Bank of Korea consumer credit panel: a new statistical initiative for financial stability[1]

Mira Kim,
Bank of Korea

---

[1] This paper was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

BOK Consumer Credit Panel:[1]

A New Statistical Initiative for Financial Stability

Mira Kim[2]

Abstract

Since the 2007-2008 financial crisis, the statistical value of personal credit information has been increasing and major countries have been compiling statistics that make use of personal credit information. The BOK also uses personal credit information to create new financial statistics to analyze systemic risk and solve the household debt problem. With this motivation, we created the BOK Consumer Credit Panel (BOK CCP), which includes detailed information on consumer debt and credit.

This panel can help to measure and analyze the levels of and changes in individual liabilities, derived from consumer credit reports, to track individuals' and households' access to debt and credit every quarter. The data collected from 2012Q1 comprises identifying information (age, sex, income, address, etc.), loans, credit cards, payments in arrears, and defaults for a random sample of one million persons (2.4% of the total population).

The BOK CCP includes individuals' financial information, which can be sensitive in terms of data security. Careful consideration should be given to privacy protection issues and appropriate data sharing. Information is managed by stages from collecting to reporting and sharing. In addition, efforts are made to reduce the misuse of data and to increase the reliability and accuracy of data.

This paper explains the motivation for setting up the BOK CCP, the sample design, the BOK CCP content, data confidentiality protection and database usage in the BOK for financial stability.

Contents

1. Introduction

With microdata being needed to address data gaps after the 2007-2008 financial crisis, the statistical value of personal credit information has been increasing. The interest in microdata has increased in recent years, and due to its popularity, the availability of microdata also has gradually increased during that time. Microdata has been expected to play an important role in financial stability analysis, including analysis of the vulnerability of the household sector, which is the main cause of financial crises.

The current financial statistics are compiled as part of the conduct of monetary policy but are not sufficient quantitatively or qualitatively for the purpose of macro-prudential policies. This is the main reason that microdata has been spotlighted. It is increasingly recognized that statistically edited microdata derived from individual credit information held by credit registers can help generate new financial statistics (household credit by age, credit score or income) and also strengthen statistical bases that can be utilized for macro-prudential assessment and research.

In the United States, the FRBNY created a new quarterly panel dataset called the Consumer Credit Panel (CCP) in 2010 based on personal credit information to analyze household-level debt and credit. The CCP is used for research and its value is highly evaluated, although the United States has similar statistics such as the Survey of Consumer Finance and Flow of Funds.

Referring to the FRBNY CCP as an example, the Bank of Korea (BOK) established the Bank of Korea Consumer Credit Panel, (BOK CCP). We were tasked with a financial stability role by the revised Bank of Korea Act of 2011, and with this responsibility, we found it necessary to obtain individual credit data for conducting macro-prudential policies and addressing the household debt problem.

Besides traditional financial statistics and processed aggregated data, the use of micro-information can be also useful in terms of its diversity. The aggregated data derived from the balance sheets of financial institutions offers only limited information.

Moreover, the BOK CCP can help to measure and analyze the levels of and changes in individual liabilities to track individuals' and households' access to debt and credit every quarter. This makes it possible to acquire the same individual's longitudinal information in the long term. The data collected from 2012Q1 comprises identifying information (age, sex, income, address, etc.), loans, credit cards, payments in arrears, and defaults for a random sample of one million persons (2.4% of the total population).

The BOK CCP includes individuals' financial information, which can be sensitive in terms of data security. Careful consideration should be given to privacy protection issues and appropriate data sharing. Information is managed by stages from collecting to reporting and sharing. In addition, efforts are made to reduce the misuse of data and to increase the reliability and accuracy of data.

This paper is organized as follows: Sections 2 and 3 describe the sample design and the BOK CCP content. Section 4 explains data confidentiality protection and database usage in the BOK. Finally, section 5 provides conclusion.

2. Sample design

2-1. Target population

The target population of the BOK CCP is all residents at least 18 years of age with credit histories. Most individuals start to build their credit histories when taking out loans, obtaining and using credit cards or retail cards, and so on.

The population excludes minors, some seniors without credit histories, those who do not perform credit activities in their own name. For those over 90 years of age, only those who maintain currently valid credit transactions are included in the target population, while those who have past credit histories but no currently valid transactions are excluded.

The target population also includes individuals with thin-files, who have had few loans or credit cards, while it excludes consumers with no-file and with inquiry-only files. Many individuals with inquiry only files have been registered by going though credit inquiry when taking out a loan or opening a cellular phone.

The table 2.1 below provides the information on individuals by credit activities. As of the end of 2014, there were 41 million individuals who have credit histories, which is the target population. It was approximately 6 million for those who did not have sufficient credit histories such as individuals with no-file or inquiry-only files.

