Federal Reserve Bank of New York
Staff Reports

An Introduction to the FRBNY Consumer Credit Panel

Donghoon Lee
Wilbert van der Klaauw

Staff Report no. 479
November 2010

An Introduction to the FRBNY Consumer Credit Panel

Donghoon Lee and Wilbert van der Klaauw

Federal Reserve Bank of New York Staff Reports, no. 479
November 2010
JEL classification: D14, E01, C83, C82

Abstract

In this paper, we introduce the FRBNY Consumer Credit Panel, a new longitudinal database with detailed information on consumer debt and credit. The panel uses a unique sample design and information derived from consumer credit reports to track individuals' and households' access to and use of credit at a quarterly frequency. In any given quarter ranging from the first quarter of 1999 to the present, the panel can be used to compute nationally representative estimates of the levels and changes in various aspects of individual and household liabilities. In addition to describing the sample design, the use of sample weights, and the credit report information included in the database, we provide some comparisons of population statistics and consumer debt estimates derived from our panel with those based on data from the American Community Survey and the Flow of Funds Accounts of the United States.

Key words: household credit, consumer debt, mortgage liabilities, panel data

---

Lee, van der Klaauw: Federal Reserve Bank of New York (e-mail: donghoon.lee@ny.frb.org, wilbert.vanderklaauw@ny.frb.org).

1. Introduction

The primary sources for current information on consumer debt and credit consist of aggregated data from the Board of Governor's Flow of Funds, and loan level data on mortgages from databases such as HMDA, First American CoreLogic and LPS. While containing detailed individual loan-level information on loan, borrower and property characteristics, the latter databases have several important shortcomings for analyzing mortgage loan behavior. First, they often only contain information for a selected sample of loans, such as securitized non-prime loans. Second, they do not longitudinally link individual-specific mortgage loans such as those associated with a refinance or a move-induced sale and purchase of a new home, nor do they link at the individual level across mortgage loans including all first and second mortgages and home equity lines of credit. Third, these databases only provide information on an individual's mortgage debt, while one ideally would like to connect these to other types of debts, including debt on credit cards, auto loans and student loans. Fourth, while consumer income and assets are commonly measured and analyzed at the household level, consumer debt traditionally has been measured at the individual level. For studying household economic behavior, and in particular for understanding life-cycle household finance, it would in fact be more informative to examine debt and credit aggregated at the household level as well as to consider the distribution of debt within households.

To address these shortcomings of existing data sources and to meet a rising need for up-to-date high quality information on household debt and credit at an ongoing basis, a research team at the Federal Reserve Bank of New York, with support from colleagues at the Board of Governors and the Federal Reserve Bank of Philadelphia, set out to create a new and unique quarterly panel dataset based on information contained in individual credit reports.[2] In addition to all mortgage loans held by individuals, this new database was to also include detailed measures of non-mortgage debt. Moreover, our main objective was to create a longitudinal panel of individuals that tracks their access and use of credit at a quarterly frequency from 1Q 1999 to the present, with future updates added each quarter. The goal was for the panel to constitute a nationally representative random sample of individual consumers in any given quarter. This requires a sampling approach that generates the same entry and exit behavior as present in the population, with young individuals and immigrants entering the sample and deceased individuals and emigrants leaving the sample each quarter at the same rate as in the U.S. population. Finally, we required our panel to provide the ability to analyze consumer debt at the household level. In fact, it was our aim to implement a sample design that would result in a longitudinal panel of households, which for any given quarter could be used to compute nationally representative estimates of household-level debt and credit.

[2] Bob Avery, Ken Brevoort, Glenn Canner, Larry Cordell and Bob Hunt all provided valuable suggestions, as did our colleagues Andy Haughwout and Joe Tracy.

In what follows we first discuss the sample design we developed to achieve the goals outlined above. We define the target population, discuss our sampling procedure and appropriate sample weights, and provide some comparisons of population statistics derived from the panel with those based on data from the American Community Survey. This is followed by a brief description of the credit report information included in our database and a comparison of overall U.S. household debt levels computed using the panel with those computed as part of the Flow of Funds Accounts.

2. Sample Design

2.1 Target Population of Individuals

Credit reporting agencies compile and maintain credit histories for all U.S. residents who have applied for or taken out a loan. Credit bureaus continuously collect information on individual consumers' debt and credit from lenders and creditors. Most individuals begin building a credit history when they first obtain and use a credit or retail card or take out a student loan, usually when they are at least 18 years of age. New immigrants with little or no credit history from their home country are often older when a credit file is first created for them.

Our use of credit report data implies that the target population for our analysis consists of all US residents with a credit history. In addition to most individuals younger than 18, who had little need or opportunity for credit activity, the target population excludes individuals who have never applied for or qualified for a loan.[3] While the target population excludes no-file consumers, it does include individuals with so-called thin files (containing only one or two trades or accounts) as well as individuals whose credit file only consists of a collection or public record item (such as bankruptcy) or only contains authorized user accounts or closed accounts.[4] However, we exclude individuals with inquiry-only files. Many inquiry-only files are associated with incomplete or invalid information on the individual's identity, which therefore could not be linked to existing files.

