residual at time t

Using the trend function determined above, the estimation problem can be written as:

$$cr_t = \theta_0 + \frac{F_t(.)}{V_t} \cdot \sum_j \alpha_j \cdot ma_{j,t} + \frac{A_t(.)}{V_t} \cdot r_{1,t} + \eta_t$$
,

subject to $\alpha_j \geq 0, \forall j$ and $\sum_j \alpha_j = 1$.

where

$j \in \{1,3,6,...,60\}$: Maturities of buckets in months

cr_t : Customer's rate at time t

V_t : Total balance at time t

$F_t(\cdot)$: Trend balance at time t;

$A_t(\cdot)$: Balancing at time t;

$ma_{j,t}$: Moving average interest rate with maturity j at time t

$r_{1,t}$: 1-month rate at time t

θ_0, α_j : Parameters to be estimated

η_t : Residual at time t

This can be seen as an optimization problem. The optimal weights of the buckets minimize the volatility of the spread between the yield of the replicating portfolio and the saving deposits.

Once the replication portfolio is constructed, the Bank can purchase and hold the products in the portfolio. To achieve the purpose of interest rate risk management, the market instruments used to construct the replicating portfolio should be frequently traded and have no credit or liquidity risk. Therefore, government bonds or LIBOR related products are recommended.

Pros and Cons

Replicating portfolio model can be used to hedge against the interest rate risk for the banks by purchasing and holding *appropriate* shares of products calculated in the model. By doing this, this model considers the impact of interest rates on the balance of deposits.

The drawbacks of replicating portfolio are summarized as below:

- The feasibility of replicating portfolio relies on the liquid products available in the market such as coupon bonds and zero coupon bonds.

- Replicating portfolio fails to deal with the asymmetric fluctuation of deposit rate.

- The use of replicating portfolio requires the application of judgmental factors. As we know, different horizon of data will lead to different strategies. Therefore the investing strategy needs to be adjusted frequently.

3.1.3.2. Data Requirement

The data requirements for replicating portfolio are listed as below:

Data	Details
Historical Balance	Historical balance of the saving deposits under study
Historical Customer Rates	Historical interest rates offered to customers of saving deposits
Historical Interest Rates	Historical yield of the financial instruments used to construct the replicating portfolio (such as LIBOR rates, swap rates).

3.1.4. *Model Comparison*

We analyzed the pros and cons of above behavioral models for saving deposits and they are shown as follows:

Model	Pros	Cons
Volatility Analysis	▶ Objective and minimal management assumption ▶ Easy to back-test and validate ▶ Easy to understand and explain ▶ Consistent with current approach used by BOCHK	▶ Factors behind the fluctuation are not analysed ▶ Growth trend is filtered out in a less sophisticated way
Time series regression Analysis	▶ Objective and minimal management assumption ▶ Relatively easy to back-test and validate ▶ Widely used and more sophicated model to decompose growth trend and irregular components	▶ More complex mathematical computation involved ▶ Higher requirement on the length of historical data for the time series analysis
Replicating Portfolio	▶ Consider the impact of interest rate on historical pattern of balance ▶ Widely used in interest rate risk management	▶ Need to apply management assumption in the optimization process ▶ Difficult to explain the results ▶ More commonly used in interest rate modeling rather than liquidity

		risk management

3.2. Term Deposits

Besides savings account, term deposit also takes up a large proportion of the Bank’s funding portfolio, and therefore is quite important to liquidity risk management. Until now, the market still does not have behavioral models for term deposits that are widely adopted. Only some academic paper mentioned a few simple models of term deposits, such as:

- In 1996, Rigsbee Ayaydin and Richard presented <Implementing ‘Value at Risk’ in Balance Sheet Management – Using the Option-adjusted Spread Model> mentioned OAS (Option-adjusted Spread) model, which forecasted early-withdraw behavior based on historical data.

Generally, the behaviors of depositors of term deposits fall into the following three categories:

- Rollover: The depositor may choose to roll over the maturing term deposit by the same tenor or a different tenor upon maturity.
- Withdraw at contractual maturity date: The depositor withdraws all deposit at the contractual maturity date and closed the transaction.
- Early withdrawal: The depositor may withdraw part or all of the term deposit before the contractual maturity date.

Therefore in liquidity risk management of term deposits, rollover and early withdrawal will be the major consideration in behavioral modeling.

3.2.1. *Rollover and Early Withdrawal Analysis*

3.2.1.1. Model Methodology

Rollover and early withdrawal analysis is an analysis intends to estimate the rollover and early withdrawal portion of the whole term deposit portfolio. The analysis is usually implemented by segment of the term deposit’s contractual tenor. It means that term deposits with the same tenor will be grouped and analyzed together.

Taking the 1-month term deposits for example, the analysis is carried out step by step as below:

- Identify all 1-month term deposits during the observation window which is from 8 Aug 2011 to 29 Feb 2012. Since we need to analyze the early withdrawal behavior of term deposit, we will further exclude those whose full life is not within the observation window. That is to say, only 1-month term deposits which are originated after 8 Aug 2011 and mature before 29 Feb 2012 are used as modeling data.

- Each transaction is traced daily from its origination date till the contractual maturity date so as to identify which kind of behavior it belongs to.
- Once all sample deposits are categorized into the three groups (i.e. early withdrawal, withdraw at maturity date and rollover), the aggregated balance and relevant proportion of each group can then be generated as below:

	Aggregated Balance	Count	Proportion (by balance)
Early Withdrawal	352,154,046	297	0.04%
Withdrawal	48,981,395,692	66,853	9.87%
Rollover	213,603,417,790	610,506	90.09%
Total	262,936,967,528	677,656	100.00%

- It means that per $100 1-month deposit, $0.04 is withdrawn before maturity date, $9.87 is withdrawn at maturity date and $90.09 will be rolled over.
- Similarly, we can get the three rates for term deposits of other tenors.
- Frequency of roll-over of term deposits is also analyzed by counting the number of roll-over. By tracing the deposit balance on and after each of the fixed deposit's rolled over maturity date, the number of roll-over can be counted and summarized.
- For other types of fixed deposits such as flexi deposits, structured deposits, odd term deposits, staff term deposits, no behavioral assumption is applied in the projection of cash flow. Only contractual tenor is applied in the projection. More details on the principal of contractual cash flow projection can be referred to Section 4 of the cash flow model implementation report.

Pros and cons

The rationale behind rollover and early withdraw analysis stated in the above section is easy and straightforward. Besides, it could be directly applied in cash flow projection without further adjustment.

However, there is one often criticized problem that if the rollover and withdrawal behaviors is not time homogeneous, the analysis might be biased in estimation. Also the switching between term deposits and saving deposits cannot be considered in the model as there is data limitation that

3.2.1.2. Data Requirement

The rollover and early withdraw analysis requires the following information:

Data	Details
Observation Date	Date of observation
Account Number	The account number of the observed deposit account
Branch No.	The branch No. indicates the location of deposit
Currency	Currency of the deposit
Current Balance	The balance of deposit as of observation date
Origination Date	The date on which the deposit is originated
Maturity Date	The contractual maturity date of the deposit

3.2.2. Linear Regression

3.2.2.1. Model Methodology

Linear regression is used as a supplementation and improvement of the rollover and early withdrawal analysis in Section 3.2.1. As stated before, the rollover and early withdrawal analysis may be biased if the behavior is not time homogeneous. As an improvement, linear regression is performed to see whether the three rates are highly correlated and driven by some macro factors such as interest rate level.

Same as rollover and early withdrawal analysis, the sample deposits are categorized by the three possible behaviors. Instead of aggregating the deposit balance of each group, we calculate the three rates using one month's data on a moving window basis.

From	To	Early Withdrawal	Withdrawal	Rollover
5 Sep 2011	4 Oct 2011	0.16%	14.08%	85.76%
20 Sep 2011	19 Oct 2011	0.15%	14.88%	84.97%
5 Oct 2011	4 Nov 2011	0.27%	13.48%	86.25%
20 Oct 2011	19 Nov 2011	0.32%	14.85%	84.83%
5 Nov 2011	4 Dec 2011	0.22%	16.07%	83.70%
20 Nov 2011	19 Dec 2011	0.13%	14.26%	85.61%
…	…	…	…	…

Then we regress early withdrawal rate and rollover rate on a set of macro factors using multinomial regression.

Pros and cons

Linear regression improves the rollover and early withdrawal analysis by taking into account the macro factors and thus solves the problem of non-time homogeneous pattern of customer behaviors.

However, the implementation and validation of multinomial regression is complicated. Moreover, if the behavioral pattern is time homogeneous, rollover and early withdrawal analysis is enough for estimation and prediction.

3.2.2.2. Data Requirement

Besides the information required in rollover and early withdrawal analysis, linear regression also needs data of macro factors.

