STAFF
REPORTS

NO. 1114
AUGUST 2024

Consumer Credit Reporting Data

Christa Gibbs | Benedict Guttman-Kenney | Donghoon Lee
Scott Nelson | Wilbert van der Klaauw | Jialan Wang

FEDERAL RESERVE BANK of NEW YORK

Consumer Credit Reporting Data

Christa Gibbs, Benedict Guttman-Kenney, Donghoon Lee, Scott Nelson, Wilbert van der Klaauw, and Jialan Wang

Federal Reserve Bank of New York Staff Reports, no. 1114
August 2024
https://doi.org/10.59576/sr.1114

Abstract

Since the 2000s, economists across fields have increasingly used consumer credit reporting data for research. We introduce readers to the economics of and the institutional details of these data. Using examples from the literature, we provide practical guidance on how to use these data to construct economic measures of borrowing, consumption, credit access, financial distress, and geographic mobility. We explain what credit scores measure, and why. We highlight how researchers can access credit reporting data via existing datasets or by creating new datasets, including by linking credit reporting data with surveys and external datasets.

JEL classification: D10, D82, E21, G50, H31
Key words: consumer credit reporting data, credit bureaus, measurement, credit scores, asymmetric information, household finance

Lee, van der Klaauw: Federal Reserve Bank of New York (emails: donghoon.lee@ny.frb.org, wilbert.vanderklaauw@ny.frb.org). Gibbs: Consumer Financial Protection Bureau (email: christa.gibbs@cfpb.gov). Guttman-Kenny: Rice University, Jones Graduate School of Business (email: benedictgk@rice.edu). Nelson: University of Chicago, Booth School of Business (email: scott.nelson@chicagobooth.edu). Wang: University of Illinois at Urbana-Champaign, Gies College of Business, NBER (email: jialanw@gmail.com). The authors thank David Romer (the editor), five anonymous referees, Aly Brown, Amy Quester, Andrés Shahidinejad, Breno Braga, Brian Bucks, Evan White, Jonah Kaplan, Lance Lochner, Matthew Notowidigdo, Michael Varley, Pavneet Singh, ASSA and NBER conference participants for their feedback improving this paper, as well as Jehoon Chung for research assistance. They shared an earlier draft with representatives from Equifax, Experian, TransUnion, FICO, and VantageScore and are grateful for their feedback. Prior to circulation, this paper was reviewed in accordance with the Federal Reserve Bank of New York review policy and the Consumer Financial Protection Bureau policy on independent research. Gibbs works for the Consumer Financial Protection Bureau, which implements and enforces Federal consumer financial laws, including those discussed in this paper. Guttman-Kenney acknowledges support from the NBER's PhD Dissertation Fellowship on Consumer Financial Management funded by the Institute of Consumer Money Management, and Wang acknowledges support from the Gies College of Business.

1 Introduction

Consumer credit reporting data—also known as credit files, credit records, or credit bureau data—are a market response to fundamental economic challenges of information asymmetry between borrowers and lenders (e.g., Jaffee and Russell, 1976; Stiglitz and Weiss, 1981). The market has generated a system where tens of thousands of firms voluntarily share information each month to produce data containing a history of consumers' borrowing and repayment behaviors for roughly nine-in-ten adults in the US (Brevoort, Grimm and Kambara, 2015), primarily recorded by three consumer reporting agencies (CRAs)—Equifax, Experian, and TransUnion. These data on borrowing, repayment, and other interactions with credit markets are the main information source for millions of lending decisions.

Consumer credit reporting data are primarily designed to cover the liabilities side of a consumers' balance sheet. These data contain monthly information about consumers' outstanding balances and repayments on credit accounts, bankruptcy and other public records, applications for credit, debts in collection, and personally identifying information. These are commonly organized across several different data files as described in Table 1 and discussed more in Section 3. The CRAs use these files to create consumer-level aggregated datasets that include geographic location, credit scores, and many other consumer-level variables. With access to credit reporting data, researchers can measure not only consumers' financial behaviors, but also their consumption, intra-household behaviors, and geographic mobility in order to answer a broad range of research questions.

Consumer credit reporting data have also contributed significantly to economics research. These data came to research prominence in helping to understand the 2007–2008 US financial crisis (e.g., Mian and Sufi, 2009, 2011, 2014), and soon after were used to study a wide array of topics in finance, and especially in household finance (e.g., Bhutta, Goldin and Homonoff, 2016; Di Maggio et al., 2017; Ganong and Noel, 2020).

Our paper is designed to inform a general audience about the wide-reaching potential of consumer credit reporting data. Indeed, these data have now been used for economic research covering every *Journal of Economic Literature (JEL)* code from C to R. These data are advancing our understanding of macroeconomics including the transmission of monetary policy (e.g., Beraja et al., 2019; Berger et al., 2021) and consumption behavior (e.g., Mian, Rao and Sufi, 2013; Benmelech, Meisenzahl and Ramcharan, 2017). These data have also been used across a diverse range of microeconomics topics, including the fields of environmental, health, industrial economics, labor, marketing, public, and urban economics. For example, researchers have studied abortion (Miller, Wherry and Foster, 2023), advertising (Bertrand et al., 2010), eviction (Collinson et al., 2024), geographic migration (Howard and Shao, 2023), health insurance reforms (Finkelstein et al., 2012), hospital admissions (Dobkin et al., 2018), human capital (Hampole, 2024), intergenerational co-residence (Dettling and Hsu, 2018), the minimum wage (Aaronson, Agarwal and French, 2012), natural disasters (Gallagher and Hartley, 2017), non-standard preferences (Meier and Sprenger, 2010), and traffic fines (Mello, 2023). For readers interested in the use of these data within a particular economic field, the Online Appendix provides a more detailed literature review, with papers grouped by *JEL* codes.

In this paper, we aim to explain the credit reporting processes and content of credit reporting data, provide practical guidance that helps standardize best practices, reduce barriers to entry for new researchers, support the work of journal editors and reviewers, outline frontiers for future research, and generally promote greater understanding among researchers about the challenges and opportunities of using these data.

To better understand how to use credit reporting data, it is helpful to understand why these data exist and how they are generated. The first subsection of Section 2 reviews theoretical work on information economics and credit market structure, as well as related institutional details, to help to understand the existence and roles of credit reporting data. The second subsection then introduces further institutional details of US credit reporting data and their implications for research. Prior work on the use of consumer credit reporting data for research only covers the early emergence of these data (Furletti, 2002; Avery et al., 2003; Miller, 2003) or a specific credit panel (Lee and Van der Klaauw, 2010), with much having changed since.

Section 3 is the heart of the paper and uses examples from the literature to provide practical guidance for researchers considering using credit reporting data to define relevant populations to study and to construct economic measures of borrowing, consumption (auto purchases, credit card spending, and cash-out equity from mortgage refinancing), credit access (including the amount of new credit and the costs of borrowing), financial distress (using a variety of approaches), intra-household behaviors, and geographic mobility. We provide overarching guidance on using these data and discuss how different data structures may affect these measures.

Credit scores are often used in research as an outcome or covariate. In Section 4, we provide a general introduction to what credit scores are, what information they are based on, and differences across different types of credit scores.

Section 5 provides guidance on how researchers can access credit reporting datasets. We discuss how to access existing panels, create new datasets, use credit reports as a sampling frame for surveys, and link credit report data with other datasets —including established links with administrative mortgage data.

Finally, Section 6 briefly concludes with a discussion that includes exciting avenues for research.

While this paper largely focuses on US credit reporting data, there are many similarities between US data and analogous data in other countries. Differences between US and international data are discussed further in the Online Appendix and in Djankov, McLiesh and Shleifer (2007); Miller (2003); International Finance Corporatation (2012) and World Bank (2012).

This paper is designed for a general readership of users. To accompany this paper, we provide a set of resources in our online appendix to cater to a variety of more detailed interests, including a collection of sample code, general coding recommendations, and excerpts of especially relevant code from publicly available replication packages. The Online Appendix provides more detailed information on the structure of credit reporting data—including details specific to each different type of credit account (including home-based loans, credit cards, auto, and student loans), existing panels, and details of how to construct datasets. We would encourage users of these data to consult these additional online materials.

2 Credit Reporting: Economic Research and Practice

2.1 The Economics of Credit Reporting

Whereas much of this article focuses on the use of consumer credit reporting data for *measurement*, this section examines the economics of the data themselves: why these data emerge in equilibrium, why these data have economic use, and why these data are the subject of considerable regulation. We highlight topics where the themes in theoretical work line up closely with issues emphasized in actual practice. We also highlight topics where the overlap between theory and practice is less perfect, presenting opportunities for further work.

Information Asymmetries

Many readers will recognize that credit reporting data can help address information asymmetries. A rich body of research has helped formalize this idea. In one early contribution, Pagano and Jappelli (1993) show why lenders may choose endogenously to share information with each other about their borrowers when facing adverse selection. In their model, information sharing is particularly helpful for screening a set of borrowers who "migrate" from other banks. When the propensity for migration is sufficiently high, it becomes privately optimal for banks to join a credit bureau, even if the credit bureau only has partial coverage across banks and borrowers. Shaffer (1998) likewise emphasizes adverse selection but focuses on potential borrowers whose applications get rejected by one or more lender, generating a winner's curse for lenders who ultimately approve a previously rejected borrower; credit bureaus can expand credit supply by partially obviating these winner's curse concerns.

A separate theoretical literature emphasizes credit reporting data's role in reducing moral hazard. Papers in this area tend to share the Diamond (1989) insight that reputational incentives discipline moral hazard in debt markets with repeated interaction. There are several variations on this idea, however, that are specific to credit bureaus: Padilla and Pagano (2000) note that credit bureaus are most effective at disciplining moral hazard when only negative information (e.g., non-repayment), rather than positive information (e.g., a history of successful repayment) is recorded in the bureau; Vercammen (1995) analyzes the optimality of finite-memory credit histories in disciplining moral hazard when types are sufficiently persistent[1]; Padilla and Pagano (1997) note that

[1]See also Elul and Gottardi (2015); Bhaskar and Thomas (2019) and Kovbasyuk and Spagnolo (2024) on the optimal length of credit bureau memory.

the formation of a credit bureau helps discipline moral hazard in part because it commits *banks* not to extract future rents from borrowers who exert effort to repay. Another form of moral hazard that credit bureaus may help address is sequential banking (Bizer and DeMarzo, 1992; De Giorgi, Drenik and Seira, 2023), where a lender may be concerned that her borrower will take on other debt at a later date and thus raise the default risk on her original loan.[2]

This theoretical emphasis on the importance of information asymmetries appears to be borne out in the real world, though historically speaking, adverse selection may have emerged as an important force earlier than did moral hazard. Lauer (2017)'s history emphasizes adverse selection as a key driver of the formation of early CRAs in the late 1800s and early 1900s in the US. These local and regional CRAs' typical function was to "*quarantine* poor credit risks" (emphasis added) and "purge dishonest debtors," rather than to discipline borrowers to repay; one early CRA representative warned that a lender who does not subscribe to the CRA will become "a dumping ground for undesirables." Discussion of encouraging repayment (and disciplining moral hazard) appears to have only come later, as CRAs become more established and well-known among consumers. By the late 1920s, one contemporaneous commentator described CRAs as having a "splendid moral influence in the community," and some retail lenders found their loans were repaid more quickly after they advertised joining a CRA (Lauer, 2017).

Overall, on the topic of information asymmetries we see substantial overlap between what is emphasized in theoretical work and what is seen as relevant by practitioners. One of CRAs' main roles is to reduce information asymmetries, and information asymmetries are also a key driver of why CRAs emerge in equilibrium.

Market Structure

The emergence of CRAs also depends crucially on lender market structure. In theory work on this subject, many papers draw on the insight in Petersen and Rajan (1995): lenders can extract information rents when they know more about their own borrowers than their competitors know, and it may not be privately or socially optimal for lenders to share this information with each other. Similar analyses are developed in Sharpe (1990) (see also a correction by Von Thadden (2004)), Dell'Ariccia and Marquez (2004), and Dell'Ariccia (2001), and discussed in Hunt (2005). Pagano and Jappelli (1993) develop this formally in the context of credit bureau formation, showing that credit bureaus may be less likely to emerge when incumbent banks face more threat of competitor entry. Similarly, Marquez (2002) shows how large banks may have less incentive to join a credit bureau than small banks, given their inherent information advantage in lending to a larger share of the potential borrower pool. The economic forces here are often subtle, however: Hauswald and Marquez (2003) study how privately and publicly available information together affect credit

[2]While it can be ambiguous whether subsequent credit access raises or lowers default risk (Hunt, 2005), this channel is often a motivation for the use of credit reporting data: Bennardo, Pagano and Piccolo (2015) illustrate how credit reporting data address a sequential banking problem, and Bar-Isaac and Cuñat (2014) develop this idea by studying "hidden lenders" who may create a sequential banking externality that limits credit supply.

supply, while Bouckaert and Degryse (2006) analyze how the interaction between market structure and information sharing depends crucially on the severity of adverse selection in the market.

Turning to practice, these theoretical insights also lend insight to how CRAs might *evolve* in the face of recent changes in market structure and technology. The rapid development of open banking (which allows secure sharing of financial data between banks and third party service providers), the emergence of large (non-CRA) data monopolists, and growing concentration among traditional banks coupled with the proliferation of FinTech and shadow-bank competitors, all raise interesting questions about the role of CRAs in the future. Can the data sharing required by open banking laws substitute in some ways for CRAs' traditional roles in overcoming information asymmetries? How does this interact with ongoing changes in lender market structure? Are the data sold by non-CRA data aggregators substitutes or complements for traditional CRA data? Work by Babina et al. (2024); He, Huang and Zhou (2023); Rishabh (2024) makes early progress on these questions, and we see abundant opportunities for future work bridging between practice and theory.

A related question concerns the advantages and disadvantages of having privately owned and operated CRAs, vs. publicly owned and operated ones. (Public CRAs are sometimes called credit registries.) Djankov, McLiesh and Shleifer (2007) review the existence of private and public CRAs across 129 countries and argue that public CRAs may be particularly valuable when legal institutions are weak, while private CRAs may offer more comprehensive services, such as bundling with credit scores. Miller (2003) notes that public CRAs may have better coverage when lenders' data sharing with private CRAs is voluntary. In the US, antitrust arguments were advanced in the 1920s that private CRAs should be granted exclusive territories, similarly to regulated utilities (e.g., electric and gas providers), but courts ruled instead that CRAs' data are a "commodity" and should be provided competitively (Lauer, 2017). We see questions about the costs and benefits of private CRAs as integral to future research and policy related to CRAs and lender market structure.