Finally, for reasons described below, our primary target population will be further restricted to individuals whose credit file contains the individual's social security number (SSN). Some individuals do not have a social security number (for example, some resident aliens are not eligible to obtain one) or did not report it to any of their lenders. We will discuss the likely

[3] The latter may include currently or previously married individuals who rely on their spouses to manage the household finances and have never borrowed money in their own names. It also includes groups who are culturally averse to credit use, including retirees and ethnic groups that distrust banks and other credit granters. According to Fair Isaac Corporation (FICO) in 2006 approximately 22 million of the nation's 220 million adults did not have a credit history (Jacob and Schneider, 2006).

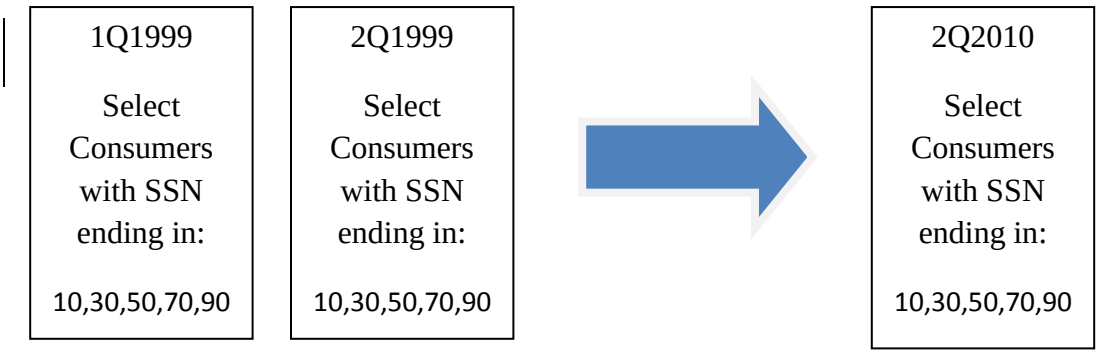
[4] Closed accounts remain on credit reports for up to 7 to 10 years after their closing. Therefore, our panel includes those with no recent credit activity, such as in the past 24 months, but with credit activity in the past 10 years.

proportion of these individuals in the population, as well as their characteristics in more detail below.

2.2 Sampling Procedure

2.2.1 Random Sample of Individuals

Our sampling procedure is based on the fact that the last 4 digits of the social security number are serial numbers which are assigned sequentially to new social security number applicants (in chronological order as applications are processed) and can therefore be treated as randomly assigned.[5] More specifically, we asked the credit bureau to only include individuals in our sample whose SSN ended with one of 5 arbitrarily picked two digit combinations. For illustration purposes, consider the two-digit combinations 10, 30, 50, 70 and 90. This sampling procedure in each quarter generates a 5% random sample that is representative of all individuals in the US who have a credit history and whose credit file includes the individual's social security number. By using the same set of five two-digit combinations in each quarter (a digital selection pattern that remains constant), we obtain a panel that satisfies one of our other main objectives: observing the same individuals in each period over time. At the same time, our sample remains representative of the target population in each quarter as our sampling approach automatically captures the new flows into and out of the target population. It thereby circumvents a major shortcoming of static panels in which a fixed cohort or set of cohorts is longitudinally tracked, and which therefore over time become increasingly non-representative of the overall population.[6]



We verified the randomness of the last two digits of the social security numbers in several ways. First, we confirmed that the sampling procedure resulted in a sample that constituted 5% of the credit bureau's total population of eligible individuals (with credit file that

[5] Note that this is not the case for the first 5 digits, which are assigned based on the location and date at which the social security number was assigned.

[6] In our panel, the only attrition that occurs is due to an individual's death or when someone changes his/her social security number, with the sampling scheme automatically generating representative and appropriately sized refreshment samples.

contains a social security number). Second, we verified that the samples corresponding to each 2 digit combination represented equal proportions (each 1% of total) and had the same average characteristics (in terms of age, number of accounts, total balances etc). Sampling based on the last two or four digits of the social security number has previously been used by the Social Security Administration to produce the 1-percent Continuous Work History Sample and to create random samples from Supplemental Social Security Records (Pickett and Scott, 1996).[7] The randomness of the 4-digit serial number has also been used in a number of economic, medical and epidemiological studies.[8]

2.2.2 Random Sample of Households

In the second step of our sampling design, we drew the credit files of all individuals residing in the same households as the primary members (individuals belonging to the original 5% sample).[9] This was implemented by including records on all other consumers living at the same address as the primary sample members, where individuals were matched based on the street name, street number, apartment number and ZIP Code. The matching process takes account of different spellings and transpositions (such as Main St. and Main Street). Contrary to the way in which we pulled the primary sample members (the original 5% sample), in drawing records of other household members we did not require these individuals to have a social security number on file and we also included inquiry-only files.[10] The resulting match of approximately 12.0 million primary sample members to their household members generates a sample that includes records of approximately 38.1 million individuals in each quarter, representing about 11.2 million households. As discussed below, a simple weighting procedure can be applied to obtain nationally representative estimates of household-level debt and credit. Note that the target population of *households* in our case includes all US households in which at least one household member has a credit history and a social security number on file. Given our sampling scheme for including household members, household size and household composition will be measured based on the number of individuals in a household with credit histories, who are typically at least 18 years of age.

By repeating the matching process in each quarter, it is possible to track total credit and debt of US households over time. Household-level variation over time will in part reflect changes in household composition, such as those associated with marriages, separations, deaths, the departure and return of adult children to the household, and the establishment of credit

[7] Statistical Policy Working Paper 6 - Report on Statistical Uses of Administrative Records, Prepared by Subcommittee on Statistical Uses of Administrative Records, Federal Committee on Statistical Methodology 1980.