Data	Details
Observation Date	Date of observation
Account Number	The account number of the observed deposit account
Branch No.	The branch No. indicates the location of deposit
Currency	Currency of the deposit
Current Balance	The balance of deposit as of observation date
Origination Date	The date on which the deposit is originated
Maturity Date	The contractual maturity date of the deposit
<Macro Factor>	Macro factors that might have correlation with the behaviors of term deposits, such as interest rate level, stock market level, unemployment rate, etc.

3.2.3. *Model Comparison*

The pros and cons of rollover and early withdrawal analysis and linear regression are summarized in the following table:

Model	Pros	Cons
Rollover and Early Withdrawal Analaysis	► Easy to implement and understand ► Can be directly used in cash flow projection	► Estimations might be baised if the behavior pattern is not time homogeneous
Linear Regression	► Can deal with non-time homogeneous behavoir pattern	► Implementation process is complicated

3.3. Mortgages

The mortgage loans are one of the most important assets on the Banks' balance sheet, hence their cash flow characteristics need to be studied comprehensively from an asset and liability management's perspective. One of the main characteristic of mortgage loans is the prepayment behavior, which changes the maturity dates of mortgage loans or subsequent monthly repayment, causing unexpected cash inflows and consequently advancing the timing of cash inflows which lowers the liquidity risk at the expense of reinvestment risk.

CPR ("Conditional Prepayment Rate") is a widely used ratio in the market to measure the prepayment speed of mortgage loans. It is an annualized ratio of prepayment amount to outstanding mortgage principal. As mortgage loans are usually repaid monthly, the CPR rate is transformed into a monthly prepayment rate-SMM ("Single Monthly Mortality"), which is defined by the following formula:

$$SMM = \frac{\text{Prepayment amount of the month}}{(\text{Outstanding balance at the begining of the month - Scheduled payment of the month})}$$

As the outstanding balance at the end of the month equals to the outstanding balance at the beginning of the month minus the sum of the scheduled payment of the month and the prepayment amount the above formula can also be written in the following form:

$$SMM = \frac{\text{Prepayment amount of this month}}{(\text{Outstanding balance at the end of the month} + \text{Scheduled payment of this month})}$$

The transformation formula between CPR and SMM is:

$$CPR = 1 - (1 - SMM)^{12}$$

Prepayment of mortgage loans will have significant impact on banks' asset and liability management. The weighted average life of a mortgage loan may reduce substantially if prepayment is considered. Hence, prepayment of mortgage loans will change liquidity risk management, interest rate risk management, and FTP management. However, the complexity of prepayment option in the structure of mortgage loans brings challenges for the banks.

In this report, we elaborated and analyzed the feasibility of the following three models which are commonly used by peers in the market:

- OTS Prepayment Model;
- Linear Regression Model;
- Logistic Regression Model;

3.3.1. *OTS Prepayment Model*

3.3.1.1. Model Methodology

Some supervision organizations require standard prepayment models to be used. One representative prepayment model is the OTS model proposed by Office of Thrift Supervision. As the OTS model was originated from the Goldman Sachs model developed by Richard and Roll, it is also called the Modified Goldman Sachs model.

OTS prepayment model was used in the liquidity risk measurement for fixed rate mortgages and floating rate mortgages. In the model, prepayment of a mortgage loan is affected by three factors which are years after issuance, seasonality and refinancing:

Years after issuance: newly issued mortgages normally have the lowest prepayment rate as the borrowers don't have enough cash flows at that time and the probability of selling out the property for prepayment is also low.

Seasonality: prepayment rate varies from month to month. For example, prepayment rates usually are lower during the year end period as people are consuming more that time.

Refinancing: for fixed rate mortgage loans, the mortgagors potentially prefer to refinance when market interest rate is low. It will drive up the prepayment rate. Refinancing factor has the similar impact on adjustable rate mortgage loans.

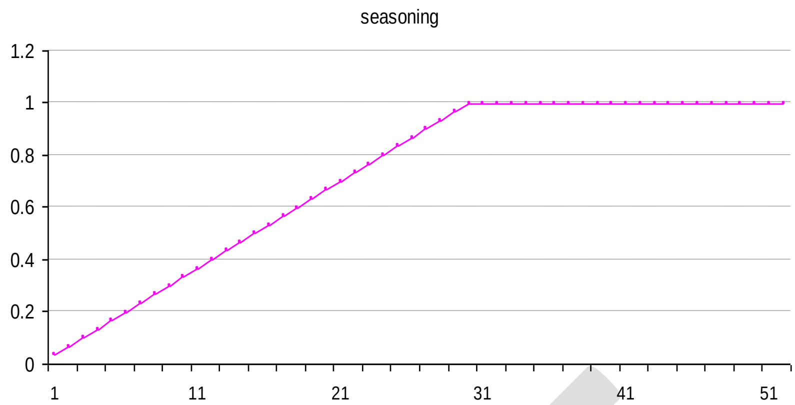
OTS model uses the product of the above three factors to estimate the prepayment rate:

$$cpr_{n,t} = seasoning_t \cdot seasonality_t \cdot refi_{n,t}$$

$cpr_{n,t}$ represents the annualized prepayment rate. $seasoning_t$, $seasonality_t$ and $refi_{n,t}$ represent the years after issuance factor, the seasonality factor and refinancing factor respectively. We take a 30-year fixed-rate mortgage loan for example to illustrate how OTS model measures prepayment rate.

For a 30-year fixed-rate mortgage loan, the seasoning factor is a broken-line function. The original prepayment speed is set as 0.03333 at issuance. Then the prepayment speed increases linearly with time until the 30th month when it reaches 1. It will keep at 1 afterwards.

$$seasoning_i = \max(i \cdot 0.0333, 1)$$

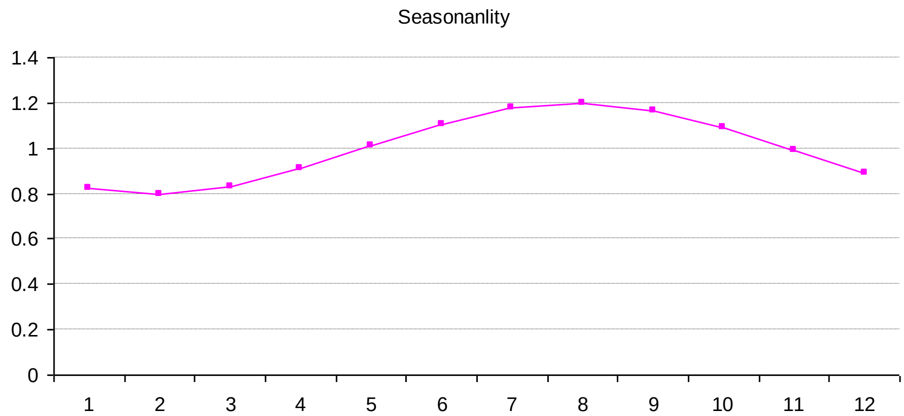


(The X-axis represents the months after issuance and the Y-axis represents the value of seasoning factor)

Seasonality factor is taken into account with a trigonometric function in the following form:

$$seasonality_t = 1 + 0.2000 \cdot \sin\{1.571 \cdot [\frac{(month + t - 3)}{3}] - 1\}$$

Its impact is shown in the following graph:



(The X-axis represents the month and the Y-axis represents the value of seasonality factor)

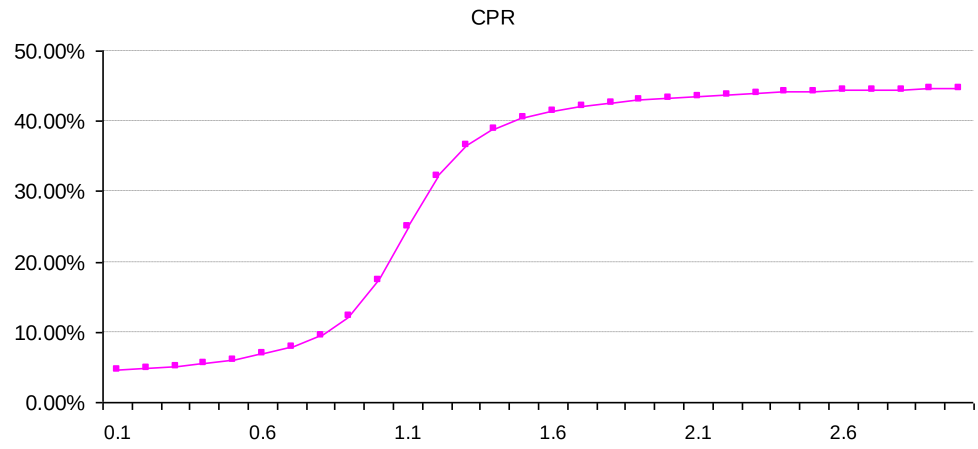
From the above graphic, the seasonality factor is at its peak in June, July, August and September.