Public Policy

The policy debate in the US over regulating CRAs has largely focused on privacy, fairness, and data quality. Interestingly, these are topics with which economic research on CRAs has engaged relatively little. In turn, theoretical work on topics such as the design of a scoring system (Bonatti and Cisternas, 2020; Frankel and Kartik, 2022), or the trade-off between the inherent informativeness of, and the manipulability of, different datapoints in credit reporting data (Ball, 2024), have seen less play in related policy debates. We review some of this policy debate next, and then discuss implications for future research.

We first consider the policy debate's emphasis on data quality. Theoretical work typically assumes CRAs can validate information at relatively low cost (for example, Padilla and Pagano, 1997), but in practice this cost might not be low, or CRAs may otherwise elect to not fully ensure the quality of their data absent regulatory intervention. In the 1960s, for example, there was evidence that some creditors, like the Federal Housing Administration, found enough accuracy issues with some credit reports that they created their own preferred lists of reliable CRAs. Additionally, consumers typically could not review their own reports, while others without legitimate purposes were perceived as accessing credit records too easily (Lauer, 2017). This led to a first attempt at regulating credit reporting data, the introduction of "A Bill to Protect Consumers Against Arbitrary or Erroneous Credit Ratings, and the Unwarranted Publication of Credit Information" (National Consumer Law Center, 2022).

This policy debate about data quality soon dovetailed with a policy debate about a second issue, consumer privacy. The aforementioned bill focused on "Arbitrary or Erroneous Credit Ratings" was followed by a series of Congressional hearings about growing concerns with privacy, as CRAs began to computerize and amass more information on consumers (National Consumer Law Center, 2022; Miller, 2003; Lauer, 2017). This debate culminated in the passage of the Fair Credit Reporting Act (FCRA) in 1970. As stated by Congress, the FCRA was originally enacted to "require that consumer reporting agencies adopt reasonable procedures [...] with regard to the confidentiality, accuracy, relevancy, and proper utilization" of credit record information.[3] The market may have eventually resolved some of these issues without the introduction of the FCRA, although continued issues with data accuracy resulting in later amendments to the FCRA in 1996 and 2003 suggest otherwise (see Table 2). Likewise, Hunt (2005) argues that the ability and incentive to correct different types of errors differ for lenders, credit bureaus, and consumers, which may result in the under-provision of data accuracy and suggests a role for regulation.

A third issue emphasized in the policy debate over CRAs is concern about discrimination and credit market disparities across protected groups such as race and gender. The 1974 Equal Credit Opportunity Act (ECOA), for example, requires that when two spouses both use or are liable for an account, a lender that reports the account to a CRA must do so for both spouses, in order to ensure that both spouses receive the benefit of the payment history on an account. ECOA initially only covered gender and marital status as a prohibited basis for discrimination but was amended in 1976 to also cover race, national origin, religion, age, and other bases.

There are, of course, sizeable economic literatures on the topics this regulatory debate has emphasized: see, for example, Charles and Guryan (2011) and Small and Pager (2020) on discrimination and its remedies, Bergemann and Bonatti (2019) on what products are sold by information intermediaries such as CRAs, and Goldfarb and Tucker (2012) and Acquisti, Taylor and Wagman (2016) on privacy and consumer demand for it. However there remain important gaps between what is known in economics research and what appears to be important for regulatory debates in the CRA context—suggesting opportunities for future work. Are the consumers who value privacy the same as those who gain, or lose, in pecuniary terms from coarsened credit reporting data? Does demand for privacy arise more from concerns about certain private matters being *knowable*, or concerns about how that knowledge will be *used* (Nissenbaum, 2020)? Is CRA regulation the most efficient way to combat discrimination? Why can market forces not on their own provide the

[3] FCRA §602(b), 15 U.S.C. §1681a(b)

levels of privacy, data quality, or non-discriminatory data features that consumers demand? On the topic of CRA policy and regulation, there is opportunity for both researchers and policymakers to learn from what each other's work has emphasized.

2.2 Regulatory Overview

We next turn to the practicalities of how regulatory changes and pressures have affected how these data are constructed. As discussed in the prior subsection, most of the regulatory and legal changes that have shaped credit reporting data are the result of concerns around privacy, discrimination, and data inaccuracies related to the apparent inability of individuals to resolve these issues in a market where they are neither the buyer nor seller but are directly affected by these data. Separately, concerns about fairness have led to laws and regulations that place restrictions on what can be included in credit scoring models and what types of information appear on credit records. While the prior subsection already introduced the basics of the FCRA and ECOA, this subsection provides greater detail that emphasizes how the regulatory environment affects the content of credit reporting data.

The Fair Credit Reporting Act (FCRA) in 1970 became the first and primary federal law in the United States regulating credit record data, CRAs, those who report credit information to CRAs ("furnishers"), and those who use the data ("users") (Table 2). The law mandates that adverse information such as delinquencies and collection accounts could generally only remain on a credit report for up to seven years, but some information can remain on a report for longer, such as bankruptcies (Table 3).[4] The FCRA was further amended by the Consumer Credit Reporting Reform Act in 1996, where Congressional testimony interestingly argued that, because consumers are not the CRAs' customers, "market incentives" do not effectively address consumers' concerns (National Consumer Law Center, 2022).

In addition, settlements and agreements with CRAs can change reporting standards. The National Consumer Assistance Plan (NCAP), for example, was the result of an investigation of the three major CRAs initiated by the New York Attorney General following consumer complaints about credit reporting errors. Under NCAP, several changes were made to improve data accuracy, especially relating to collection accounts, public records, and authorized user accounts. This ultimately led to many changes including the deletion of tax liens, some collections, and many civil judgments from credit reports. Finally, some changes in reporting, like the $500 minimum for reporting medical collections in 2023, have been voluntarily implemented by the CRAs following extensive public discussion of the issue.

Despite these regulatory changes and settlements to address ongoing problems, issues with credit reporting continue. For example, the Federal Trade Commission conducted a series of reports

[4] Congress lowered the reporting duration for bankruptcies appearing on credit reports from 14 years to ten years in 1978. Since then, Congress has occasionally proposed further reductions in obsolescence thresholds, but none have been enacted (National Consumer Law Center, 2022). Obsolescence periods vary by country. For example the threshold is three years in Sweden and ten years in Greece (Bos and Nakamura, 2014).

reviewing credit report errors and estimated in 2012 that 5% of consumers' credit reports contained errors that meaningfully adversely affected their credit access (Federal Trade Commission, 2012). Similarly, credit reporting problems persistently top the Consumer Financial Protection Bureau's (CFPB) consumer complaints database.[5]

Notably, the FCRA does not require CRAs to furnish data on their lending agreements to any CRAs,[6] but it does impose accuracy requirements when information is furnished, and it specifies some information that must be reported if a furnisher provides any credit information (see the Online Appendix for more information). Some fields, like actual payment amount, are selectively not reported by some furnishers in some markets trying to maintain their advantage with asymmetric information on more profitable accounts (Guttman-Kenney and Shahidinejad, 2024).

To aid compliance and reduce coordination costs, the Consumer Data Industry Association (CDIA), the primary trade association for the credit reporting industry, established and manages a format for furnishing data. These formatting rules, known as the "Metro2" format, are not legally required, but were developed to help data furnishers comply with legal requirements while also offering benefits to furnishers, CRAs, and credit data users from greater consistency in formats and definitions. These formats are updated over time to reflect credit market developments (e.g., codes for buy now, pay later (BNPL) products were added in 2022 and, codes for rent furnishing were added in 2023).

Finally, we conclude this section with a high-level overview of how these market and regulatory dynamics influence the flow of information and participation by different firms. As previously discussed, there are benefits to lenders to furnish information to the CRAs, such as reducing moral hazard and adverse selection (e.g. Liberman et al., 2019), and most large lenders choose to furnish. While there are no direct costs paid to the CRAs to submit data, there are several indirect costs involved. Furnishers must submit their data in a specified format and comply with the requirements of the FCRA and ECOA. Sharing information on their consumers also reduces their information advantage over other lenders. As a result of these costs, some lenders may choose to not furnish at all or to selectively furnish to a subset of CRAs, on a subset of product lines, or using a subset of data fields (see the Online Appendix for more discussion of these issues).

The CRAs, for their part, face similar compliance costs plus additional concerns over accuracy in trying to synthesize information from different furnishers into individual consumer credit reports. CRAs first validate the data they receive primarily by identifying inconsistencies, but the CRAs "generally rely on furnishers to report information on consumers that is complete and accurate" (Consumer Financial Protection Bureau, 2012). If the CRA identifies any issues, the data are rejected and need to be resubmitted by the furnisher. CRAs then use their proprietary algorithms to determine which accounts belong on each record (and what is a separate consumer record). These

[5] https://www.consumerfinance.gov/data-research/consumer-complaints/

[6] Other laws or federal rules, however, may require that some types of credit are reported. For example, in 2008 the Higher Education Act was amended to require CRAs to furnish information on all federal student loans they service to limit credit record differences for borrowers due to their servicer or lender (20 U.S.C. §1080a).

reports can then be queried for a cost by creditors, potential employers, and others in response to applications submitted by consumers. Consumers can also access their reports to monitor for fraud and file disputes, which the CRAs are obligated by the FCRA to investigate. The CRAs also supplement their traditional credit record data by acquiring alternative credit reporting bureaus, and developing other products and capabilities based on their data (e.g., datasets designed for marketing purposes) to sell back to lenders.

For more details on the credit reporting process, including potential sources of measurement error, see the Online Appendix.

3 Constructing Economic Measures

Credit reporting data enable researchers to quantify centrally important economic statistics. How many consumers reside in a geographic area? Where are consumers moving to and from, and how frequently? How much are they consuming? How much credit can they access and how much debt do they have? What type of debt and at what cost? Are they repaying that debt or are they in financial distress? These are all questions credit reporting data can be used to answer.

In this section, we explain how researchers can construct a variety of economic measures from these data. Section 3.1 provides overarching guidance to researchers for using consumer credit reporting data. Section 3.2 explains how to define populations of consumers and accounts in credit reports, and de-duplicate these for accurately calculating aggregated population statistics. Sections 3.3 to 3.8 then explain how to construct various measures of economic interest, highlighting approaches used in prior literature and making some recommendations to encourage greater standardization of what we consider best practices. Section 3.3 describes how to measure consumer borrowing. Section 3.4 shows a variety of measures of financial distress: bankruptcy, collections, delinquency, and other approaches. Section 3.5 explains how to construct measures of credit access: new accounts, credit limits, inquiries, and borrowing costs. Section 3.6 covers measures of consumption: auto purchases, credit card spending, and cash-out equity from mortgage refinancing. Section 3.7 discusses how to use these data to measure geographic mobility, while Section 3.8 discusses measuring intra-household and intergenerational behaviors.

3.1 Overarching Guidance

Our overarching recommendation for researchers using credit reporting data is to be clear and precise on how they construct their measures. At a minimum, we suggest researchers should clearly state four things. (1) Which credit reporting datasets the measures are being calculated from. Credit reporting data consists of a variety of datasets, summarized in Table 1 (see the Online Appendix for details): for example, the “Tradeline File” contains account-level information about consumers’ outstanding balances and repayments on credit accounts, while the ‘Inquiries File” includes applications for credit. (2) The frequency at which the measures are calculated from (e.g., monthly, quarterly, annually). (3) Which data restrictions are applied (e.g., criteria for excluding inactive accounts, low-quality credit records (discussed more below), or deceased consumers). (4) Whether the researcher calculates a measure themselves, and if so, the formula used for calculation including whether any inference is made for missing data. This clarity is especially important as credit reporting data were not originally created for research purposes and therefore lack standard conventions for defining economic measures.

CRAs have data available at the tradeline-level—supplied to the CRAs by furnishers—and aggregated to the consumer-level by the CRAs ("Consumer-Level Aggregated Datasets" in Table 1). Which should researchers use? If tradeline-level data are available, this will allow researchers the most flexibility to transparently construct measures closest to the target objects of interest. Nonetheless, for many researchers, consumer-level aggregated data will be sufficient and cost effective. The general downsides of consumer-level aggregated data are that these variables can be opaquely defined by the CRAs, their definitions may vary across CRAs, and they are primarily designed as inputs to scoring models which may not match researchers' needs. Consumer-level aggregated data is often split across many modular datasets designed to cater to CRAs' heterogeneous client base, necessitating careful review of data dictionaries before purchase. A particular challenge of consumer-level aggregated data is that changes in reporting practices at the tradeline-level can give a false impression of changing real behavior in the aggregated data, whereas with tradeline data such changes in reporting are directly observed and can be adjusted for. When introducing economic measures below, we consider how constructing measures using granular tradeline data may compare to constructing them using consumer-level aggregated data.

3.2 Populations

First, we explain how to define various populations or sample frames of interest. Answering a seemingly-straightforward question like "how many consumers have a credit record?" would vary depending on how a "consumer" (or "credit record") is defined in these data. Different approaches generate answers that differ by tens of millions of consumers (Brevoort, Grimm and Kambara, 2015). Similarly straightforward statistical exercises, like measuring how many credit products or how much debt a consumer holds, can vary by several multiples depending on which types of variables in the data are used.

3.2.1 Populations of Consumers

Containing information on consumers' ages and geographic locations over time, credit reporting data have the key strengths of high coverage, frequency, and size. These strengths enable researchers to study populations and sub-populations split by characteristics such as age, across granular geographies, and over time. This can complement official public data sources—most notably Census and Internal Revenue Service (IRS) data—as well as commercial data products such as address histories collected from other sources including telephone directories and subscription services (Phillips, 2020). A relative disadvantage of credit reports is that consumers without credit reports (so-called "credit invisibles") are unobserved, a group that disproportionately includes some racial and ethnic minorities, younger consumers, and unbanked consumers.

Credit records do not have a perfect one-to-one correspondence with people. Credit records are assembled by CRAs by matching information from different furnishers based on proprietary algorithms and the identifying information they receive. If the furnished data contain errors or are incomplete, the resulting records may sometimes result in extraneous information that can either be included on a record that should be matched to another file, or stored in "fragment" records. These low-quality fragment records occur when one or more tradelines, inquiries, or public records for an individual cannot be correctly consolidated into the same credit file. Instead, one individual may have multiple unlinked credit reports for some periods. Fragmented records are especially likely to occur for credit records with lower quality identifying information (e.g., without social security numbers or SSNs), for individuals who move frequently, or who have common names. Ultimately, this means that there often are more credit records than adults in a population.

Data furnishers and CRAs do not always have timely and accurate death information, leading to the continued reporting and updating of credit reports of individuals following their death. To help address this, researchers typically remove consumers with a missing date of birth or a birth date that is unrealistic (e.g., implied ages over 100).[7] Some researchers may be interested in stricter age restrictions (e.g., working-age consumers) to reduce the number of unobservably deceased consumers in their sample.