[8] Some examples include Johnson, Parker and Souleles (2006), Broda and Parker (2008), Kopczuk, Saez and Song (2009), Xue et al (2006) and Camp et al (2001).

[9] If a household contained more than one primary member, each household member was included only once.

[10] The reason for including household members with inquiry-only files was to be able to get a more accurate count of the number of adults in the household. However, as some of these files may be associated with fraudulent or erroneously reported information, in some analyses one may want to exclude such cases.

reports as children take out their first loan or are added as authorized users of their parents credit cards.

2.3 Sample Weights

2.3.1 Individual consumer level analysis

Analysis at the individual consumer level does not require use of weights if the sample is restricted to primary sample members, as this group represents a 5% random sample of the target population. Estimates of population aggregates can be obtained simply by multiplying counts by 20 (weight equal to 1/0.05).

Alternatively, one could extend the sample to all individuals in the household-matched database. In this case the target population is slightly larger than when using only primary sample members. As discussed earlier, when using the 5% sample, the target population consists of all individuals with a credit file who also have a social security number on file. When including all individuals in the household-matched database, instead the target population consists of all individuals with a credit file who live in a household in which at least one member (not necessarily the individual him/herself) has a social security number on file.

When using this larger sample, account should be taken of the way individuals are sampled, with the probability of being included in the sample being equal to the probability of this person's household being selected. The latter equals the probability that the household includes a primary sample member (one whose social security number ends in one of the 5 chosen two-digit combinations). As persons without a social security number can never be primary sample members, only the number of individuals in the household with a social security number on file are relevant for computing this probability. For households containing one individual with a social security on file, the probability that the household is selected is 0.05. For households in which two members have a social security number on their credit report the probability that the household is selected equals the probability that either is a primary sample member, which is 1-0.95*0.95. More generally, for a household that contains N individuals with SSN-inclusive credit reports, the probability that the household is included in the sample equals $1-0.95^N$. Therefore, the appropriate individual sample weight when using the entire sample of individuals in the household-matched dataset equals $1/(1-0.95^N)$.

2.3.2 Household level analysis

Analyses that take the household as unit of observation need to take account of the fact that our sampling approach oversamples large households. The larger the number of individuals in the household with a SSN-inclusive credit report, the greater the likelihood that the household

contains a primary sample member and thus the higher the chances of being included in our sample. To make the sample representative of the target population of households and to obtain unbiased estimates of population characteristics therefore requires the use of weights. The appropriate weight for household-level analysis again equals the inverse of the probability that a given household in the target population is included in our sample. The latter is exactly what we computed in the previous section. Thus the appropriate household weight equals $1/(1-0.95^N)$.

2.4 Comparability with ACS

To analyze the comparability of our target population of individuals to the overall US populations, we first contrast the age distribution in the most recent American Community Survey (ACS), which was conducted in 2008, with that in the Q4 2008 wave of the consumer credit panel. Given our sample design where we only sample individuals with a credit history, we restrict the ACS sample to individuals who are at least 18 years, and at least 20 years old respectively. In Table 1, we compare the age distribution in the ACS sample with that in our 5% sample for the US as a whole, for the state of New York, for New York City, and for Manhattan. Similarly, Table 2 shows the same comparison for the *extended* panel sample, which in addition to the 5% sample of primary sample members includes their household members.

Table 1. Comparison of 2008 Age Distributions Based on 5% Sample

Age	US			NY State			NYC			Manhattan		
	ACS Age≥18	ACS ≥20	FRBNY 5%	ACS Age≥18	ACS ≥20	FRBNY 5%	ACS Age≥18	ACS ≥20	FRBNY 5%	ACS Age≥18	ACS ≥20	FRBNY 5%
18-24	13.1	9.5	**8.5**	13.3	9.7	**8.4**	12.6	9.3	**8.2**	10.5	7.9	**7.1**
25-34	17.6	18.3	**16.8**	16.8	17.5	**16.8**	19.2	19.9	**21.2**	21.6	22.2	**24.2**
35-44	18.5	19.3	**18.5**	18.6	19.4	**19.0**	20.5	21.2	**20.9**	23.4	24.0	**21.2**
45-54	19.3	20.1	**19.9**	19.4	20.2	**20.2**	18.2	18.8	**19.5**	16.6	17.1	**17.3**
55-64	14.7	15.3	**15.8**	14.7	15.4	**15.5**	13.6	14.1	**14.4**	12.7	13.0	**14.0**
65-74	8.7	9.1	**9.7**	8.8	9.2	**9.3**	8.3	8.6	**8.1**	8.1	8.3	**8.0**
75-84	5.7	5.9	**6.9**	5.8	6.1	**6.7**	5.4	5.6	**5.0**	5.2	5.3	**5.0**
85+	2.4	2.5	**4.0**	2.6	2.7	**4.1**	2.3	2.4	**2.9**	2.3	2.4	**3.3**
Total (millions)	230.2	221.1	**239.6**	15.1	14.5	**14.7**	6.4	6.2	**5.6**	1.4	1.3	**1.3**

American Community Survey figures are 1-yr estimates for 2008 from table B11016 Household Type by Household Size. The FRBNY figures are based on the Q4 2008 wave in the consumer credit panel. All counts are in millions.

