Consumer Credit Reporting Data

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

July 22, 2024

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: D82, E21, G51

Keywords: Consumer credit reporting data, credit bureaus, measurement, credit scores, asymmetric information, household finance

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

*Last Updated: July 22, 2024. Gibbs: Consumer Financial Protection Bureau, christa.gibbs@cfpb.gov. Guttman-Kenney: Rice University, Jones Graduate School of Business, benedictgk@rice.edu. Lee: Federal Reserve Bank of New York, donghoon.lee@ny.frb.org. Nelson: University of Chicago, Booth School of Business, scott.nelson@chicagobooth.edu (corresponding author). van der Klaauw: Federal Reserve Bank of New York, wilbert.vanderklaauw@ny.frb.org. Wang: University of Illinois at Urbana-Champaign, Gies College of Business & NBER, jialanw@gmail.com. We are extremely grateful to 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. We shared an earlier draft with representatives from Equifax, Experian, TransUnion, FICO, and VantageScore and are grateful for their feedback. Guttman-Kenney acknowledges support from the NBER's PhD Dissertation Fellowship on Consumer Financial Management funded by the Institute of Consumer Money Management. The views expressed are those of the authors and do not necessarily reflect those of the Consumer Financial Protection Bureau, the Federal Reserve Bank of New York, the Federal Reserve System, or the United States.

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 coresidence (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 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,

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

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 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

[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.

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 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.

[3] FCRA §602(b), 15 U.S.C. §1681a(b)

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 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

[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).

(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 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

[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).

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).

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

[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.

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 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-

[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.

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).

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

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.

$$s_t = \{\, b_t - b_{t-1} + p_t if \geq 00 otherwise \tag{2}$$

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

[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 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 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.

[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.

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

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.

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

[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."

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 Report 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

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

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 coresidence 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.

Online Appendix:
Consumer Credit Reporting Data

Christa Gibbs[*] Benedict Guttman-Kenney[†] Donghoon Lee[‡]
Scott Nelson[§] Wilbert van der Klaauw[¶] Jialan Wang[||]

July 22, 2024

Contents

[*]Consumer Financial Protection Bureau, christa.gibbs@cfpb.gov
[†]Rice University, Jones Graduate School of Business, benedictgk@rice.edu
[‡]Federal Reserve Bank of New York, donghoon.lee@ny.frb.org
[§]University of Chicago, Booth School of Business, scott.nelson@chicagobooth.edu (Corresponding Author)
[¶]Federal Reserve Bank of New York, wilbert.vanderklaauw@ny.frb.org
[||]University of Illinois at Urbana-Champaign, Gies College of Business & NBER, jialanw@gmail.com

Introduction to Online Appendix

This online appendix provides a variety of more detailed resources to assist users of consumer credit reporting data in conducting research, and to enable potential users to evaluate the potential to do so.

Section A provides a summary of the literature using credit reporting data, highlighting both key contributions as well as the wide range of economic fields in which they have been used. The first subsection briefly describes examples in the literature that use credit reporting data to study various topics. The second subsection focuses on the use of these data to understand issues pertaining to credit reporting itself. The third subsection provides a more exhaustive list of relevant papers, without descriptions, with citations grouped by *Journal of Economic Literature* (JEL) codes.

Section B provides detail on the credit reporting process.

Section C provides details on each of the credit reporting files, summarized in Table 1 of the main paper.

Section D provides details on the debt products contained in tradeline consumer credit reporting data. It includes subsections on each of the main classes of debt products: Mortgages & Home Equity Lines of Credit (HELOCs), Credit Card Accounts, Auto Loans, Student Loans, and Other Loans.

Section E provides additional details on economic measures. Subsections E.2 and E.3 provide additional details on measuring delinquency and new account openings respectively. Subsection E.4 provides additional measures of credit card spending as consumption measure to complement that presented in the main paper. For each of these measures, we note their limitations.

Section F provides additional details on credit scoring.

Section G provides additional details on accessing credit reporting data. This includes subsection G.1 that expands on Table 4 in the main paper by providing a summary of each of the consumer credit reporting panels potentially available to researchers. These are accurate at the time of writing; the institutions with panels and the contents of such panels change over time.

Section H, provides a summary of code available to assist researchers in working with these data. We highlight some published papers that have public code and provide some additional examples for common tasks using these data.

Finally, references to all papers cited across this Online Appendix are at the end, in Section I.

A Literature Review

Here we provide a summary of the literature using credit reporting data, highlighting both key contributions as well as the wide range of economic fields in which they have been used.

The first subsection briefly describes examples in the literature that use credit reporting data to study various topics. The second subsection focuses on the use of these data to understand issues pertaining to credit reporting itself. The third subsection provides a more exhaustive list of relevant papers, without descriptions, with citations grouped by JEL codes.

A.1 Review of Literature Using Consumer Credit Reporting Data

The earliest well-known research using credit reporting data are studies of the 2007–2008 US financial crisis, for example work by Mian and Sufi (2009). Research on the financial crisis then expanded from this early work using aggregated credit reporting data to explore the lessons from individual-level data (e.g., Mian and Sufi, 2011; Adelino et al., 2020) and has shed light on the role of labor markets in the crisis (Mian and Sufi, 2014). Since the crisis, additional work in ***macroeconomics*** has also shown the value of credit reporting data in areas including monetary economics, fiscal policy, consumption behavior, and the study of business cycles.

In the study of monetary policy, this is particularly true with respect to the role of home mortgage borrowing, which leads to path dependence in the effectiveness of monetary policy (e.g., Berger et al., 2021) and regional heterogeneity in monetary policy's implications for inequality (e.g., Beraja et al., 2019). This work also highlights the importance of equity extraction and mortgage refinancing (e.g., Bhutta and Keys, 2016; Di Maggio et al., 2020).

Complementing this work on monetary policy, macroeconomic studies of the effectiveness of fiscal policy have also benefited from credit reporting data and have focused on loan products beyond mortgages (e.g., Mian and Sufi, 2012). Similarly, macroeconomists have used these data to study consumption and overall borrowing behavior (e.g., Mian et al., 2013; Benmelech et al., 2017; Chatterjee et al., 2023). Researchers have also used credit reporting data to study the drivers and dynamics of the business cycle (e.g., Gross et al., 2020).

A large body of ***finance*** research also uses these data. The ability to observe the portfolio of debt held by consumers over time enables an understanding of household finances and measurement of how policy changes can affect credit access and financial distress.[1] Research using these data has studied lending and borrowing via auto loans (e.g., Chakrabarti and Pattison, 2019), credit cards (e.g., Keys and Wang, 2019), mortgages (e.g., Bhutta et al., 2022), student loans (e.g., Di Maggio et al., 2023), payday loans (e.g., Gathergood et al., 2019a), and FinTech (e.g., Fuster et al., 2019). As examples of the effects of specific policy interventions, Butcher and Munoz (2017) and Conway et al. (2023) evaluate the impact of the Community Reinvestment Act on consumer credit access and outcomes.

The use of credit reporting data in research goes well beyond the macroeconomics and finance fields. Using credit reporting data in ***health economics*** to better understand the effects of health policies and events is a relatively new use of these data that saw significant growth starting in the 2010s. Several studies have used geographic or birth year information to show reductions in financial distress following expansions of health insurance coverage (Mazumder and Miller, 2016; Hu et al., 2018; Brevoort et al., 2020; Batty et al., 2022). Others have used credit data linked to health-related data sources to document the financial consequences of health events such as hospital admissions (Dobkin et al., 2018), abortions (Miller et al., 2023) and Alzheimer's diagnosis (Nicholas et al., 2021; Gresenz et al., 2024). Meanwhile, the growing use of medical credit cards and financing plans remains largely unexplored using credit record data.

Credit reporting data has also been used to inform studies in ***labor economics***. For example, studies linking credit and census data have advanced understanding of labor search and entrepreneurship (e.g., Herkenhoff et al., 2023, 2021), as have studies using a CRA's wage data from payroll records (e.g., Di Maggio et al., 2022). For example, Dobbie et al. (2020), Corbae and Glover (2018), Bartik and Nelson (2022), and Braxton et al. (2024b) study the interaction between credit histories and labor market outcomes. Several analyses have relied on credit data to study the impact of minimum wage increases on spending, debt, and access to credit (e.g., Aaronson et al., 2012; Cooper et al., 2020; Gopalan et al., 2021b). Similarly, several studies have analyzed the determinants and consequences of participation in the gig economy using credit data (e.g., Buchak, 2024; Fos et al., 2024). Relatively little work has explored intra-household and inter-generational behavior, but there is great potential in this avenue (e.g., Dokko et al., 2015; Bleemer et al., 2017; Benetton et al., 2022; Bach et al., 2023).

Additionally, the coverage of these data—including nearly all US adults and following their movements over long periods of time—makes them well-suited to studying issues in ***environmental economics*** and ***urban economics***. For example, several studies have

[1] For reviews of the field of household finance research see (Guiso and Sodini, 2013; Beshears et al., 2018; Gomes et al., 2021) in which these data have proven valuable.

investigated the effects of natural disasters on credit (e.g., Gallagher and Hartley, 2017; Billings et al., 2022) and non-credit outcomes such as migration (e.g., Bleemer and van der Klaauw, 2019; DeWaard et al., 2020). Gallego and Meisenzahl (2022) study internal migration patterns following the Financial Crisis. Differences in credit profiles between renters and home owners were analyzed by Li and Goodman (2016), while the impact of tuition and student debt on home ownership was studied using credit data by Mezza et al. (2020) and Bleemer et al. (2021). These data can also be used to document regional disparities (e.g., George et al., 2019) and to help inform whether these reflect place-based or person-based factors (e.g., Keys et al., 2023).

There are many other fields where credit reporting data have only made small inroads so far, but where there is still a wealth of potential for their application by researchers. For example, there is work in ***public economics*** studying the impacts of fiscal stimulus, as with the cash for clunkers program (Mian et al., 2010), and public policies such as the moving-to-opportunity program (Miller and Soo, 2021), housing vouchers (Davis et al., 2021), EITC (Caldwell et al., 2023), and traffic fines (Mello, 2023). At the same time, there is little work studying the relationship between debt and different retirement saving systems; exceptions are Beshears et al. (2022, 2024)'s analyses of the effects of pension auto-enrollment on debt, tax changes, and borrowing decisions.

