A Guide to Consumer Credit Reporting Data

Christa Gibbs[1], Benedict Guttman-Kenney[2], Donghoon Lee[3], Scott Nelson[2], Wilbert van der Klaauw[3], & Jialan Wang[4]

[1]CFPB, [2]Chicago Booth, [3]FRBNY, [4]UIUC

AEA 2023, 7 January 2023

Disclaimer: All views are authors' own, do not necessarily reflect position of the Consumer Financial Protection Bureau (CFPB), Federal Reserve Bank of New York (FRBNY), or the Federal Reserve System.

Today's Panel

1. What are Credit Reporting Data? (Scott)
2. How to Access Credit Reporting Data? (Wilbert)
3. Constructing Economically-Important Measures (Ben)
4. Open Issues (Jialan)
5. Q + A

Slides: www.benedictgk.com

Paper: In-progress

AEA 2023 Session on Consumer Credit Reporting Data

Scott Nelson

January 7, 2023

Chicago Booth

Traditional Credit Report Data

What we see:

Loan balances

Delinquency history

Credit limits

Applications/inquiries

Debt in collection

Bankruptcies

Civil judgments

Traditional Credit Report Data

What we see:

- Loan balances
- Delinquency history
- Credit limits
- Applications/inquiries
- Debt in collection
- Bankruptcies
- Civil judgments

What we don't:

- Income
- Expenditure / consumption
- Credit card spending
- Interest rates / prices
- Demographics and education
- Lender (vs. servicer)
- "Alternative" financial products
- .....

What is reported and why?

- **Three forces** shape credit reporting practices:
    - Industry's "self-regulation": CDIA (trade association) and Metro 2 (reporting guidelines)
    - Legislation/regulation: FCRA, ECOA
    - Just tradition / path-dependence: payday lenders vs. credit card lenders
- **Long history** of US consumer credit reporting
    - First grocers/butchers, then department stores, then "National Association of Credit Men" (1880s - 1920s)
    - See Lauer (2017), Hyman (2011)
    - Today: $>$1b loans monitored among $>$200m consumers
- **Coverage:** Not all adults are in credit report data; $>$20m are "credit invisible" (Brevoort et al., 2015)

"Raw" Data Structure

Four types of data files:

1. Tradelines
2. Inquiries
3. Public Records
4. Header Files

Also available: aggregates derived from these (consumer-level "roll-ups" or geography-level summary statistics)

"Raw" Data Structure

Four types of data files:

1. Tradelines $\iff$ "loans"
2. Inquiries $\iff$ "applications"
3. Public Records $\iff$ "court data"
4. Header Files $\iff$ "consumers"

Also available: aggregates derived from these (consumer-level "roll-ups" or geography-level summary statistics)

Data Structure, ctd.: Tradelines

- **Tradeline = loan** (roughly)
  - Exception: collections tradelines
- **Who is "the lender"?** servicers vs. originators
- Difficulties with **panel identifiers**: transfers, lost cards, defaults and modifications
- **Stocks vs. flows** of delinquent trades
- Some data richness is lost after **account closure** $\implies$ importance of panel data
- **Obsolescence rules:** 7 years for negative information, 10 years for positive information (after account closure)

Data Structure, ctd.: Inquiries

- **Hard vs. soft** (two different permissible purposes under FCRA)
  - Availability of historical data on soft inquiries varies by CRA
- Hard inquiry = application (roughly)
- De-duplicated vs. non-de-duplicated
- **Furnishing patterns:**
  - Mortgages (relatively complete) vs. credit cards (relatively incomplete)
  - Implications for estimating approval rates and search behavior
- **Obsolescence rule:** 2 years

Data Structure, ctd.: Public Records

- **What are public records?** Bankruptcies, foreclosures, civil judgments, tax liens
  - N.b. implied bankruptcy prevalence may be inconsistent with tradeline-level flags
- **Data quality:** depends on local (town/county) institutions and data collection (sometimes by hand)
- **Availability over time:** NCAP (June 2017, March 2018) $\Longrightarrow$ fewer (but higher-accuracy) public records observed
- **Obsolescence rules:** typically 7 years (10 years for Ch. 7)