Table 1 shows that across the four geographic areas the age distribution based on the 5% sample is very similar to that based on the ACS sample, especially when the latter is restricted to individuals 20 years of age or older. The panel generally has a slightly higher proportion of individuals 85 and older, and a slightly lower proportion under age 25.[11] The population counts estimated from both samples are also similar. There are several reasons for why the age distributions and population size estimates could differ between the ACS and our panel. First, as discussed above, the target population based on our 5% sample only includes individuals with SSN-inclusive credit files, while that for the ACS it includes the entire population of individuals at least 18 or 20 years of age. As not all individuals have a credit history or a social security number (or one listed on their credit report), one would have expected population counts based on the panel to be somewhat lower that those based on the ACS.

As shown in Table 2, while the age distributions are again very similar, population counts based on the *extended* sample of individuals that includes other household members are in fact somewhat larger than that in the ACS. It is likely that the ACS will miss some individuals living in informal or illegally converted apartments or rural mobile homes. There is also potential for some double counting in the panel, where the same individual may be associated with two separate credit files.[12] Moreover, when including other household members the potential for double counting will be higher as we include household members with inquiry-only files.[13]

There are also slight differences between the dates at which data was collected and in the way an individual's age is measured. The figures for the consumer credit panel refer to the end of December of 2008, while those for the ACS are computed based on interviews that occurred throughout the whole of 2008. In our consumer credit panel age is simply computed as 2008 minus the year of birth, while in ACS it is the person's age at the time of the survey. Differences also exist in the way an individual's residential location is measured and in which people are sampled. The ACS, which is a residence/address based survey, counts the number of individuals who have lived at the address for more than two months at the time of the interview. The selection of primary sample members (5% sample) in the consumer credit panel is instead

[11] A lower proportion of younger individuals may reflect the time it takes for younger individuals to take out a first loan. A higher proportion of older individuals may also be due to some delay in the removal of deceased individuals' credit files from the database.

[12] Occasionally, there is insufficient information to confidently link a loan to a person in the database, leading to the creation of a new credit file. One possible indicator for the presence of so-called fragmented files is a non-negligible number of credit files that share the same social security number as that included in another credit files. The same SSN is reported at least twice in 2.4% of all credit files in the Q4-2008 5% sample. Note that while the presence of fragmented files leads to an upward bias in the estimated population counts, it should not affect the calculation of total debt balances and other aggregates across individuals, either for the US as a whole or for different regions or age groups.

[13] When individuals with inquiry-only files or files without any trades during the past two years are omitted, population totals are more comparable. For the US, the extended-sample based estimate of the population size drops from 266.2 to 244.2 million, while the corresponding estimate based on the 5% sample falls from 239.6 to 224.6 million.

individual-based, and the individual's address corresponds to the primary mailing address listed for that person on his/her credit report.

Table 2. Comparison of 2008 Age Distributions Based on Extended Sample

Age	US			NY State			NYC			Manhattan		
	ACS		FRBNY	ACS		FRBNY	ACS		FRBNY	ACS		FRBNY
	Age≥18	≥20	All	Age≥18	≥20	All	Age≥18	≥20	All	Age≥18	≥20	All
18-24	13.1	9.5	**9.4**	13.3	9.7	**9.1**	12.6	9.3	**8.9**	10.5	7.9	**7.4**
25-34	17.6	18.3	**17.0**	16.8	17.5	**17.1**	19.2	19.9	**21.6**	21.6	22.	**24.4**
35-44	18.5	19.3	**18.4**	18.6	19.4	**18.9**	20.5	21.2	**20.8**	23.4	24.0	**21.2**
45-54	19.3	20.1	**19.6**	19.4	20.2	**19.9**	18.2	18.8	**19.2**	16.6	17.1	**17.0**
55-64	14.7	15.3	**15.5**	14.7	15.4	**15.3**	13.6	14.1	**14.2**	12.7	13.0	**13.9**
65-74	8.7	9.1	**9.5**	8.8	9.2	**9.1**	8.3	8.6	**7.8**	8.1	8.3	**8.0**
75-84	5.7	5.9	**6.7**	5.8	6.1	**6.6**	5.4	5.6	**4.8**	5.2	5.3	**4.9**
85+	2.4	2.5	**4.0**	2.6	2.7	**4.0**	2.3	2.4	**2.8**	2.3	2.4	**3.2**
Total (millions)	230.2	221.1	**266.2**	15.1	14.5	**16.6**	6.4	6.2	**6.8**	1.4	1.3	**1.7**

American Community Survey figures are 1-yr estimates for 2008 from tables B11016 Household Type by Household Size. The FRBNY figures are based on the Q4 2008 wave in the consumer credit panel. All counts are in millions.

In table 3 we compare the distribution of household size in our panel with that in the ACS. Given our sampling procedure, for the consumer credit panel we only count as household member the number of individuals living at the same address who have a credit report. As this excludes children, it is not surprising that the proportions of households with 1 or 2 individuals generally are higher in the FRBNY panel (although they are very similar to the ACS proportions in Manhattan), while the proportions of households with more than 3 members is lower than in the American Community Survey. Comparing across geographic areas, the credit panel captures the higher proportion of single individuals living in New York City and particularly in Manhattan. The total number of households computed from the consumer panel, which corresponds to the number of unique addresses, corresponds reasonably closely to the number of households, similarly identified by addresses, in the ACS.