Likewise, there is only a small existing ***political economy*** literature using these data (e.g., Mian et al., 2010; Brown et al., 2019). However, the wide geographical coverage that can be shared down to a fine granularity (e.g., zipcode, census tract, or census block group) makes these data well suited to studying this topic by exploiting spatial variation; in principle, voter registration data and election participation data in some states may be linkable with credit reporting data.

These data have also been used in ***behavioral economics*** frameworks to, for example, better understand credit card borrowing (e.g., Meier and Sprenger, 2010; Ponce et al., 2017; Gathergood et al., 2019b). ***Industrial organization*** and ***marketing*** research has used versions of these data merged with marketing offers to study consumer demand (e.g., Agarwal et al., 2010; Bertrand et al., 2010; Stango and Zinman, 2016; Han et al., 2018) or optimal regulation under imperfect competition (e.g., Galenianos and Gavazza, 2022; Nelson, 2022), but there is considerable untapped potential to extend industrial organization and marketing research using these data. Finally, these data can be useful in informing topics of ***economic measurement***, especially for researchers looking for "big data" to take their machine learning and AI methods to (e.g., Albanesi and Vamossy, 2019; Blattner and Nelson, 2022; Blattner et al., 2021).

A.2 Review of Literature Studying Credit Reporting

In the previous subsection, we provide an overview of how consumer credit reporting data has been used to study topics across economic fields. In this subsection, we complement this by specifically reviewing additional literature studying consumer credit reporting.

Diamond (1984) and Ramakrishnan and Thakor (1984) provide theoretical rationales for firms to form coalitions to share information, and to delegate monitoring to an intermediary such as a consumer credit reporting agency (CRA). Following Pagano and Jappelli (1993), a series of studies understand the formation of information sharing regimes across domains (e.g., Brown et al., 2009; De Janvry et al., 2010; Doblas-Madrid and Minetti, 2013; Brennecke, 2016; Liberti et al., 2022). Brown and Zehnder (2007) provide experimental evidence to understand the circumstances when firms voluntarily share data and its implications for lending. Closely related, in addition to the studies referenced in the main paper, there are also a information economic theory literature on information sharing (e.g., Raith, 1996) and the economics of data (e.g., Bergemann and Bonatti, 2019; Acemoglu et al., 2022) including potential social gains from sharing given data are non-rivalry (e.g., Jones and Tonetti, 2020). Einav and Levin (2014) discuss the gains to researchers from new big datasets becoming available.

For readers interested in credit reporting around the world, we refer them to Jappelli and Pagano (2002); Djankov et al. (2007); Miller (2003); International Finance Corporatation (2012) and World Bank (2012). There is international and historical variation in what information is recorded in CRAs, often distinguished by "negative-only," that only shows delinquencies, and "positive" that also includes other information such as balances and credit limits. Jappelli and Pagano (2002) provide cross-country evidence showing countries with credit bureaus have more lending and lower defaults. They document that public credit registers are more common in countries where creditor rights are less protected and where private credit reporting agencies (CRAs) have not naturally developed. Corroborating evidence on the importance of creditor rights is also provided in La Porta et al. (1997); Djankov et al. (2007). Bruhn et al. (2013) show a credit bureau is less likely to emerge in economies with a high bank concentration as sharing information would reduce the large incumbents' informational rents. Mian (2012) makes the case for public credit registers. Early studies of US credit bureaus show the value of observing such data on consumers and businesses (e.g., Avery et al., 1996; Barron et al., 2000; Barron and Staten, 2003; Kallberg and Udell, 2003).

A series of papers study relationship lending and related competitive issues in business credit and consumer credit markets (e.g., Petersen and Rajan, 1994, 1995, 2002; Bouckaert and Degryse, 2004; Hauswald and Marquez, 2006; Gehrig and Stenbacka, 2007; Schenone, 2010; Sutherland, 2018; Bank et al., 2023; De Giorgi et al., 2023). Dell'Ariccia and Marquez (2006) show how information sharing may not arise endogenously and mandating information sharing may increase lending volume but increase the probability of a banking crisis.

Researchers have examined credit reports (from private credit bureaus and public credit registers) across the world including Argentina (e.g., Hertzberg et al., 2011), Canada (e.g., Agarwal et al., 2020; Allen et al., 2022; Xu, 2023), Chile (e.g., Foley et al., 2022; Madeira, 2024), India (e.g., Fiorin et al., 2023; Ghosh and Vats, 2023), Mexico (e.g., Seira et al., 2017; Castellanos et al., 2022), Peru (e.g., Lee et al., 2024b), South Africa (e.g., Bertrand et al., 2010), South Korea (e.g., Hahm and Lee, 2011), Sweden (e.g., Bos et al., 2018), and the UK (e.g., Gathergood et al., 2019a; Adams et al., 2022; Guttman-Kenney et al., 2023; Beshears et al., 2024). A variety of empirical studies have examined the effects of adding information to credit reports. Hertzberg et al. (2011) show lending decisions become more coordinated when information is made public. Foley et al. (2022) show the competitive effects of sharing ("positive") information that covers information on non-defaulted credit cards. Guttman-Kenney and Shahidinejad (2024) show how mandating sharing information on credit card limits affects credit access and competition. Guttman-Kenney and Shahidinejad (2024) also show the value of actual payments information for predicting profitability and the fragility of voluntary information sharing to innovations enabling targeting marketing.

Credit reports contain many sources of exogenous variation for researchers to use to study credit and non-credit behaviors. For example, research has examined the removal of past delinquencies (e.g., Bos et al., 2018; Liberman et al., 2019; Blattner et al., 2022; Guttman-Kenney, 2024; Madeira, 2024), bankruptcies (e.g., Musto, 2004; Dobbie et al., 2020; Gross et al., 2020; Herkenhoff et al., 2023, 2021; Jansen et al., 2023), public records (e.g., Fulford and Nagypál, 2023), medical debts in collections (e.g., Batty et al., 2022), and the exclusion of inquiries from credit scores (e.g. Madeira, 2024). Other sources of variation include geographic moves (e.g., Keys et al., 2023) and exposure to geography-based treatments (e.g., Gallagher and Hartley, 2017; Bleemer and van der Klaauw, 2019; Billings et al., 2022), and variation in firms' policies.

A variety of work studies credit scores. Avery et al. (2009) analyze how credit scoring has affected the availability and affordability of credit. Meier and Sprenger (2012) show time discounting predicts credit scores. Israel et al. (2014) show credit scores also predict cardiovascular health. Homonoff et al. (2021) show that when consumers receive information about their credit score, they make fewer late payments. A handful of studies examine the effects of fraud (e.g., Mikhed and Vogan, 2018; Blascak et al., 2019; Mohr and Kohli, 2024). Brevoort et al. (2013) consider the impacts on credit scores of including accounts held by authorized users (i.e., users of an account, such as a credit card, who are different from but are authorized by a primary account holder).

A variety of studies examine the value of alternative data sources in predicting consumer defaults. Khandani et al. (2010); Norden and Weber (2010); Puri et al. (2017); Tobback and Martens (2019); Lee et al. (2024b) show the value of bank transactions data. Alexandrov et al. (2023) shows cashflow data measured in survey data predicts default. Djeundje et al. (2021) show the value of email usage, psychometrics, and demographic variables. Consumer Financial Protection Bureau (2014) examine remittance histories. Björkegren and Grissen (2018, 2020) study mobile phone data. Wei et al. (2016) study social media data and Lin et al. (2013) study social networks. Berg et al. (2020); Fu et al. (2020); Chioda et al. (2024) examine digital footprints. These alternative data sources can be especially important for evaluating credit risk in countries where banking systems are less developed (e.g., Burlando et al., 2024; Robinson et al., 2023).

An emerging literature is studying the implications of open banking, where a consumer can grant permission for their banking data to be shared with other institutions (e.g., Babina et al., 2024; He et al., 2023; Rishabh, 2024). Lenders around the world appear to be increasingly using such data instead of or as well as traditional credit reporting data.

There are many other related literatures implicated in the regulation of credit reporting data. For example, work on discrimination and policy remedies for it (e.g., Charles and Guryan, 2011; Small and Pager, 2020), the literature on design of a scoring system (e.g., Bonatti and Cisternas, 2020; Frankel and Kartik, 2022; Liang et al., 2021), and the literature on consumer demand for privacy (e.g., Goldfarb and Tucker, 2012; Acquisti et al., 2016; Nissenbaum, 2020). There is a substantial computer science and operations research on the methods for constructing credit risk models (e.g., Hand and Henley, 1997; Thomas, 2009).

A.3 Summary of Literature, By JEL Code

This subsection groups papers by their *Journal of Economic Literature* (JEL) codes, and topics within these. This list is not intended to be comprehensive. We assign papers to a single JEL code but many could be regarded as being relevant to multiple JEL codes. In case of future updates to this online appendix, readers who know of papers not included in this list are welcome to email those to the corresponding author.

- **C: Mathematical and Quantitative Methods:**

Machine Learning - Albanesi and Vamossy (2019); Blattner and Nelson (2022); Blattner et al. (2021); Bono et al. (2021); Bartlett et al. (2022); FinRegLab et al. (2022).

- **D: Microeconomics:**

  Behavioral Economics - Meier and Sprenger (2010, 2012); Ponce et al. (2017); Gathergood et al. (2019b); Agarwal et al. (2020); Gopalan et al. (2023).

  Information, Knowledge, and Uncertainty - Chava et al. (2021); Kovrijnykh et al. (2023).

- **E: Macroeconomics and Monetary Economics:**

  Consumption - Musto and Souleles (2006); Fulford and Schuh (2017); Di Maggio et al. (2017); Demyanyk et al. (2017); Berger et al. (2018); Gross et al. (2020); Ganong and Noel (2020); Agarwal et al. (2023a, 2018); Athreya et al. (2019); Lee and Maxted (2024).

  Great Recession - Mian and Sufi (2009, 2011, 2012); Mian et al. (2013); Mian and Sufi (2014); Avery and Brevoort (2015); Bhutta (2015); Bhutta and Keys (2016); Benmelech et al. (2017); Bhutta et al. (2017); Mian and Sufi (2017); Foote et al. (2021); Piskorski and Seru (2021); Albanesi et al. (2022); Mian and Sufi (2022); Pinto and Steinbaum (2023).

  Monetary Policy Beraja et al. (2019); Di Maggio et al. (2020); Berger et al. (2021).

- **G: Financial Economics:**

  Auto Loans - Chakrabarti and Pattison (2019); Yannelis and Zhang (2023); Butler et al. (2023a); Argyle et al. (2023).