Data Structure, ctd.: Header Files

- **Header files:** the *only* "raw" consumer-level datafile
- Age, name(s), SSN
- Address...or *likely* addresses... and potentially *lagged* addresses
- **Gramm-Leach-Bliley Act** (1999) $\implies$ CRAs have a data quality advantage for mobility data
- **Many demographics missing**...at least in the raw data

Key Variables

Several types of variables to be aware of when purchasing credit report data:

- Payment "grids," "strings," or histories
- Statuses
- Subscriber-based fields
- Credit limits
- Balances
- "Actual" payment amount
- Scores...(next slide)

Easy to get overwhelmed by the many permutations of these in roll-ups or attribute files. Remember the raw data these come from!

Scores!

- **Credit scoring 101:** affine transformation of a log odds of default
- **Many variants:**
  - FICO vs. Vantage
  - Input data (which bureau)
  - Which outcome is being predicted? On new or existing trades?
  - Which model version?
- **Common misconceptions:**
  - Business cycle risk
  - Quantiles
  - Updated models vs. scores used in practice

Tradeline-specific issues

- **Mortgages:** modifications, forbearance, ambiguous foreclosures...
- **Credit Cards:** transfers, missing credit limits, revolvers vs. transactors...
- **Student Loans:** deferment, transfers ...
- **Auto Loans:** repossessions, deficiency judgments, furnishers...

FEDERAL RESERVE BANK of NEW YORK

Assessing Credit Bureau Data

Donghoon Lee and Wilbert van der Klaauw, Federal Reserve Bank of New York

AEA meetings, New Orleans, January 7 2023

Assessing Credit Bureau Data

Credit Bureaus provide <u>anonymized</u> credit report data as

- Aggregated data (using 100% of data)– by geography – ZIP, Census Tract, Census block group (e.g. Mian et al 2010)
- Customized samples
  - Matched to prespecified research dataset of individuals or loans
  - Nationally representative sample of individuals or loans: cross-section and panels
  - Credit bureaus often facilitate linking of credit report sample to second data source – e.g. allowing use as sampling frame for subsequent survey (CFPB Consumer Credit Panel as frame for CFPB's *Making Ends Meet* survey)

Creating consumer credit panels

Panels usually drawn by

- Drawing a random representative cross-section (*Archive*). Going forward, track individuals over time, while including refreshment samples of individuals with first-time/new credit reports
- Drawing based on last 4 digits of SSN particularly useful for drawing longitudinal panels in a way that guarantees that panel remains representative over time (NYFed CCP; Lee and van der Klaauw 2010)

Using prespecified dataset of individuals or loans
(FCRA-compliant merges of credit reports to a research sample)

Many papers have matched credit report data to prespecified sample. Some examples:

- Credit report data linked to tax return data from set of tax filers (Meier and Sprenger 2010)
- TransUnion data to Oregon Medicaid applicants (Finkelstein et al 2012)
- Equifax data to student records from WV (Scott-Clayton and Zafar 2019)
- Experian data to MTO records from HUD (Miller and Soo 2021)
- Credit bureau data to Dept of Defense payroll & Dept of the Army personnel data (Beshears et al 2022)
- Equifax credit bureau data to Survey of Consumer Payment Choice data (Stavins 2020)
- Credit bureau data to records on bankruptcy filings (Dobbie et al 2017; Argyle et al 2020)
- Equifax credit reports in the NYFed CCP to Medicare claims data (Nicholas et al 2021)
- TransUnion data linked to LEHD matched employer-employee employment records (Herkenhoff et al 2016, 2021)

Matching to credit bureau data

- Common approach for linking datasets while maintaining anonymity: Triparty linking to merged data user C:
  - Partner A provides data D_A and anonymous identifier I_A to C
  - Partner B provides data D_B and anonymous identifier I_B to C
  - Partner A provides crosswalk between person identity (SSN, name, address, birthdate) and I_A to B
  - Partner B sends to C crosswalk between I_A and I_B
- Important features: Partners A and B never see each other's data. C never sees person identities
- Some additional examples: NYFed CCP linked to payday loan data (Bhutta et al 2015), NSC education records (Chakrabarti et al 2020), Medicare data (Nicholas et al 2021)