Table 3. Comparison of 2008 Household Size Distributions

Size	US			NY State			NY City			Manhattan		
	ACS #	ACS %	FRBNY All (%)	ACS #	ACS %	FRBNY All (%)	ACS #	ACS %	FRBNY All (%)	ACS #	ACS %	FRBNY All (%)
1	31,406	27.8	**34.3**	2,119	29.7	**36.3**	1,023	33.5	**41.5**	377	50.3	**49.6**
2	37,599	33.2	**35.7**	2,199	30.8	**32.0**	839	27.5	**27.6**	203	27.1	**29.6**
3	17,907	15.8	**17.1**	1,150	16.1	**16.3**	499	16.3	**14.4**	86	11.5	**12.1**
4	15,143	13.4	**7.5**	953	13.3	**8.0**	375	12.3	**7.4**	54	7.2	**4.6**
5	6,900	6.1	**3.0**	441	6.2	**3.6**	186	6.1	**3.9**	21	2.8	**1.8**
6	2,588	2.3	**1.2**	164	2.3	**1.7**	73	2.4	**2.1**	5	0.7	**0.8**
7+	1,558	1.4	**1.3**	112	1.6	**2.1**	59	1.9	**3.2**	3	0.4	**1.5**
Total (1000)	113,101		**117,105**	7,137		**6,994**	3,055		**2,789**	749		**733**

American Community Survey figures are 1-yr estimates for 2008 from tables B11016 Household Type by Household Size. The FRBNY figures are based on the Q4 2008 wave in the consumer credit panel. All counts are in thousands.

Finally, before turning to a description of the individual-level consumer debt and credit information contained in our panel, we take a look at the number of non-primary sample household members for whom no social security number was reported on their files. Because our 5% sample is based on individuals with a SSN-inclusive credit file, analysis based on the 5% sample is only representative of the population of individuals with SSN-inclusive credit files. When including all household members (including those for whom no SSN was on file), our sample becomes representative of all individuals with a credit file and living in households containing at least one individual with an SSN-inclusive credit file. Both samples therefore omit some individuals for whom no SSN was recorded in their credit files.

In the Q42008 wave of our panel of the 38.1 million individuals included in our total sample, for 9.4% of individuals no SSN was recorded. Accounting for sample weights, the proportion of such individuals in the target population is 9.0%. If r represents the proportion of individuals in the population with a credit file for whom no SSN is recorded, and if individuals are randomly matched into households of different sizes according to the proportions shown in Table 3, then given our sampling design a sample proportion of 9.0% implies a population proportion of r=16.6%. That is, a rough estimate is that 16.6% of all credit files in the entire original database contain no SSN. This relatively high number suggests that a relatively large proportion of these files may be files with insufficient information to link them to a person, leading to multiple files existing for the same person. Indeed, for almost 75% of these credit files

the person's birth year is missing and some 23% of these files contain no trades at all and only include inquiries. As will be shown later, total debt associated with non-SSN files is very small. Therefore for many analyses using the extended sample with household members, it may be acceptable or advisable to only include household members with a SSN-inclusive credit file.

3. Consumer Credit Panel Content

A comprehensive overview of the specific content of consumer credit reports is provided in Avery, Calem, Canner and Bostic (2003). Consumer credit reports contain information that can be broadly categorized into loan account data, public record and collection agency data and individual background information. Loan or credit account data can be further subdivided by loan type: mortgage accounts, home equity revolving accounts, auto loans, bank card accounts, student loans and other loan accounts. The credit report data included in our panel primarily consists of information on accounts that have been updated by the creditor within 3 months of the date that the credit reports were drawn each quarter. Thus, accounts that are not currently updated are excluded.[14] Such accounts may be closed accounts with zero balances, dormant or inactive accounts with no balance, or accounts that when last reported had a positive balance. The latter accounts include accounts that were either subsequently sold, transferred, or paid off as well as accounts, particularly derogatory accounts, that are still outstanding but on which the lender has ceased reporting. According to Avery et al (2003), the latter group of non currently reporting accounts, with positive balances when last reported, accounted for approximately 8% of all credit accounts in their sample. For the vast majority of these accounts, and particularly for mortgage and installment loans, additional analysis suggested they had been closed (with zero balance) or transferred.[15] Our exclusion of the latter accounts is comparable to some 'stale account rules' used by credit reporting companies, which treat non currently reporting revolving and non-revolving accounts with positive balances as closed and with zero balance. In addition to these inactive trades, our data excludes authorized user trades, disputed trades, lost/stolen trades, medical trades, child/family support trades and commercial trades.

Our database includes **account level information** on all *mortgage installment and revolving accounts*. The former includes mortgage installment loans such as first mortgages and

[14] Note that the inclusion criteria for account level information differ from the file exclusion criteria discussed earlier. While a record will be included in our panel for all individuals with some credit activity on their credit reports over the past 7 years or so (and, in case of primary sample members, who did not have inquiry-only files), records will only include information on recently updated accounts.