  Buy Now Pay Later - Zeballos Doubinko and Akana (2023); Shupe et al. (2023); Papich (2023).

  Credit Cards - Fulford (2015); Debbaut et al. (2016); Keys and Wang (2019); Fulford and Schuh (2023); Nelson (2022); Adams et al. (2022); Guttman-Kenney et al. (2023); De Giorgi et al. (2023); Chava et al. (2023a); Xu (2023).

  Mortgages - Bhutta and Canner (2013); Brevoort and Cooper (2013); Piskorski et al. (2015); Bayer et al. (2016); Chan et al. (2016); Bond et al. (2017); Fuster et al. (2018); Gupta (2019); Lambie-Hanson and Reid (2018); Abel and Fuster (2021); Laufer and Paciorek (2022); Hossain et al. (2023); Gao et al. (2024); Zhang (2023a); Pal (2024).

Student Loans - Di Maggio et al. (2023); Black et al. (2020); Yannelis and Zhang (2023); Herbst (2022); Chakrabarti et al. (2023); Hampole (2024); Sauers (2022); Dinerstein et al. (2024); Chava et al. (2023b); Hamdi et al. (2024); Gallagher et al. (2023a).

Payday Loans - Bhutta (2014); Bhutta et al. (2015, 2016); Carter and Skimmyhorn (2017); Desai and Elliehausen (2017); Gathergood et al. (2019a); Miller and Soo (2020); Fulford and Shupe (2021b); Xie et al. (2023); Correia et al. (2024); Di Maggio et al. (2024).

Debt Collection - Brevoort et al. (2020); Fedaseyeu (2020); Kluender et al. (2021); Guttman-Kenney et al. (2022); Romeo and Sandler (2021); Cheng et al. (2021); Fonseca (2023); Kluender et al. (2024).

FinTech - Fuster et al. (2019); Berg et al. (2020); Di Maggio and Yao (2021); Jagtiani and Dolson (2021); Ben-David et al. (2022); Mishra et al. (2022); Balyuk (2023); Zhang (2023b); Chioda et al. (2024).

Credit File Forbearance - Cherry et al. (2021); Allen et al. (2022); Kim et al. (2022); Guttman-Kenney (2024); Xie and Moulton (2023).

Credit Reporting - Avery et al. (2009); Brevoort et al. (2013); Brown et al. (2015); Haughwout and van der Klaauw (2015); Garmaise and Natividad (2017); Arya et al. (2013); Mikhed and Vogan (2018); Blascak et al. (2019); Fulford and Nagypál (2023); Jansen et al. (2023); Blattner et al. (2022); Foley et al. (2022); Guttman-Kenney and Shahidinejad (2024); Burke et al. (2023); Madeira (2024); Mohr and Kohli (2024).

Credit Unions - Shahidinejad (2024).

- **H: Public Economics:** Mian et al. (2010); Demyanyk et al. (2019); Davis et al. (2021); Dupor et al. (2022); Mello (2023); Fulford and Shupe (2021a); Miller and Soo (2021); Beshears et al. (2022); Bornstein and Indarte (2022); Fulford and Nagypál (2023); Zhong et al. (2023); Andre et al. (2024); Beshears et al. (2024).

- **I: Health, Education, and Welfare:** Finkelstein et al. (2012); Mazumder and Miller (2016); Brown et al. (2016); Bhole (2017); Hu et al. (2018); Dobkin et al. (2018); Nicholas et al. (2021); Argys et al. (2020); Goldsmith-Pinkham et al. (2021); Batty et al. (2022); Blascak and Mikhed (2023); Miller et al. (2023); Smith et al. (2020); Frisancho (2023); Dooley and Gallagher (2024); Bruhn et al. (2023); Butler et al. (2022); Andre et al. (2023); Gresenz et al. (2024).

- **J: Labor and Demographic Economics:** Aaronson et al. (2012); Dokko et al. (2015); Ghent and Kudlyak (2016); Herkenhoff et al. (2023); Bos et al. (2018); Bleemer et al.

(2017); Dettling and Hsu (2018); Dobbie et al. (2020); Ballance et al. (2020); Braxton et al. (2024b); Cooper et al. (2020); Mezza et al. (2020); Bellon et al. (2021); Gopalan et al. (2021a); Herkenhoff et al. (2021); Fos et al. (2024); Fritsch and Heimer (2020); Gopalan et al. (2021b); Benetton et al. (2022); Buchak (2024); Cortés et al. (2022); Di Maggio et al. (2022); Bach et al. (2023); Flamang and Kancherla (2023); Moulton et al. (2023); Braxton et al. (2024a); Butler et al. (2023b).

- **K: Law and Economics:** Bankruptcy - Musto (2004); Dobbie et al. (2017); Albanesi and Nosal (2018); Gross et al. (2021); Nagel (2024).
- **L: Industrial Organization. & M: Business Administration and Business Economics; Marketing; Accounting; Personnel Economics:** Agarwal et al. (2010); Bertrand et al. (2010); Stango and Zinman (2016); Han et al. (2018); Galenianos and Gavazza (2022); Jiang et al. (2021, 2023); Jiang (2022); Mayer (2024); Chan et al. (2022); Granja and Nagel (2024).
- **O. Economic Development, Innovation, Technological Change, and Growth:** Seira et al. (2017); Castellanos et al. (2022); Fiorin et al. (2023); Ghosh and Vats (2023); Agarwal et al. (2023b).
- **P. Political Economy and Comparative Economic Systems** - Brown et al. (2019); Mian et al. (2010).
- **Q. Agricultural and Natural Resource Economics; Environmental and Ecological Economics** - Gallagher and Hartley (2017); Roth Tran and Sheldon (2017); Bleemer and van der Klaauw (2019); DeWaard et al. (2020); Billings et al. (2022); Benjamin et al. (2022); Cookson et al. (2022); Gallagher et al. (2023b); Cookson et al. (2023).
- **R: Urban, Rural, Regional, Real Estate, and Transportation Economics:** Brevoort (2011); Haughwout et al. (2011); Whitaker (2018); DeWaard et al. (2019); Bleemer et al. (2021); Howard and Shao (2023); Keys et al. (2023); Mabille (2023); Fonseca and Liu (2024); Liebersohn and Rothstein (2024).

B Details on Credit Reporting Processes

Below, we further explain the practicalities of how credit reporting data are furnished, or transferred from consumer-facing firms to consumer reporting agencies that aggregate and standardize the data before they are shared with researchers. Understanding this data generation process enables researchers to better anticipate and mitigate challenges for their research designs (e.g., confusion between stocks and flows can easily result from misunderstanding the furnishing process). This section also briefly explains potential sources of measurement error, such as incomplete coverage of debts and people, fragmented records, reporting lags, and stale information.

Reporting Changes over Time

As discussed in the main text, the FCRA is the primary law regulating US consumer credit reporting information, and it has been amended several times over the last 50 years by Congress to address persistent issues with accuracy and the ability of consumers to access remedies. For example, the 1996 amendment added a new 30-day limit for CRAs to respond to consumer-initiated disputes, imposed new requirements about the deletion and potential re-insertion of disputed information, and placed obligations on data furnishers for the first time, primarily regarding data accuracy. A 2003 amendment, meanwhile, added more protections to help those affected by identity theft, among other changes.

More recent changes have arisen due to rulemakings by federal agencies, and some have focused on the information reported on credit records. For example, a 2009 rule issued pursuant to the 2003 amendments generally mandated the reporting of credit limits, which some lenders had chosen to not report. Regulators stated the omission of this field could create a misleading assessment of a consumers' creditworthiness.

Other changes have arisen in response to changes in various credit markets. For example, following the Great Recession, the US government introduced the Home Affordable Modification Program (HAMP) in 2009 to help homeowners under stress, but the existing credit reporting system at that time had no means to accommodate this new program and reported them as "making partial payment," harming credit scores. After the US Treasury recommended that the industry address the issue, the Consumer Data Industry Association (CDIA) created a new code designed to signify participation in the Making Home Affordable program including HAMP. Prior to the COVID-19 pandemic, payment deferrals and loan modifications were typically ad-hoc and varied by market and over time. By contrast, at the start of the pandemic, the Coronavirus Aid, Relief, and Economic Security (CARES) Act amended the FCRA to define pandemic-related accommodations and outlined how the payment status should be reported for accounts with an accommodation (15 U.S.C. §1681s-2).

Credit Reporting Processes

Credit reports only exist for individuals with a credit record, which are a subset of adults in the population (discussed in Section G.1). Researchers therefore need to consider the implications for their study of individuals unobserved in credit reports (also referred to as "credit invisibles" in Brevoort et al., 2015).

Even when an individual has a credit file, sometimes this file is a "fragment" record whereby the CRA is unable to consolidate an individual's credit data 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. As a result of these fragments, there are more credit records than adults in the US,[2] and not accounting for this results in average debts per credit file to be lower than average debts per person.

Fragment records may merge into older records as the CRA receives new or corrected information, or if the CRA changes its matching algorithm. Existing records may also split into different records via the same processes when the CRA determines parts of a record belong to another record. These changes can make it difficult to properly define a panel of consumers over time. (Section G.1 below provides guidance on this issue.)

Additionally, credit file data used by researchers are not real-time data. Researchers will typically analyze credit reports in the form of "archives" or "retros" which recreate how a credit file appeared at a point-in-time (typically at the end of a calendar month). Instead of reflecting consumers' real-time credit outcomes as of a given point in time, a credit "archive" reflects the best available information furnished by lenders as of that date. While furnishing broadly operates at a monthly frequency, with new data being furnished by different lenders throughout the month, some lenders do not report all new credit activity within a calendar month, leading to reporting lags. That is, a given archive may contain information for different calendar months for different credit products and different consumers. Researchers should keep in mind that this may create issues with research designs that require precise timing.

Reporting lags are especially likely when accounts are opened, transferred, or severely delinquent. Such gaps complicate the tracking of loans over time as well as the computation of aggregate outstanding debt, often requiring imputation of debt balances that remain outstanding. There can be lags between an account opening and when it first

[2] A small number of these records may be credit records that correspond to children under the age of 18 but where the birthyear is not reported. This is more likely in earlier years before changes in reporting standards that increased the reporting of birthyear.

appears in a credit file, and these lags vary across lenders, asset classes, and over time. For example, new credit card originations tend to appear much faster than new mortgage originations. It is also not uncommon for large transfers of accounts to disappear from the credit record data for a few months before reappearing with a new furnisher. Accounts in collections or charge-off are also sometimes more likely to not be updated regularly by the furnisher. These reporting lags can result in "stale" trades whereby tradelines not furnished with updated information (e.g., account closed, updated balance, or delinquency status) persistently remain on credit reports. These delays raise issues and require special attention when relating individual and aggregate-level activities on credit reports to high frequency events.