Table 2.1: Individuals by Credit Acitivity[1]   Table 1

Individuals (50 million)		
Under 18 years of age	Over 18 years of age (47 million)	
	With credit history[2]	With insufficient credit history
(3 million)	(41 million)	(6 million)

Notes: ① As of end-2014

② It is the target population

Sources: NICE Information Service

An individual-level credit database obtained from the NICE Information Service, a credit bureau, is used for the sampling frame. Table 2.2 shows the similarity of the target population and the estimated population projected by Statistics Korea. As you can see in the table, the target population from the credit bureau is 99.5% of the estimated population of Statistics Korea at the end of 2014. The estimated population by age group differs somewhat between the credit bureau and Statistics Korea. It appears that the number of database registrants in the 20-69 age groups is slightly larger than the estimated populations for those age groups of Statistics Korea, mainly due to Koreans living overseas.

For the estimated population compiled by Statistics Korea, people residing overseas for 90 days or longer due to emigration are classified as non-residents. However, since the BOK CCP includes long-term credit histories, non-residents are not immediately excluded. This is the main reason for the difference between the target population and the estimated population.

Table 2.2: Comparison of age groups

between the target population and estimated population Table 2

Age group	Credit history holders[1] (A)	Estimated population[2] (B)	Ratio(A/B)
18-19	493 (1.2)	1,300 (3.2)	37.9
20-29	6,913 (16.8)	6,775 (16.4)	102.0
30-39	7,985 (19.4)	7,779 (18.9)	102.6
40-49	9,007 (21.9)	8,514 (20.6)	105.8
50-59	8,346 (20.3)	7,949 (19.3)	105.0
60-69	4,830 (11.8)	4,584 (11.1)	105.4
Over 70	3,495 (8.6)	4,359 (10.6)	80.2
Total	41,070 (100.0)	41,260 (100.0)	99.5

Notes: ① NICE Information Service (as of end-2014)

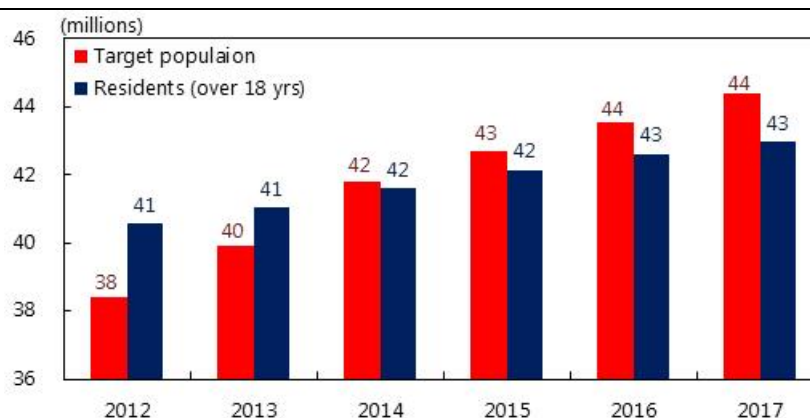② Statistics Korea (as of July 1, 2014)

③ Figures in ( ) represents proportions

④ thousand, %

Figure 2.1 illustrates the increase of the target population from 38 million in 2012 to 44 million in 2017, showing a faster increase than the estimated population growth. To be specific, the target population, represented by red bar, has exceeded the estimated population represented by blue bar since 2014. Meanwhile, the target population, the red bar, had been lower in 2012 and 2013.

Figure 2.1:

Comparability between target population and estimated population[1] Figure 1

Notes: ① The estimates are based on the 2015 Census(1.11.2015) and reflect changes to the July 1, 2015 population

Sources: the BOK CCP, Statistics Korea

The target population has increased sharply and has exceeded the number of residents aged 18 or older since 2014. This is mainly due to Koreans residing overseas. Koreans residing overseas are excluded from the estimated population from Statistics Korea after 90 days, while they are still included in the BOK CCP approximately for five years.

2-2. Sampling procedure

The sampling procedure is carried out using a method called "simple random sampling". The advantages of this sampling method are that it can guarantee that each individual is chosen entirely randomly with the same probability to extract and its simplicity can also make sampling easier. For these reasons, simple random sampling is suitable for big data management.

We generated the longitudinal data by selecting a specific number of random numbers created using a person's birth date from the individual information and extracting the sample; panel data can be obtained to track individuals' and households' access to debt and credit every quarter. In other words, we are able to track the same individuals in each quarter. On average, 99% of the sample remains unchanged from the previous quarter.