Nationally representative samples

Evaluation sample Federal Reserve Board (Avery et al 2003)

- One out of 657 nationally representative sample of individuals as of June 1999.
- Full anonymized credit records from credit reporting company. Cross-section

NYFed Consumer Credit Panel (Lee and van der Klaauw 2010)

- 5% nationally representative sample of individuals, quarterly since 1999; from Equifax
- Also represents representative random samples of mortgages and student loans (now expanded to include auto loans, credit cards by Philadelphia Fed)
- Includes credit records of all individuals at same residence

Nationally representative samples (continued)

Credit Risk Insight Servicing McDash (CRISM) (Beraja et al 2015)

- Matched entire McDash's mortgage servicing records (from Black Knight Financial Services) to individual-level Equifax credit data, since 2005
- Loan-level data. McDash adds income, original property value, DTI. Merge permits computation of CLTV

CFPB Consumer Credit Panel

- 1-in-48 nationally representative sample of individuals, starting as annual in 2001, quarterly in 2004, monthly since 2013
- Constitutes a panel of individuals. More tradeline data than NYFed CCP
- Also includes deidentified record of co-borrowers or joint account holders
- CFPB CCP was used as sampling frame for National Survey of Mortgage Borrowers (NSMB) - quarterly survey of borrowers with new mortgages

Nationally representative samples (continued)

University of California Consumer Credit Panel (UC-CCP)

- 2% nationally representative sample of individuals, and 100% sample of Californians with credit reports. Quarterly since 2004
- Includes records from consumers who shared an address or an account with those in primary sample; more tradeline data than NYFed CCP
- Sampling based on “consumer pin” (assigned sequentially) ending in one of two two-digit numbers

Other new datasets:

- Ohio State University nationally representative, randomly selected 1% sample 2019-2020 Experian (Brown et al 2022)
- University of Chicago nationally representative, randomly selected 10% panel of persons with credit reports maintained by TransUnion (Guttman-Kenney et al 2022)
- University of Illinois at Urbana-Champaign – Gies Consumer and Small Business Credit Panel (GCCP) - 1% random sample with Experian credit report linked to business credit report data (Fonseca and Wang 2022)

Considerations in creating new datasets

- What sampling frame? What is population of interest: population with credit reports?
- How to sample? Want panel that remains nationally representative?
  - Using last 4 digits of SSN has great benefits: cost is restricting samples to individuals with an SSN
  - Are consumer pins/IDs sufficiently stable for creating panel?
  - Drawing a cross-section and supplementing with refreshment samples can be complicated
- What information to include: entire file with all tradeline data (can be overwhelming) or roll-ups for some?
- Include thin/fragmented files and files with only authorized user accounts?
- Dealing with deceased individuals, new immigrants and those taking out first loan
- Include household members/co-borrowers? How to define households? Want to track households?

Considerations in creating new datasets (cont.)

- Primary shortcoming of credit report data is lack of demographics
  - Can link to ACS census tract level and IRS ZIP income data
  - Credit bureau supplied imputed demographics data often not any better
- Dangers of linking data over time and matching data sets:
  - Need stability of main individual identifier
  - Increased risk of identification when linking in demographic information
  - Linking/matching errors. While match rates appear very high when linking based on name and SSN, still likely to have considerable matching errors
    - even when small percentage can have big effect on research findings (e.g. migration)

Considerations in creating new datasets (cont.)

- Comparing and matching with other population-based data
  - By FCRA, credit bureaus exclude individuals below age 18
  - Remember only individuals with loans have credit reports, debit cards excluded, hence credit bureau data under-represent younger individuals (18-24) and lower income areas. Aggregate debt for those groups are ok, but average debt should be adjusted.
  - Credit bureaus don’t have good information on who are deceased, some of the older individuals in the data might have been deceased.
  - Having birth year of individuals is very useful in this regard.

Considerations in creating new datasets (cont.)

- Issues with inquiry only files
  - Credit reports are typically newly created by inquiries by the prospective lenders, but inquiry only files are often premature to indicate a real person is associated with the file.
  - NYFed CCP excludes inquiry only files and requires some accounts/public records/collections to be included.