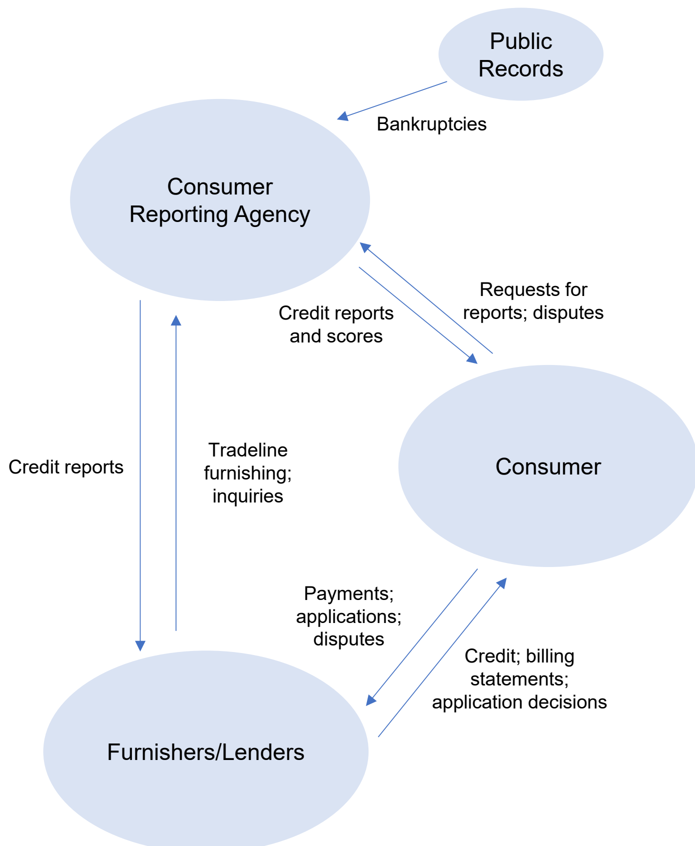
In addition to reporting gaps and delays, another common feature of credit reports is the continued reporting and updating of credit reports of deceased individuals. Data furnishers and CRAs do not always have timely and accurate death information. Failure by researchers to account for inactive individual credit records will naturally lead to incorrect population counts and per-capita debt calculations.[3] Importantly, the inclusion of deceased-person credit reports implies a divergence between credit reports-implied population counts and other population benchmarks that is strongly increasing with age, with a relatively large number of credit reports associated with individuals over age 70. While credit record data do typically include a deceased flag, these flags tend to be sparsely populated, especially in earlier years, and this can vary across CRAs. Patterns in the data suggest these flags can feature both type-1 and type-2 errors in measuring deceased status; in particular, deceased flags can be observed to turn "on" and "off" for some consumers over time. Lee et al. (2024a) have proposed an algorithm for removing inactive records likely associated with deceased individuals, based on the absence of outstanding debt balances, account flags, public filings, and credit inquiries. After implementing this adjustment, their primary sample has an age distribution that looks like the Census target.

Figure B1 is adapted from Consumer Financial Protection Bureau (2012b) and displays the flow of information to create credit reporting data. This shows how lenders (and other furnishers of data) send tradeline data to the CRAs. The CRAs then collate this, along with public records information such as bankruptcies, and return credit reports for lenders and other customers to use. Consumers interact with the system by filing

[3] In Q3 2016, for example, the Equifax-data-based New York Fed Consumer Credit Panel (NYFed CCP), which only includes records with a SSN (and should, therefore, exclude many fragment records), implied a total adult population of 264.9 million, which is well above the Census Bureau's adult population estimate of 249.5 million in 2016, despite the fact that many adults do not appear in credit record data because they do not have formal credit records.

disputes with the lenders or CRA, and by applying for credit which trigger hard inquiries. For more details on the US credit reporting system, see Consumer Financial Protection Bureau (2012b).

Figure B1: The Consumer Credit Reporting System



Flow of information between entities in consumer credit market. *Source: Consumer Financial Protection Bureau (2012b).*

C Details on Credit Reporting Datasets

In this section we explain the structure of consumer credit reporting data itself. We begin with a high-level discussion of the general content of credit reports, and of the type of information typically extracted from them when pulling samples. We differentiate between traditional types of credit reporting data (tradelines, collections, public records, inquiries, and attributes) as well as newer types that have emerged in the last ten years (e.g., alternative credit data, trended data, and non-credit data). We also discuss the types of consumer debts historically missing from credit reports.

Some important household finance information is *not* found on credit reports. Notable missing information includes interest rates and prices, the identity of the lender (as opposed to furnisher), checking or savings account data, assets, 401(k) loans (loans from oneself), stock margin loans, individual and household income, most expenditures, and individual transactions on a credit card.

In discussing the information available at the tradeline-level, we focus in this section on generic issues that researchers encounter in its use. Credit reporting data cover a broad set of credit products with heterogeneous structures and idiosyncrasies in reporting. Later in this online appendix, we highlight key features of these data, interpretation issues, and best practices as exemplified by leading papers from the literature separately for each product type. For home-based loans this includes a discussion of home equity lines of credit (HELOCs), mortgages, loan modifications, refinances and forbearances. For credit cards we cover different card types and the challenges in differentiating revolvers and transactors. For auto loans we comment on differentiation between lender types and the reporting of repossessions. We also include an extensive discussion of idiosyncratic aspects of student loan reporting, including servicer transfers, reporting of delinquencies, federal versus private loan differentiation, deferments, forbearances, refinances, and consolidations. We strongly encourage researchers working with tradeline data on specific types of accounts to review these sections for important further details.

The primary components of credit reports are the header file—containing personal information of the consumer—tradelines, public records, inquiries, and collections. We discuss each of these in more detail next.

C.1 Header File

Each credit report includes a header file which contains identifying information, such as the person's social security number, date of birth, name (including alternate spellings), phone number(s), current address (including state, county, and zip code), and previous addresses. For those with joint accounts, the names of co-borrowers may also be listed. Header files available to researchers are redacted of personal information.

Addresses listed on the credit report are typically the mailing addresses reported by creditors. The type of residence associated with an address may further include a flag for single family or apartment complex, and, for some individuals, the address can be a post office box. When an individual moves and provides his/her new residential address to creditors, the new address gets reported to the CRAs when the lender updates the account information. Using their own proprietary algorithms, CRAs then update the main mailing address associated with an individual, usually made after the end of the billing cycle (some 30 to 45 days after the new address is reported). The algorithm will consider all recently reported addresses associated with all of the individual's reported account as well as the reliability of each source to determine whether there is sufficient evidence that the borrower has moved to a new location.

C.2 Tradeline File

Credit reports include tradeline (i.e., account) data for each revolving and installment credit account that belongs to an individual. Revolving tradelines include credit cards and lines of credit such as HELOCs, while installment tradelines include mortgages, auto loans, student loans, and personal loans. Each tradeline includes specific information about the account provided by the lender, including the current account balance, type of debt, type of account (e.g., revolving, installment), and the so-called ECOA designator (i.e., whether the individual's legal responsibility over the account is as an authorized user, joint account holder, individual account holder, or co-signed account holder). Whether accounts are individual or jointly held is reported to comply with ECOA requirements (and so this information is sometimes referred to as the "ECOA code"). ECOA only requires the reporting of this information for spouses, but as noted in Brevoort et al. (2013), in practice furnishers provide this information for all associated borrowers regardless of marital status. In addition, the tradeline data include information about the lender (name and address), (partial) account number, current 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.

For confidentiality reasons, samples containing tradelines pulled from credit reports usually exclude lender names but do often include product or industry codes indicating whether the lender is a bank, credit union, finance company, or some type of specialized lender. A few important exceptions are studies that have analyzed credit report records of individuals who took out loans with a specific lender or group of lenders, discussed later in Section G.1.

Tradeline payment history is usually reported as a payment "pattern" or "grid" showing between 24 and 84 months of payment history as a sequence or string of payment status codes. Payment or delinquency status varies between current (paid as agreed), 30-days late (between 30 and 59 days late; not more than 2 payments past due), 60-days late (between 60 and 89 days late; not more than 3 payments past due), 90-days late (between 90 and 119 days late; not more than 4 payments past due), 120-days late (at least

120 days past due; 5 or more payments past due), and a number of categories that indicate the loan is charged-off or otherwise in some "severely derogatory" terminal state of default (e.g., foreclosure, repossession, collections, etc.). The payment status may also indicate that the account was included in a bankruptcy filing by the credit recipient. Not all creditors provide updated information on payment status, especially after accounts have been "derogatory" for a long period of time.[4] Thus, the payment performance profiles obtained from credit reports will to some extent reflect different reporting practices of creditors. Typically, these payment histories are not retroactively changed, but there are occasionally exceptions.

The scheduled payment amount for each account is the required payment amount. In case of a mortgage account (and installment loans more generally) it represents the required payment between payment cycles. For revolving accounts, the scheduled payment amount typically represents the minimum payment amount required as displayed on a periodic account statement. Highest credit is a field with varying meanings depending on the type of loan: for revolving loan products, such as HELOCs or credit cards, it is the credit limit (if reported) or the highest balance ever reported; for installment loans, it is typically the original loan principal; for other accounts it is typically the highest balance reported during the history of the loan.

The reporting of delinquencies on credit reports differs in an important way from delinquencies as conventionally reported by industry. The latter typically remove any accounts that have already been charged off from their delinquency statistics. However, after lenders charge off non-performing balances from their books, the borrower's credit report will have a past-due balance until the debt is repaid or sold to a third-party debt collector, or the lender gives up attempting to collect. As long as the furnisher continues to report and update these outstanding debts, they typically will be included in credit-data-based household debt delinquency measures. Haughwout et al. (2019) show that dropping charged-off debts that continue to report to CRAs yield revised delinquency stock measures that are very comparable to industry measures.

While discharged private loans of different types will eventually stop being reported and may show up instead as collection accounts, this is not the case for delinquent federal student debt, which cannot be charged off and will typically continue to be reported to CRAs until the debt can no longer be reported under the requirements of the FCRA and Higher Education Act. In the case of a moratorium or forbearance of debt payments, such as during the early phases of the pandemic, while CRAs stopped recording such

[4]Researchers can align payment status with calendar time by using the reported balance or status date. See H.2.14 for an example.

loans as being delinquent, industry statistics typically continue to include them as past due amounts.