The refinancing factor is incorporated by an arctangent function as below:

$$refi_{n,t} = 0.2406 - 0.1389 \cdot \arctan[5.952 \cdot (1.089 - \frac{c}{m_{n,t-3}})]$$

where n is the term of the mortgage, $m_{n,t-3}$ is the three month lagged market mortgage rate. c is the coupon rate of the mortgage.

The curve of the refinancing factor demonstrates an S-shape as shown in the following graphic:



(The X-axis represents the refinancing factor which is the ratio of mortgage loan rate to market rate and the Y-axis represents the refinancing factor)

Pros and Cons

Given the parameters, OTS Prepayment Model is easy to apply, as only the mortgage rate, market mortgage rate and the time after issuance are required. Further, primary factors affecting prepayment rates are taken into account. In addition, the refinancing factor in the model can reflect S-shape of prepayment rates. Finally, OTS Prepayment Model can be applied to individual mortgage to estimate the prepayment rate.

The main disadvantages of OTS Prepayment Model include:

- The parameters need to be calibrated if the model is to be applied in a different market other than US, and the calibration process is very complicated.
- Calibration of the coefficient parameters is complicated.

3.3.1.2. Data Requirement

The data required for OTS Prepayment model is simple. To perform the model on a given mortgage, the following data are needed:

Data	Details
Effective Date	Effective date of the mortgage loan
Maturity Date	Maturity date of the mortgage loan
Coupon Rate	The interest rate of the mortgage loan.
Market Mortgage Rate	The prevailing market interest rate of mortgages.

3.3.2. *Linear Regression Model*

3.3.2.1. Model Methodology

Linear Regression model tries to build a direct equation between the prepayment rate and the driving factors behind by utilizing regression models on time series with certain macro economic factors. This model is based on a portfolio level, which means that the sample data points are the historical prepayment rates of a specific pool. If the sample data for the dependent variable is a series of historical monthly prepayment rate of a portfolio of mortgages, the sample data of dependent variable and the explanatory factors (such as the macro economic factors) are time series, hence, *unit root test* for stationary of these time series are needed before running the regression models. If the time series are not stationary, then it could lead to a spurious regression causing the test statistics to fail.

Linear regression is one of the frequently used regression models in which the relationship between the dependent variable and one or more explanatory variables is linear. The multivariate linear regression model takes the following form:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, i = 1,..., N$$

where $(x_{i1},..., x_{ip})$ is the i-th sample data from a population of $(x_1,..., x_p)$, which are the explanatory variables affecting the dependent variable y. N is the number of observations (samples). And ε_i is the residual error for the i-th sample, which is a random variable capturing all the idiosyncratic factors other than $(x_1,..., x_p)$, and $(\beta_1,..., \beta_p)$ are the parameter coefficient, which measure the sensitivity of the dependent factors to corresponding the explanatory factors. With sample value $(y_i, x_{i1},..., x_{ip})$ where $i = 1,2,..., N$, we have $p+1$ parameters to be determined, i.e., $(\beta_0, \beta_1,..., \beta_p)$. In order to estimate the parameters, the above regression model can be re-written to a matrix form:

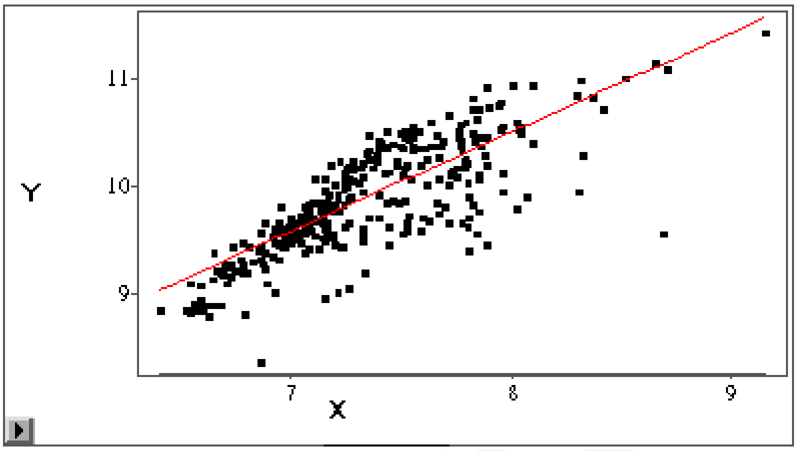
$$Y = X \cdot \beta + \varepsilon$$

where Y is a vector of the sample data of $(y_1,..., y_N)'$, ε is a vector of the residual error terms $(\varepsilon_1,..., \varepsilon_N)'$, and the matrix X is the array of sample explanatory variables:

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Np} \end{pmatrix}$$

For ordinary linear regression, Ordinary Least Square ("OLS") method can be used to estimate the coefficient parameter β, and the OLS estimator of the coefficient β is given by:

$$\hat{\beta} = (X'X)^{-1}X'Y$$



Picture: Linear regression illustration with one independent variable

The OLS estimator given by the above equation is BLUE ("Best Linear Unbiased Estimator") if Gauss-Markov assumptions are satisfied:

- $E(\varepsilon_i) = 0$, for all i.
- $Var(\varepsilon_i) = \sigma^2 < \infty$, for all i.
- $COV(\varepsilon_i, \varepsilon_j) = 0$, for all i and i.

Auto-correlation is a common problem in time series regression models, especially for the economic factors. If the time series of residual exit auto-correlation, then the basic assumptions in regression models are violated, causing the following consequences:

- The OLS estimator is not BLUE (e.g., the variance of OLS estimator is not the least in all the linear estimators);
- The standard deviation calculated by OLS formula is not correct;
- If there are lagged independent variables in the right of the regression equation, then the OLS estimator is biased and not consistent.

Generally speaking, if auto-correlation exists in the residual series, it could results to false regression results. T-testing and F-testing statistics are not reliable any more. Hence, we need to test if the residual series of ordinary linear regression model are auto-correlated. The common testing methods include *Ljung-Box Q* testing and *Breush-Godfrey LM* testing. If auto-correlation does exist, auto-regression model $AR(p)$ to amend the original models.

Different structures of auto-regressive residuals lead to different linear regression models. The most popular practice is the first-order residual auto-regression linear model:

$$\begin{cases} Y = X'\beta + \varepsilon \\ \varepsilon_i = \rho\varepsilon_{i-1} + v_i, i = 2, \cdots, n \\ E(v_i) = 0, E(v_i^{\ 2}) = \sigma^2, E(v_i v_j) = 0, i \neq j \end{cases} \qquad (1)$$

As one can see, the residuals in the above model do not satisfy the assumption of independence in OLS, however, we can build a true OLS model on ε_i .

when $|\rho| < 1$， the first-order regression process is stationary.

$$\begin{aligned} \varepsilon_i &= v_i + \rho\varepsilon_i = v_i + \rho v_{i-1} + \rho^2 \varepsilon_{i-2} \\ &= v_i + \rho v_{i-1} + \rho^2 v_{i-2} + \rho^3 v_{i-3} + \cdots \end{aligned}$$

then

$$E(\varepsilon_i) = \sum_{k=0}^{\infty} \rho^k E(v_{i-k}) = 0$$

$$Var(\varepsilon_i) = \sum_{k=0}^{\infty} \rho^{2k} Var(v_{i-k}) = \sum_{k=0}^{\infty} \rho^{2k} \sigma^2 = \sigma^2 / (1 - \rho^2) = \sigma_\varepsilon^2$$

$$E(\varepsilon_i \varepsilon_j) = \rho E(\varepsilon_{i-1}^2) + E(\varepsilon_{i-1} v_i) = \rho\sigma^2 / (1 - \rho^2)$$

For the same reason， $E(\varepsilon_i \varepsilon_{i-s}) = \rho^s \sigma^2 / (1 - \rho^2), s = 1, 2, ...$

Notice that all of them have the same factor σ^2, as a result, we can obtain the covariance matrix of residual series:

$$\Phi = \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\ \rho & 1 & \rho & \ldots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \ldots & \rho^{n-3} \\ & & \ldots & & \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & 1 \end{pmatrix}$$

If denoting $\Phi = \sigma^2 \Psi$， the first-order residual auto-regression model can be re-written to:

$$\begin{cases} Y = X'\beta + \varepsilon \\ E(\varepsilon) = 0, Var(\varepsilon) = \sigma^2 \Psi \end{cases} \qquad (2)$$