Imposing data restrictions beyond age involve more trade-offs. We generally recommend researchers do not include consumers who only have inquiries on their credit files. Inquiry-only consumers are often fragmented records. Researchers may wish to restrict their analysis to consumers with SSNs / ITINs; these are less likely to be fragment files, but this choice may also remove some groups of consumers of particular interest. Researchers may also wish to restrict to consumers based on the number of observed tradelines, as credit reports with more tradelines may be less likely to be fragment files. For example, a researcher could restrict to consumers who have held at least one credit product over the last ten years, or consumers that appear continuously with an open tradeline over a sustained period of time.[8] Using this approach typically produces an aggregate number of consumers which is plausible given the size of the US adult population. Finally, researchers may also follow the approach of Brevoort, Grimm and Kambara (2015) and keep only records that persist in the data for at least four years (or some other threshold).

[7]Birth dates are not missing for data from 2009 onwards due to the Fair and Accurate Credit Transactions Act, which requires the use of birth dates to ensure more accurate matches between tradelines and consumers. However, birth dates are missing for a third of data from the early 2000s (Federal Reserve Board, 2007; Lee and Van der Klaauw, 2010).

[8]Some research may warrant including consumers who only have collections or public records, though doing so leads to the inclusion of additional fragment files, as evidenced by sample sizes that imply an implausibly large US adult population, especially prior to the changes introduced by NCAP.

3.2.2 Populations of Active Accounts

Credit accounts remain on credit reports long after they are no longer in use or have been closed. If using consumer-level aggregated variables, the criteria a CRA uses for including inactive accounts may be unclear. While the inclusion of zero-balance closed accounts will not affect the computation of debt aggregates, their inclusion will affect analyses focused on the number and types of accounts or account-holders.

If a researcher is using tradeline-level data, then they can specify their own criteria for defining inactive accounts. In particular, researchers may want to remove accounts for which updated records have not been recently furnished. Accounts that are not recently furnished may have been closed, have different balances, or have become inactive. Different CRAs have different guidance on how to do so, ranging from removing accounts not updated in the last 1 to 12 months. Researchers may wish to check the time series to ensure that removing "inactive" accounts does not generate artificial jumps in the volume or value of accounts, and loosen the threshold as needed. Regardless of the approach, we recommend researchers be clear on what criteria they use.

Inactive credit cards that are open but not used by consumers are difficult to define but greatly affect the number of accounts measured in credit reports. Researchers interested in studying credit card behaviors may want to focus on accounts in use. Historically, once a credit card account has a zero statement balance for every month in the last year, it rarely gets used in the near future.

Although the largest furnishers typically furnish information to all three nationwide CRAs, there is typically no legal requirement that a furnisher do so, and some smaller firms do not furnish to all three. Even some large firms have occasionally furnished to only one CRA (e.g., Harney, 2003). This means that credit file data from any single CRA does not contain all debts for all people, and some consumers may appear in one CRA's data but not in another CRA's data. For example, Guttman-Kenney and Hunt (2017) find differences across CRAs in the credit reports of UK payday lending customers.

Because not all debts appear on credit records, the relative size of a credit market according to credit record data may differ from that in other sources. For example, the Federal Reserve Board's G.19 data show student loan debt as the second largest form of household debt and auto debt as the third largest as of 2023, but credit reporting data suggest the ranking is reversed. Brown et al. (2015) find aggregate debt estimates from credit reporting data align with estimates from the Survey of Consumer Finances.

A related issue is that some types of debt appear in credit reporting data rarely, if at all. Argyle et al. (2021) label debt not observed in credit reports "shadow debt" and find that in their sample of bankruptcy filers, 7.4% of total debts are not observed in credit reports from one CRA at the time of filing. Similar estimates for non-bankrupt consumers are difficult to find, as there are few comprehensive sources for this information. Shadow debt may include some subprime loans not typically furnished to CRAs (e.g., some subprime auto loans and payday loans), most unpaid utility, business, and rent bills. Credit reports do not include information on a number of other financial products, including most BNPL loans, many business credit cards and loans, cash advance apps, car title loans, pawnshop loans, and tax refund anticipation checks. Informal lending (e.g., via family, friends, illegal lenders) is also never observed in credit reports. Finding new data sources to study these unreported, but economically important, debts is an important challenge for researchers.

3.2.3 De-duplicating for Aggregated Population Statistics

Researchers may need to de-duplicate accounts that are jointly-held or have authorized users. De-duplication is required to calculate accurate aggregated population statistics—such as outstanding balances, number of accounts, or delinquency rates—at a national-level or at more granular geographic groups, but are not required for using disaggregated data (e.g., account-level or consumer-level). To avoid double-counting jointly held or cosigned accounts, weights are commonly assigned to accounts not reported as held by a single individual. For example, individual accounts may received a weight of one; jointly held or cosigned accounts may receive a weight of one-half, and authorized user accounts may receive a weight of zero. The de-duplication approach can vary depending on the sampling methodology used to create a credit reporting dataset (see Lee and Van der Klaauw, 2010, for details). Researchers with credit reporting data on other individuals living at the same address can also aggregate to household-level statistics. For further details about household-level statistics, see the Online Appendix.

3.3 Borrowing

Borrowing is naturally a central component of household balance sheets (e.g., Zinman, 2015), and data on borrowing can lend insights in fields as broad as macro, labor, health, and finance. Credit reports show outstanding balances in both tradeline-level and consumer-level aggregated datasets. Researchers may measure total borrowing by summing all outstanding balances across all active tradelines, or separately by loan type. For example, Beshears et al. (2022) study the effects of auto-enrollment in a retirement plan on borrowing. In their analysis, increased borrowing may have different welfare implications depending on whether it is in the form of mortgage debt or credit card debt.

Researchers interested in studying borrowing via alternative financial services—such as a payday or auto title loans—not contained in traditional credit reports, may explore using several leading "alternative credit datasets": Clarity (acquired by Experian), DataX (acquired by Equifax), and FactorTrust (acquired by TransUnion). See Miller and Soo (2020); Blattner and Nelson (2022); Fonseca (2023); Correia, Han and Wang (2024); Di Maggio, Ma and Williams (2024) for examples using such data. Researchers using such alternative data should be aware that coverage changes substantially over time, and further back in time these data are more similar to a dataset recording credit inquiries than recording borrowing.

3.4 Financial Distress

Capturing heterogeneity across consumers is increasingly recognized as critical to understanding economic behavior. There is wide variation in the ability of households to insure against adverse events, with financial distress potentially being an observable result for those that do not. Financial distress is both an important source of heterogeneity (e.g., Gross, Notowidigdo and Wang, 2020; Pfäuti, Seyrich and Zinman, 2024) and an important economic outcome delivering welfare losses (e.g., Olafsson, 2016). Financial distress has large, persistent disparities across geographies (e.g., Keys, Mahoney and Yang, 2023) and can be persistent over the life cycle (e.g., Athreya, Mustre-del Río and Sánchez, 2019).

A broad set of measures of financial distress can be constructed from credit reporting data. In this section we summarize some of these measures, and in the last subsection also discuss measures of financial fragility. Often researchers will want to study a handful of measures to capture different stages of financial distress. For example, Finkelstein et al. (2012) study financial distress by measuring bankruptcy, debts in collection, and delinquency. Or, researchers may also construct their own composite measures of financial distress as in Miller, Wherry and Foster (2023).

3.4.1 Bankruptcy

One symptom of financial distress is bankruptcy. A large literature studies the economic decision of consumers to file for bankruptcy (e.g., Fay, Hurst and White, 2002; White, 2007). Consumer bankruptcy is typically filed under either Chapter 7 or Chapter 13 of the bankruptcy code. The public records file of credit reporting data (Table 1) show when and whether a bankruptcy is filed, dismissed, or discharged (e.g., Keys, Mahoney and Yang, 2023). The timing of bankruptcy can also be precisely observed using a variable in consumer-level aggregated datasets that records the number of months since bankruptcy. When measuring bankruptcy, researchers need to be aware of changes to bankruptcy laws over time, for example the 2005 Bankruptcy Abuse Prevention and Consumer Protection Act (BAPCPA), which made filing for Chapter 7 personal bankruptcy more difficult and led to a sharp increase in filing just prior to the reform followed by a large decline (Gross et al., 2021).

3.4.2 Debt in Collections

A direct measure of financial distress is debts in collections. These are contained in the "Collections File" and also summarized in consumer-level aggregated datasets (Table 1).[9] Approximately half of debt in collections reported by third-parties are medical debts (Keys, Mahoney and Yang, 2023), and medical debt is often a special focus in research (e.g., Batty, Gibbs and Ippolito, 2022; Kluender et al., 2021, 2024) and policy. As with bankruptcy, debt in collections can be measured using

[9] Collection accounts may appear in the "Tradelines File" instead of a separate "Collections File" depending on the practices of the CRA and the furnisher.

consumer-level aggregated data, although richer analysis more precisely isolating the timing, value, and type of collections is possible using tradeline data as done in Keys, Mahoney and Yang (2023).

We recommend researchers measure the flow of the number or value of new accounts in collections because debt in collections are infrequently updated and therefore the stock may be out-of-date. We recommend researchers also report the stock of collections debt, while noting there are meaningful differences in the persistence of different types of collections (Consumer Financial Protection Bureau, 2014). Researchers should note that the reporting of medical debt in collections has substantially declined since 2017, due to a series of nationwide changes in reporting practices detailed Table 2. As of 2024, the CFPB has proposed banning reporting of such debt. See the Online Appendix for details.

3.4.3 Delinquency

A broadly-used measure of financial distress is accounts in delinquency. Measures of delinquency can be calculated in a variety of ways. Researchers may study whether a consumer has any delinquent accounts, the number of delinquent accounts, the value of delinquent balances, or delinquent accounts or balances as a share of the consumer's outstanding accounts or balances respectively, as well as flows of new delinquencies. Delinquency before 30 days past due (i.e., two consecutive missed monthly payments) is not observed in credit bureau data. Researchers can use different definitions for stages of delinquency depending on how many days past due the debt is (e.g., 30+, 60+, 90+, 120+, 150+, 180+) and foreclosures (Piskorski and Seru, 2021).

We recommend researchers generally use one of two delinquency measures: (1) number of trades measured as 30+ days past due, (2) number of trades measured as 90+ days past due. Depending on their goal, a researcher may want to examine delinquencies as a stock, or study the flow of new transitions into delinquencies. The 30+ measure is useful as it captures any form of financial distress including early-stage financial distress that may not lead to charge-offs. The 90+ measure is useful as it closely relates to the binary outcome that main credit scoring models are trained on—whether a consumer has any trades 90+ days past due over 24 months—and will capture more severe financial distress. Depending on a researcher's focus, these may be calculated by aggregating all credit accounts held by a consumer, or aggregating particular types of credit (e.g., credit cards). See Online Appendix for more details.

Between March 2020 and August 2023, delinquency was under-reported because many accounts received COVID-19 pandemic-related accommodations that allowed late payments to be reported as non-delinquent in credit reporting data. Researchers therefore may define a broader measure of delinquency that includes accommodations for auto loans, mortgages, and revolving accounts (Cherry et al., 2021).

Consumer-level aggregated variables record delinquency measures, though how CRAs generate these aggregates from tradeline-level data is not always clear. For researchers using tradeline-level data, each tradeline includes a field reporting the last 84 months of delinquency statuses in one array, often enabling a historical time series of delinquency to be reconstructed (see Gross, Notowidigdo and Wang, 2020; Gross et al., 2021)—even if the lender does not furnish data every month. This adjustment can make a difference when using data from the early 2000s, when some tradelines are only furnished with new information once per quarter (or less frequently), but less so if using data since 2010s (as the overwhelming majority of tradelines are furnished with new information each month). However, these 84-month arrays can become less reliable after accounts enter severe delinquency, if updates are no longer furnished. See the Online Appendix for more details on measuring delinquency.

3.4.4 Financial Fragility

Researchers may also be interested in measures of financial fragility which predict financial distress. This may be of interest given broader work on the cognitive constraints of consumers with scarce resources (e.g., Mullainathan and Shafir, 2013) and the adverse macroeconomic impacts of high household leverage (e.g., Mian and Sufi, 2011). Researchers can construct measures of financial fragility such as debt-to-income or payment-to-income ratios. We recommend using public income data (e.g., IRS Statistics of Income data by zip code) or linking in individual-level income measures. CRAs also construct estimates of income from credit data and other sources—see Blattner and Nelson (2022) for a comparison to income data from mortgage applications, and Mello (2023) for comparisons to IRS and payroll data. Equifax observes employment information for a selected subset of consumers, having acquired a payroll data provider ("The Work Number"). See Albanesi, DeGiorgi and Nosal (2022); Kalda (2020); Gopalan et al. (2021) for examples using these data. Other CRAs also observe some employment information in their alternative credit datasets. Linking credit and payroll data enables researchers to better understand life-cycle consumption (e.g., Garber et al., 2024).

Other datasets can also enhance research on financial distress or fragility. The CRAs' alternative credit datasets include data on how consumers are managing other financial obligations (e.g., utility, telecommunications, and rent payment histories, as in Cochran, Stegman and Foos, 2021), for a selected subset of consumers. Linking household financial transactions data (Baker and Kueng, 2022) such as checking and savings account data with credit reports can reveal overdraft and non-sufficient funds (NSF) use and liquid cash balances (e.g. Alexandrov, Brown and Jain, 2023; Guttman-Kenney et al., 2023) that more directly relate to how liquidity constraints and hand-to-mouth consumers appear in heterogeneous agent consumption models.

3.5 Credit Access

The ability of consumers to access credit can play an important role in consumption smoothing. See Kovrijnykh, Livshits and Zetlin-Jones (2023) for theory on how consumers build their credit access. Credit access can increase welfare by enabling consumers to purchase houses, vehicles, other goods, and to fund human capital accumulation. Credit access can have negative impacts if it leads behavioral consumers to overconsume (Beshears et al., 2018). On the other hand, some consumers may not take out credit even though they can access and would benefit from doing so, with one explanation being debt aversion (e.g., Gopalan et al., 2023; Martínez-Marquina and Shi, 2024).

How to measure credit access? A credit score is often used as a summary statistic for credit access. This is useful but paints an incomplete picture because credit access depends on more than just credit scores (see, e.g. Agarwal et al. 2018; Dobbie et al. 2020; Laufer and Paciorek 2022, who study the relationships between credit scores and other measures of credit access). In this section, we cover several measures of credit access: new accounts, credit limits, credit inquiries, and the costs of borrowing.

3.5.1 New Accounts

Credit access can be measured along both the extensive and intensive margins: the number of new accounts a consumer has opened and how much new credit is granted (e.g., origination amount, credit limit).