This sampling procedure creates a 2.4% random sample, which is representative of the target population. The sample size is automatically held constant by adjusting the number of individuals who enter and exit the target population. The sampling procedure maintains the representativeness and stability of the sample by automatically adjusting the flows in and out of the population. In the sample, the flows in and out of the total population are the same. Figure 2.3 gives that the sample can be seen to be increasing in proportion to growth in the target population.

Figure 2.3: Target population and sample size

Figure 2
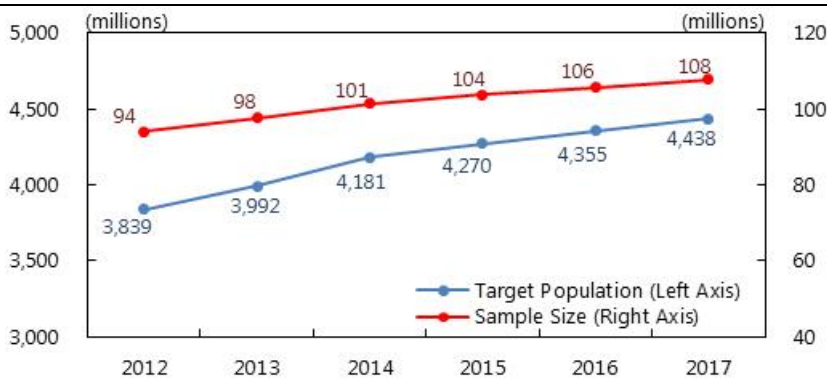
Sources: NICE Information Service, the BOK CCP

In addition, the target population and sample of the BOK CCP include those who have their own credit histories, but are not necessarily borrowers. Even if an individual does not presently have any loans, he can remain in the population with his credit history by having a history of past overdue payments or borrowing. At the end of 2017, the sample size is 108 million individuals, while borrowers with outstanding household loans are 46 million, which is 42.3%

Figure 2.4: Sample size and borrowers

Figure 3
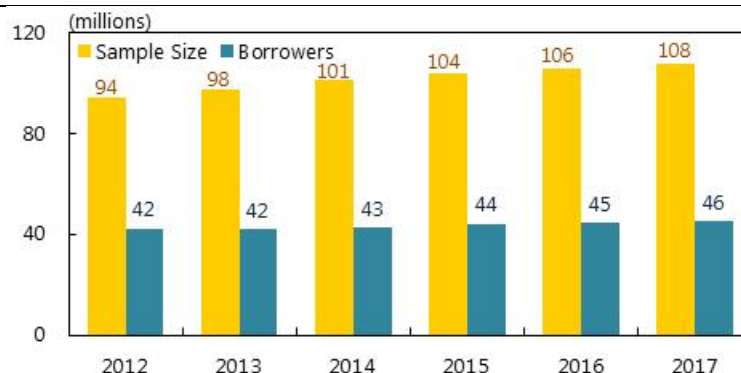
Sources: the BOK CCP

2-3. Sampling results

The sampling procedure provided the sample which represents 2.4% of the individuals with credit histories. In order to verify the adequacy of the procedure, we examined the results by observing the same characteristics of the sample and the target population in terms of age, region and credit score.

The sample proportion (2.4%) appeared very similar to the target population in terms of key variables (age, region and credit score), confirming that the sampling procedure resulted in a representative random sample of individuals.

Table 2.3: Sample proportions by age, region and credit rating

Table 3

Age		Region		Credit Rating	
20~29	2.5	Seoul	2.4	1st decile	2.4
30~39	2.4	Busan	2.4	3rd decile	2.4
40~49	2.5	Daegu	2.4	5th decile	2.5
50~59	2.4	Incheon	2.4	7th decile	2.4
60~69	2.4	Gwangju	2.4	9th decile	2.4
Total	2.4	Total	2.4	Total	2.4

Notes: ① 1ST decile includes individuals with the highest credit score.

② %

Sources: NICE Information Service (as of end-2014)

As the BOK CCP is a panel database that samples 2.4 % of the target population with credit histories, it can be used to track the credit activities of individuals and to identify changes in distribution by characteristics of borrowers. However, since the sum of samples cannot be used as they are, a parameter estimation process is definitely necessary. The simple method for parameter estimation is to multiply the statistics by a constant (the reciprocal of the sample proportion, 1/0.024). Estimates of population aggregates can be acquired simply in this way.

We compared the estimates of population obtained by multiplication with the target population, and found that the differences were not significant.