While credit reports pulled at a specific date yield useful measures of debt stock delinquencies by indicating the amount of debt at various stages of delinquency, observing loan-level longitudinal panel data will reveal richer detail on delinquency transition rates by showing the amount of debt transitioning into and out of various stages of delinquency.

If a consumer closes an account, that account will typically remain on the credit report as a tradeline for seven years, though in some cases the account can fall off the report sooner. Tradelines with a negative history are generally dropped within seven years of the reported delinquency, while account closures following full payment (positive information) generally remain on credit reports up to ten years.

As most revolving and open non-revolving accounts with a positive balance require monthly payments if they remain open, a sudden halt in reporting of an account often indicates that it has been closed. Derogatory accounts can remain unchanged for a long time when the borrower has stopped paying and the creditor may have stopped trying to collect on the account. Avery et al. (2003) report that some of these accounts in fact appear to have been paid off. However, sometimes, typically due to some servicer transfers, accounts without reported updates for more than three months, are later reported again, which necessitate the data user to fill in the intervening months/quarters to make up the disappearing accounts. These gaps were more frequent in earlier periods of data, such as in the early 2000s, but even now those gaps and lapses do occur.

C.3 Public Records File

Credit reports contain data on some public records. Public record information is sourced from county, state, and federal courts, and historically included bankruptcies, foreclosures, civil judgments, and state and federal tax liens. How long such information is reported on credit reports varies by the type of record.

Bankruptcy information includes the filing date and the form of bankruptcy, called "chapters", according to chapters in bankruptcy law. The most common types of non-business bankruptcy for consumers are chapter 7 and chapter 13 bankruptcies. Chapter 7 is the most common among consumers and allows borrowers who cannot afford to make payments to discharge all eligible debts. Chapter 13 bankruptcies instead are structured as a repayment plan that lasts between three and five years. They are used by individuals with regular income who are not eligible for Chapter 7, as well as individuals who want to retain certain assets or to get caught up on their mortgage payments. As regulated by the FCRA, chapter 7 bankruptcy filings generally remain on credit reports for up to ten years, while Chapter 13 bankruptcies generally drop off credit reports after seven years. In addition to a bankruptcy flag, credit reports usually include information on which debts were discharged or included in the bankruptcy filing. Once discharged, such accounts generally show with zero balance. Accounts included in a bankruptcy will usually drop off from credit reports after seven years, while the bankruptcy itself may remain up to ten years.

Other public records historically included civil judgments and tax liens, collected from city, state and federal courthouses by third-party vendors. Information included is the amount of the judgment or amount due, filing date and status. Civil judgments are court filings in favor of a creditor, often a debt collector trying to recover unpaid debts. Tax liens instead are legal claims against a person's property (home, car, bank account) made by the government when a person fails to pay taxes, such as income or property taxes.

Most civil judgments and tax liens remain on a person's credit report for up to seven years after they are filed with the court. However, due to the National Consumer Assistance Plan (NCAP) settlement reached in 2015 between CRAs and 31 state attorneys general, there has since 2017 been a reduction in the number of public records added to credit reports due to new policies adopted by CRAs (for details see Clarkberg and Kambara (2018)). The new policies limit the inclusion of public records to those containing, at a minimum, the consumer's name, address, and Social Security number or date of birth. The public record information must be updated/verified (with a courthouse visit) at least once every 90 days. As a result of the change, civil judgments and tax liens are generally no longer included in credit reports since 2018, though CRAs may still include these when data archived from prior 2018 are used for research (e.g., Fulford and Nagypál, 2023).

C.4 Inquiries File

Credit reports also include information on credit inquiries, which log 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 also by employers, landlords, and the person him/herself. The only information included on credit reports for inquiries is the date of the inquiry and identity of the company or person who requested a copy of the credit report. In anonymized data available to researchers, the information may be coded as a type of business for the company and a type of loan application the inquiry is for.

There are two types of credit inquiries, corresponding to two different permissible purposes under FCRA: so-called hard and soft pulls. Hard pulls are usually triggered by an application for a new loan or, in some cases, for an apartment rental. Hard pulls generally have a modestly negative effect on a consumer's credit score, and a large number of hard inquiries within a short time has a more substantial negative effect, as this type of "credit-seeking" behavior can be predictive of later default. An exception to this is when a large number of hard inquiries are for the same type of loan in a short window. Because this might indicate shopping for a single loan, for example a mortgage or auto loan, CRAs typically have a de-duplicated version where the multiple inquiries are collapsed into a single inquiry for use in credit scoring models. The availability of raw versus de-duplicated inquiry data for researchers may vary across CRAs.

By contrast, soft pulls or soft credit checks typically occur when someone (such as an employer or utility company) checks a person's credit as part of a background check or when someone requests a copy of his/her own report. Since July 2020 new phone and internet service inquiries, which used to count as hard inquiries, are classified as soft inquiries. Soft inquiries do not affect credit scores, and they also are generally not displayed on credit reports provided to third parties. Recently, some lenders have also offered soft pulls to consumers for an initial credit application and only initiate a hard pull if the consumer chooses to proceed.

Each CRA only has information on the inquiries that are submitted to that specific CRA. As a result, when a lender only pulls a credit report from one or two of the major consumer reporting agencies (as is common for most non-mortgage credit inquiries), researchers will not observe all credit inquiries in their data and which ones they observe may vary over time.

Credit inquiries only show one part of a consumers' search process. It does not show consumers who searched but did not reach the credit application stage (e.g., they expected to be rejected or were deterred from applying). Similarly, if credit is not taken out, it is not clear whether this is because the consumer was rejected, changed their mind, or rejected the terms presented. To learn more generally about consumer search behaviors, users may want to examine other complementary datasets, such as the National Survey of Mortgage Originations (Avery et al., 2017; Durbin et al., 2021).

C.5 Collections File

Credit reports include a dataset of debts in collections (third-party collections or collection tradelines). These 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. Debt collectors' reporting practices are not uniform and not all delinquent accounts appear on credit reports. A recent CFPB report (Consumer Financial Protection Bureau, 2023) found that collection agencies collecting debt for a fee primarily furnish medical collections as well as telecommunications and utilities accounts, while the owners of delinquent debt primarily furnish financial and retail collection tradelines. The report found large declines in the aggregate number of collections over the past five years which primarily reflected a decline in the reporting of collection tradelines, not in actual collection activities themselves. It also found collection tradelines to largely be low-balance, non-financial accounts, with medical collections representing the majority of collection tradelines.

We recommend researchers study both the flow and the stock of debt in collections. The flow is generally a more accurate measure because of low persistence in collections accounts on credit records. For some research, such as studies of medical debt in collections, the stock itself may be important.

Medical accounts (and on-time payments on them) are otherwise not regularly reported to CRAs, so these accounts often appear for the first time as collections tradelines. Most collections firms do not report paid medical debts, or unpaid medical debts under $500. There are changes over time in the reporting of medical collections debt to be aware of: for example, in July 2022, CRAs stopped adding new, unpaid medical collections debts until they are one year old—up from the six months imposed under the 2017 National Consumer Assistance Plan (NCAP) settlement—and also stopped reporting paid medical collections debts (Kluender et al., 2021); in April 2023, CRAs stopped reporting unpaid medical collections debt less than or equal to $500. The No Surprises Act of 2022 prohibits surprise medical bills for emergency services and therefore debts arising from such events no longer appear in credit reports. For more details on these changes see Sandler and Nathe (2022) and Brown and Wilson (2023). In September 2023, the CFPB proposed a policy to remove medical debt entirely from credit reports. In recent years, many states have debated or enacted laws to ban medical debts in collections from appearing on credit reports. For example, Colorado's House Bill 23-1126 and New York's Fair Medical Debt Reporting Act both passed in 2023. Finding new data sources to study unreported medical debt is an increasingly important challenge for researchers.

Another type of information included on credit reports for some individuals is unpaid child support, alimony, and separate maintenance payments under a divorce decree or separation agreement. In many states, the state or local child support enforcement agency is required to report unpaid child support debts once they reach $1,000, but they may also report lesser amounts. However, many states do not report to all three nationwide CRAs. Unpaid child support may show up on a credit report as a collection account, court judgment (initiated by either the child support enforcement agency or custodial parent), or as a separate tradeline. Unpaid child support or alimony payments can remain on credit reports for up to seven years.

C.6 Trended Data

Beginning in 2013, the credit reporting agencies developed a new product referred to as "trended data." Prior to this development, credit reporting data used by lenders was based only on the latest cross-section available. Trended data combines this cross-section with a panel dimension of characteristics from a consumer's credit report from roughly the prior two years, though the time range varies by CRA.

The distinction here is subtle. Each cross-section of credit reporting data contains backwards-looking variables—for example, bankruptcy filings from up to 10 years prior. However, in standard consumer-level aggregated data, some data fields such as credit card utilization are only observed contemporaneously. Trended data can be thought of as "lags" of what were previously only contemporaneously observed data fields.

By combining information across archives, credit reporting agencies create new variables that show trends such as whether balances, utilization, and credit risk are trending over time. Interestingly, because the panel dimension of the data can be necessary for inferring how a loan is amortizing, trended data also include estimated borrowing costs: estimated interest rates for mortgages, and estimated effective APRs for auto loans, credit cards, and unsecured loans. These estimated borrowing costs are calculated based on undisclosed proprietary algorithms since the underlying data does not contain a tradeline's pricing, and so these estimates may be measured with error. Trended data also include estimated credit card spending and repayment behaviors, such as which consumers pay their balance in full each month and which instead "revolve" a balance on the card.

C.7 Alternative Data

In recent years CRAs have started to collect additional financial data beyond the traditional sources listed above and have begun to use such new data in some of their credit scoring models. Such alternative credit data, also known as expanded FCRA-regulated data, can be used to evaluate an individual's creditworthiness but is not included in traditional credit reports.

To comply with the FCRA, alternative credit data must be displayable, disputable, and correctable. These may include alternative financial services data on small-dollar installment loans, auto title loans, rent-to-own agreements, and point-of-sale financing, including information provided by at least one of the four largest BNPL lenders. Alternative credit data also includes user-permissioned bank statements, utility and telecommunications bill payments, and rent payment history (Cochran et al., 2021), as well as payroll income, gig economy income, and insurance and childcare payments.