To solve the first-order residual auto-regression model, we should

First, solve the OLS estimator of β in the regression equation (1) using original data,

$$\hat{\beta} = (X'X)^{-1} X'Y$$

calculate the estimator of residuals

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i, i = 1, \ldots, n$$

and get the estimator of ρ using $\hat{\varepsilon}_i$ in the auto-regression equation of residuals.

$$\hat{\rho} = \sum_{i=2}^{n} \hat{\varepsilon}_i \hat{\varepsilon}_{i-1} / \sum \hat{\varepsilon}_i^2$$

Second, compute GLS ("Generalized Least Square") estimator of β in equation (2):

$$\hat{\hat{\beta}} = (X'\Psi^{-1}X)^{-1} X'\Psi^{-1}Y$$

$$\Psi^{-1} = \begin{pmatrix} 1 & -\rho & 0 & \ldots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \ldots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \ldots & 0 & 0 \\ & & & \ldots & & \\ 0 & 0 & 0 & \ldots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \ldots & -\rho & 1 \end{pmatrix}$$

It can be proven that there exists a lower triangle matrix so that

$$P'P = \Psi^{-1}$$

The lower triangle matrix is:

$$\begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \ldots & 0 & 0 \\ -\rho & 1 & 0 & \ldots & 0 & 0 \\ 0 & -\rho & 1 & \ldots & 0 & 0 \\ & & & \ldots & & \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & -\rho & 1 \end{pmatrix}$$

Doing transformation $Y^* = PY, X^* = PX$, then model (2) is changed to:

$$\begin{cases} Y^* = X^*\beta + \varepsilon^* \\ E(\varepsilon^*) = 0, Var(\varepsilon^*) = \sigma^2 I \end{cases} \qquad (3)$$

Now the model satisfies the basic assumptions of OLS, as a result, its OLS estimator can be obtained by:

$$\hat{\hat{\beta}} = (X^{*'} X^*)^{-1} X^{*'} Y$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(Y^* - X^*\hat{\hat{\beta}})'(Y^* - X^*\hat{\hat{\beta}})$$

3.3.2.2. Data Requirement

In the linear regression model, the dependent (responsive) variable is the monthly prepayment rate for the portfolio under study. The sample data of the prepayment rate for the portfolio is monthly prepayment observed at each month for a specified historical period. For the explanatory variables, their values as of the corresponding month of the observed prepayment rate are needed. Hence, to summarize, each sample data point in the linear regression should include the following data field:

Field	Data Type	Description	Note
Monthly Prepayment Rate	NUM	The prepayment rate of a portfolio at a given month	Dependent variable
Observation Month	NUM	The month of the Monthly Prepayment Rate	
Loan Age	NUM	The weighted average number of years passed since origination of the mortgages in the pool.	Explanatory variable
Percentage Outstanding	NUM	The ratio of current balance to the initial balance of the pool	Explanatory variable
Seasonality	NUM	A transformed factor to describe different prepayment rate at different months.	Explanatory variable
Interest Rate Shape	NUM	The spread between the long term swap yield and short-term yield.	Explanatory variable
Unemployment Rate	NUM	The latest observed unemployment rate of the observation month (1 month lagged)	Explanatory variable
Property Sales Volume	NUM	The latest observed turnover of housing sales of the observation month (1 month lagged)	Explanatory variable
Property Index	NUM	The latest observed property index of the observation month (1 month lagged)	Explanatory variable

3.3.3. *Logistic Regression Model*

3.3.3.1. Model Methodology

Logistic Regression Model is based on individual mortgage level, which means that the sample data point is the prepayment behavior of each mortgage in the portfolio of similar prepayment characteristics. Normally, there are two dimensions to acquire sample data. One is the prepayment info of each mortgage at a specified month. The other is the monthly prepayment info of a single mortgage over a historical period. In practice, sample data is obtained from both dimensions.

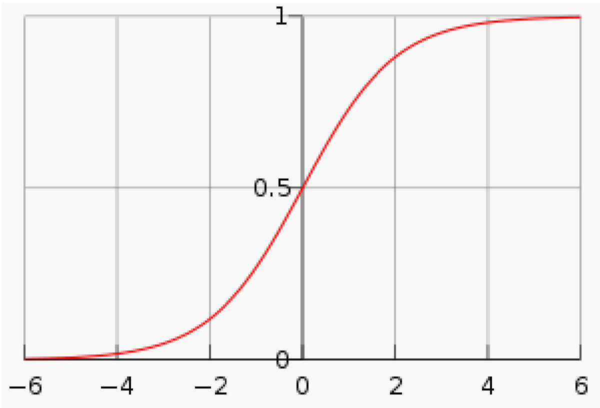
Another difference compared to linear regression model is that logistic regression uses a dichotomous variable as the dependent variable. This type of variable is called a Bernoulli variable. For modeling the prepayment behavior of mortgages, the event could be if prepayment happens. Let Y_i as the Bernoulli variable with $(Y_i = 1)$ denoting there is a prepayment and $(Y_i = 0)$ if there is no prepayment and $p = P(Y = 1 | X_1, ... X_y)$ is the probability of prepayment conditional on explanatory factors $X_1, X_2, ..., X_n$. Then the multivariate logistic regression model takes the following form:

$$\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_t X_t$$

From the above equation, we can solve p as

$$p = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + .\cdots + \beta_t X_t)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + .\cdots + \beta_t X_t)}$$

We can observe from above equation that, the conditional probability p ranges from 0 to 1, hence $\ln(\frac{p}{1-p})$ ranges from $-\infty$ to $+\infty$. On the left side of the regression equation $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + .\cdots + \beta_t X_t$ also ranges from $-\infty$ to $+\infty$.



Picture: Illustration of Single Factor Logit Function

Given sample data $(y_i, x_{i1}, x_{i2},..., x_{it})$, $i = 1,2,.., n$, we have t+1 parameters to be determined, $\beta_0, \beta_1,...\beta_t$. In this case, ordinary least square ("OLS") estimates can not be used as the linear regression model. Instead, the Maximum Likelihood Estimator ("MLE") is usually used to estimate the coefficient parameters in logistic regression. Let $p_i = P\{y_i = 1 | x_{i1},..., x_{it}\}$ as the conditional probability of y_i equaling to 1 given $(x_{i1}, x_{i2},..., x_{it})$; so the conditional probability of y_i equaling to 0 given $(x_{i1}, x_{i2},..., x_{it})$ is $P\{y_i = 0 | x_{i1},..., x_{it}\} = 1 - p_i$. As a result, for each observation,

$$P(y_i | x_{i1}, x_{i2},..., x_{it}) = p_i^{y_i} \cdot (1 - p_i)^{1-y_i} = [\frac{1}{1+e^{-(\beta_0+\beta_1 X_{i1}+\cdots+\beta_t x_{t1})}}]^{y_i} [1 - \frac{1}{1+e^{-(\beta_0+\beta_1 X_{i1}+\cdots+\beta_t x_{t1})}}]^{1-y_i}$$

Assuming all observations are independent, their multivariate distribution is the product of all marginal distribution:

$$P(y_1,..., y_n | x_{i1}, x_{i2},..., x_{it}, i \in (1, n)) = \Pi_{i=1}^{n} [\frac{1}{1+e^{-(\beta_0+\beta_1 X_{i1}+\cdots+\beta_t x_{t1})}}]^{y_i} [1 - \frac{1}{1+e^{-(\beta_0+\beta_1 X_{i1}+\cdots+\beta_t x_{t1})}}]^{1-y_i}$$

Then we obtain the likelihood function $L(\beta_0, \beta_1,..., \beta_t)$:

$$L(\beta_0, \beta_1,..., \beta_t) = \Pi_{i=1}^{n} [\frac{1}{1+e^{-(\beta_0+\beta_1 X_{i1}+\cdots+\beta_t x_{t1})}}]^{y_i} [1 - \frac{1}{1+e^{-(\beta_0+\beta_1 X_{i1}+\cdots+\beta_t x_{t1})}}]^{1-y_i}$$

According to the theory of maximum likelihood, the coefficient parameters $\beta_0, \beta_1,...\beta_t$ should maximize the above likelihood function. The estimators are usually computed by maximizing the natural logarithm of likelihood function. This is done by taking the partial derivatives to $\beta_0, \beta_1,...\beta_t$, and solve the equations by letting the partial derivatives equal to zero:

$$\begin{cases} \dfrac{d\ln[L(\beta_0,\beta_1,...,\beta_t)]}{d\beta_0} = 0 \\ \dfrac{d\ln[L(\beta_0,\beta_1,...,\beta_t)]}{d\beta_1} = 0 \\ .... \\ \dfrac{d\ln[L(\beta_0,\beta_1,...,\beta_t)]}{d\beta_t} = 0 \end{cases}$$

Deviance

For logistic regression, the deviance (also known as residual deviance) is used to assess the fit of the overall model. The deviance for a logistic model can be likened to the residual sum of squares in ordinary regression. The smaller the deviance the better is the fit of the model. The deviance can be compared to a chi-square distribution, which approximates the distribution of the deviance. This is an asymptotic result that requires large sample sizes. The deviance for the combined turtle data is 14.863 on 3 degrees of freedom. The deviance statistic is:

$$D = -2\ln(likelyhood(fitted))$$

Essentially we are using the above statistics to test H_0 : fit is good. For example, a p-value of 0.0019 indicates that the deviance left after the fit is too large to conclude that the fit is good. Thus, there is room for improvement in the model.