Table 2.4:

Comparability between sample statistics multiplied and population　Table 4

	Sample statistics (multiplied, A)	Population (B)	Ratio (A/B)
Total (trillion)	1,120.9	1,126.1	0.995
Borrowers(thousand)	17,655	17,629	1.001
(Seoul)	7,811	8,064	0.969
(Age; 40~49)	9,156	9,007	1.016
(Credit rating; 5th decile)	7,110	6,996	1.016

Sources: NICE Information Service (as of end-2014)

3. The BOK CCP content

The data is collected from 2012Q1 every quarter. The BOK CCP is broadly composed of five categories: individual background information, loans, accounts, cards, and overdue payments and defaults. The key variables in these categories are described in detail below.

Table 3.1: Structure of the BOK CCP

Table 5

Categories	Key variables
Individual background information	Consumer identification number, age, birth, sex, region, credit score, income (estimate), etc.
Loans	· Household debt: financial institution, loan type, amount, total number of lenders that borrowers have taken out loans from, total number of accounts, origination, etc. · Commercial loans: industry, financial institution, loan type, purpose of loan, amount, maturity, etc.
Accounts	Financial institution, loan type, purpose of loan, payment method, first date of borrowing, amount, maturity, etc.
Cards	Credit limit, total amount, revolving balance, cash advance, etc.
Overdue payments and defaults	Amount registered in arrears, amount pending or in default, origination registered in arrears, etc.

Sources: the BOK CCP

Income (estimate)

The income information is generated using the evidence that individuals submitted to the financial institution when they applied for or took out a loan. If there is little evidence available, the credit bureau (NICE Information Service) estimates income. The estimate is based previous evidence of income, card records, occupation and real estate information.

Credit score

The credit score is a credit rating that evaluates an individual's credit risk by assessing repayment capacity and measuring the probability of default. A credit score of 0 occurs only for those who are not subjects of a credit rating (under 18 years or over 90 years of age) or for those whom it is difficult to evaluate due to insufficient credit information. However, the ratio of consumers with a credit score of 0 to the total sample is only 0.54% in 2018Q1.

Credit rating is independently determined by the credit rating agency (in the case of the BOK CCP, NICE information Service), which considers various information such as repayment history information (late payments, etc.), level of debt (loans and cards), and type of credit information (financial institution and product), etc.

Residence/address

Residential information is the address listed on an individual's credit report, and it offers detailed information on small administrative divisions, showing the first 3 digits of the postal code. Figure 3.1 describes the distributions of individuals and household debt by region. The first graph shows the similarity between individuals from the BOK CCP and residents (at least 18 years old).

The second graph illustrates the distribution of household debt by region, showing the difference between the statistics from the BOK CCP and the aggregated statistics; Credit to households. The difference between two statistics is larger in Incheon&Gyeonggi area and Seoul. This is because many individuals, dwelling in Incheon or Gyeonggi area, take part in economic activities in Seoul, the capital of Korea. The statistic on 'Credit to Households' provides regional information based on the location of financial institution branches. Therefore, if the address where an individual resides and the location of the branch at which the individual borrowed are not consistent, the total loan balance by region can be different. You also can confirm that, in Incheon&Gyeonggi, the household debt from the BOK CCP is larger than the aggregate, while it is even lower in Seoul.

Figure 3.1: Individuals and household debt by region[1]

Figure 6
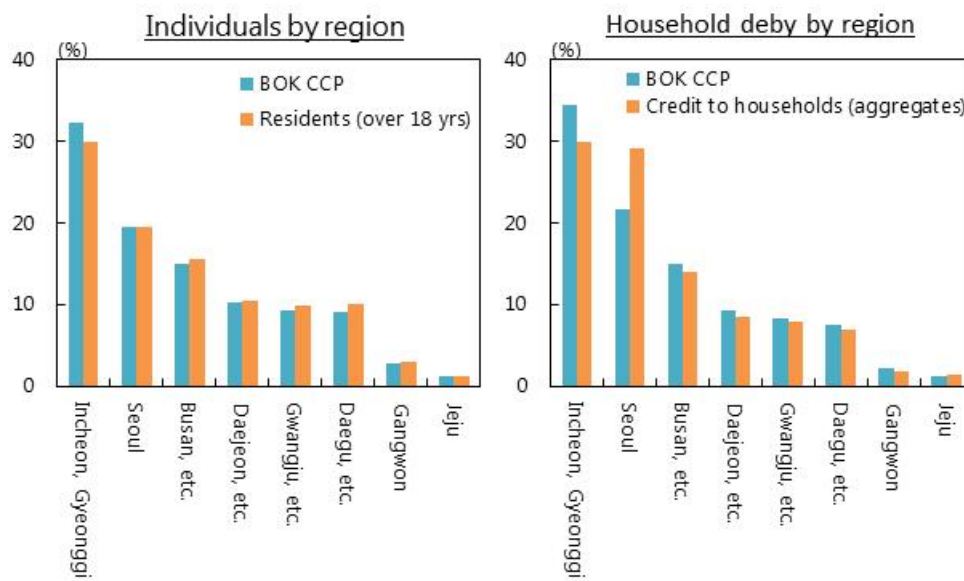
Notes: ① as of 2018Q1
Sources: the BOK CCP, Statistics Korea, the Bank of Korea

Loan amount

We can acquire data on loan amount with 9 digits, but since the unit value is 100,000 won, the amount can go up to 100 trillion won, meaning that there is, in effect, no cap. Amounts of less than 100,000 won will be written as "0". Thus, you cannot interpret "0" as actually zero.