Considerations in creating new datasets (cont.)

- Joint accounts and authorized user credit card accounts
  - About half of mortgages are joint accounts typically held by 2 individuals, and individual/joint account information should be used to properly adjust the aggregate/average balances.
  - For credit cards, authorized user accounts should be dropped or indicated as such.

Researcher access to data

- Access to each dataset follows separate rules and requirements set by each institution (and in agreement with credit bureau)

  - Access to Federal Reserve System credit panels generally requires co-authorship with Fed researcher, requires signing of user agreement and data can only be accessed on FRS servers

- Acquisition of credit report data directly from a credit bureau has become easier and less costly

Thank you!

Constructing Economically-Important Measures from Consumer Credit Reporting Data

Christa Gibbs[1], **Benedict Guttman-Kenney[2]**, Donghoon Lee[3], Scott Nelson[2], Wilbert van der Klaauw[3], & Jialan Wang[4]

[1]CFPB, **[2]Chicago Booth**, [3]FRBNY, [4]UIUC

AEA 2023, 7 January 2023

Disclaimer: All views are authors' own, do not necessarily reflect position of the Consumer Financial Protection Bureau (CFPB), Federal Reserve Bank of New York (FRBNY), or the Federal Reserve System.

Economically-Important Measures

1. Populations
2. Financial Distress
3. Credit Access
4. Consumption
5. Mobility

Illustrated with examples from literature

1. Populations

How Many People, Accounts, & Debt?

- **Population of Consumers**
  - Remove deceased.
  - Restrict by age (e.g. missing ages, 20 - 80)
  - Restrict by geography (e.g. exclude non-US)
  - Restrict by data quality (e.g. consumers with tradeline data, with SSN/ITIN)

How Many People, Accounts, & Debt?

- **Population of Consumers**
  - Remove deceased.
  - Restrict by age (e.g. missing ages, 20 - 80)
  - Restrict by geography (e.g. exclude non-US)
  - Restrict by data quality (e.g. consumers with tradeline data, with SSN/ITIN)
- **Population of Credit Accounts**
  - De-duplicate joint accounts & authorized users
  - Remove accounts closed or not recently updated (1, 3, 6, 12 months)
  - Accounts in dispute may want to remove observations (but these may also be of interest)

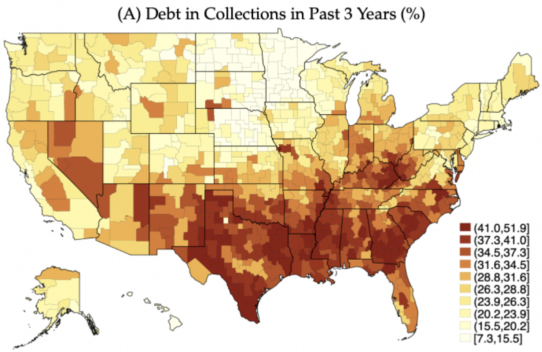
How Many People, Accounts, & Debt?

- **Population of Consumers**
  - Remove deceased.
  - Restrict by age (e.g. missing ages, 20 - 80)
  - Restrict by geography (e.g. exclude non-US)
  - Restrict by data quality (e.g. consumers with tradeline data, with SSN/ITIN)
- **Population of Credit Accounts**
  - De-duplicate joint accounts & authorized users
  - Remove accounts closed or not recently updated (1, 3, 6, 12 months)
  - Accounts in dispute may want to remove observations (but these may also be of interest)
- **Value of Outstanding Debt**
  - See above adjustments
  - Credit card balances are statement balances *not* debt
  - Debt with or without debt in collections
  - Some subprime credit and student loans unobserved

2. Financial Distress

Measuring Financial Distress

Example: Keys, Mahoney, & Yang (*Review of Financial Studies, 2022*)