If only consumer-level aggregated data are available, a researcher may, for example, use an increase in auto loan balances as a proxy for a new auto loan being taken out. This approach is only applicable for installment loans, such as auto loans, mortgages, and unsecured personal loans. See Agarwal et al. (2023*b*) for an example of such an approach, explained in more detail in our Online Appendix. Consumer-level aggregated data may also contain CRA-created variables for the number of new accounts opened within a given window of time, which is sufficient for many users.

Using tradeline data ensures the timing and amount of new account openings are more precisely measured, and can be useful for event study designs that rely on the exact timing of shocks affecting the consumer (e.g., Bhutta and Keys, 2016; Gross, Notowidigdo and Wang, 2020). This is because there is a lag between when a loan is originated and when a loan first appears on a credit report. For installment loans, we recommend using the origination amount, rather than the outstanding balance in the month when the loan is first observed, and the origination date, rather than the date on which the loan is first observed. For lines of credit, we recommend researchers also use the origination amount, but use the first non-zero credit limit value on the account as the best estimate of the credit limit at origination (e.g., Gross, Notowidigdo and Wang, 2020; Laufer and Paciorek, 2022). Such measures can be computed by researchers who have lower-than-monthly frequency of tradeline data (e.g., annual or quarterly), because one archive of tradeline data includes historical origination details for a consumer's opened and closed accounts going back several years from the archive date.

3.5.2 Credit Limits

Consumers can also access credit through their existing accounts. Credit cards and home equity lines are the most common credit lines a consumer can potentially access to flexibly draw from.

These credit limits can increase and decrease over time. See Gross, Notowidigdo and Wang (2020) and Fulford (2015) for examples studying the changes in a consumer's credit card limits.

The amount of credit limits constructed from consumer-level aggregated variables can differ depending on how cards are classified as active. To address this we recommend, where possible, researchers calculate the total available credit card limits from tradeline-level data, using all open credit card tradelines instead of imposing filters on which cards are active versus inactive.

In the 1990s and early 2000s, not all lenders reported credit limits, but from 2010 onward, credit limits are required to be reported under an amendment to the FCRA. If cards do not have credit limits, then we suggest either using the variable showing the highest balance recorded on the account or, if limits are later observed on those accounts, backfilling the missing limits.

Researchers may also wish to examine the amount of available credit: credit limits on open accounts less outstanding balances on those accounts. Such measures are often used to measure consumer liquidity (e.g., Gross and Souleses, 2002). Utilization rates, defined as the sum of balances on revolving accounts divided by the sum of credit limit on these accounts, are also used as a measure of credit constraints. Utilization rates above 90 percent are generally regarded as binding, and some consumers may even exceed or overdraw their credit limits (Athreya, Mustre-del Río and Sánchez, 2019). More generally, utilization rates are a key input to credit scores, with high utilization strongly predicting default and providing an early indicator of financial stress.

3.5.3 Credit Inquiries

Credit inquiries data have been used to provide a measure of credit demand (e.g. Han, Keys and Li, 2018), the difficulty of accessing credit (e.g. Romeo and Sandler, 2021), and rejected applications (e.g. Blattner and Nelson, 2022).

Romeo and Sandler (2021) provide an example of how to use inquiries data. They create a binary measure where an inquiry is successful if a new account is opened within 14 days, and unsuccessful if no new account is opened. Blattner and Nelson (2022) use a window of three quarters for inferring whether a mortgage application translates into a new opening. Researchers may also use a ratio of new account openings to inquiries as a measure of credit supply. The CFPB's credit tightness index is similar to that used by Romeo and Sandler (2021), but uses different search windows for different product types and aggregates to a national or subgroup level while keeping the composition of credit scores constant over time. This tightness index therefore reflects changes attributable to lender policies not changes due to varying credit scores of applicants.

Credit inquiries are often only observed in consumer-level aggregated datasets. However, some researchers may have access to the more granular Inquiries File (Table 1). An important caveat for researchers to be aware of when using inquiries data is that individual CRAs have incomplete coverage of credit inquiries, whereas originated loans are more commonly furnished to all CRAs. For many credit applications, lenders will only conduct inquiries via one or two CRAs. An exception to this is mortgage applications where lenders typically conduct inquiries across all three CRAs.

Only “hard” inquiries that relate to applications for credit are typically observed by researchers, whereas “soft” inquires, that correspond to getting quotations for expected credit terms (and a variety of other functions including marketing and background checks) are not typically observable to researchers (see Ballance, Clifford and Shoag (2020) for an exception). More generally, hard credit inquiries are just one part of a consumer’s search process and therefore researchers may benefit from examining other data sources to more fully understand consumer search.

3.5.4 Costs of Borrowing

Credit reports do not contain variables showing the costs of borrowing. However, researchers are increasingly able to estimate these from tradeline data. Researchers may also purchase consumer-level or tradeline-level variables estimating borrowing costs, but it may be unclear to the user how the CRA estimates these.

For fixed-rate installment loans, such as auto loans and unsecured personal loans, once a researcher observes the principal origination amount (P), origination term (n), and scheduled monthly payment amount (A), they can calculate the interest rate (i) at origination using a root-solver shown in Equation 1 (Yannelis and Zhang, 2023). For cases where $P \geq A \times n$, loans are assumed to have zero-percent interest rates. If a researcher is interested in the realized effective interest rate, to capture costs changing post-origination, researchers can calculate this using multiple observations after origination (Conkling and Gibbs, 2019).

$$A = \frac{P \times i}{1 - (1+i)^{-n}} \quad \text{if} \quad P < A \times n \tag{1}$$

For mortgages, the above calculation does not work because the scheduled payment amount may include taxes, insurance escrow, and other fees such as home owner association fees. Shahidinejad (2024) develops an algorithm using changes in outstanding balances over time to estimate interest rates and verifies its accuracy against market data.

Separately from installment loans, Guttman-Kenney and Shahidinejad (2024) develop a methodology for estimating financing charges on credit cards. Their methodology’s intuition is that credit card minimum payments are a deterministic function of statement balances, following a generic formula structure. With sufficient data, a researcher can estimate each credit card furnisher’s minimum payment formula, and can then recover financing charges.

3.6 Consumption Measures

3.6.1 Auto Purchases

Auto purchases are an important component of consumption and can be used as an indicator for macroeconomic conditions. In credit reporting data we observe autos purchased on finance (“auto loans”)— representing over 80% of new auto purchases (Benmelech, Meisenzahl and Ramcharan, 2017). Newly opened auto loans can be calculated as previously explained in Section 3.5.1. Some subprime auto lenders do not appear in credit reports, and therefore credit report measures do not fully cover auto purchases by this segment (Low, Clarkberg and Gardner, 2021). Benmelech, Meisenzahl and Ramcharan (2017) and Di Maggio et al. (2017) verify the accuracy of this consumption measure, showing that auto loan originations in credit reports match up to external data and track total sales in the time series, including those with and without loan financing.

3.6.2 Credit Card Spending

Credit cards are broadly used by US consumers, with high coverage across geography and credit scores. Approximately 30% of all consumer payments are made via credit cards, and this share is growing over time, whereas the share of cash and checks are declining over time (e.g. Cubides and O'Brien, 2023). The large volume of spending on credit cards therefore makes them well-suited as a measure of consumption. A strength of CRA data is that the data do not under-report credit card balances, unlike relevant survey data (Brown et al., 2015). When calculating credit card spending, we generally recommend combining general-purpose credit cards with private-label retail credit cards, the latter of which can only be used at one or a small group of merchants.

The economic object of interest—"credit card spending" (s_t)—is the total value of new purchases on a credit card at time t. Our preferred measure of credit card spending (s_t) is shown in Equation 2, as used in Ganong and Noel (2020). This measure takes the changes in statement balances and adds payment amounts (p_t). If the measure produces a negative number, it is bounded at zero.[10] Measuring credit card spending relies on the researcher being able to observe the actual payment amount variable at the tradeline-level over time. If using this measure, it is important to study only the cards of furnishers who consistently report the actual payment amounts. For example, Ganong and Noel (2020) exclude furnishers where over 90% of card months have zero or missing payment amounts. From 2014 to at least 2023, credit card actual payment amounts are only observed for a small, selected subset of credit card lenders (Consumer Financial Protection Bureau, 2020; Guttman-Kenney and Shahidinejad, 2024). We recommend that researchers who want to use this measure confirm the reporting coverage of the actual payment amount variable for the time period they are planning to study before using or purchasing data. It is also possible to produce estimates of credit card spending without observing actual payment amounts using other methodologies discussed in the Online Appendix.

$$s_t = \begin{cases} b_t - b_{t-1} + p_t & \text{if } \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

We note that the concepts of credit card debts and balances are related to, but distinct from,

[10] This contains some measurement error as it includes financing charges (the sum of interest and fees). The Online Appendix shows how Guttman-Kenney and Shahidinejad (2024) address this by estimating and deducting financing charges.

credit card spending. While the literature has not defined these terms consistently, we suggest for clarity that researchers refer to outstanding credit card statement balances as "credit card balances," refer to *new* credit card expenditures as described in the previous paragraph as "credit card spending," and reserve the terms "credit card debt" or "revolving debt" to describe the balances that are carried over into the next statement.

"Credit card debt" (d_t) can be measured in credit reports by taking the preceding month's statement balance (b_{t-1}) less actual payments made since (p_t), observed this month, and, if $d_t < 0$, setting it to zero. This approach recognizes that the credit card actual payment amount observed in credit report archive t corresponds to the payment made against the statement balance and scheduled payment observed in archive $t-1$. This methodology typically requires studying the subset of credit card lenders that consistently report actual payment amount (p_t). Bornstein and Indarte (2022) and Lee and Maxted (2024) provide examples using CRAs' estimates of revolving debt. Fulford and Schuh (2023) provide an example of a machine learning approach to estimate which balances likely represent debt vs. new expenditures.

3.6.3 Cashed-Out Home Equity

Researchers can use credit reporting data to estimate the amount of equity a consumer extracts from a home when refinancing, for so-called "cash-out" refinances. Researchers without access to linked mortgage data can use the approach in Bhutta and Keys (2016) to identify equity extractions in credit reports. This approach identifies increases to consumers' outstanding mortgage debt by more than 5% over a one year period among those who didn't move, with a minimum increase of $1,000, while inferring lien status from tradeline data. Mian and Sufi, 2022 use a similar methodology to identify refinances in credit bureau data. Beraja et al. (2019) and Berger et al. (2021) use a similar method with a mortgage origination dataset linked to credit reports (see the discussion of CRISM in Section 5.1.2), and verify this method against external data. The intuition behind their methodology is to first find loans recorded as refinances in mortgage origination data, and compare the difference between the value of a new mortgage originated to the mortgage(s) previously outstanding, in order to isolate the amount of equity cashed-out. See the online appendices of Beraja et al. (2019) and Berger et al. (2021) for the detailed methodology.

3.7 Geographic Mobility

The geographic mobility of consumers is important as it affects local labor markets, housing costs, and the accessibility of amenities (e.g., healthcare, schools) (see Jia et al. 2023 for a review). While there are large potential gains for some consumers moving, financial and non-financial frictions can be a barrier to such gains being realized. It is therefore important to measure geographic mobility and understand what affects it.

Geographic mobility can be measured usefully in credit reports, as these contain information on a consumer's primary address and track a large panel of consumers over a long period of time. Whitaker (2018) and DeWaard, Johnson and Whitaker (2019) validate this data source for measuring geographic mobility against other sources of data, and Bleemer and van der Klaauw (2019) provide an example of using mobility data, studying the long-run effects of Hurricane Katrina on consumer changes of address, county, and state. Other examples of use include Keys, Mahoney and Yang (2023), who use geographic mobility for identification of person versus place-based factors in credit markets, and Howard and Shao (2023), who construct a gravity model of migration. Molloy and Shan (2013) analyzes the post-foreclosure residential destinations of households.

A caveat to using these geographic mobility measures is that they rely on a CRA's view of a consumer's primary address. The CRA may only update a consumer's primary address with a lag, due to delays in information arrival and in determining whether a new address is primary. The timing of address changes in credit reports can also depend on when and whether a consumer chooses to update their address with their financial institutions. Given this caveat, researchers may wish to examine address changes quarterly (e.g. Keys, Mahoney and Yang, 2023) or at annual (or longer) horizons (e.g. Bleemer and van der Klaauw, 2019).

Moreover, an apparent residential move in credit reporting data may be the CRA reassigning the consumer's primary address; especially for some demographic groups, this may not indicate an actual move. Students often have multiple concurrent addresses (e.g., their parents' address and a college address), and consumers with multiple homes can make it difficult to establish which is their primary residence. For individuals who have multiple first-lien mortgages, without matching external data, it is not observed which mortgage is associated with the current mailing address or the addresses of other properties owned.

CRAs' algorithms for identifying primary addresses have considerably improved since the early 2000s, with fewer cases of moves between locations A and B appearing as multiple moves back and forth; see Mian and Sufi (2022) for related analysis when identifying housing speculators and Varley (2024) for how to account for spurious moves.

Measuring geographic mobility is difficult in general, so—despite the caveats above—credit reporting data likely offer one of the most promising opportunities for researchers to study the causes and consequences of geographic mobility. This is especially true given these data's large sample and long panel dimension.

3.8 Intra-household and Intergenerational Behaviors

A small but growing set of papers use credit report data to explore intra-household and intergenerational behavior. Credit report data can identify individuals either sharing an address or sharing responsibility for an account, as described further in Section 5. Such linkages are often otherwise difficult to identify outside of survey data and tax data. Using such linkages, researchers have studied household formation (Dokko, Li and Hayes, 2015), co-habitation between adult children and parents (Bleemer et al., 2017; Dettling and Hsu, 2018), intergenerational wealth transmission (Benetton, Kudlyak and Mondragon, 2022), and correlation between parents' and children's credit scores (Ghent and Kudlyak, 2016; Bach et al., 2023). Other studies use credit report data while identifying intra-household linkages in other merged data (e.g., Braxton et al., 2024). Given the empirical challenges with studying intergenerational and especially intra-household behavior using other data sources, we see great potential for future work in this area.

4 Credit Scores

One of the primary uses of credit reporting data in the marketplace is for the generation of credit scores. Lenders use credit scores as measures of a consumer's credit risk to enable them to decide whether to accept a credit application, and if so, the contractual terms to offer.

Credit scores are often available to researchers as part of credit reporting data, and are sometimes interpreted as a sufficient summary statistic for "financial well-being." Given this, it is important for researchers to understand the basic features of credit scores to effectively use and interpret them. In this section, we explain what credit scores are (in subsection 4.1) and what goes into their construction (in subsection 4.2). We then provide practical guidance to researchers on how to choose which scores to use and how to interpret them (in subsection 4.3).