[15] Avery et al (2003) found that for many non-reported mortgage accounts a new mortgage account appeared around the time the account stopped being reported, suggesting a refinance or that the servicing was sold. Most revolving and open non-revolving accounts with a positive balance require monthly payments if they remain open, suggesting the accounts had been closed. Non currently reporting derogatory accounts can remain unchanged and not requiring updating for a long time when the borrower has stopped paying and the creditor may have stopped trying to collect on the account. Avery et al report that some of these accounts in fact appeared to have been paid off.

home equity installment loans (HEL), both of which are closed-end loans. Home equity revolving accounts (also known as Home Equity Line of Credit or HELOC), unlike home equity installment loans, are home equity loans with a revolving line of credit where the borrower can choose when and how often to borrow up to a given credit limit. Some care should be taken in using the mortgage installment account classification. In addition to lender and account information, the credit bureau from which we obtained our data uses the loan origination balance to classify a mortgage as a first or second (HEL). As a result, relatively small first mortgage loans (such as those for mobile homes) may be misclassified as home equity installment loans, while some larger installment loans are sometimes incorrectly classified as a first mortgage. Additional loan information included in our panel can often be used to reclassify such loans.[16] Users should also be aware that the classification of mortgage loans that was applied by the credit reporting agency does not immediately provide the position of the lien. For example, for a consumer with a HEL but no first mortgage, the home equity installment loan would sit in the first position.

The consumer credit panel also includes **individual-level information** on all other loan types, including auto loans, bankcard loans, student loans and other loans. *Auto loans* are loans taken out to purchase a car, including Auto Bank loans provided by banking institutions (banks, credit unions, savings and loan associations), and Auto Finance loans, provided by automobile dealers and automobile financing companies. *Bankcard accounts* (or credit card accounts) are revolving accounts for banks, bankcard companies, national credit card companies, credit unions and savings & loan associations. *Student loans* include loans to finance educational expenses provided by banks, credit unions and other financial institutions as well as federal and state governments. The *Other* category includes Consumer Finance (sales financing, personal loans) and Retail (clothing, grocery, department stores, home furnishings, gas etc) loans. For all non-mortgage loan categories, information is available at the individual-level information, aggregated over accounts of that loan type.

In addition to loan account information provided by creditors, our consumer credit panel contains information from *public records*, including records of bankruptcy and tax liens. It also contains information reported by collection agencies on actions associated with credit accounts and noncredit-related bills. Finally, our database contains a limited number of individual characteristics, including the year of birth, and the individual's consumer credit score at the end of each quarter. The latter is like the FICO score and while based on a different algorithm, predicts the same likelihood of severe delinquency over the next 24 months. The consumer credit score ranges from 280-850, with a higher score being viewed as a better risk than someone with a lower score. In addition to the consumer credit score, we know the state, county, ZIPcode, and

[16] As GSEs do first liens almost exclusively, loans owned by these Agencies that in our panel are identified as home equity installment loans can be reclassified as first mortgage loans. This can be done using the narrative codes associated with each mortgage loan, attributes that are discussed below.

census tract and block associated with the address on file.[17] The address listed is generally that to which most of the individual's mail is sent by creditors. An address change occurs when the majority of data providers report a new address, where more reliable data providers are given more weight.[18] Finally, a household identifier also allows one to derive the number of household members with credit reports living at the same address.

Finally, it is important to note that all individuals included in our database are anonymous: we do not know their names, street addresses or social security numbers. Individuals in our data are distinguished and can be linked over time through a consumer identification number (CIN), while individuals living at the same address can be linked using a household identification number (HHID). Note that while the CIN is fixed over time, HHID is not. They cannot be used for linking households over time. Linking a primary sample member's household records over time requires linking the HHIDs associated with that person's CIN across quarters.

3.1 Description of Attributes

A record for a given individual in the Consumer Credit Panel consists of a set of custom attributes, including characteristics of each mortgage account listed in their credit report, and two sets of derived variables at the account and consumer level. The latter include several hundred variables summarizing various aspects of an individual's credit behavior. In what follows we will focus primarily on some of the custom attributes.

The custom attributes consists of account-specific attributes and a number of aggregate variables, rolled-up across all accounts of a specific type. More specifically, for each of up to 14 different most recent first mortgages, 5 most recent home equity installment accounts and 5 most recent HELOCs, the database includes the loan origination date, origination amount, current balance, requested payment amount or term of the loan, credit limit (on HELOCs), individual/joint account type and payment status.[19] There are also up to two narrative codes associated with each account which provides additional information regarding the type of mortgage account, including whether it was guaranteed by one of the GSEs, whether the mortgage was for a mobile home, or a second mortgage and whether the account was included in

[17] FIPS and census blocks codes for all quarters are those as defined in 2000. In each quarter, for a little over 6% of individuals their address corresponds to a Post Office Box.

[18] The process used by the credit reporting agency from which we acquired the data to update the main address has changed over time. Before 2003 the criteria used for a change in address were less restrictive, so generally the address reported was that associated with the most recently processed account update at the time data records were pulled for the panel at the end of each quarter.

[19] Note that this includes closed mortgage trades with a zero balance that are still being reported by creditors. When linking individual loans across quarters this provides a confirmation that the loans were indeed paid off and closed and not disappeared for other reasons.

a bankruptcy or foreclosure.[20] Finally, for each mortgage loan there is an indicator of the industry the lender belongs to, which for example allows one to distinguish between banks, credit unions and mortgage companies.

The payment amount listed for each mortgage account represents the scheduled payment between payment cycles. For revolving accounts, payment amount typically represents the minimum payment amount required as displayed on the statements. **High credit** is the credit limit on HELOCs and is the highest balance ever reported during the history of the loan on installment accounts. Whether a mortgage account was a **joint or individual account** is important for avoiding the double counting of mortgages listed on two different individual's credit reports. **Payment or delinquency status** varies between current (paid as agreed), 30-day late (between 30 and 59 day late; not more than 2 payments past due), 60-day late (between 60 and 89 days late; not more than 3 payments past due), 90-day late (between 90 and 119 days late; not more than 4 payments past due), 120-day late (at least 120 days past due; 5 or more payments past due) or collections, and severely derogatory (with reports of a repossession, charge off to bad debt, foreclosure or a defaulted student loan). Not all creditors provide updated information on payment status, especially after accounts have been derogatory for a longer period of time. Thus the payment performance profiles obtained from our data will to some extent reflect reporting practices of creditors.