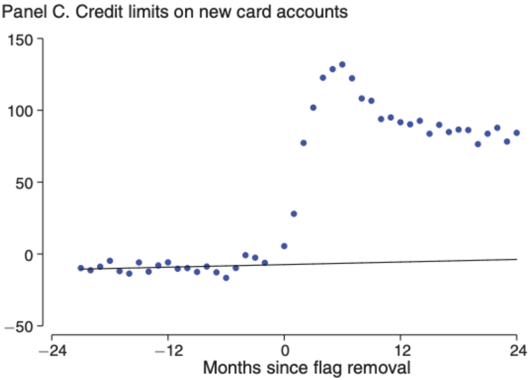
Measures

- Credit Score
- Delinquency
  - 30+ / 60+ / 90+
  - If studying COVID-19 period include accommodations (e.g. forbearance)
- Debt in Collections
  - Flow typically better measure than stock due to low persistence in reporting
  - Medical & non-medical but reporting practices change over time / x-states
- Bankruptcy

3. Credit Access

Measuring Credit Access

Example: Gross, Notowidigdo, & Wang (*American Economic Journal: Macro*, 2022)



Measures

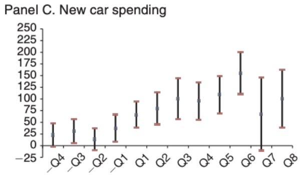
- Credit Score
  (e.g. Laufer, *American Economic Journal: Economic Policy*, 2021)
- Credit Card Limits and HELOC Limits
  - Outstanding, New Originations
- Account Openings / Inquiries
  (e.g. Romeo & Sandler, *Journal of Public Economics*, 2021)
- Inferred borrowing rates
  (e.g. Shahidinejad, 2022 WP; Yannelis & Zhang, 2022 WP)

4. Consumption

Consumption Measures

Auto Loans

Example: Di Maggio, Kermani, Keys, Piskorski, Ramcharan, Seru, & Yao (*American Economic Review, 2017)*



- New auto loans originations #, amounts
- Change in auto loan balances

Caveats: Unobserved

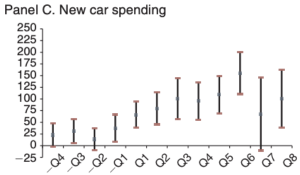
(i) All autos purchased with cash

(ii) Some subprime auto loans

Consumption Measures

Auto Loans

Example: Di Maggio, Kermani, Keys, Piskorski, Ramcharan, Seru, & Yao (*American Economic Review, 2017)*



- New auto loans originations #, amounts
- Change in auto loan balances

Caveats: Unobserved

(i) All autos purchased with cash

(ii) Some subprime auto loans

Credit Card Spending

Example: Ganong & Noel (*American Economic Review,* 2020)

$$spend_t = balance_t - balance_{t-1} + payment_t$$

- Bound at zero
- Restrict to credit card lenders reporting payments data

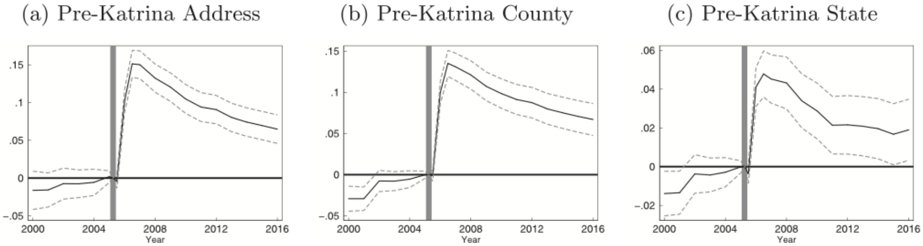
Caveat

Poor, highly selected coverage post-2015 (Guttman-Kenney & Shahidinejad, 2022 WIP)

5. Mobility

Mobility

Example: Bleemer & van der Klauuw (*Journal of Urban Economics*, 2019)



Guidance

- Best for 'permanent' moves
  (e.g. not good if move address every month, but good if you rarely move)
- Caveats
  - Can take time to register
  - Tricky establishing 'primary' residence if multiple addresses (e.g. students, 2nd homes)
  - What's possible varies with data structure

Thank you! All slides from today are posted on my website

www.benedictgk.com

benedict@chicagobooth.edu

@gk_ben

Questions?

Slides:

www.benedictgk.com

Citation:

Gibbs, Guttman-Kenney, Lee, Nelson, van der Klauuw, & Wang (2022) A Guide to Consumer Credit Reporting Data.