4.1 What Are Credit Scores?

Fundamentally, credit scores are designed to evaluate a consumer's creditworthiness by predicting the consumer's future default risk based on the credit history observed in credit reports. One consumer does not possess a single credit score. Credit scores have many versions (e.g. FICO 8.0, FICO 9.0, etc.) that arise from refining the scoring algorithm's predictive model over time, and there are a range of credit scores catering to heterogeneous client needs by building models on specific populations and targeting specific outcomes. Sophisticated lenders typically create their own proprietary in-house credit scoring models.

This section describes the basic features common to the two major, widely available credit scoring models used for credit risk in the United States: FICO and VantageScore. [11] FICO and VantageScore arise from logit models of 24-month forward-looking default risk. The definition of "default" is typically four consecutive payments below the minimum contractual payment, also termed the "90-day" default rate or "90 days past due" (Federal Reserve Board, 2007). Scoring models use the information available at time t on an individual's credit report with one of the three major CRAs to predict default between dates $t + 1$ and $t + 24$ months. These scores are affine transformations of the log odds of default based on the logit models, mapped to an integer scale typically ranging from 300 to 850 (Thomas, 2009), although some versions have slightly different ranges. Because scores are linear in log odds, a given absolute change in credit scores has different

[11] While the generic term "credit score" often refers to FICO or VantageScore in the United States, the CRAs also have their own credit risk scores, such as the Equifax Risk Score observed in the Federal Reserve Bank of New York's credit panel. There are many other types of scores, including created by lenders, for a variety of purposes; see the Online Appendix for details.

implications for default risk at different ranges. For example, a 100-point score decrease from 800 to 700 corresponds to a much smaller change in predicted default rate than a decrease from 600 to 500.[12] Because the ranking of consumers stays relatively stable over the business cycle, credit scores can be thought of as an ordinal ranking of credit risk across consumers.

While the exact formulas used in commercial credit scores are proprietary, and researchers are generally prohibited contractually from attempting to reverse-engineer these exact formulas, the basic ingredients of these logit models are well-known and publicly disclosed by CRAs and score providers.[13] By following the guidelines described below, researchers with access to credit reporting data can build their own credit models that are highly correlated with commercially available models without knowing their exact formulas.

4.2 What Goes Into Credit Scores?

The logit models underlying credit scores typically take attributes derived from credit reports as inputs, and various measures of default as the outcome being predicted. The major types of attributes that are included as inputs into credit scoring models include payment history (e.g., 90+ day delinquency on various types of tradelines, collections trades, public records such as bankruptcies), amount owed and utilization (e.g., total debt, balances as a fraction of available credit lines), length of credit history (e.g., age of oldest account), credit mix (the variety of different trade types on a consumer's record), and new credit (e.g., the number of credit inquiries within the last year, number of new accounts).[14]

Regulations do not mandate the existence of credit scores or specify their exact nature or formulas, and their evolution has largely been driven by market forces (see Online Appendix for a history). Lenders update and modify scoring models subject to the ECOA, and other relevant laws, and protect the exact formulas and training datasets behind their proprietary models as trade secrets. However, once FICO and other traditional credit scores became the industry standards, they have been integrated into regulations such as those governing mortgage lending backed by government-sponsored enterprises (GSEs).

An under-explored aspect of credit scores is the potential ability for consumers to game their credit reporting data and the inputs to credit scores. Ball (2024)'s theoretical work cites Mark Zandi of Moody's as saying, "The [credit] scoring models may not be telling us the same thing that they have historically, because people are so focused on their scores and working hard to get them up."[15] It is an interesting challenge to ensure credit scoring models are suitably transparent but also robust to the incentives they create for consumers.

[12] See the VantageScore RiskRatio tool for an illustration https://www.vantagescore.com/lenders/risk-ratio/

[13] https://www.myfico.com/credit-education/whats-in-your-credit-score and https://www.vantagescore.com/consumers/how-credit-scores-work/

[14] See, for example https://www.myfico.com/credit-education/whats-in-your-credit-score.

[15] In addition to Ball (2024), Frankel and Kartik (2019) also help formalize the idea of agents "working hard to get [their scores] up."

An economic feature of credit scores is that the reliance on payment history does not distinguish between idiosyncratic and systematic drivers of default. That is, consumers who enter delinquency during recessions or due to mass layoffs, health shocks, or other arguably exogenous factors are treated the same way as those who become delinquent due to moral hazard or personal events such as divorce or entrepreneurship. Thus, credit scores reduce the insurance value of credit with respect to many types of shocks consumers face (Avery et al., 1996). Potentially important avenues of research include studying the economic causes of defaults and also developing credit scoring systems that can better distinguish bad luck from bad types. In early work in this direction, Chatterjee et al. (2023) and Blattner, Hartwig and Nelson (2022) empirically analyze the hidden type processes that may underlie US credit scores and histories. There is growing interest in research on the distributional consequences of different credit scoring approaches across socio-economic groups, such as race (e.g., Fuster et al., 2022).

Many factors researchers may think would affect default risk, such as income and liquid assets, are not included in standard credit scoring models by practice or due to technical limitations. Regulation B which implements ECOA governs that information related to sex, race, and other protected classes is not allowed to be included in credit scoring models in an effort to prohibit lending discrimination on these bases. Although it may seem intuitive to researchers to include income as a predictor of credit risk, verified information on income has not historically been collected by CRAs as a part of their standard data furnishing formats, and hence is not part of the set of attributes available to credit scoring models (see Beer, Ionescu and Li, 2018 for an example studying the relationship between income and credit scores).

New alternative data sources are increasingly being used by lenders instead of, or in combination with, traditional credit reports, for credit scoring (and underwriting) and therefore it will be important for researchers to study their effects for the functioning of credit markets and their real effects on other markets. See the Online Appendix for some early literature studying how alternative data sources affect credit scoring.

4.3 Practical Guidance on Using Credit Scores

Credit scores can be observed by researchers in the CRAs' consumer-level aggregated datasets (Table 1). Researchers may have a choice between multiple scoring models (e.g., FICO versus VantageScore) or versions within the same model. As most credit scores are highly correlated with each other, researchers may be able to use the cheapest score available from a CRA.

Older versions of credit scoring models may also be a better choice for researchers since they were potentially the scores available to lenders in the historical time period studied. Additionally, the CRA has to pay a licence fee to FICO to sell FICO scores to researchers. Researchers may be able to use a cheaper credit score if FICO is not necessary for a specific project, but researchers may want to use a particular proprietary credit score because of their identification strategy. For example, some lenders have sharp cutoffs in their underwriting which can be used for regression discontinuity designs if the researcher observes the same score, calculated at the same time as used by the lender (e.g., Agarwal et al., 2018; Argyle, Nadauld and Palmer, 2023). If the researcher uses a different scoring model, those cutoffs will not align. Many lenders use their own proprietary scoring models which use information from the CRAs but are not available to the CRAs or researchers.

Market coverage for the different scoring models are typically not readily available. As of 2010, reportedly more than 90% of lenders used a version of FICO as part of their underwriting decisions (Consumer Financial Protection Bureau, 2011). VantageScore has increasingly been used for other decisions, and coverage continues to change over time. For example, new mortgage regulations from FHFA require lenders to use both FICO and VantageScore for the first time for GSE mortgage securitization.[16]

While credit scores have been used as proxies for financial sophistication (e.g., Agarwal, Rosen and Yao, 2016; Amromin et al., 2018; Bhutta, Fuster and Hizmo, 2021; Agarwal et al., 2023*a*) based on the rationale that credit scores are correlated with these, researchers should consider the potential sources of bias when doing so. The outcome credit scores target—default—depends on much more than financial sophistication. A sophisticated consumer may have a low credit score due to default caused by negative life events (e.g., Ganong and Noel, 2023; Low, 2023). Credit scores may conflate sophistication with the opportunities consumers have historically faced given how maps of credit scores (e.g., Keys, Mahoney and Yang, 2023) correlate with maps of historical racial inequities. High credit-score consumers assumed to be financially sophisticated may not be sophisticated in other ways such as in their choice of credit product, refinancing, or retirement saving decisions. Despite such limitations, good measures of financial sophistication are challenging to find, therefore if researchers use credit score as a proxy, we recommend checking the robustness of their results to other proxies for financial sophistication (e.g., Agarwal et al., 2009; Varley, 2024).

5 Accessing Credit Reporting Data

How can researchers access credit reporting data? Section 5.1 describes the existing datasets available to researchers and their access terms, including both established credit panels and linked mortgage datasets. Section 5.2 covers how researchers can design new credit reporting datasets including new panels, linking data, and constructing surveys from these data.

5.1 Existing Datasets

5.1.1 Established Consumer Credit Panels

Public data aggregated to the national and geographic level derived from established consumer credit panels is available from the Consumer Financial Protection Bureau's Consumer Credit Trends, the spreadsheet accompanying the Federal Reserve Bank of New York's Quarterly Re-

[16] https://www.fhfa.gov/policy/credit-scores

port on Household Debt and Credit, and the Urban Institute's Data Catalog, while the University of California's Consumer Credit Dashboard provides data for California.

A variety of established, nationally representative anonymized credit panels exist, and these are listed in Table 4. Typically direct access to these panels is restricted to employees of the organization, but this is sometimes broadly defined. For example, the Federal Reserve Bank of New York's Consumer Credit Panel is available to researchers across the Federal Reserve System (Lee and Van der Klaauw, 2010), and the University of California Consumer Credit Panel is available to researchers across the University of California system. An increasing number of universities have recently created their own credit panels, some of which (e.g., California and Ohio) include data for the full population of their state to complement their national sample.

Panels vary in whether they include only a primary sample of consumers or also include consumers who have an association with the primary sample. For example, both the NYFed CCP and the University of California Consumer Credit Panels (UC-CCP) include consumers at the same address as the primary sample, while the UC-CCP and Consumer Financial Protection Bureau's (CFPB) Consumer Credit Information Panel (CFPB-CCIP) include credit records of associated borrowers, defined as borrowers who share a credit account, irrespective of their address.

Researchers outside institutions with credit panels can co-author on research projects that use these panels but generally are not be able to access the underlying data (unless they have an employee status, such as with an internship). Under the terms of access, the CRAs will typically review outputs before publication primarily to ensure that output is sufficiently aggregated.

5.1.2 Established Linked Mortgage Data

A productive approach in research has been linking credit reports with product-level data on mortgages. This is a valuable merge as some mortgage originations data do not show mortgage repayment after origination or the other debts held by a consumer over time, though this information is observed in credit reports. Moreover, the linked data enable researchers to observe detailed mortgage product features (e.g., government-backed, securitized, property type and estimated property value) as well as a richer array of borrower demographic information.

A variety of existing linked datasets are available to researchers to purchase. The most prominent example in the literature is the Equifax/ Black Knight Financial Services Credit Risk Insight Servicing McDash (CRISM) database, an anonymous loan-level match between mortgage servicing data and Equifax credit reporting data used in Berger et al. (2021); Beraja et al. (2019); Agarwal et al. (2023*b*). Also available is Moody's Analytics data (previously known as Blackbox Logic) that links mortgage originations data with Equifax credit reports (see Piskorski, Seru and Witkin, 2015; Di Maggio et al., 2017; Gupta, 2019; Varley, 2024). Another example is a match between the credit reports and loan-level data from CoreLogic (e.g., Haughwout et al., 2011; Bhutta, Dokko and Shan, 2017). See the Online Appendix for details on links with other mortgage datasets, including the Home Mortgage Disclosure Act (HMDA) data.

5.2 Creating New Datasets

5.2.1 Creating New Consumer Credit Panels

Researchers can construct credit record panels in a variety of forms, including samples based on individuals or loans drawn from a CRA's database. Often researchers want a panel that remains representative over time, which requires dynamically updating the data to include records newly created since the start of the panel. Two of the most common ways to draw and maintain a nationally representative sample are to select the sample based on the last few digits of the SSNs on the credit records or the internal ID assigned by the CRA. These result in similar but not identical panels (because credit records without SSNs will still have internal IDs). Both approaches can be readily applied to the nearly full population of adults with a credit record or to a subset of consumer records (e.g., by age, geography, or presence of specific tradeline types, as is the case with the National Mortgage Database described in Table 4).

Off-the-shelf and customized credit panels typically include anonymized IDs for consumers (and possibly furnishers) in order to protect consumers' privacy and comply with CRA requirements. If researchers need the ability to identify specific subsets of furnishers, they may be able to work with the CRA to construct flags for these furnishers (as in Di Maggio and Yao, 2021; Granja and Nagel, 2024), but each CRA has different requirements for the types of flags they will provide and the minimum number of furnishers covered by such flags.

The main cost driver experienced by researchers creating new credit panels is usually deciding how many points-in-time to purchase data for. The cost of each point-in-time varies depending on exactly what CRA products are purchased, the sample size, and the number of potential users with access at an institution. Bulk discounts are typically available for researchers purchasing multiple points-in-time, and researchers can purchase additional points-in-time at a later date to increase the frequency or duration of their dataset. The sample size chosen by a researcher can also add costs. Some CRAs have off-the-shelf products that researchers have purchased covering a nationally-representative panel of monthly, tradeline-level data over twenty years (e.g., Equifax Analytics Dataset).

5.2.2 Linking Credit Reporting Data to Other Data Sources

Increasingly researchers have been merging other types of data to enhance existing credit panels, or creating ad hoc panels using merges with other data sources. Linking to other data sources allows researchers to enhance credit record information and analyze populations that cannot be readily identified in credit data alone, but the process to merge these data sources is often complicated by important steps to protect consumers' privacy and comply with various regulations.

Researchers can only access anonymized credit reporting data and the CRA cannot release personal information. This means the agency rather than the researcher typically matches data. Most matches use consumer names, birth dates, addresses, and/or social security numbers. Match rates are particularly high when using social security number (see, for example, Collinson et al., 2024, Dobkin et al., 2018, and Miller et al., 2021). Researchers beginning with a dataset that includes personal information may be able to send it to a CRA to link it to the credit data (as in Finkelstein et al. (2012) to medical records, and Miller and Soo (2021) to HUD Moving to Opportunity (MTO) records). When the non-credit record data contains sensitive data, such as medical information as in Miller, Wherry and Foster (2023), researchers may need to send additional records from another source to help mask from the CRA which records are in the source data.