In addition, an outlier can be observed because the sample size is large (about 1 million). If it is observed that an outlier is large enough to interfere with parameter estimation, the outliers can be excluded for accurate estimation. However, we also provide guideline on outliers for appropriate usage. Also, the data is available on request.

Meanwhile, a limited loan, such as an overdraft, can be counted on the basis of the limit and therefore differ from the actual loan balance.

Loan type

The BOK CCP provides detailed information on household debt by type, compared to the aggregated statistics provided by 'Credit to Households' and 'Flow of Funds'. In the BOK CCP, loans are classified in more details (mortgage, credit card, student loan, auto loan, etc.), while statistic on 'Credit to Households' is classified into two loan types (mortgage and others). Figure 3.2 shows the total debt balance and its composition from the BOK CCP. A secured loan included a loan in which collateral is savings, securities, etc. But, a secured loan in which collateral is a property such as home belongs to Mortgage, which makes up the largest proportion of total household debt.

4. Data confidentiality protection and database usage in the BOK

4-1. Data confidentiality protection

As interest in personal information security has increased in recent years, we felt it necessary to approach the personal identifiable information sensitively. It was carefully reviewed whether the BOK CCP data could be interpreted as personal information subject to the Personal Information Protection Act. Since it can be considered as personal credit information, special security measures are needed.

De-identification method

The important issue in the data confidentiality protection is the possibility of person identification and personal information leakage. NICE information service assigned each consumer the unique number, a Consumer identification number, instead of an individual's social security number. These consumer identification numbers in the BOK CCP are unique to each individual, who can easily be identified by combining the consumer identification numbers with other information from the credit bureau. Also, individuals can be personally identified by combining the consumer identification numbers with external information.

Therefore, de-identification measures are required. In consideration of the data needed to analyze household debt, we adopt measures such as categorization or aggregation. The goal to prevent an individual's privacy problem can be achieved by obtaining unidentified raw data using total processing, data value deletion, and categorization.

Table 4.1:

De-identification measures   Table 6

Variables		Measures
Cards	The number of card issuers, the card issue date, debt-to-limit ratio	Not obtained
Overdue	The first and last date when an individual is registered in arrears	Not obtained
Default	The first and latest date of public record (bankruptcy, tax liens, etc.)	Categorization (month, not day)
Account	The date of account creating, expiration date	Categorization (month, not day)
Region		Categorization
Debt Amount		Aggregation (100,000 won)

Sources: the BOK CCP

The BOK CCP is a large data panel that tracks credit activities such as loans, late payments, and card use of approximately 1 million individuals every quarter. Thus, in addition to de-identification measures, additional measures are also required to prevent a leak of the raw data.

As a principle, we utilize a system that does not allow for the downloading of raw data, in order to prevent a re-identification of personal information in the database. Therefore, users who want to obtain the data set for analysis are able to process and manipulate it only within the given system. It is impossible for raw data regarded as personal information to be transferred to a user's personal computer. These measures are adopted for the purpose of data confidentiality. However, even though it is impossible for users to download raw data directly to their personal computers, we do allow analysis results to be taken out of the system and transferred to users' personal computers.

4-2. Database usage in the BOK

This longitudinal data can be used in the analysis of topics such as the distribution of household liabilities and the dynamic changes of vulnerable groups, which were difficult to analyze with existing statistics. In other words, the BOK CCP can complement existing aggregated statistics to give a different method of analysis. However, the BOK CCP contains a vast amount of microdata, and it is not possible to completely avoid a variety of problems such as missing data, errors and outliers in the process of credit information registration in financial institutions. Therefore, data cleansing is necessary for usage.