At least one CRA has started to include employment information on credit reports. This information may be based on an employment verification database built from payroll records, or information provided by lenders. Some lenders may include, as part of the account information, the name of up to three employers (current and two previous), including (to extent available) employer name and location, date employed, date left, and position.

D Details on Debt Products

This section provides details on the debt products contained in tradeline consumer credit reporting data. Each subsection provides details on each of the main classes of debt products: Mortgages & Home Equity Lines of Credit (HELOCs), Credit Card Accounts, Auto Loans, Student Loans, and Other Loans.

D.1 Mortgages & Home Equity Lines of Credit (HELOCs)

At $12.4 trillion in Q1 2024, mortgage debt is the largest form of debt held by households, representing 70% of total household debt reported on credit reports. Together with Home Equity Lines of Credit (HELOCs), aggregate housing-related debt amounts to 72% of total household debt. Credit reports include account-level information on all mortgage installment and revolving accounts. The former includes mortgage installment loans such as first mortgages and home equity installment loans/home improvement loans/second mortgages (HELOANs), sometimes referred to as closed-end second liens, secured by housing collateral. Home equity revolving loans, also known as home equity lines of credit (HELOCs), are home equity loans with a revolving line of credit where the borrower can choose when and how often to borrow up to a given credit limit.

Some care should be taken in using the mortgage installment account classification. In addition to lender and account information, some CRAs may use the loan origination balance to classify a mortgage as a first or second (HELOAN). As a result, relatively small first mortgage loans (such as those for mobile homes) may be misclassified as home equity installment loans, while some larger home equity installment loans are sometimes incorrectly classified as a first mortgage. Remarks codes associated with each mortgage loan can often be used to reclassify such loans. For example, as GSEs secure first liens almost exclusively, loans securitized by GSEs can be reclassified as first mortgage loans. The same applies for FHA loans and VA loans. Users should also be aware that the classification of mortgage loans that was applied by the CRA does not immediately provide the position of the lien. For example, for a consumer with a HELOAN but no first mortgage, the home equity installment loan would sit in the first position.

It is relatively easier to identify the lien status of a mortgage loan in case of "piggyback" second mortgages, made at the same time as the main mortgage. The purpose of such loans is to allow borrowers who are not able to make a 20% down payment to borrow additional funds in order to qualify for a primary mortgage without having to pay private mortgage insurance that lenders typically require when putting less than 20% down. Such mortgages were very popular in the early to mid 2000s, when piggyback loans often permitted buying a home with very small down payment. Since the Global Financial Crisis, piggyback loans have been limited to 90% combined loan-to-value.

For each mortgage and HELOC, credit reports typically include the loan origination date (year and month), origination amount, current balance, requested payment amount, term of the loan, credit limit (on HELOCs), individual/joint account type, and payment status. This includes closed mortgage trades with a zero balance that, temporarily, continue to be reported by creditors. When linking individual loans over time, such reported trades help confirm that a loan was indeed paid off and closed and did not disappear for other reasons.

It is not always the case that an account continues to be reported with a zero balance before it stops being reporting altogether. Avery et al. (2003) examine non-reported mortgage accounts and found that for many, a new mortgage account appeared around the time an account stopped being reported, suggesting a refinance or that the servicing was sold.

Primary versus second/investor home mortgage

Unlike loan-level mortgage databases such as HMDA, McDash, CoreLogic and Black Knight (formerly LPS), credit report data do not include the intended use and occupancy status reported on mortgage applications.

Credit report data can reveal whether a given borrower has multiple first mortgages, although it does not include the locations or purchase prices of the homes. Haughwout et al. (2011) use this information to characterize borrowers with two and three or more first mortgage loans over a continuous 2-quarter period as second homeowners and investors, respectively. By linking to LPS administrative data, they were able to assess the accuracy of self-reported intended occupancy status and found extensive misreporting (see also Garcia, 2022 and Elul et al., 2023). Many mortgage borrowers who listed an intention to move into the property never did so, while often reporting holding a large number of first mortgages. They found misreporting to be especially prominent during the boom in the "sand states" (Arizona, California, Florida, Nevada, and Texas), and found such investors defaulted at much higher rate during the housing bust. This research raises concerns about the quality of such occupancy variables of traditional mortgage databases, while illustrating the value of credit reporting data.

An additional benefit of linking loan-level mortgage databases to credit report data is that they enable researchers using mortgage datasets to evaluate selection into their dataset compared to the more complete population of mortgages in credit reports, and, if desired, weight observations accordingly (Fuster et al., 2018). See H.2.8 for a code example of identifying first and second lien mortgages.

Remarks codes and joint account status

Associated with each mortgage products are usually descriptive codes (which may be referred to in different ways by each CRA). For example, Equifax credit reports traditionally include up to two narrative codes for each mortgage account (newer credit reports have up to four narrative codes) which provide additional information regarding the product type of the accounts, the security type of mortgage account, including whether it was guaranteed by one of the GSEs or FHA or VA, whether the mortgage was for a mobile home, or a second mortgage/home equity loan/home improvement loan, and whether the account was included in a bankruptcy or foreclosure.

Importantly, over the life of a loan, new narrative codes may be added. For example, a loan modification or forbearance code may replace a previous narrative code. For this information not to get lost, panel data that allows a user to track and link loans over time are useful.

Another important data field is an identifier that typically accompanies each account indicating whether the account is a joint or individual account. While it may be warranted to treat each joint account holder as responsible for repaying the entire balance in some individual-level analyses, it is important to avoid double counting of joint accounts listed on two different individuals' credit reports when computing household-level or aggregate-level debt balances. A standard way to do so is to divide joint balance amounts by two, assuming joint accounts are held jointly by roughly two persons on average (see H.2.2 for examples).

Foreclosures

Foreclosures, a legal action initiated by mortgage lenders to take control of a property when a borrower fails to keep up their mortgage payments, show up on credit reports soon after filing and often provide information on when the foreclosure proceeding has been completed (which in some states could take a year or longer). They stay on credit reports for seven years from the date of first missed payment that led to the foreclosure action (also known as the "date of delinquency").

Alternatives to foreclosure include a loan modification, short sale, and a deed in lieu of foreclosure. The latter, also called a mortgage release, is an arrangement where a mortgage servicer agrees to let the homeowner turn over the deed to the home and move out, instead of waiting for the servicer to foreclose. In exchange, the servicer will release the borrower from their mortgage obligations. A preforeclosure sale or short sale is the pre-approved sale of a property by a homeowner who has proven an inability to make mortgage payments, for less than is owed. Such borrowers may still remain responsible for making up the difference between the sale price and the outstanding mortgage balance. This could show up on a credit report as a "deficiency judgment." Both short sales and deeds in lieu are borrower-initiated and typically will remain on credit report for up to seven years. Like foreclosures, they typically are reported on the credit report through remarks codes such as "short sale" or "forfeit deed-in-lieu of foreclosure."

Modifications and refinancing

New mortgage originations appear on credit reports without an indicator for whether the loan represents a new purchase or refinance mortgage. Individual- and mortgage account-level panel data can be used to help distinguish refinances. New refinance mortgages typically follow a recently closed (prepaid) mortgage without a change in mailing address. Users may want to allow for a reporting gap of up to three quarters following the closed loan (although usually the new loan appears in one or two quarters). As an example of such an approach, Haughwout et al. (2023) calculates aggregate equity extraction from refinanced mortgages in the US since 2000.

A new address appearing on the credit report around the time of the mortgage origination, or a mortgage origination without a preceding mortgage that was paid off, instead points to a purchase origination. For this reason it is advisable when acquiring credit report data from CRAs to request inclusion of an anonymous or scrambled address identifier, or at a minimum the census block or tract corresponding to the address. See Mian and Sufi (2022) for one example of such a strategy and H.2.7 for another example. For other examples of researchers measuring refinancing activity, see Bhutta and Keys (2016); Beraja et al. (2019), and Berger et al. (2021).

Forbearances and modifications

Credit report data do not always include a direct forbearance indicator, and furnishers notate forbearance in various ways. Some of the forbearances are notated in narrative codes such as "Natural Disaster" or "Forbearance." Other forbearances appear only as a change in payment amount to zero.

As discussed in appendix Section B, modifications related to the Home Affordable Modification Program (HAMP) after the Global Financial Crisis now are identified by a new CDIA code designed to signify participation in the Making Home Affordable (MHA) program, including HAMP.

Servicer versus lender

Researchers may also use credit reporting data to understand lender, rather than consumer, behavior, and may use lender variation as a source of identification. This requires an understanding of how lenders are observed in these data. Crucially, it is the *servicer* of loans that reports the accounts, and they are not necessarily the same as the lenders or the owners of the debt. Therefore, one should avoid equating furnisher identity to the lender identity.

D.2 Credit Card Accounts

Credit cards are the most widely held formal credit product in the US and the most likely to be a consumer's first-ever tradeline. As open-end credit, cards are also a channel frequently used both as a means of payment and as a source of short-term borrowing. As of Q1 2024 aggregate credit card balances stood at $1.1 trillion. Credit cards come in a variety of forms and are used in a variety of ways, which researchers should be mindful of when using credit card CRA data. As discussed in more detail below, our definition of credit cards, as well as the total credit card balances reported, includes balances on credit cards and charge cards, but excludes store or retail cards; this choice follows how CRAs sometimes view these data, though we recognize it differs from how some regulators and lenders view the market (Consumer Financial Protection Bureau, 2021). We will discuss the latter types of cards in the section on "Other Debt."

Revolver versus transactor

There is an important distinction between "revolving" and "transacting" use of a credit card. *Transacting* refers to credit card accounts where the user (a "*transactor*") fully pays off the past month's (or billing cycle's) balance at each due date. *Revolving* refers to accounts where the user (a "*revolver*") does not. Typically about two-thirds of outstanding balances are revolving debt, and over half of credit card holders have at least one account on which they revolve at any given time, with persistence in revolving behavior over time (Keys and Wang, 2019; Consumer Financial Protection Bureau, 2021; Grodzicki and Koulayev, 2021; Board of Governors of the Federal Reserve System, 2023; Chernousov et al., 2024).[5] Except in cases where an account has a 0% interest rate, for example as a promotion offered by the card issuer, revolving typically implies a user incurs an interest charge or finance charge on their balance. Transactors who have recently transitioned from revolving also may incur interest or finance charges, typically in the first month after such a transition from revolving while their so-called grace period has not yet been restored.