Another approach to such a merge involves a three-party data agreement where the matching variables are "hashed", mapping via a one-way function to a fixed-length value, by the CRA and the third-party. This process is often further secured with a "salt," an added value that further changes the resulting hashed output to prevent a repeated input value from showing up as identical output values. For example, the CRA and the third-party data source agree on a hashing algorithm and then separately send their data with the hashed matching information to the researcher (e.g., Chakrabarti et al., 2023), or instead the CRA may provide a cross-walk between anonymous identifiers in both datasets (e.g., Gresenz et al., 2024). In this arrangement the CRA and third party will not need to share their data with one another, and the researcher does not see any identifiers to help maintain confidentiality. The researcher does not have access to the hashing algorithm and typically destroys the hashed variables after the match. Nicholas et al. (2021) develop a methodology to match Medicare data to credit reports without exchanging personal information by working out unique consumers in both datasets. For a more general toolkit for matching data with a hash, see Davis et al. (2022).

In recent years the number and range of different linkages with credit reporting data has grown rapidly. In addition to previously mentioned studies that linked to health records, payroll data, marketing offers, and HUD MTO data, credit reporting data have been linked to payday loan data (Bhutta, 2014), tax return data from a sample of tax filers (Meier and Sprenger, 2010), bankruptcy filing records (Argys et al., 2020; Dobbie, Goldsmith-Pinkham and Yang, 2017), and education records from specific universities and the National Student Clearinghouse (Scott-Clayton and Zafar, 2019; Chakrabarti et al., 2023).

What are the costs of linking data? Different CRAs have different appetites and costs to merge data, and therefore we recommend researchers obtain multiple quotes. The cost of the merge itself is largely a fixed cost irrespective of the number of records to be merged, the more substantial variable cost is the number of points-in-time a researcher requires. As of 2024, the University of California has a fee of $12,981 to merge external data to its credit panel.

5.2.3 Constructing Surveys from Credit Reporting Data

Using credit records to draw survey samples is a relatively new approach to augment credit record data, and it can be done with a new sample or an existing credit panel. Researchers can ask for a sample among consumers in a particular region or among consumers with a specific loan type.

For example, the CFPB and Federal Housing Finance Agency (FHFA) began the National Survey of Mortgage Originations in 2014 based on a 1-in-20 sample of new mortgage originations from a CRA and added another sample of existing mortgages in 2016 (Avery et al., 2017; Durbin et al., 2021). Separately, the CFPB has surveyed borrowers on their experiences with debt collection (Consumer Financial Protection Bureau, 2017), making ends meet (Fulford and Shupe, 2021), and student loan experiences by drawing survey samples from existing credit panels.

As detailed in Consumer Financial Protection Bureau (2017), this approach offers several advantages to credit data alone and to some other survey sampling strategies. First, researchers can more readily target and oversample specific populations of interest to increase sample sizes. Additionally, researchers have the full credit record for the initial sample to adjust survey weights and nonresponse bias. The credit data may also help clarify incomplete, conflicting, or uncertain responses.

When conducting a survey directly from a sample of credit records, researchers will typically need to work jointly with the CRA and potentially with a third party to field the survey to protect consumers' confidentiality. As with creating a general panel of consumer credit records, the specific constraints involved in or willingness to conduct a survey may vary by CRA. Researchers may instead match existing survey data to credit records. For example, Miller, Wherry and Foster (2023) link prior survey data from another study to credit records and help reduce privacy concerns by including additional people in the matched sample to prevent the CRA from knowing which records were part of the prior survey.

6 Conclusions

This paper provides a general overview of the economics and use of consumer credit reporting data to increase awareness of these data's research potential. We show examples of how these data can be used to answer questions across economic fields and provide advice for how to do so. We encourage *users* of these data to read the more detailed information in the Online Appendix.

We end this paper by emphasizing some especially exciting open avenues for researchers to explore. One area of great promise is linking credit reports with other datasets. Research linking data on consumers' assets, liquidity, income, expenditures, or utilities can be especially valuable for filling in important aspects of consumer cashflows and balance sheets that are missing from credit reporting data. Linking sources such as voting records or social networks can enable researchers to study links between financial and other behaviors. Few studies currently link surveys with credit reports, but doing so has great potential, for example to study the role of expectations in households' economic behavior.

There has also been exciting recent innovation in credit reporting for small and medium enterprises (SMEs). While distinct from consumer credit reporting, SME credit reporting is related in that entrepreneurs may finance SMEs through a combination of personal and business credit, and accordingly, consumer CRAs are developing datasets to track SME credit in a format similar to, and linkable with, consumer credit reporting data (see e.g., Bellon et al., 2021; Haughwout et al., 2021; Benetton, Buchak and Garcia, 2022; Fonseca and Wang, 2024). These data offer promising avenues for studying consumer and firm behaviors.

While our paper focuses on US credit reports, there is exciting untapped potential to research credit reports from other countries. Data from other countries contain variables not observed in US reports, as well as sources of variation arising from different legal structures. Studying credit reporting across international domains can help to understand fundamental issues such as the role of the financial system in enabling access to efficiently priced credit. The issues surrounding the use of "big tech" or social media data for consumer credit decisions are especially interesting to study, and with regulatory environments being internationally heterogeneous, these are issues where data from other countries can be especially fruitful.

Finally, there is a wealth of fascinating topics to explore using credit reporting data without needing to link these data to other sources. Recent methodological developments have unlocked new opportunities for studying prices and related consumer and firm behaviors within credit reporting data. Meanwhile, the longer time series of credit reporting panels that now exist enable researchers to study life cycle topics of consumer behavior.

Table 1: Summary of Contents of Consumer Credit Reporting Data

Dataset	**Contents**
Header File	The header file includes identifying information, including the individual's social security number, date of birth, name(s), phone number(s), current address (including state, county, and zip code) and previous addresses. Although personal information in the data accessible to researchers is redacted—researchers may sometimes observe full dates of birth.
Tradeline File	Tradeline account-level information are included for each revolving and installment credit account that belongs to an individual. Revolving tradelines include credit cards and home equity lines of credit, while installment tradelines include mortgages, auto loans, student loans, and personal loans. Each tradeline includes specific information about the account including the outstanding account balance, type of debt, type of account (e.g., revolving, installment), and account ECOA designator (e.g., whether the individual's legal responsibility over the account is as an authorized user, joint account, individual account, or co-signed account). In addition, it includes the payment status, date or month the account was opened, origination loan amount or credit limit, date of last activity, monthly payment, and some information about the recent payment and payment history.
Public Records File	Public record information is obtained from county, state, and federal courts, and includes bankruptcies and foreclosures and prior to NCAP included civil judgments and state and federal tax liens.
Inquiries File	Logs the views or "pulls" of the consumer's credit file over the past two years. Such reviews may be initiated by current and prospective lenders and landlords.
Collections File	This file (also known as collection tradelines) represent unpaid bills or other unpaid accounts, typically unsecured such as credit cards and personal loans, sold to or managed (for a fee) by a collection agency. These debt collection companies sometimes furnish such collection accounts to CRAs.
Consumer-Level Aggregated Datasets	The CRAs use information across the files listed above to create their own summary consumer-level aggregated dataset—also known as "attributes" or "roll-ups"—and this includes demographic information such as consumers' estimated primary residence. While the formats of the other files are fairly standardized across bureaus, each agency has their own version of consumer-level aggregated datasets with differing modules.

Table 2: US Credit Reporting Changes

Law or other change	**Year**	**Primary effects on credit reports and scores**
Fair Credit Reporting Act (FCRA)	1970	Deletion of most adverse information after seven years and after 14 years for bankruptcies; requirements for CRAs related to data accuracy, sharing information with consumers, and legitimate business purpose for accessing reports
Equal Credit Opportunity Act (ECOA)	1974	Prohibition of discrimination due to gender or marital status; reporting of joint and cosigned accounts begins to implement requirement that spouses benefit from credit history of joint and cosigned accounts
ECOA Amended	1976	Added prohibition of discrimination due to race, nationality, religion, age, or use of public assistance; Regulation B implementation prohibits inclusion of bases in crediting decisions with few exceptions
Consumer Credit Reporting Reform Act	1996	First requirements on furnishers to provide accurate data; requirements to report voluntary closures and date of first delinquency for collections and charge offs; procedures for CRAs to prevent reinsertion of deleted information from disputes; 30-day requirement for responses following disputes
Metro 2® Format	1997	Update to standardized reporting format
Fair and Accurate Credit Transactions Act (FACTA)	2003	Requirements related to identity theft (including credit file freeze flags and deletion of information); consumer right to obtain credit scores; changed standard of accuracy for furnishers
National Consumer Assistance Program (NCAP)	2015	Deletion of: many public records due to new minimum requirements for matching to credit records and standards for updates; all authorized user accounts reported without a birth year and month; all third-party collections without an original creditor classification code; and all third-party collections not updated by furnisher within prior six months; 180-day delay in reporting medical collections
Coronavirus Aid, Relief, and Economic Security (CARES) Act	2020	Accounts with an accommodation required to be reported as current (or the delinquency status prior to the accommodation) while accommodation in effect; availability of forbearance for government-backed mortgages regardless of delinquency status
Voluntary changes in medical debt reporting	2022	Removal of paid medical collections from reports; one-year delay in reporting medical collections; removal of all medical collections under $500

Notes: This table is accurate at the time of writing. Laws change over time so researchers should check the latest versions for current practices. This table only summarizes how laws and settlements changed what appears on credit reports or in credit scores. The effective date for some changes may have been up to two years after the year reported here.

Table 3: How Long Does Information Typically Remain On Consumer Credit Reports?

Credit File Information	**Typical Reporting Duration**
Hard Credit Inquiry	Up to 2 years from inquiry date
Open Credit Agreement	Indefinitely
Closed, Non-Delinquent Credit Agreement	Up to 10 years from agreement's last activity
Delinquent Credit Agreement	Up to 7 years from payment first 30 days past due
Debt in Collections (Medical and Non-Medical)	Up to 7 years
Bankruptcy - Chapter 13 - Chapters 7, 11, 12	Up to 7 years 10 years

Notes: This table is accurate at the time of writing. Laws and practices change over time so researchers should check the latest versions for current practices. Large credit transactions (currently defined as those with principal amounts of $150,000 or greater) are exempt from the FCRA's time limits, though CRAs typically still delete derogatory information for these exempt accounts following the maximum durations that the FCRA applies to smaller credit transactions.

Table 4: US Consumer Credit Reporting Panels

Credit File Panel	**Starting Year**	**Frequency**
Federal Reserve Bank of New York Consumer Credit Panel / Equifax	1999	Quarterly
University of Chicago Booth School of Business / TransUnion	2000	Monthly
Consumer Financial Protection Bureau Consumer Credit Information Panel	2002	Monthly
University of California Consumer Credit Panel	2004	Quarterly
University of Illinois at Urbana-Champaign Gies Consumer and Small Business Credit Panel / Experian	2004	Annual
Rice University Jones GSB / Experian	2004	Annual
Credit Risk Insight Servicing McDash (CRISM) / Black Knight & Equifax	2005	Monthly
National Mortgage Database	2005	Quarterly
Urban Institute	2010	Annual
Georgia Institute of Technology Scheller College of Business / Equifax	2005	Monthly
Ohio State University / Experian	2017	Quarterly

Notes: In nearly all cases, researchers with access to credit panels can have external coauthors, but external coauthors do not get data access. The Federal Reserve Bank of New York Panel is available to researchers across the Federal Reserve System. Data confidentiality agreements mean not all panels can disclose which consumer reporting agency data are sourced from. Other institutions may have access to credit panels purchased by faculty-members but we do not include cases where data are not necessarily broadly available to researchers at the institution. The CRAs offer off-the-shelf products for purchase; the names and contents of these frequently change. This table is accurate at the time of writing but contents will change over time with panels being created or no longer being updated, and with additional data added to existing panels to extend their coverage or provide more information. See the Online Appendix for additional information on these datasets.

References

Aaronson, Daniel, Sumit Agarwal, and Eric French. 2012. "The spending and debt response to minimum wage hikes." *American Economic Review*, 102(7): 3111–3139.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman. 2016. "The economics of privacy." *Journal of Economic Literature*, 54(2): 442–492.

Agarwal, Sumit, Andrea Presbitero, André F Silva, and Carlo Wix. 2023*a*. "Who Pays For Your Rewards? Redistribution in the Credit Card Market." *Finance and Economics Discussion Series Working Paper No. 2023-007.*

Agarwal, Sumit, Gene Amromin, Souphala Chomsisengphet, Tim Landvoigt, Tomasz Piskorski, Amit Seru, and Vincent Yao. 2023*b*. "Mortgage refinancing, consumer spending, and competition: Evidence from the home affordable refinancing program." *Review of Economic Studies*, 90(2): 499–537.

Agarwal, Sumit, John C Driscoll, Xavier Gabaix, and David Laibson. 2009. "The Age of Reason: Financial Decisions over the Life Cycle and Implications for Regulation." *Brookings Papers on Economic Activity*, 2009(2): 51–117.

Agarwal, Sumit, Richard J Rosen, and Vincent Yao. 2016. "Why do borrowers make mortgage refinancing mistakes?" *Management Science*, 62(12): 3494–3509.

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel. 2018. "Do banks pass through credit expansions to consumers who want to borrow?" *The Quarterly Journal of Economics*, 133(1): 129–190.

Albanesi, Stefania, Giacomo DeGiorgi, and Jaromir Nosal. 2022. "Credit growth and the financial crisis: A new narrative." *Journal of Monetary Economics*, 132: 118–139.

Alexandrov, Alexi, Alyssa Brown, and Samyak Jain. 2023. "Looking at credit scores only tells part of the story – cashflow data may tell another part." *Consumer Financial Protection Bureau Blog 26 July 2023.*

Amromin, Gene, Jennifer Huang, Clemens Sialm, and Edward Zhong. 2018. "Complex mortgages." *Review of Finance*, 22(6): 1975–2007.

Argyle, Bronson, Benjamin Iverson, Taylor D Nadauld, and Christopher Palmer. 2021. "Personal Bankruptcy and the Accumulation of Shadow Debt." *NBER Working Paper No. 28901.*

Argyle, Bronson, Taylor Nadauld, and Christopher Palmer. 2023. "Real effects of search frictions in consumer credit markets." *The Review of Financial Studies*, 36(7): 2685–2720.

Argys, Laura M, Andrew I Friedson, M Melinda Pitts, and D Sebastian Tello-Trillo. 2020. "Losing public health insurance: TennCare reform and personal financial distress." *Journal of Public Economics*, 187: 104202.

Athreya, Kartik, José Mustre-del Río, and Juan M Sánchez. 2019. "The persistence of financial distress." *The Review of Financial Studies*, 32(10): 3851–3883.