Data cleansing

Credit information is fundamentally based on the information registered individually and independently by financial institutions. Thus, the process of data cleansing is conducted in a manner that follows its source. Even after a basic verification process, there may occasionally be outliers in infrequently used datasets. In such a case, it will confirm with the credit bureau whether the data were registered incorrectly or whether they are just simple outliers. After the data is revised, a notice will be provided to users to prevent misleading analytics. Users have discretion to decide how to deal with these outliers in their analyses, and we just provide guidelines for user convenience.
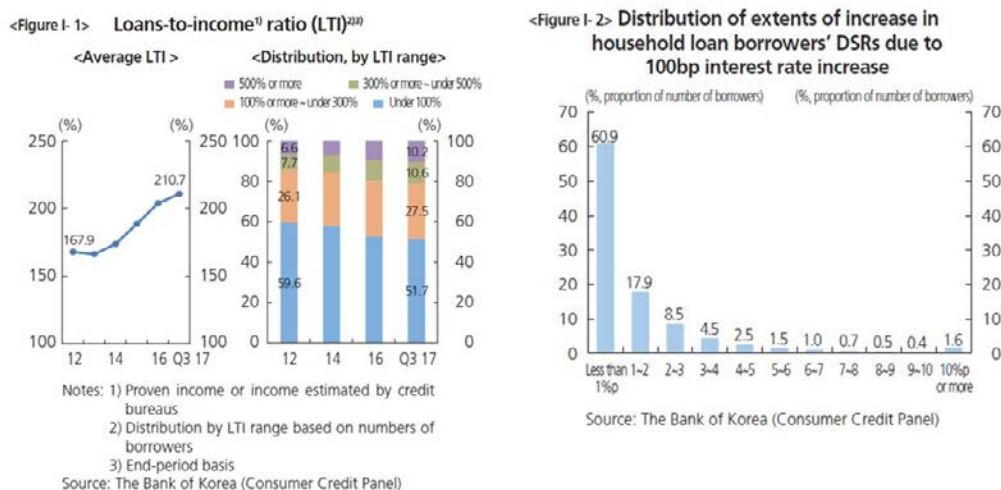
Financial Stability Report

The individual-level information enables the examination of the current distribution of total debt balance and loan origination specifically by age, income, credit rating and region. The structure of household debt can also be determined. Moreover, it is possible to identify the characteristics of the vulnerable (age, region, income, debt structure, etc.) and analyze the dynamic changes of multiple-debt-holding borrowers.

The BOK CCP is, in fact, actively used in the BOK in various ways and the most significant example is a "Financial Stability Report". The Bank of Korea has published the report on a biannual basis as part of its conduct of macro-prudential policies. Since December 2015, the Consumer Credit Panel data has been used for analysis in the report.

Microdata in the BOK CCP is used in the case of household debt, analyzing and assessing the potential risks inherent in the Korean financial system. Figure 4.1 below is a part of the Financial Stability Report (Dec. 2017) analyzing household debt using the loans-to-income ratio and DSR.

Figure 4.1: Financial Stability Report (Dec. 2017)

Figure 7

Sources: the Bank of Korea; Financial Stability Report (Dec. 2017)

5. Conclusion

Unlike macroeconomic statistics, the BOK CCP contains individual-level information, thus allowing for flexible analysis tailored to the purpose of use. In addition, it has advantages over conventional survey panels in terms of coverage, sample size, timeliness, frequency, and reliability.

However, it also has some limitations, despite the fact that the data is very timely and extensive. First of all, the information is presented not at the household level but at the level of the individual. It is important to obtain information on households, because the basic economic unit is a household rather than an individual, and household data enables the analysis of matters such as the transfer of wealth between generations.

The BOK CCP also lacks information indicating the individual's demographic characteristics. Information such as occupation, education level, and marital status is limited. Despite the information on liabilities, it is difficult to perform a comprehensive study of financial activity that takes both assets and liabilities into account. This is due to the lack of information on assets, especially housing tenure, housing types, tangible assets and financial assets.

In spite of these limitations, the BOK CCP, which has been constructed across all types and financial institutions, will be available for various analyses and will also be used for further analysis, by merging with other panel data or adding other important variables in the future.

However, the database is a major asset of the Bank of Korea and is growing in value, as it reflects accumulated data management know-how and methods. It has been actively used for analyzing and addressing the household debt issue. The more time series are accumulated and the more analysis capabilities are improved, the greater the value of the data is expected to become.

References

Donghoon Lee, Wilbert van der Klaauw (2010): An Introduction to the FRBNY Consumer Credit Panel; Federal Reserve Bank of New York Staff Report no. 479

BOK (Dec. 2015): Financial Stability Report, pages 37 to 40. Retrieved from https://www.bok.or.kr/eng/bbs/E0000737/list.do?menuNo=400042

BOK (Dec. 2017): Financial Stability Report. Retrieved from https://www.bok.or.kr/eng/bbs/E0000737/list.do?menuNo=400042

Ninth IFC Conference on "Are post-crisis statistical initiatives completed?"