For a typical revolver, part of the balance will be associated with new transactions and part will be carried-over debt. The former approximately equals the new balance minus the previous balance, plus the actual payment amount in the billing cycle. The carried-over balance approximately equals the new balance plus the actual payment amount, minus new transactions.

See Chernousov et al. (2024) for a comparison of the size of the credit card market using different data sources. Brown et al. (2015) shows credit card balances in credit reports are substantially higher than observed in surveys, such as the Survey of Consumer Finances, where there is a known issue of under-reporting (Zinman, 2009; Beshears et al., 2018).

Differentiating revolvers from transactors

It is difficult to distinguish between revolvers and transactors in credit report data. Accordingly, for transacting accounts, the balance shown in credit report data indicates a monthly flow of expenditure, whereas for revolving accounts, the balance indicates a stock of debt.

Account holders' actual payment amounts each month are sometimes, but not always, reported to CRAs. There has also been a downward trend recently in the prevalence of this reporting (see Herman et al., 2020, Guttman-Kenney and Shahidinejad, 2024, and McNamara, 2023). In cases where these data are reported, it becomes possible to infer which accounts are revolving or transacting, and these data can be used to train predictive models of which accounts are revolving or transacting to be used in cases where actual payment is not reported. For more details see section 3.6.2 in the paper.

Utilization and missing credit card limits

[5] Consumers who repay their outstanding balances before the end of the reporting cycle will have a balance of zero dollars reported, but their card use may be inferred by observing payment dates and actual payments made (if reported). See E.4 for more details.

Researchers may be interested in data features other than just the balance on the credit card. The credit limit, for example, is the total credit line that is nominally available to a consumer. In practice, some credit card issuers may approve transactions that bring a user's balance above the credit limit, which generates a nontrivial share of accounts that can be observed with utilization rates greater than 100%. Credit limits are not always reported to CRAs, especially prior to a 2009 rule mandating the reporting of credit limits. In such cases, the credit limit may appear as missing, or may reflect the "high credit," on the account, which is the highest balance ever reported to the CRA for that account. Fulford (2015) and Fulford and Schuh (2017) discuss how to address such data features when trying to measure variability in consumers' credit limits over time. Care should therefore be taken in using reported credit limits to identify whether a credit card holder is "maxed out" on a card, especially prior to 2009. Such a measure is sometimes used to measure the extent to which someone is credit constrained, and it is used by many as a component of a measure of financial distress.

There also is evidence that credit card limits may not be updated as frequently in credit report data as they are changed for the credit card account holder. For example, accounts that transition into delinquent status are sometimes observed to have a coincident increase in their credit limit, which, given that credit limit increases are unlikely for delinquent accounts, could reflect prior credit limit increases that had not been reported to the CRA.

Issuer versus servicer versus card network

Another data feature sometimes available in anonymized credit report data on credit cards is the subscriber, or furnisher, that reports a given account's data to the CRA. Furnishers are typically the entities that service a given account—that is, who receive payments from the consumer, keep track of the account status, and remit any net returns on the loan to an investor.

Credit card servicers may differ from the credit card issuers, especially in cases of small-scale credit card issuers such as small banks or credit unions. Moreover, banks that service their own credit card portfolios may use different subscriber codes for different parts of their portfolio. This makes it difficult to make inferences about market structure or about bank-consumer relationships using anonymized subscriber identifiers alone. We also note that both the issuer and the servicer are often distinct from the card network (e.g., Visa, Mastercard), though the issuer and the card network do coincide in some cases (e.g., Discover, American Express). For more background on the structure and history of card networks, see Evans and Schmalensee (2004).

Intrinsic differences across different types of cards versus semantic-only differences

Another important distinction among credit cards is between general-purpose credit cards and private-label credit cards.[6] General-purpose cards can be used at all merchants who accept cards from a given payment network. Private-label cards, also referred to as store cards or retail cards, can only be used at a limited set of stores, for example a single retailer or a family of retail brands (see Hall, 2024 for a historical study of how and why the credit card market took over the retail cards market between 1970 and 2000).[7] While credit cards started as general-purpose credit cards issued by credit card companies, banks and credit unions, and "retail cards" and "consumer finance cards" were issued by finance companies for specific stores, over time those distinctions have become less binding. Approximately 90% of outstanding credit card balances and 69% of cards are general-purpose credit cards (Consumer Financial Protection Bureau, 2021). While these two types of cards are classified differently in credit report data, a researcher may want to focus only on one of these categories, or both categories together, depending on the setting.

As retail cards, including department, furniture, and jewelry store cards, are classified differently from credit cards issued by banks and credit card companies, a transfer of card accounts between different types of lenders can lead to sudden shifts in outstanding aggregate credit card and retail card balances. For example, such a shift occurred when Walmart store cards issued by Synchrony were sold to Capital One Bank. While the loan product did not really change, its re-classification on credit reports led to a larger increase in aggregate credit card balances and a reduction in retail card balances in data from at least one CRA.

Inactive credit cards that are open but not used by consumers are difficult to define but greatly affect the number of accounts in credit reports. Researchers interested in studying credit card behaviors may want to focus on accounts actually 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 future. A credit card account may be inactive but may have a non-zero statement balance one month a year if it still charges an annual fee.

[6] Prepaid credit cards are not loans, so they are not reported to CRAs.

[7] Confusion may arise when general-purpose credit cards are co-branded, whereby a retailer's or other firm's branding is used on the card. A co-branded general-purpose card might include a card that offers rewards at a particular merchant such as an airline, while the card can still be used at all merchants in a given payment network, not just to make purchases from that airline.

D.3 Auto Loans

Auto loans are closed-end loans used by consumers to finance the purchase of a new or used auto, where the auto is used as collateral for the loan. Auto loans are generally approved with terms of three to eight years with longer terms becoming more common. These are installment loans, meaning they require equal monthly payments for a specific period of time. The credit record data also include the initial loan balance, current balance, and payment history. Car leases, though quite different from auto loans, are also usually reported to CRAs and are typically reported as leases.

Type of car loan lender

There are five categories of auto lenders with different business models. The first two are banks and credit unions which use funding from consumer deposits. The third type, auto finance companies, provide auto loans to consumers using alternative sources of funding, often through securitizing the loans they originate. The fourth type of lender, "captives," are similar to finance companies in the way they fund their lending, but they typically are owned by or affiliated with auto manufacturers to help finance purchases of their cars. Captives have a high market share among both prime and subprime consumers. Finally, there are also "buy-here-pay-here" lenders which provide loans directly for the vehicles they sell, primarily in the subprime market. Not all auto lenders furnish information to the CRAs, and that is particularly true for this last category (Low et al., 2021).

Auto loan delinquencies, even short-duration delinquencies, can lead to car repossessions, which typically show up either as a payment status or a remark code of "repossession."

D.4 Student Loans

Student loans, sometimes referred to as "education loans," are typically installment loans made to students or their families to finance higher education program enrollments. In contrast to other credit products, the federal government plays a large role as a lender in the student loan market, with federal loans making up the overwhelming majority of student loans. The role of the government as a large lender in this market, along with the large share of loans made to borrowers with limited or no income at the time of origination, leads to some unique patterns and reporting for student loans. For example, originations of student loans tend to track school activities and academic years and thus exhibit a seasonal pattern, although interest rates also drive trends in refinancing federal student loans into private student loans and the consolidation of some federal student loans to lock in lower interest rates. Additionally, most borrowers typically have multiple student loans if they use multiple types of loans or borrow for multiple school years.

Credit record data include both federal and private student loans. Federal student loans include loans originated by the government through the Federal Direct Student Lending (Direct) Program, federally guaranteed loans made by private lenders through the Federal Family Education Loan (FFEL) Program, and federally subsidized Perkins loans made by schools.[8]

Despite the inclusion of both federal and private student loans, total outstanding balances reported in credit record data were $1.6 trillion as of Q1 2024, slightly below the amount reported by the Department of Education. We believe one ofthe primary reasons for this discrepancy is the nonreporting of older defaulted federal loans that dropped off from credit reports but are still included in Department of Education. In compliance with FCRA and the Higher Education Act, these older defaulted loans are not reported to CRAs, although borrowers still owe these debts.

As with other debts, defaulted student loans drop off credit records after seven years, although the date that period is measured from may be later. Federal student loans can be reported with a negative payment history for seven years from the time of default (rather than the initial delinquency that lead to default) under the Higher Education Act. This is true for both Direct and FFEL loans. For private loans, the loans will only appear for up to seven years from the intial delinquency. Defaulted federal student loans are also subject to wage and tax-refund garnishments, but it is unclear how reliably this information appears on credit records. Some federal student loans are discharged or forgiven, but there are no special codes to identify when this occurs.[9] When the Department of Education forgives or discharges a student loan, the balance drops to zero, and the loan is reported as paid and closed, the same way a loan repaid by the borrower directly would be reported. For more on the differences between federal and private student loans in credit record data, see below.

Delinquencies and defaults

The Department of Education has special requirements for the reporting of delinquent

[8] All non-Perkins federal loans originated since June 30, 2010 have been made by the government under the Direct Program. Prior to this, private lenders could also make federally guaranteed loans under the (FFEL) Program. The Perkins loan program ended in 2017 and there have been no disbursements since 2018.

[9] See https://studentaid.gov/manage-loans/forgiveness-cancellation and https://studentaid.gov/manage-loans/forgiveness-cancellation/closed-school for more information on the requirements for forgiveness and discharge.

federal student loans that do not apply to private student loans. Specifically, federal student loans cannot be reported as delinquent to the CRAs until they are at least 90 days past due. As a result, delinquent federal loans will often be reported as "current" and then "90 days past due" or more with no intermediate delinquency (e.g., 30 or 60 days past due) observed. Federal loans which fall further behind are categorized as in "default" after 270 or 360 days of delayed payments depending on the loan type and may be reported as a "government claim" on credit records. Defaulted federal loans are then transferred to another servicer, either a guaranty agency or a collections agency depending on the type of loan. As a result, defaulted loans often move between furnishers and may have changes in reported tradeline or account numbers depending on how the CRA assigns these numbers. Defaulted loans which are rehabilitated and brought current are then transferred again to a new servicer.[10] By contrast, private student loans may be reported as delinquent at 30 or more days past due and may be reported as "charged off" when severely delinquent.