Likelihood Ratio Test

Once the multivariate logistic regression model has been fitted, the first step is to assess the overall significance of the explanatory variables in the model using likelihood ratio test, which is analogous to the global F-test in linear regression analysis. The hypothesis test:

$H_0 : \beta_i = 0, i = 1,2,...,t$

The test statistic G equals to:

$$G = -2\ln(L_0 / L_1) = -2(L_0 - L_1)$$

where L_0 is the likelihood of the model without explanatory variables and L_0 is the likelihood with the explanatory variables. Under null hypothesis, G will have a chi-square distribution with t degrees of freedom.

Wald Test

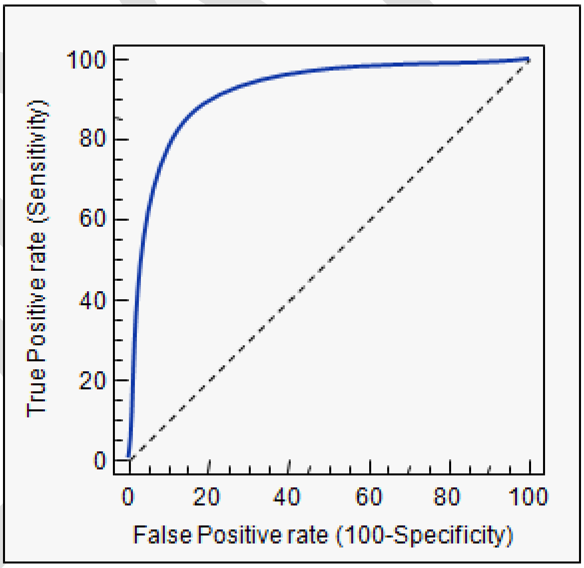
Similar to linear regression, the Wald test is used to test the statistical significance of each coefficient β_i in the logistic model. A Wald test calculated a Z statistic which is:

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

Under the null hypothesis that the coefficient is equal to zero, this statistics follows a standard normal distribution. Alternatively, this value is squared and will yield a chi-square distribution with 1 degree of freedom. Either way, it gives equivalent results.

ROC Curve Analysis

Another method to evaluate the logistic regression model makes use of ROC curve analysis. In a logistic regression, a classification table can be created for any cut-off value of the fitted probability and hence the sensitivity and specificity are then available for this particular table. The fraction calculated as count of predicted positives divided by the actual total of positives is the sensitivity and the fraction calculated as the count of predicted negatives divided by the total negatives will be the specificity.



The AUC, sometimes referred to as the C-statistic (or concordance index), is a value that varies from 0.5 (discriminating power not better than chance) to 1.0 (perfect discriminating power). AUC-statistic of 0.84, for example, means that a randomly selected individual from the positive group has a test value larger than that for a randomly chosen individual from the negative group in 84% of the time. When the variable under study cannot distinguish between the two groups, i.e. where there is no difference between the two distributions, the area will be equal to 0.5 (the ROC curve will coincide with the diagonal). When there is a perfect separation of the values of the two groups, i.e. there no overlapping of the distributions, the area under the ROC curve equals 1 (the ROC curve will reach the upper left corner of the graph).

Odds Ratio

The odds ratio OR is defined as the ratio of the odds for $x=1$ and the odds for $x=0$ and is given by the equation:

$$OR = \frac{odds_1}{odds_0} = \frac{e^{\beta_0+\beta}}{e^{\beta_0}} = e^{\beta}$$

Hence, for a logistic regression with a dichotomous explanatory variable, coded zero and one, the relationship between the odds ratio and regression coefficient is:

$OR = e^{\beta}$

The odds ratio can be interpreted as having a harmful or protective effect upon the subject event depending on how far it deviates from 1 (i.e., no effect). Odds ratios whose confidence limits exclude 1 are statistically significant.

Transformation of Probability to SMM

The SMM of a given month can be calculated by the by the following equation:

$$SMM = \frac{\sum_{i}^{N} p_i^P \cdot \overline{M} \cdot (B_{0,i} - P_{0,i}^S) + \sum_{i=1}^{N} p_i^F \cdot (B_{0,i} - P_{0,i}^S)}{\sum_{i=1}^{N} B_{0,i} - \sum_{i=1}^{N} P_{0,i}^S}$$

$$\sum_{i}^{N} p_i^P \cdot \overline{M} \cdot (B_{0,i} - P_{0,i}^S) = \text{Total Partial Repaid Amount}$$

$$\overline{M} = \frac{\text{Total Partial Repaid Amount}}{\sum_{i}^{N} p_i^P \cdot (B_{0,i} - P_{0,i}^S)}$$

$\overline{M}$ is the average ratio of prepayment amount to the beginning outstanding balance minus scheduled installment in the month given there is a partial prepayment. It is calibrated with historical data with equating the estimated partial repaid amount to the actual partial repaid amount;

p_i^P is the probability of partial prepayment for the i-th mortgage. It is estimated by the logistic regression model;

p_i^F is the probability of full prepayment for the i-th mortgage. It is estimated by the logistic regression model;

$B_{0,i}$ is the initial outstanding balance of the i-th mortgage in the month;

$P_{0,i}^S$ is the scheduled payment for the i-th mortgage in the month

In the logistic regression model, we want to find out the probability of prepayment for each mortgage loan given the information of the mortgage and the macro economic factors. Then the probability of prepayment of each mortgage in the portfolio is used to calculate to SMM of the portfolio. Hence the biggest difference as compared with the linear regression model is that logistic regression is performed at individual mortgage level instead of at portfolio level for the linear regression model.

To satisfy the need of regression at deal level, each sample data point is collected at individual mortgage level. The data info for each sample point includes:

Field	Data Type	Description	Note
Prepayment Indictor	BIN	Prepayment signal with "1" implying prepayment occurs and "0" implying no prepayment occurs during the observed month.	Dependent variable
Observation Month	NUM	The month for the Prepayment Indictor.	
Mortgage Number	NUM	The unique mortgage ID of the observed Prepayment Indictor	
Loan Age	NUM	The number of years passed since the origination of the Mortgage Loan	Explanatory variable
Origination Maturity	NUM	The original tenor (number of years) for the Mortgage Loan	Explanatory variable
Time to Maturity	NUM	The number of years remaining for the Mortgage Loan	Explanatory variable
Percentage Outstanding	NUM	The ratio of current book balance to the initial book balance	Explanatory variable
Repayment History	BIN	"1" implies the borrower has partially repaid the mortgage loan in some period before (e.g. within 30 days), "0" implies the borrower has never partially repaid the loan in that period	Explanatory variable

Outstanding Balance	NUM	The outstanding balance of the Mortgage Loan	Explanatory variable
Initial Balance	NUM	The initial balance of the Mortgage Loan	Explanatory variable
Loan-to-Value Ratio	NUM	The loan-to-property value ratio of the Mortgage Loan	Explanatory variable
Refunding Spread	NUM	The spread between the prevailing mortgage interest rate and 10-year swap yield.	Explanatory variable
Interest Rate	NUM	Market prevailing interest rate (for instance, the HIBOR, prime rate as of the observation Month	Explanatory variable
HSI Return	NUM	The alternative return compare to the property market	Explanatory variable
Unemployment Rate	NUM	The latest observed unemployment rate of the observation month (1 month lagged)	Explanatory variable
Property Sales Volume	NUM	The latest observed turnover of housing sales of the observation month (1 month lagged)	Explanatory variable
Property Index	NUM	The latest observed property index of the observation month (1 month lagged)	Explanatory variable

3.3.4. ***Models Comparison***

The OTS model is standard model required by authorization. The model is easy to understand in that the prepayment rate equals to a multiple of three factors. Further, it is designed to be able to capture the S-shape of prepayment rate.

The primary shortcomings of OTS model are:

- The model coefficients may be appropriate in areas outside US;
- The implementation of OTS is usually through Monte Carlo simulation of interest rate models

Both the Linear Regression Model and Logistic Regression Model are used by large commercial banks around the world. As the two models deal with prepayment at different levels with linear regression model at portfolio level and logistic regression model at individual mortgage level respectively, their performance are different.