The custom attributes also include for each other loan type (auto, student, credit card, and other loans) the total balance, high credit and payment amount aggregated across joint accounts of that loan type. The latter attributes are again important to avoid double counting of debt associated with joint, shared or co-signed accounts, but also of interest in their own right as well and for calculating overall household debt.

3.2 Trend in Total Household Debt and Comparison with Flow of Funds

Figure 1 shows the trends in total household mortgage and non-mortgage debt since the first quarter of 1999.[21] In computing total debt, account was taken of the joint or individual nature of various loan accounts. For example to minimize biases due to double counting, in computing individual-level total balances, 50% of the balance associated with each joint account was attributed to individuals owning such an account. As shown in Figure 1, overall U.S. consumer debt at the end of 2010Q2 was estimated to be $11.7 trillion, which is $812 billion (6.5%) below its peak value reached at the end of 2008Q3. Household mortgage indebtedness

[20] Since the second quarter of 2010 it also may include a narrative code for loan modifications.

[21] The figures in the chart are based on the 5% random sample of individuals. To reduce processing costs, we drew a 2% random subsample of these individuals, meaning that the results presented here are for a 0.1% random sample of individuals with credit reports, or approximately 230,000 individuals as of 2010Q2. Total balances exclude accounts in bankruptcy.

was $9.4 trillion at the end of 2010Q2, 6.3% below its peak, while non-mortgage debt stood at $2.3 trillion, 8.4% below its 2008Q4 peak.

How do these estimates based on the FRBNY Consumer Credit Panel compare to household debt estimates reported in the Board of Governors' Flow of Funds (FoF) Accounts? It turns out that the debt estimates are surprisingly similar. In comparing our aggregate measures of household debt to those included in the FoF, there are several important considerations. First, the household debt measures in the Flow of Funds are not based on direct data but instead are derived as residual amounts. Among the different components included in the FoF household debt measure, our measures are most directly comparable to two of its components: home mortgage debt and consumer credit. While total mortgage debt and non-mortgage debt in the second quarter of 2010 were respectively $9.4 and $2.3 trillion, the comparable amounts in the FoF for the same quarter were $10.2 and $2.4 trillion, respectively.[22] Given that both measures were obtained in entirely different ways, it is remarkable that they are so similar.

What accounts for the differences? Unlike our measure, the FoF measure of household mortgage debt includes some mortgage debt held by nonprofit organizations (churches, universities, etc.). On the other hand, our debt totals exclude some debt held by individuals without social security numbers. However, debt held by household members who had no SSN listed on their credit file on average was only 2.6% of debt held by those with SSN-inclusive credit files. Assuming that overall about 16.6% of credit files associated with different individuals do not contain a SSN, and assuming that these individuals are similar to those we observe in our panel, a rough estimate is that they would add at most another 0.5% to our debt totals, suggesting that this is not an important factor. There may be differences in the speed at which changes in various types of debt are recorded, where new mortgage accounts usually appear on credit reports with some delay, making direct comparisons difficult. Clearly, a complete accounting for the differences between the debt measures from both data sources would require a more detailed breakdown and documentation of the computation of the FoF measures.

4. Summary

In this paper we discuss the sample design and content of a new longitudinal database with detailed information on consumer debt and credit. The panel overcomes several important shortcomings of alternative loan-level data for analyzing mortgage loan behavior. First, it allows

[22] Flow of Funds Accounts of the United States, Flows and Outstandings, Second Quarter 2010, Board of Governors, Table L.100.

one to longitudinally link consumer-specific mortgage loans over time including those associated with a refinance or home transactions. Second, it permits linking at the individual consumer level across all mortgage loans including first and second mortgages and home equity lines of credit. Third, unlike most existing databases which only provide information on mortgage debt, our panel allows one to connect mortgage loans to other types of debts, including debt on credit cards, auto loans and student loans. Fourth, the panel allows one to measure debt at the household level, which is more relevant for analyzing household balance sheets and household financial behavior.

In addition to its direct use in general research relating to consumer loans and debt, the FRBNY Consumer Credit Panel can be used to compute nationally representative estimates of various measures of levels and changes in consumer and household level liabilities. A good overview of some of the information contained in the panel can be found in the *Quarterly Report on Household Debt and Credit* which is accessible on FRBNY's public access website.[23] In addition to the descriptive analyses contained in the report, we provide access on our website to some of the underlying data.

[23]The *Quarterly Report on Household Debt and Credit* is available at http://data.newyorkfed.org/creditconditions/

References

Avery, R.B., P.S. Calem, G.B. Canner and R.W. Bostic, "An Overview of Consumer Data and Credit Reporting", *Federal Reserve Bulletin*, Feb. 2003, pp 47-73.

Broda, C. and J.A. Parker (2008), "The Impact of the 2008 Tax Rebate on Consumer Spending: Preliminary Evidence", mimeo University of Chicago, GSB, July 2008.