Basel, 30-31 August 2018

Bank of Korea consumer credit panel:
a new statistical initiative for financial stability[1]

Mira Kim,
Bank of Korea

BOK Consumer Credit Panel

- A New Statistical Initiative for Financial Stability

MiRa, Kim

The Bank of Korea

Aug. 2018

THE BANK OF KOREA

Overview

- Introduction

- Sample design

- The BOK CCP content

- Database usage in the BOK

- Conclusion

Introduction

- With microdata being needed to address data gaps after the 2007-2008 financial crisis, the statistical value of personal credit information has been increasing.

- The current financial statistics (aggregated data) are compiled as part of the conduct of monetary policy, but are not sufficient quantitatively or qualitatively for the purpose of macro-prudential policies.

- Moreover, individual-level information has been expected to play an important role in financial stability analysis, including analysis of the vulnerability of the household sector, which is the main cause of financial crises.

- With this motivation, we established the BOK Consumer Credit Panel, which can help to measure and analyze the levels of and changes in individual liabilities to track individuals' and households' access to debt and credit.

Sample design

Target population

- The target population of the BOK CCP is all residents at least 18 years of age with credit histories.

Individuals by Credit Acitivity[1]

Individuals (50 million)		
Under 18 years of age	Over 18 years of age (47 million)	
	With credit history[2]	With insufficient credit history
(3 million)	(41 million)	(6 million)

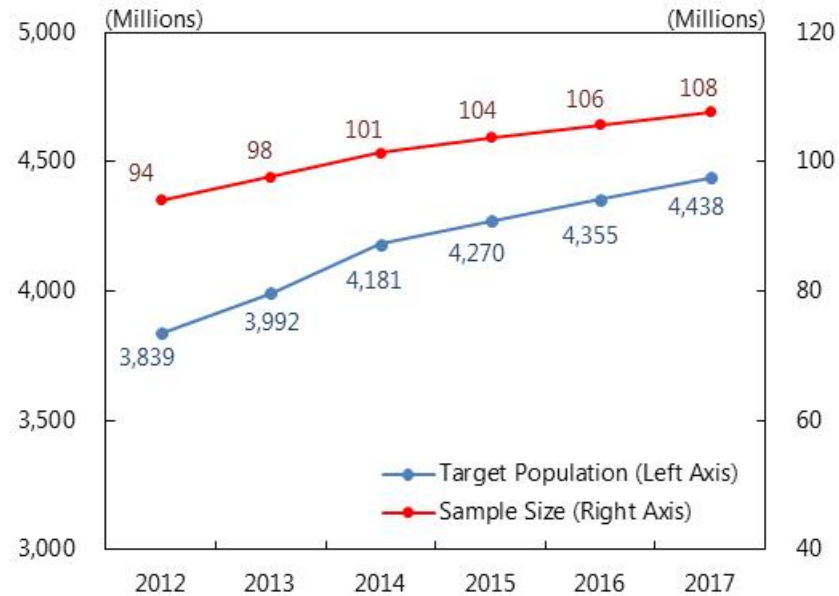Notes: ① As of end-2014

② Target population

Sources: NICE Information Service
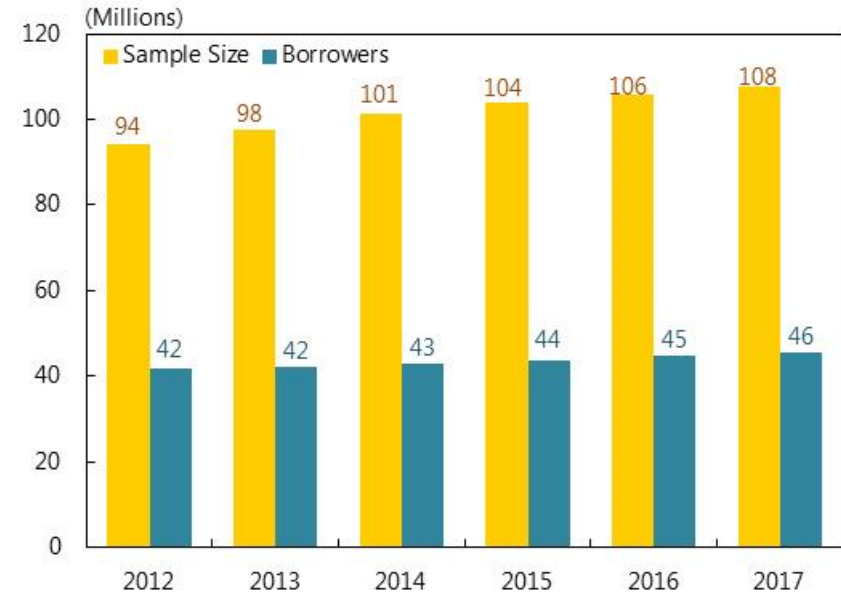
Sample design

Sampling procedure

- Simple random sampling; by selecting a specific number of random numbers generated using a person's birth date from the individual information.