Avery, Robert B, Mary F Bilinski, Brian K Bucks, Christine Chai, Tim Critchfield, Ian H Keith, Ismail E Mohamed, Forrest W Pafenberg, Saty Patrabansh, Jay D Schultz, and Claudia E Wood. 2017. "A profile of 2013 mortgage borrowers: Statistics from the National Survey of Mortgage Originations." *Technical Report No. 16-01.*

Avery, Robert B, Paul S Calem, Glenn B Canner, and Raphael W Bostic. 2003. "An overview of consumer data and credit reporting." *Federal Reserve Bulletin No. 89.*

Avery, Robert B, Raphael W Bostic, Paul S Calem, and Glenn B Canner. 1996. "Credit risk, credit scoring, and the performance of home mortgages." *Federal Reserve Bulletin No. 82.*

Babina, Tania, Saleem Bahaj, Greg Buchak, Filippo De Marco, Angus Foulis, Will Gornall, Francesco Mazzola, and Tong Yu. 2024. "Customer data access and fintech entry: Early evidence from open banking." *Stanford University Graduate School of Business Research Paper.*

Bach, Helena, Pietro Campa, Giacomo De Giorgi, Jaromir Nosal, and Davide Pietrobon. 2023. "Born to Be (sub) Prime: An Exploratory Analysis." *AEA Papers and Proceedings*, 113: 166–171.

Baker, Scott R, and Lorenz Kueng. 2022. "Household Financial Transaction Data." *Annual Review of Economics*, 14: 47–67.

Ballance, Joshua, Robert Clifford, and Daniel Shoag. 2020. ""No more credit score": Employer credit check bans and signal substitution." *Labour Economics*, 63: 101769.

Ball, Ian. 2024. "Scoring strategic agents." *American Economic Journal: Microeconomics*, Forthcoming.

Bar-Isaac, Heski, and Vicente Cuñat. 2014. "Long-term debt and hidden borrowing." *The Review of Corporate Finance Studies*, 3(1-2): 87–122.

Batty, Michael, Christa Gibbs, and Benedic Ippolito. 2022. "Health insurance, medical debt, and financial well-being." *Health Economics*, 31(5): 689–728.

Beer, Rachael, Felicia Ionescu, and Geng Li. 2018. "Are income and credit scores highly correlated?" *FEDS Notes August 13, 2018.*

Bellon, Aymeric, J Anthony Cookson, Erik P Gilje, and Rawley Z Heimer. 2021. "Personal wealth, self-employment, and business ownership." *The Review of Financial Studies*, 34(8): 3935–3975.

Benetton, Matteo, Gregory Buchak, and Claudia Robles Garcia. 2022. "Wide Or Narrow?: Competition and Scope in Financial Intermediation." *Working Paper*.

Benetton, Matteo, Marianna Kudlyak, and John Mondragon. 2022. "Dynastic Home Equity." *Working Paper*.

Benmelech, Efraim, Ralf R Meisenzahl, and Rodney Ramcharan. 2017. "The real effects of liquidity during the financial crisis: Evidence from automobiles." *The Quarterly Journal of Economics*, 132(1): 317–365.

Bennardo, Alberto, Marco Pagano, and Salvatore Piccolo. 2015. "Multiple bank lending, creditor rights, and information sharing." *Review of Finance*, 19(2): 519–570.

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra. 2019. "Regional heterogeneity and the refinancing channel of monetary policy." *The Quarterly Journal of Economics*, 134(1): 109–183.

Bergemann, Dirk, and Alessandro Bonatti. 2019. "Markets for information: An introduction." *Annual Review of Economics*, 11: 85–107.

Berger, David, Konstantin Milbradt, Fabrice Tourre, and Joseph Vavra. 2021. "Mortgage prepayment and path-dependent effects of monetary policy." *American Economic Review*, 111(9): 2829–78.

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman. 2010. "What's advertising content worth? Evidence from a consumer credit marketing field experiment." *The Quarterly Journal of Economics*, 125(1): 263–306.

Beshears, John, James J Choi, David Laibson, and Brigitte C Madrian. 2018. "Behavioral household finance." In *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 1, 177–276. Elsevier.

Beshears, John, James J Choi, David Laibson, Brigitte C Madrian, and William L Skimmyhorn. 2022. "Borrowing to save? The impact of automatic enrollment on debt." *The Journal of Finance*, 77(1): 403–447.

Bhaskar, V, and Caroline Thomas. 2019. "Community enforcement of trust with bounded memory." *The Review of Economic Studies*, 86(3): 1010–1032.

Bhutta, Neil. 2014. "Payday loans and consumer financial health." *Journal of Banking & Finance*, 47: 230–242.

Bhutta, Neil, and Benjamin J Keys. 2016. "Interest rates and equity extraction during the housing boom." *American Economic Review*, 106(7): 1742–1774.

Bhutta, Neil, Andreas Fuster, and Aurel Hizmo. 2021. "Paying too much? Borrower sophistication and overpayment in the US mortgage market." *Working Paper.*

Bhutta, Neil, Jacob Goldin, and Tatiana Homonoff. 2016. "Consumer borrowing after payday loan bans." *The Journal of Law and Economics*, 59(1): 225–259.

Bhutta, Neil, Jane Dokko, and Hui Shan. 2017. "Consumer ruthlessness and mortgage default during the 2007 to 2009 housing bust." *The Journal of Finance*, 72(6): 2433–2466.

Bizer, David S, and Peter M DeMarzo. 1992. "Sequential banking." *Journal of Political Economy*, 100(1): 41–61.

Blattner, Laura, and Scott Nelson. 2022. "How Costly is Noise? Data and Disparities in Consumer Credit." *Working Paper.*

Blattner, Laura, Jacob Hartwig, and Scott Nelson. 2022. "Information Design in Consumer Credit Markets." *Working Paper.*

Bleemer, Zachary, and Wilbert van der Klaauw. 2019. "Long-run net distributionary effects of federal disaster insurance: the case of Hurricane Katrina." *Journal of Urban Economics*, 110: 70–88.

Bleemer, Zachary, Meta Brown, Donghoon Lee, Wilbert Van der Klaauw, et al. 2017. "Tuition, jobs, or housing: What's keeping millennials at home." *Federal Reserve Bank of New York Staff Reports*, 700.

Bonatti, Alessandro, and Gonzalo Cisternas. 2020. "Consumer scores and price discrimination." *The Review of Economic Studies*, 87(2): 750–791.

Bornstein, Gideon, and Sasha Indarte. 2022. "The Impact of Social Insurance on Household Debt." *Working Paper.*

Bos, Marieke, and Leonard I Nakamura. 2014. "Should defaults be forgotten? Evidence from variation in removal of negative consumer credit information." *FRB of Philadelphia Working Paper No. 14-21.*

Bouckaert, Jan, and Hans Degryse. 2006. "Entry and strategic information display in credit markets." *The Economic Journal*, 116(513): 702–720.

Braxton, J Carter, Nisha Chikhale, Kyle F Herkenhoff, and Gordon M Phillips. 2024. "Intergenerational Mobility and Credit." *NBER Working Paper No. 32031.*

Brevoort, Kenneth P, Philipp Grimm, and Michelle Kambara. 2015. "Credit Invisibles." *Bureau of Consumer Financial Protection Data Point Series No. 15-1.*

Brown, Meta, Andrew Haughwout, Donghoon Lee, and Wilbert Van Der Klaauw. 2015. "Do we know what we owe? A comparison of borrower- and lender-reported consumer debt." *Economic Policy Review*, 21(1): 19–44.

Chakrabarti, Rajashri, Vyacheslav Fos, Andres Liberman, and Constantine Yannelis. 2023. "Tuition, Debt, and Human Capital." *Review of Financial Studies*, 36(4): 1667–1702.

Charles, Kerwin Kofi, and Jonathan Guryan. 2011. "Studying discrimination: Fundamental challenges and recent progress." *Annual Review of Economics*, 3(1): 479–511.

Chatterjee, Satyajit, Dean Corbae, Kyle P Dempsey, and José-Víctor Ríos-Rull. 2023. "A quantitative theory of the credit score." *Econometrica*, 91(5): 1803–1840.

Cherry, Susan, Erica Xuewei Jiang, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2021. "Government and Private Household Debt Relief During Covid-19." *Brookings Papers On Economic Activity*, Fall Edition.

Cochran, Kelly Thompson, Michael Stegman, and Colin Foos. 2021. "Utility, Telecommunications, and Rental Data in Underwriting Credit." *Urban Institute, December.*

Collinson, Robert, John Eric Humphries, Nicholas Mader, Davin Reed, Daniel Tannenbaum, and Winnie van Dijk. 2024. "Eviction and Poverty in American Cities." *The Quarterly Journal of Economics*, 139(1): 57–120.

Conkling, Thomas S, and Christa Gibbs. 2019. "Borrower Experiences on Income-Driven Repayment." *Consumer Financial Protection Bureau Data Point Series No. 19-5.*

Consumer Financial Protection Bureau. 2011. "The Impact of Differences Between Consumer- and Creditor-Purchased Credit Scores." *Consumer Financial Protection Bureau Report (July).*

Consumer Financial Protection Bureau. 2012. "Key dimensions and processes in the US credit reporting system." *Consumer Financial Protection Bureau Report (December).*

Consumer Financial Protection Bureau. 2014. "Consumer credit reports: A study of medical and non-medical collections." *Consumer Financial Protection Bureau Report (December).*

Consumer Financial Protection Bureau. 2017. "Consumer experiences with debt collection: Findings from the CFPB's Survey of Consumer Views." *Consumer Financial Protection Bureau Report (January).*

Consumer Financial Protection Bureau. 2020. "Payment Amount Furnishing & Consumer Reporting." *Consumer Financial Protection Bureau Office of Research Report No. 20-13.*

Correia, Filipe, Peter Han, and Jialan Wang. 2024. "The Online Payday Loan Premium." *Working Paper.*

Cubides, Emily, and Shaun O'Brien. 2023. "2023 Findings from the Diary of Consumer Payment Choice." *The Federal Reserve Bank of San Francisco, Fed Notes May 5 2023.*

Davis, Charles, Samantha Fu, Aparna Ramesh, Jesse Rothstein, and Evan White. 2022. "Connecting Families to Benefits Using Linked Data: a Toolkit." California Policy Lab.

De Giorgi, Giacomo, Andres Drenik, and Enrique Seira. 2023. "The extension of credit with non-exclusive contracts and sequential banking externalities." *American Economic Journal: Economic Policy*, 15(1): 233–71.

Dell'Ariccia, Giovanni. 2001. "Asymmetric information and the structure of the banking industry." *European Economic Review*, 45(10): 1957–1980.

Dell'Ariccia, Giovanni, and Robert Marquez. 2004. "Information and bank credit allocation." *Journal of Financial Economics*, 72(1): 185–214.

Dettling, Lisa J, and Joanne W Hsu. 2018. "Returning to the nest: Debt and parental co-residence among young adults." *Labour Economics*, 54: 225–236.

DeWaard, Jack, Janna Johnson, and Stephan Whitaker. 2019. "Internal migration in the United States: A comprehensive comparative assessment of the Consumer Credit Panel." *Demographic Research*, 41: 953.

Diamond, Douglas W. 1989. "Reputation acquisition in debt markets." *Journal of Political Economy*, 97(4): 828–862.

Di Maggio, Marco, Amir Kermani, Benjamin J Keys, Tomasz Piskorski, Rodney Ramcharan, Amit Seru, and Vincent Yao. 2017. "Interest rate pass-through: Mortgage rates, household consumption, and voluntary deleveraging." *American Economic Review*, 107(11): 3550–88.

Di Maggio, Marco, and Vincent Yao. 2021. "FinTech borrowers: Lax screening or cream-skimming?" *The Review of Financial Studies*, 34(10): 4565–4618.

Di Maggio, Marco, Angela T Ma, and Emily Williams. 2024. "In the red: Overdrafts, payday lending and the underbanked." *Journal of Finance*, Forthcoming.

Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer. 2007. "Private credit in 129 countries." *Journal of Financial Economics*, 84(2): 299–329.

Dobbie, Will, Paul Goldsmith-Pinkham, and Crystal S Yang. 2017. "Consumer bankruptcy and financial health." *Review of Economics and Statistics*, 99(5): 853–869.

Dobbie, Will, Paul Goldsmith-Pinkham, Neale Mahoney, and Jae Song. 2020. "Bad credit, no problem? Credit and labor market consequences of bad credit reports." *The Journal of Finance*, 75(5): 2377–2419.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J Notowidigdo. 2018. "The economic consequences of hospital admissions." *American Economic Review*, 108(2): 308–52.

Dokko, Jane, Geng Li, and Jessica Hayes. 2015. "Credit scores and committed relationships." *FEDS Working Paper No. 2015-081.*

Durbin, Erik, Greta Li, David Low, and Judith Ricks. 2021. "Characteristics of Mortgage Borrowers During the COVID-19 Pandemic." *Consumer Financial Protection Bureau Office of Research Reports Series No. 21-3.*

Elul, Ronel, and Piero Gottardi. 2015. "Bankruptcy: Is it enough to forgive or must we also forget?" *American Economic Journal: Microeconomics*, 7(4): 294–338.

Fay, Scott, Erik Hurst, and Michelle J White. 2002. "The household bankruptcy decision." *American Economic Review*, 92(3): 706–718.

Federal Reserve Board. 2007. "Report to the congress on credit scoring and its effects on the availability and affordability of credit."

Federal Trade Commission. 2012. "Report to Congress Under Section 319 of the Fair and Accurate Credit Transactions Act of 2003(December 2012)."

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. 2012. "The Oregon health insurance experiment: evidence from the first year." *The Quarterly Journal of Economics*, 127(3): 1057–1106.

Fonseca, Julia. 2023. "Less Mainstream Credit, More Payday Borrowing? Evidence from Debt Collection Restrictions." *Journal of Finance*, 78(1): 63–103.

Fonseca, Julia, and Jialan Wang. 2024. "How Much Do Small Businesses Rely on Personal Credit?" *Working Paper.*

Frankel, Alex, and Navin Kartik. 2019. "Muddled information." *Journal of Political Economy*, 127(4): 1739–1776.

Frankel, Alex, and Navin Kartik. 2022. "Improving information from manipulable data." *Journal of the European Economic Association*, 20(1): 79–115.

Fulford, Scott, and Cortnie Shupe. 2021. "Consumer use of payday, auto-title, and pawn loans: Insights from the Making Ends Meet survey." *Consumer Financial Protection Bureau Office of Research Reports Series No. 21-4*.

Fulford, Scott, and Scott Schuh. 2023. "Revolving versus Convenience Use of Credit Cards: Evidence from US Credit Bureau Data." *Journal of Money, Credit and Banking*, 55(7): 1667–1701.

Fulford, Scott L. 2015. "How important is variability in consumer credit limits?" *Journal of Monetary Economics*, 72: 42–63.

Furletti, Mark J. 2002. "An overview and history of credit reporting." *Federal Reserve Bank of Philadelphia Payment Cards Center Discussion Paper No. 02-07*.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. 2022. "Predictably unequal? the effects of machine learning on credit markets." *Journal of Finance*, 77: 5–47.