Camp,S., K. P. Srinath, S. Arday, and E. Axelrod, "Two Phase Sampling Design for Selecting Representative Samples from Large Administrative Databases", *Proceedings of the Annual Meeting of the American Statistical Association,* Aug 2001).

Jacob, K. and R. Schneider. *Market Interest in Alternative Data Sources and Credit Scoring*. The Center for Financial Services Innovation, December 2006.

Johnson, D.S. , J.A. Parker and N.S. Souleles, ""Household Expenditure and the Income Tax Rebates of 2001," *American Economic Review,* 2006, vol. 96(5), pages 1589-1610.

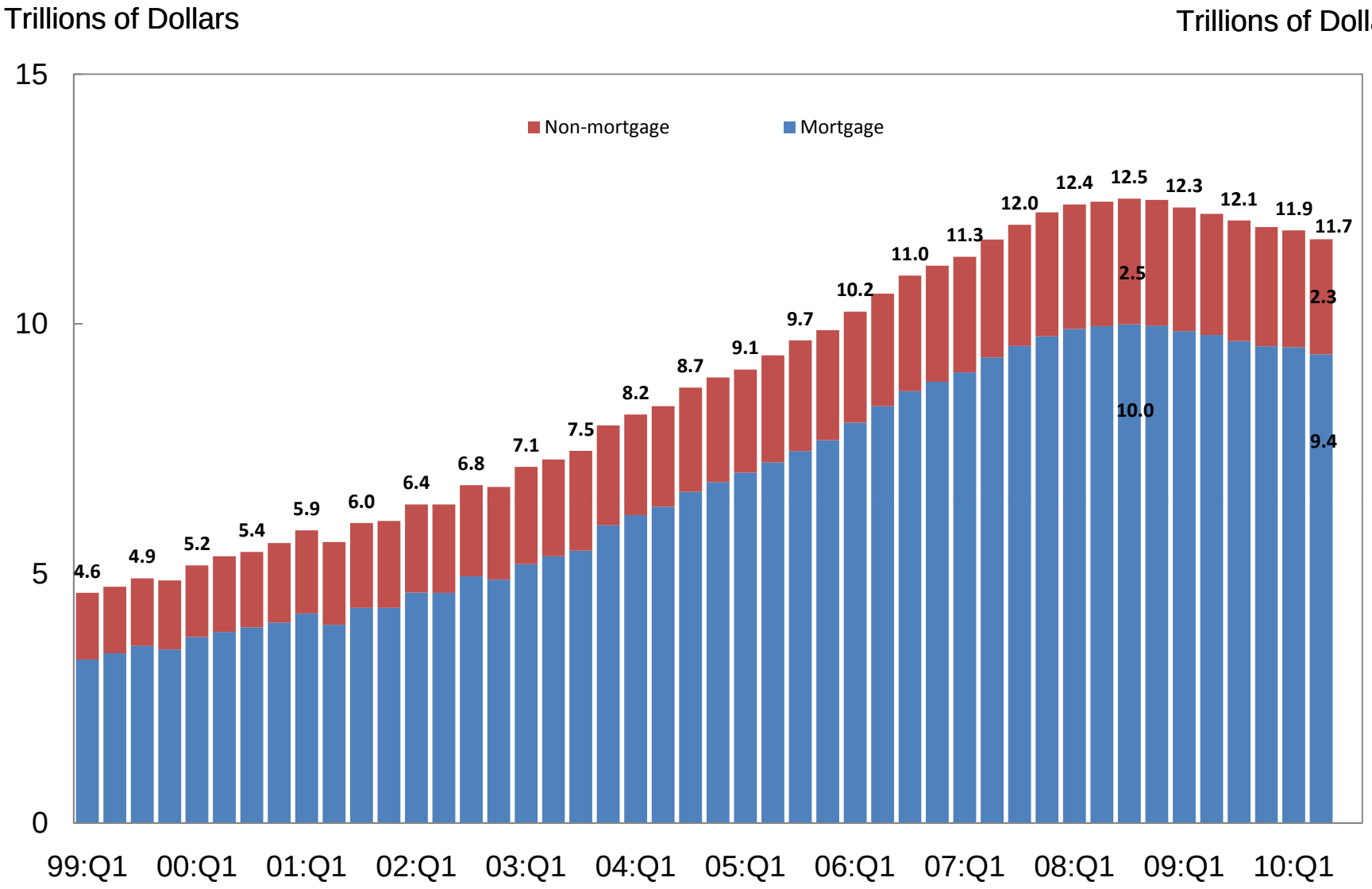
Kopczuk, W. , E. Saez and J. Song, "Earning Inequality and Mobility in the United States: Evidence from Social Security Data since 1937", *Quarterly Journal of Economics*, 2010, 125(1), 91-128.

Pickett, C.D. and C.G. Scott, "Reinventing SSI Statistics: SSA's New Longitudinal File", *Social Security Bulletin*, 1996, 59(2): 31-56.

Quarterly Report on Household Debt and Credit, Research and Statistics, FRBNY. http://data.newyorkfed.org/creditconditions/

Xue, J.L., F. Daniels, R. A. Star, P.L. Kimmel, P.W. Eggers, B.A. Molitoris, J. Himmelfarb, and A. J. Collins, "Incidence and Mortality of Acute Renal Failure in Medicare Beneficiaries, 1992 to 2001", *Journal of the American Society of Nephrology*, 2006, 17: 1135-1142.

Figure 1. Total Debt Balance and its Composition



Source: FRBNY Consumer Credit Panel