Target population and sample size

Sample size and borrowers

The BOK CCP content

| Categories	Key variables
Individual background information	Consumer identification number, age, birth, sex, region, credit score, income (estimate), etc.
Loans	· Household debt: financial institution, loan type, amount, total number of lenders that borrowers have taken out loans from, total number of accounts, origination, etc. · Commercial loans: industry, financial institution, loan type, purpose of loan, amount, maturity, etc.
Accounts	Financial institution, loan type, purpose of loan, payment method, first date of borrowing, amount, maturity, etc.
Cards	Credit limit, total amount, revolving balance, cash advance, etc.
Overdue payments and defaults	Amount registered in arrears, amount pending or in default, origination registered in arrears, etc.

Sources: The BOK CCP

Data usage in the BOK

- The BOK CCP can complement existing aggregated statistics to give a different method of analysis.
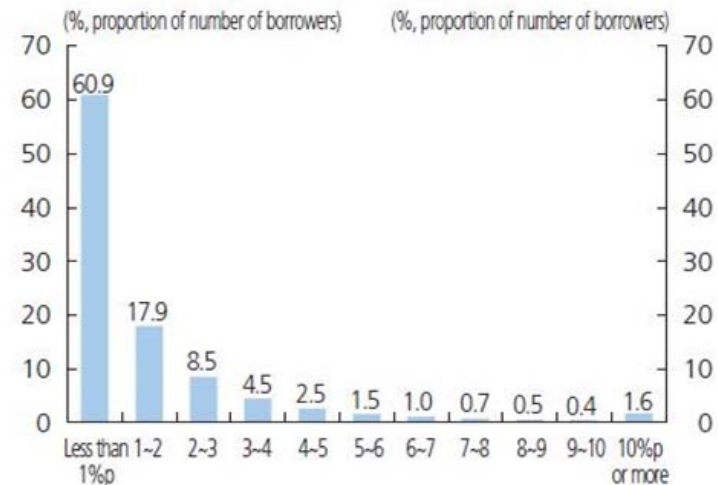
Financial Stability Report

<Figure I-1> Loans-to-income[1] ratio (LTI)[2][3]

<Average LTI> <Distribution, by LTI range>
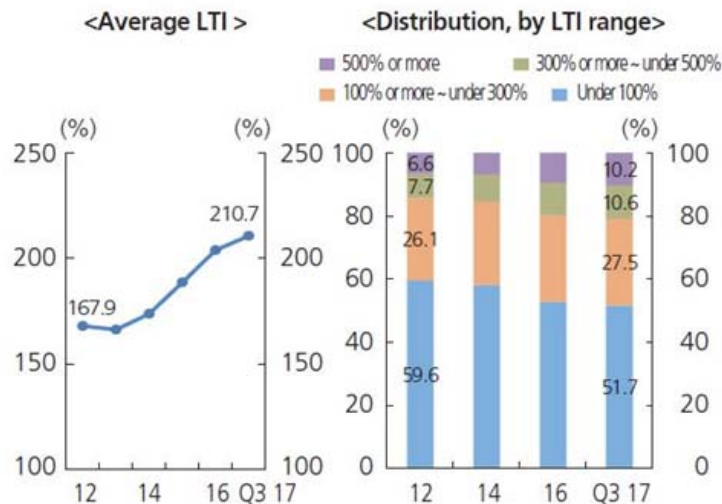
Notes: 1) Proven income or income estimated by credit bureaus
2) Distribution by LTI range based on numbers of borrowers
3) End-period basis
Source: The Bank of Korea (Consumer Credit Panel)

<Figure I-2> Distribution of extents of increase in household loan borrowers' DSRs due to 100bp interest rate increase

Source: The Bank of Korea (Consumer Credit Panel)

Conclusion

- Unlike macroeconomic statistics, the BOK CCP contains individual-level information, thus allowing for flexible analysis tailored to the purpose of use.

- However, it also has some limitations, despite the fact that the data is very timely and extensive. First of all, the information is presented not at the household level but at the level of the individual.

- The BOK CCP also lacks information indicating the individual's demographic characteristics such as occupation, education level, and marital status.

- In spite of these limitations, the BOK CCP, which has been constructed across all types and financial institutions, will be available for various analyses and will also be used for further analysis.

- The more time series are accumulated and the more analysis capabilities are improved, the greater the value of the data is expected to become.