For linear regression model:

- It builds a relationship of the prepayment rate of a pool of mortgages to certain macro economic factors directly. The model is relatively easy to understand and calculation speed is fast with considerable less sample data than logistic regression model.
- The required data is relatively lower than the logistic regression model which takes a lot of customer information into account. This is useful for mortgage pools the customer info of which is not available.

The primary limitations of linear regression model are summarized as below:

- Linear regression model is performed at portfolio level, hence the implementation of its results shall be applied also at portfolio level, which lead to that the prepayment analysis for individual mortgage is not available.
- Linear regression model fails to consider the factors of individual mortgage info, especially the customer info. In reality, the customer's education, occupation and etc may have effect on the prepayment behavior. This limitation also can be viewed as a shortcoming as a consequence of a portfolio-level analysis.

Logistic regression model has several advantages over the linear regression model:

- Logistic regression is performed at individual mortgage level. This allows that the probability of each individual mortgage can be calculated from the results of the model updated as required;
- Logistic regression model can easily incorporate the customer info into the regression as opposed to the linear regression model, thus improving the results accuracy;
- Logistic regression model can capture the S-shape of prepayment rate behavior.

The primary drawbacks of logistic regression model are summarized as below:

- Logistic regression puts a high requirement on the data. Hence the data availability may not be satisfied for some banks.
- The parameter estimation procedure of logistic regression relies heavily on having an adequate number of samples for each combination of independent variables, small sample sizes can lead to widely inaccurate estimates of parameters. On the other side, large sample size reduces the computation speed.

The pros and cons of are summarized in the following table:

Model	Pros	Cons
OTS Model	▶ Easy to understand ▶ Can Capture S-shape of prepayment rate	▶ Model coefficient may not applied to areas outside US ▶ Implementation is through Monte Carlo simulation of interest rate models
Linear Regression Model	▶ Direct relationship of prepayment rate to economic factors ▶ Relatively easy to understand ▶ Less sample data and fast computation speed	▶ Results can't be applied to individual mortgage ▶ Customer info are not taken into account
Logistic Regression Model	▶ Probability of each mortgage can be calculated ▶ Customer info is taken into account ▶ Commonly used in credit risk and liquidity risk management	▶ Data requirement is high ▶ More appropriate to use in the dynamic balance sheet analysis

As logistic regression model has a better statistical explanatory ability compared with linear regression model. Hence it is achieving more popularity among industry. As a result, we apply the logistic regression model in the behavior modeling of mortgages.

3.4. Commitments

Usually there is a large portion of off-balance-sheet items, such as loan commitment, financial guarantee, and letter of credit in the banks, which make banks exposed to contingent drawdown risk in the future. This contingent drawdown could be very volatile causing large cash outflows. Hence, to manage the cash flow banks need to pay attention to instruments of this type and the estimate of potential cash flow drawdown from these off-balance-sheet instruments is needed.

In industry, the following three methods are commonly used to analyze the future drawdown of off-balance-sheet instruments. Each method is designed to focus on different perspectives:

- Static Utilization Ratio Analysis
- Life Cycle Utilization Ratio Analysis
- Peaks and Valleys Analysis
- Maximum Utilization Ratio Analysis
- Portfolio Drawdown Rate Analysis

We introduce in details the methodology of the above five methods in the following sections.

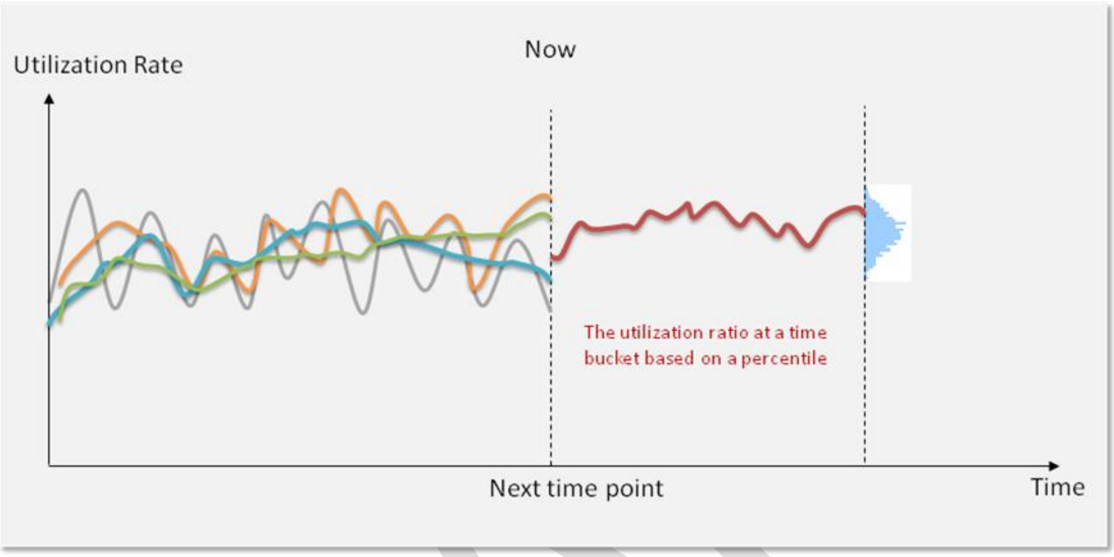
3.4.1. Static Utilizaton Ratio Analysis

Regarding the static utilization ratio approach, the first step is to obtain the historical utilization ratio based on the analysis of historical data, and then the future utilization ratio can be developed by mathematical methods in accordance with the management's risk appetite. The essence of this method is to estimate the impact of off-balance-sheet items on the on-balance-sheet cash flow based on a static utilization ratio, which is defined as the ratio of the amount drawdown to the total limit of a product with the same tenor.

On a certain balance sheet reporting date, four key numbers are associated with the off-balance-sheet items: the commitment amount (credit line), utilized amount (the amount drawn down and accounted for on the balance sheet), utilization ratio and timeline. The four numbers will be used in the analysis with the following basic steps:

- Segment the commitments by “revolving” or “non-revolving” and terms of them. An analysis is performed on each segment.
- Calculate the utilization ratio at each time bucket (i.e. every day/week/monthly end) for all the commitments in the segment under analysis.

- The changes of the utilization ratio and the future utilization ratio are estimated by the static ratio method, which is to assign one or more future static utilization ratios, in line with the principle of prudence, based on the change trend of historical utilization ratios, historical average utilization ratios and combined with experience judgment. For example, one can use a quantile of the historical utilization ratios as the estimator of the future utilization ratio.



The data requirement for the static utilization rate analysis includes the following data of each sample commitment:

Data	Data Type	Description	Note
Observation Date	DATE	The observation date	
Facility Code	CHAR	The unique ID of the sample product	Key Word
Product Type	CHAR	The product type: e.g. Undrawn Loan, Trade Guarantee	
Revolve Indicator	CHAR	Revolving or non-revolving	
Revocable Indicator	DATE	Revocable or irrevocable	
Effective Day	DATE	Effective date of the facility	
Expire Day	DATE	Expire date of the facility	
Tenor	NUM	The tenor of the sample product	
Surviving Days	NUM	Number of days from Effective Date to Observation Date	
Total Limit	NUM	Total Limit of the sample product	
Drawdown Amount	NUM	Used Limit as of Observation Date	
Utilization Ratio	NUM	Ratio of Drawdown Amount to Total Limit	

3.4.2. *Life Cycle Utilizaton Rate Analysis*

Life cycle utilization ratio analysis which goes beyond the static utilization rate analysis for off-balance-sheet products is an investigation of the utilization ratio of the products for the life cycle. It tries to link the utilization rate with existing time of products. The utilization ratio, which will be used to forecast future cash flow, is defined as the ratio of the amount drawdown to the total limit of this instrument.

The analysis is usually performed by segment of tenor and revolving feature. This means that products of the same tenor and revolving feature are grouped together and analysis is performed on each group. As samples from each group probably will be of different effective dates, hence their lives are overlaid as if the products are based on the same effective date. Specifically, the following are the steps in performing Life Cycle Utilization Rate Analysis:

- The model granularity is determined. For example, non-revolving commitments with 1-year tenor are grouped together to build a model, and revolving commitments with 1-year tenor are grouped together to build model. Other grouping criterions are also applicable, but a granularity is preferred to be at least based on tenor and revolving or non-revolving.