Gallagher, Justin, and Daniel Hartley. 2017. "Household finance after a natural disaster: The case of Hurricane Katrina." *American Economic Journal: Economic Policy*, 9(3): 199–228.

Ganong, Peter, and Pascal Noel. 2020. "Liquidity versus wealth in household debt obligations: Evidence from housing policy in the Great Recession." *American Economic Review*, 110(10): 3100–3138.

Ganong, Peter, and Pascal Noel. 2023. "Why do borrowers default on mortgages?" *The Quarterly Journal of Economics*, 138(2): 1001–1065.

Garber, Gabriel, Atif Mian, Jacopo Ponticelli, and Amir Sufi. 2024. "Consumption smoothing or consumption binging? The effects of government-led consumer credit expansion in Brazil." *Journal of Financial Economics*, 156: 103834.

Ghent, Andra C, and Marianna Kudlyak. 2016. "Intergenerational Linkages in Household Credit." *Federal Reserve Bank of San Francisco Working Paper No. 2016-31*.

Goldfarb, Avi, and Catherine Tucker. 2012. "Shifts in privacy concerns." *American Economic Review*, 102(3): 349–353.

Gopalan, Radhakrishnan, Barton H Hamilton, Ankit Kalda, and David Sovich. 2021. "State minimum wages, employment, and wage spillovers: Evidence from administrative payroll data." *Journal of Labor Economics*, 39(3): 673–707.

Gopalan, Radhakrishnan, Barton H Hamilton, Jorge Sabat, and David Sovich. 2023. "Aversion to student debt? Evidence from low-wage workers." *Journal of Finance*, 79(2): 1249–1295.

Granja, Joao, and Fabian Nagel. 2024. "Current Expected Credit Losses and Consumer Loans." *Working Paper.*

Gresenz, Carole R, Jean Marie Mitchell, Belicia Rodriguez, R Scott Turner, and Wilbert van der Klaauw. 2024. "The Financial Consequences of Undiagnosed Memory Disorders." *Federal Reserve Bank of New York Staff Report No. 1106.*

Gross, David B., and Nicholas S. Souleles Souleses. 2002. "Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data." *The Quarterly Journal of Economics*, 117(1): 149–185.

Gross, Tal, Matthew J Notowidigdo, and Jialan Wang. 2020. "The marginal propensity to consume over the business cycle." *American Economic Journal: Macroeconomics*, 12(2): 351–84.

Gross, Tal, Raymond Kluender, Feng Liu, Matthew J Notowidigdo, and Jialan Wang. 2021. "The economic consequences of bankruptcy reform." *American Economic Review*, 111(7): 2309–41.

Gupta, Arpit. 2019. "Foreclosure contagion and the neighborhood spillover effects of mortgage defaults." *The Journal of Finance*, 74(5): 2249–2301.

Guttman-Kenney, Benedict, and Andrés Shahidinejad. 2024. "Unraveling Information Sharing in Consumer Credit Markets." *Working Paper.*

Guttman-Kenney, Benedict, and Stefan Hunt. 2017. "Preventing financial distress by predicting unaffordable consumer credit agreements: An applied framework." *FCA Occasional Paper No. 28.*

Guttman-Kenney, Benedict, Paul Adams, Stefan Hunt, David Laibson, Neil Stewart, and Jesse Leary. 2023. "The semblance of success in nudging consumers to pay down credit card debt." *NBER Working Paper No. 31926.*

Hampole, Menaka. 2024. "Financial Frictions and Human Capital Investments." *Working Paper.*

Han, Song, Benjamin J Keys, and Geng Li. 2018. "Unsecured credit supply, credit cycles, and regulation." *The Review of Financial Studies*, 31(3): 1184–1217.

Harney, Kenneth R. 2003. "Sallie Mae Policy Could Hurt Some Loan Applicants." *Washington Post.*

Haughwout, Andrew, Donghoon Lee, Joelle Scally, and Wilbert van der Klaauw. 2021. "Small Business Owners Turn to Personal Credit." *Liberty Street Economics Blog May 19, Federal Reserve Bank of New York.*

Haughwout, Andrew, Donghoon Lee, Joseph S Tracy, and Wilbert Van der Klaauw. 2011. "Real estate investors, the leverage cycle, and the housing market crisis." *Federal Reserve Bank of New York Staff Report No. 514.*

Hauswald, Robert, and Robert Marquez. 2003. "Information technology and financial services competition." *Review of Financial Studies*, 16(3): 921–948.

He, Zhiguo, Jing Huang, and Jidong Zhou. 2023. "Open banking: Credit market competition when borrowers own the data." *Journal of Financial Economics*, 147(2): 449–474.

Howard, Greg, and Hansen Shao. 2023. "The Dynamics of Internal Migration: A New Fact and its Implications." *Working Paper.*

Hunt, Robert M. 2005. "A century of consumer credit reporting in America." *Federal Reserve Bank of Philadelphia Working Paper No. 05-13.*

International Finance Corporatation. 2012. *Credit Reporting Knowledge Guide.* World Bank.

Jaffee, Dwight M, and Thomas Russell. 1976. "Imperfect information, uncertainty, and credit rationing." *The Quarterly Journal of Economics*, 90(4): 651–666.

Jia, Ning, Raven Molloy, Christopher Smith, and Abigail Wozniak. 2023. "The economics of internal migration: Advances and policy questions." *Journal of Economic Literature*, 61(1): 144–180.

Kalda, Ankit. 2020. "Peer financial distress and individual leverage." *The Review of Financial Studies*, 33(7): 3348–3390.

Keys, Benjamin J, Neale Mahoney, and Hanbin Yang. 2023. "What Determines Consumer Financial Distress? Place-and Person-Based Factors." *Review of Financial Studies*, 36(1): 42–69.

Kluender, Raymond, Neale Mahoney, Francis Wong, and Wesley Yin. 2021. "Medical Debt in the US, 2009-2020." *JAMA*, 326(3): 250–256.

Kluender, Raymond, Neale Mahoney, Francis Wong, and Wesley Yin. 2024. "The Effects of Medical Debt Relief: Evidence from Two Randomized Experiments." NBER Working Paper No. 32315.

Kovbasyuk, Sergey, and Giancarlo Spagnolo. 2024. "Memory and Markets." *Review of Economic Studies*, 91(3): 1775–1806.

Kovrijnykh, Natalia, Igor Livshits, and Ariel Zetlin-Jones. 2023. "Building Credit Histories." *Working Paper.*

Lauer, Josh. 2017. *Creditworthy: A history of consumer surveillance and financial identity in America.* Columbia University Press.

Laufer, Steven, and Andrew Paciorek. 2022. “The effects of mortgage credit availability: Evidence from minimum credit score lending rules.” *American Economic Journal: Economic Policy*, 14(1): 240–76.

Lee, Donghoon, and Wilbert Van der Klaauw. 2010. “An introduction to the FRBNY Consumer Credit Panel.” *Federal Reserve Bank of New York Staff Report No. 479.*

Lee, Sean Chanwook, and Peter Maxted. 2024. “Credit Card Borrowing in Heterogeneous-Agent Models: Reconciling Theory and Data.” *Working Paper.*

Liberman, Andres, Christopher Neilson, Luis Opazo, and Seth Zimmerman. 2019. “The equilibrium Effects of Information Deletion: Evidence from Consumer Credit Markets.” *NBER Working Paper No. 25097.*

Low, David. 2023. “What triggers mortgage default? New evidence from linked administrative and survey data.” *Review of Economics and Statistics*, 1–26.

Low, David, Jasper Clarkberg, and Jack Gardner. 2021. “Subprime Auto Loan Outcomes by Lender Type.” *Consumer Financial Protection Bureau Office of Research Reports Series No. 21-12.*

Marquez, Robert. 2002. “Competition, adverse selection, and information dispersion in the banking industry.” *The Review of Financial Studies*, 15(3): 901–926.

Martínez-Marquina, Alejandro, and Mike Shi. 2024. “The opportunity cost of debt aversion.” *American Economic Review*, 114(4): 1140–1172.

Meier, Stephan, and Charles Sprenger. 2010. “Present-biased preferences and credit card borrowing.” *American Economic Journal: Applied Economics*, 2(1): 193–210.

Mello, Steven. 2023. “Fines and financial wellbeing.” *Working Paper.*

Mian, Atif, and Amir Sufi. 2009. “The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis.” *The Quarterly Journal of Economics*, 124(4): 1449–1496.

Mian, Atif, and Amir Sufi. 2011. “House prices, home equity-based borrowing, and the US household leverage crisis.” *American Economic Review*, 101(5): 2132–56.

Mian, Atif, and Amir Sufi. 2014. “What explains the 2007–2009 drop in employment?” *Econometrica*, 82(6): 2197–2223.

Mian, Atif, and Amir Sufi. 2022. “Credit supply and housing speculation.” *The Review of Financial Studies*, 35(2): 680–719.

Mian, Atif, Kamalesh Rao, and Amir Sufi. 2013. “Household balance sheets, consumption, and the economic slump.” *The Quarterly Journal of Economics*, 128(4): 1687–1726.

Miller, Margaret J. 2003. *Credit Reporting Systems and the International Economy.* Mit Press.

Miller, Sarah, and Cindy K Soo. 2020. "Does increasing access to formal credit reduce payday borrowing?" *NBER Working Paper No. 27783.*

Miller, Sarah, and Cindy K Soo. 2021. "Do neighborhoods affect the credit market decisions of low-income borrowers? evidence from the moving to opportunity experiment." *The Review of Financial Studies*, 34(2): 827–863.

Miller, Sarah, Laura R Wherry, and Diana Greene Foster. 2023. "The economic consequences of being denied an abortion." *American Economic Journal: Economic Policy*, 15(1): 394–437.

Miller, Sarah, Luojia Hu, Robert Kaestner, Bhashkar Mazumder, and Ashley Wong. 2021. "The ACA Medicaid Expansion in Michigan and financial health." *Journal of Policy Analysis and Management*, 40(2): 348–375.

Molloy, Raven, and Hui Shan. 2013. "The Postforeclosure Experience of U.S. Households." *Real Estate Economics*, 41(2): 225–254.

Mullainathan, Sendhil, and Eldar Shafir. 2013. *Scarcity: Why having too little means so much.* Macmillan.

National Consumer Law Center. 2022. "Fair Credit Reporting."

Nicholas, Lauren Hersch, Kenneth M Langa, Julie PW Bynum, and Joanne W Hsu. 2021. "Financial presentation of Alzheimer disease and related dementias." *JAMA Internal Medicine*, 181(2): 220–227.

Nissenbaum, Helen. 2020. *Privacy in context: Technology, policy, and the integrity of social life.* Stanford University Press.

Olafsson, Arna. 2016. "Household financial distress and initial endowments: evidence from the 2008 financial crisis." *Health economics*, 25: 43–56.

Padilla, A Jorge, and Marco Pagano. 1997. "Endogenous communication among lenders and entrepreneurial incentives." *The Review of Financial Studies*, 10(1): 205–236.

Padilla, A Jorge, and Marco Pagano. 2000. "Sharing default information as a borrower discipline device." *European Economic Review*, 44(10): 1951–1980.

Pagano, Marco, and Tullio Jappelli. 1993. "Information sharing in credit markets." *The Journal of Finance*, 48(5): 1693–1718.

Petersen, Mitchell A, and Raghuram G Rajan. 1995. "The effect of credit market competition on lending relationships." *The Quarterly Journal of Economics*, 110(2): 407–443.

Pfäuti, Oliver, Fabian Seyrich, and Jonathan Zinman. 2024. "Bad luck or bad decisions? Macroeconomic implications of persistent heterogeneity in cognitive skills and overconfidence." NBER Working Paper No. 32305.

Phillips, David C. 2020. "Measuring housing stability with consumer reference data." *Demography*, 57(4): 1323–1344.

Piskorski, Tomasz, Amit Seru, and James Witkin. 2015. "Asset quality misrepresentation by financial intermediaries: Evidence from the RMBS market." *The Journal of Finance*, 70(6): 2635–2678.

Piskorski, Tomasz, and Amit Seru. 2021. "Debt relief and slow recovery: A decade after Lehman." *Journal of Financial Economics*, 141(3): 1036–1059.

Rishabh, Kumar. 2024. "Beyond the Bureau: Interoperable Payment Data for Loan Screening and Monitoring." *Working Paper.*

Romeo, Charles, and Ryan Sandler. 2021. "The effect of debt collection laws on access to credit." *Journal of Public Economics*, 195: 104320.

Scott-Clayton, Judith, and Basit Zafar. 2019. "Financial aid, debt management, and socioeconomic outcomes: Post-college effects of merit-based aid." *Journal of Public Economics*, 170: 68–82.

Shaffer, Sherrill. 1998. "The winner's curse in banking." *Journal of Financial Intermediation*, 7(4): 359–392.

Shahidinejad, Andrés. 2024. "Are (nonprofit) banks special? The economic effects of banking with credit unions." *Working Paper.*

Sharpe, Steven A. 1990. "Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships." *The Journal of Finance*, 45(4): 1069–1087.

Small, Mario L, and Devah Pager. 2020. "Sociological perspectives on racial discrimination." *Journal of Economic Perspectives*, 34(2): 49–67.

Stiglitz, Joseph E, and Andrew Weiss. 1981. "Credit rationing in markets with imperfect information." *American Economic Review*, 71(3): 393–410.

Thomas, Lyn C. 2009. *Consumer credit models: Pricing, profit and portfolios.* OUP Oxford.

Varley, Michael. 2024. "Contractual Lock-In: Mortgage Prepayment Penalties and Mobility." *Working Paper.*

Vercammen, James A. 1995. "Credit bureau policy and sustainable reputation effects in credit markets." *Economica*, 461–478.

Von Thadden, Ernst-Ludwig. 2004. "Asymmetric information, bank lending and implicit contracts: the winner's curse." *Finance Research Letters*, 1(1): 11–23.

Whitaker, Stephan D. 2018. "Big data versus a survey." *The Quarterly Review of Economics and Finance*, 67: 285–296.

White, Michelle J. 2007. "Bankruptcy law." *Handbook of law and economics*, 2: 1013–1072.

World Bank. 2012. *Global financial development report 2013: Rethinking the role of the state in finance. Chapter 5: The Role of the State in Financial Infrastructure.* The World Bank.

Yannelis, Constantine, and Anthony Lee Zhang. 2023. "Competition and selection in credit markets." *Journal of Financial Economics*, 150: 103710.

Zinman, Jonathan. 2015. "Household debt: Facts, puzzles, theories, and policies." *Annual Review of Economics*, 7(1): 251–276.