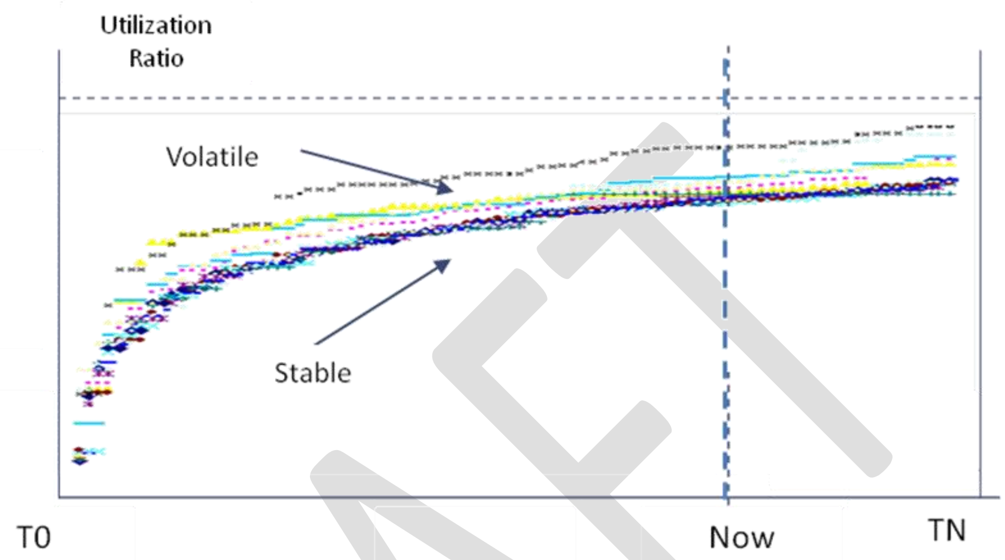
- Samples are overlaid to have an identical effective date. For example, if the effective date of one 1- year commitment is 8 May 2010, and another 1-year commitment is 23 Oct 2010, then in the analysis of utilization rates, the effective dates of the two commitments are overlaid as if they are effective from the same date as we only care the relationship of between the utilization rate and the time passage after effective date.

- For every sample, the accumulative drawdown balance (repayment is taken into account) from the effective date to maturity date are recorded and utilization rates are calculated as shown in the following table of an example.

Age (days)	Sample Commitment 1	Sample Commitment 2	Sample Commitment 3
1	0.14	0.07	0.07
2	0.11	0.16	0.16
3	0.22	0.08	0.08
4	0.20	0.07	0.07
5	0.33	0.12	0.12
6	0.19	0.21	0.21
7	0.27	0.22	0.22
8	0.31	0.11	0.11
9	0.38	0.15	0.15
…	…	…	…
357	0.70	0.42	0.42
358	0.71	0.38	0.38
359	0.70	0.37	0.37
360	0.70	0.39	0.39
361	0.71	0.39	0.39
362	0.73	0.40	0.40
363	0.69	0.38	0.38

364	0.68	0.36	0.36
365	0.72	0.39	0.39

- The average utilization ratio for each age for all the samples is calculated. For example, the average utilization ratio at the first day after effective is (0.14+0.07+0.07)/3.



As we have stated above, the purpose of utilization rate analysis is to find out the fluctuation utilization rate, which reflect the loan drawdown and debt repayment, during the life cycle of off-balance-sheet products. In liquidity management, it is common to estimate utilization ratio Y by fitting the average utilization ratio with some functional forms such as exponential function, depending on the shape of the curve:

$$\hat{Y} = a_0 + a_1 \cdot \ln(t) + Z(t)$$

Where t and Y denotes the age of the product, which is the passage of time in years after effective date, and the corresponding utilization ratio at age t respectively. $a_i, i = 0,1,2,3$ are the parameters to be estimated.

$Z(t)$ denotes a volatile term, which follows a Brownian movement and takes the following form:

$$\frac{\delta Z}{Z} = \mu_V \delta t + \sigma_V \varepsilon \sqrt{\delta t}$$

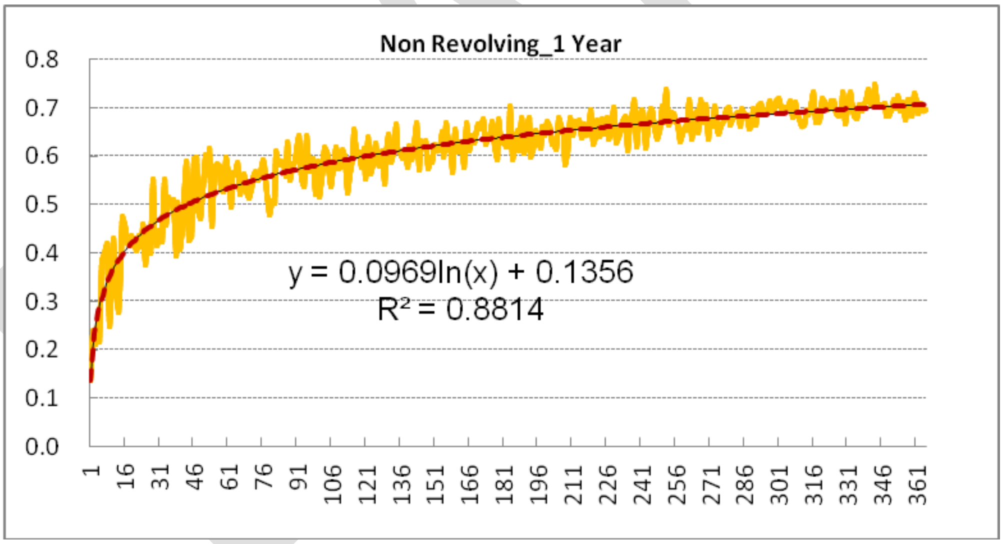
where μ_v refers to the mean of historical rates of utilization change, σ_v refers to the volatility of the historical utilization rates, δ_t refers to the time interval for the changes (i.e. day/week/month), and ε refers to the random variable following the standard normal distribution.

In the above models, the first part represents a stable utilization portion which is a linear function of age; $Z(t)$ represents a volatile utilization. For the interpretation of stable utilization and volatile utilization,

- Stable utilization refers to the line closer to the bottom in the above graph (defined by certain percentile). Stable utilization can be seen as a stable and determined utilization. It linearly relates to the age of the product.

- Volatile utilization refers to a random utilization factor on the top of stable utilization. Through the adjustment of the volatile utilization, the actual utilization, which equals to sum of the stable utilization and volatile utilization, will fluctuate upward or downward around the stable utilization.

Based on the above model, we can estimate the stable portion and volatile portion of utilization.



The data requirement for Utilization Rate Analysis includes the following data of each sample commitment:

Data	Data Type	Description	Note
Observation Date	DATE	The observation date	
Facility Code	CHAR	The unique ID of the sample product	Key Word
Product Type	CHAR	The product type: e.g. Undrawn Loan, Trade Guarantee	
Revolve Indicator	CHAR	Revolving or non-revolving	
Revocable Indicator	DATE	Revocable or irrevocable	
Effective Day	DATE	Effective date of the facility	

Expire Day	DATE	Expire date of the facility	
Tenor	NUM	The tenor of the sample product	
Surviving Days	NUM	Number of days from Effective Date to Observation Date	
Total Limit	NUM	Total Limit of the sample product	
Drawdown Amount	NUM	Used Limit as of Observation Date	
Utilization Ratio	NUM	Ratio of Drawdown Amount to Total Limit	

3.4.3. *Peaks and Valleys Analysis*

A common method for the analysis of the line of credit card is the peaks and valleys model. The peaks and valleys of the utilization ratio of overdraft balance in credit cards occur periodically. The valley value of utilization ratio always appears at the repayment date in which the repayment reaches a maximum, while the peak value appears half a month after the valley. The dates for the peaks and valleys and the related credit line utilization ratios are almost stable. Thus, some banks in the industry use peaks and valleys model to deal with the future cash flow for credit card.

Based on the analysis of the historical peak and valley values, the date on which the peak and valley appear in the next 30 days as well as the proportional relationship between the peaks and that between the peaks and the valleys can be determined. Further, the next 30 days can be divided into two time separate sub-periods, corresponding to the ascending and descending of the overdraft balances. The increasing amplitude and decreasing amplitude would be distributed to each day in each sub-period. And then, the peaks and valleys curve is simulated repeatedly until after 12 months.

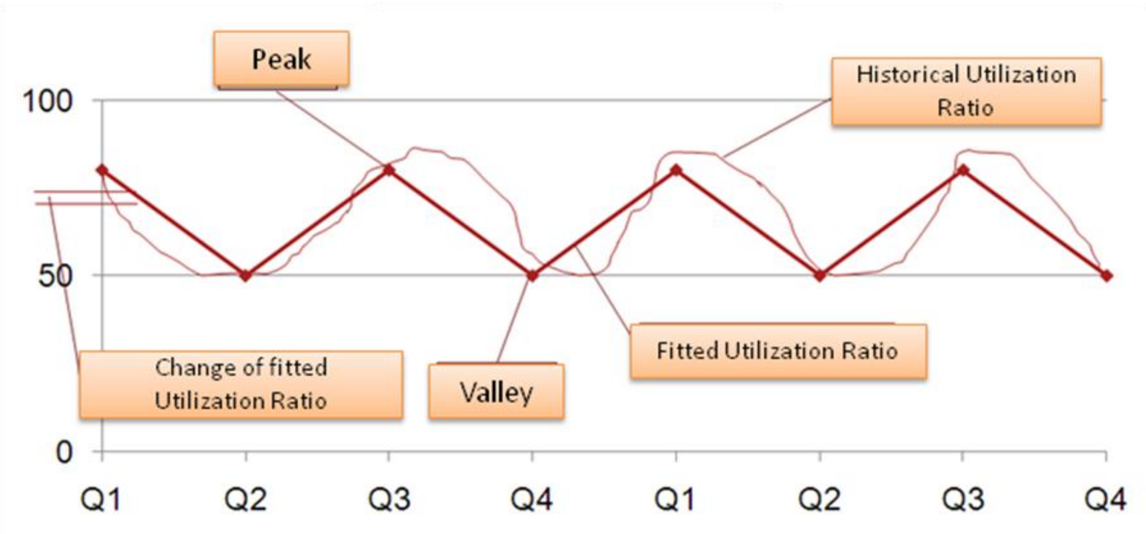
The model parameters obtained are the following: the mode of the dates the peaks and valleys appear and the historical average utilization ratios:

Date	Utilization Ratio for Peaks
D1	X%
Date	Utilization Ratio for Valleys
D2	Y%

Then, the period is divided into two time intervals for ascending and descending according to the days

Calendar Date	1	2	3	…	15	16	17	18	…	30	Total
Size of Fluctuation	%	%	%	%	%	-%	-%	-%	-%	-%	%

The following is the schematic plot for the historical and fitted utilization rate illustrating the peak and valley method:

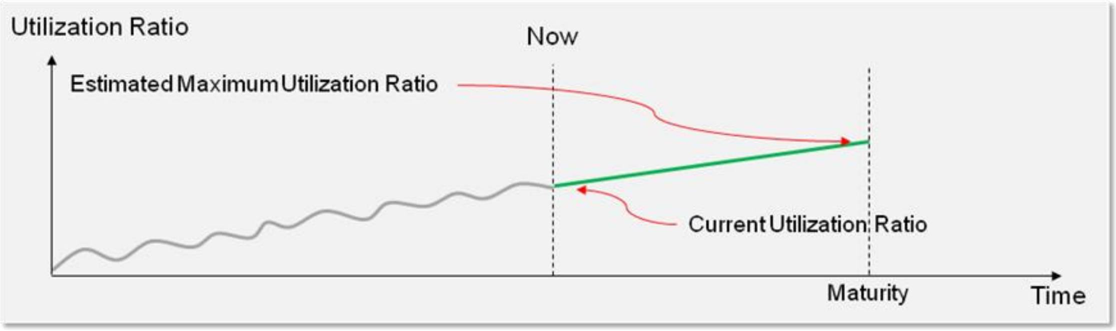


The data requirement for the peaks and valleys analysis includes the following data of each sample commitment:

Data	Data Type	Data Description	Note
Observation Date	DATE	The observation date	
Facility Code	CHAR	The unique ID of the sample product	Key Word
Product Type	CHAR	The product type: e.g. Undrawn Loan, Trade Guarantee	
Revolving Indicator	CHAR	Revolving or non-revolving	
Revocable Indicator	DATE	Revocable or irrevocable	
Expire Date	NUM	The expire date of the facility	
Tenor	NUM	The tenor of the sample product	
Total Limit	NUM	Total Limit of the sample product	
Drawdown Amount	NUM	Used Limit as of Observation Date	
Utilization Ratio	NUM	Ratio of Drawdown Amount to Total Limit	

3.4.4. ***Maximum Utilization Ratio Analysis***

Maximum utilization ratio analysis is similar to the static utilization ratio analysis. The first step is to obtain the historical maximum utilization ratio of each account based on the analysis of historical data, and then the future maximum utilization ratio can be estimated by mathematical methods in accordance with the management's risk appetite. This is to account for the impact of off-balance-sheet items on the on-balance-sheet cash flow based on a maximum utilization ratio, which is defined as the maximum ratio of the drawdown amount to the total limit of an account within its life.



The data requirement for the peaks and valleys analysis includes the following data of each sample commitment:

Data	Data Type	Data Description	Note
Observation Date	DATE	The observation date	
Facility Code	CHAR	The unique ID of the sample product	Key Word
Product Type	CHAR	The product type: e.g. Undrawn Loan, Trade Guarantee	
Revolving Indicator	CHAR	Revolving or non-revolving	
Revocable Indicator	DATE	Revocable or irrevocable	
Effective Day	DATE	Effective date of the facility	
Expire Date	NUM	The expire date of the facility	
Tenor	NUM	The tenor of the sample product	
Surviving Days	NUM	Number of days from Effective Date to Observation Date	
Total Limit	NUM	Total Limit of the sample product	
Drawdown Amount	NUM	Used Limit as of Observation Date	
Utilization Ratio	NUM	Ratio of Drawdown Amount to Total Limit	

3.4.5. *Portfolio Drawdown Rate Analysis*

Portfolio drawdown rate analysis is also similar to the static utilization ratio analysis. The first step is to obtain the historical drawdown rate of the portfolio of each date based on the analysis of historical data, and then the future portfolio drawdown rate can be estimated by mathematical methods in accordance with the management's risk appetite. This is to account for the impact of off-balance-sheet items on the on-balance-sheet cash flow based on a portfolio drawdown rate, which is defined as the maximum ratio of the drawdown amount to the total limit of a portfolio of products without considering the tenor of each account.

Since the date of historical data is considered in this model, we can do more statistical analysis on the data. Besides getting statistics from the distribution of these historical data, a linear regression model can also be used to estimate future drawdown rate, subject to the significance of the model and parameters.

Linear regression is one of the frequently used regression models in which the relationship between the dependent variable and one or more explanatory variables is linear. For the details of linear regression model, refer to Section 3.3.2.

The data requirement for the portfolio drawdown rate analysis includes the following data of each sample commitment:

Data	Data Type	Data Description	Note
Observation Date	DATE	The observation date	
Facility Code	CHAR	The unique ID of the sample product	Key Word
Product Type	CHAR	The product type: e.g. Undrawn Loan, Trade Guarantee	
Revolving Indicator	CHAR	Revolving or non-revolving	
Revocable Indicator	DATE	Revocable or irrevocable	
Undrawn Limit	NUM	Undrawn Total Limit of the sample product	
Monthly Drawdown Amount	NUM	Used Limit in the month as of Observation Date	
Drawdown Rate	NUM	Ratio of Monthly Drawdown Amount to Undrawn Limit	
Macro Economic Factors	NUM/CHAR	The explanatory factors to the drawdown rate in the linear regression model. The factors include macro economy index such as interest rate, GDP, and etc.	

3.4.6. ***Model Comparison***

The advantage of static utilization ratio analysis approach is that the analysis process is simple and straightforward with limited data requirement, as only four key data for each commitment account are needed: the commitment amount (credit line), utilized amount (the amount drawn down for each account), utilization ratio and timeline. The limitation of this method is that it only uses statistics to forecast the future utilization rate which limits the accuracy.

The life cycle utilization ratio approach explore deeper than the static utilization approach by dividing the total utilization into stable one and volatile one and forecasting the future utilization with regression on the age of the commitments. Compared with static portfolio utilization ratio approach, its forecast for the future utilization and cash outflow is more dynamic and with higher predictive power. However, the analysis process is more complicated and the computation of utilization and cash outflow with time scale also brings enormous computing workload. More importantly, the assumption of consistent customers' behavioral pattern along the life cycle has to be verified before using this model.

Peak and valley model fits those products such as credit card and overdraft with clear periodic pattern of drawn down and repayment. For those without such pattern, the model may not be valid.

The maximum utilization ratio analysis approach is the simplified version of the life cycle utilization ratio without considering the pattern of the drawdown time. When comparing to static utilization ratio analysis approach which takes statistics of all the data, this method captures the maximum of every account’s history and then takes the statistics of this resultant data. This result will not be affected by the noise such as intermediate changes of drawdown amount and the length of the loan being drawn. The limitation of this method is that it only uses statistics to forecast the future utilization rate which limits the accuracy and it assumes the amount of drawdown and time of drawdown among clients are evenly distributed.

The advantage of portfolio drawdown rate analysis approach is that the result can be applied on portfolio level without considering the tenor or life of each account. And a regression model can be used to project a forward-looking drawdown rate instead of an unchanged rate. The limitation is that it does not differentiate between commitments of different tenors and the time of life, and does not consider the behavior of each individual but as a whole in portfolio level.