Defaulted federal student loans can be cured if the borrower repays the loan in full, consolidates the loan (see below), or rehabilitates the loan. In the event the borrower successfully rehabilitates the loan, the default status is deleted from consumer's credit record, and the payment history is replaced with a '-' in months where the default was reported. When the borrower consolidates a defaulted loan, the prior default will still appear on the credit record (as a closed loan), and the consolidated loan will appear as a new loan.[11]

Federal versus private loans

Federal and private student loans are not typically directly distinguishable in the credit record data without access to the names of the furnishers, and those may still leave some ambiguity. Private education loans are reported much the same way as federal student loans and some furnishers have both types of loans in their portfolios which can make it difficult to distinguish between them.

Users can try to infer loan types based on some remarks codes or loan characteristics. For example, users can try to leverage differences in term lengths or interest rates for federal and private loans, but users need to remember that federal loans may have atypical term lengths or interest rates due to income-driven repayment (IDR) plans, ex-

[10] In late 2022, the Department of Education implemented a program called "Fresh Start" to give borrowers with defaulted federal student loans an opportunity to access benefits to help get and stay out of default. As a result of this program, all federal student loans reported as in default in credit data were newly reported as current; this happened in late 2022 for defaulted Direct loans and in early 2023 for defaulted FFELP loans and will continue for one year (Gibbs, 2023).

[11] For more on federal student loan default, see https://studentaid.gov/manage-loans/default.

tended repayment plans, consolidations, and differences across federal loan types which may complicate these distinctions.

Additionally, certain remarks or narrative codes or other indicators only apply to certain types of loans. For example, a cosigner on the account indicates a private student loan and is typically reported for the life of the loan unless a borrower obtains a release from the lender for the cosigner. In contrast, users can have a designation of "permanently assigned to the government" or "government claim" to identify defaulted federal loans, but these codes are only used when the loan is in default. The CARES Act and subsequent administrative actions provide a unique opportunity to help classify loans into federal and private. Through the CARES Act, all Direct federal student loans went into an automatic payment suspension and interest rates were lowered to 0 percent for more than two years starting in March 2020.

Both private loans and privately owned federal loans were not covered by the CARES Act. As a result, users can infer that a loan is federal based on scheduled monthly payments during the pandemic, but some loans that continued to have scheduled monthly payments of zero may still be federal loans under the FFEL Program. Overall, users may be able to classify many loans as federal or private, but it is difficult to confidently categorize all loans and users should be aware that their estimates will likely be noisy as a result. For some a code example identifying different types of loans, see H.2.13.

Income-driven repayment

IDR plans for federal student loans offer alternative repayment plans for borrowers and have become increasingly common. There are no remarks codes that specify whether a loan is enrolled in an IDR plan, so users must infer enrollment based on other reported information such as loan term, balance amount, scheduled payment amount, and changes in these measures. For example, some loans are reported with $0 scheduled monthly payments (but not in deferment) or they have scheduled monthly payments that would imply a negative or improbably low interest rate. These changes should be reported for a year, since IDR plans have a one-year enrollment period and typically require re-certification to maintain lower payments, but borrowers can resubmit documentation early. In general, reported loan terms should be the maximum number of months for repayment (including accounting for potential forgiveness outside of Public Service Loan Forgiveness), but users should expect that this may not be consistent, especially with older data. For further discussion on identifying loans enrolled in IDR, see Conkling and Gibbs (2019).

Deferments and forbearances

Payment deferments and forbearances are not necessarily indicators of financial distress for student loans. Most student loans are put into a deferred payment status when originated if the student is still in school. This is automatic for federal loans borrowed by the student and is followed by an automatic six-month grace period once the student's enrollment drops below at least half time.[12] These loans may re-enter deferment if the borrower returns to school. These deferments and grace periods may be reported with a remarks code of "payment deferred" or "account in forbearance," depending on the furnisher and these codes have sometimes been used interchangeably. More recently, servicers of federal student loans have been told to furnish loans in deferment, grace, or forbearance as in deferment to avoid sending potentially negative signals to lenders.

Meanwhile, private student loan borrowers may have options available such as deferred payment while in school, the loan put into an "interest only payment" with principal loan payments deferred until the student leaves school, or their loan may be classified as "in repayment" as soon as the loan is originated.

Forbearances, meanwhile, may occur due to borrower distress or for administrative reasons. Borrowers, for example, may request a temporary suspension of payments due to a hardship such as job loss. Borrowers of federal student loans may also be placed in a temporary administrative forbearance while a servicing issue is resolved. See H.2.16 for a code example to identify forbearances in credit record data.

To provide relief to borrowers during the pandemic, payments on all federally held student loans were paused through the CARES Act and subsequent administrative actions but without any narrative code indicating a payment accommodation. From March 2020 through September 2023, all non-defaulted federal loans owned by the Department of Education were reported with a $0 scheduled monthly payment. Additionally, the payment status for all delinquent non-defaulted loans were changed to current and no new delinquencies were reported for federally held loans. The Department of Education also instituted a 12-month "on-ramp" for borrowers so that delinquencies on federally held student loans were not reported for another year after the end of the payment pause.

Servicer versus lender

While the Department of Education owns most student loans, they do not service any of their portfolio. Instead, servicing is split across several companies, which may service other student loans not owned by the Department of Education. Some federal loans (FFEL Program loans) are serviced by the owner of the loans (either the original private lender or another private lender who has purchased the loans since origination) or a third-party servicer if a lender does not service their own loans or in the case of federally-held

[12] For Perkins loans, the grace period is nine months. For Parent PLUS loans, the deferment is not automatic but is currently available to all Parent PLUS borrowers.

FFEL Program loans.[13] Prior to 2013, all Direct loans were serviced and furnished by one company, but the Department of Education has since revised its servicing contracts, and all Direct loans were transferred to other servicers. Over the last several years, some of these servicers have left the system triggering additional large transfers of student loans which can sometimes make it difficult to link individual loans over time. Some of these servicers also furnish information on FFEL loans (made by themselves or other lenders they provide servicing for) and private student loans. Large transfers of student loans may be the result of a change in federal contracting, contracting by private lenders who do not service their loans in-house, or by private lenders selling off their portfolios. As a result, users cannot typically separate loans types by relying on furnisher codes, though it is possible some servicers report different types of loans under different sub-furnishers.

Refinancing and consolidations

In addition to new loans to immediately finance education, student loan originations may also be refinances or consolidations of existing loans. Both federal and private student loans can be refinanced into new private student loans typically in the pursuit of a lower interest rate. Consolidations, meanwhile, combine existing federal student loans into a single new federal loan. A consolidated loan has a new interest rate that is the weighted average of the rates on the prior loans and the new loan may have a longer term, depending on the total loan amount.[14] Consolidations are also an option to help borrowers rehabilitate federal student loans in default which can make it difficult to track some loans over time. Federal consolidation loans also have specific maximum repayment terms ranging from 10 to 30 years based on the total loan amount. The relationship between loan term and loan amount and the weighted interest rate structure of consolidated loans can help researchers distinguish between consolidations and refinances when researchers have loan-level data.

Servicer transfers and reporting gaps

As previously noted, furnishers occasionally stop reporting accounts temporarily. This is often, though not always, associated with a servicer transfer. Most gaps due to transfers are three months or shorter, but there are exceptions. Data users in these cases may need to fill in the intervening periods to account for the missing tradelines. These gaps have been particularly frequent in recent years in reporting by student loan servicers because of the large number of federal servicing transfers. A common practice by some

[13] Several private lenders, for example, sold off their FFEL portfolio to the government during the Great Recession or to other lenders (Wells Fargo, for example, sold their portfolio to Navient). SoFi is an example of a private lender that outsources its servicing to another company, MOHELA.

[14] Older variable rate loans are changed to fixed rate loans during consolidation.

researchers has been to repeat the most recently reported status of the loan (or interpolate the missing periods based on the statuses in the surrounding periods) in cases where there is a simultaneous large drop in reported loans by a specific anonymized furnisher. See H.2.15 for an example with missing student loans in 2011–2012.

D.5 Other Loans

Other loans are, by definition, a residual catch-all category not captured by the main product categories explained in preceding sections. As a result it can contain a broad variety of product types. However, they can be generally considered as revolving accounts for consumer products not captured in the credit cards category or installment loans.

There may be differences in how these accounts are characterized across datasets and projects. For example, some researchers group retail cards (see section D.2 above) into one category while others, like the NY Fed, group them into a larger category of "other" loans.

Some of these loans have remarks codes that provide specific types of relatively small product categories such as "recreational merchandise loans" and "agricultural loans." Still some other loans are included in this residual category due to a lack of identifying description of the nature of the loan. The nominal amount of outstanding debt in this category is fairly unchanged from 2003 to 2024 in the FRBNY-CCP: measured at \$0.49 trillion in 2003, troughing at \$0.30 trillion in 2013 and reaching \$0.54 trillion in Q1 2024.

Given the heterogeneity within this product category and the smaller market sizes, these loans are less frequently the focus of research. Sometimes, however, researchers are able to use institutional knowledge, such as information on the servicer or loan characteristics, to isolate the subset of accounts they are interested in studying. For example, Di Maggio and Yao (2021) identify loans provided by FinTech lenders. In general, it is more common for researchers to only examine this as one disaggregation of a household's debt or as an input to a predictive model. Classifications of loans within this category may change over time as new products develop and reporting categories are generated. For example, CRAs are developing new ways to classify buy now pay later (BNPL) loans.

E Details on Constructing Economic Measures

In Section 3 of the main paper we show how consumer credit reporting data can be used by researchers to construct a variety of economic measures. The subsections of this section of the appendix provide additional details for users of these data who are interested in measuring delinquency, new account openings, or spending. Before this, we provide a high-level evaluation of consumer credit reporting data.

E.1 An Ideal Dataset?

We provide a high-level evaluation of the strengths and weaknesses of consumer credit reporting data, based on adapting the framework of the ideal household finance dataset Campbell (2006) laid out. Although that analysis focused on assets rather than liabilities, many of the themes remain useful. Campbell (2006) writes: *"The ideal data set would have at least five characteristics. First, it would cover a representative sample of the entire population. It is particularly important to have good coverage by both age and wealth, since many aspects of financial behavior vary with these characteristics. Second, for each household the data set would measure both total wealth and an exhaustive breakdown of wealth into relevant categories. Third, these categories would be sufficiently disaggregated to distinguish among asset classes, and ideally would capture specific individual assets so that one could measure household diversification within asset classes. Fourth, the data would be reported with a high level of accuracy. Finally, the data set would follow households over time; that is, it would be a panel data set rather than a series of cross-sections."* How does consumer credit reporting data do?

First, **data representativeness**. By definition, consumer credit reports contain information for all consumers who have credit reports. This covers roughly nine-in-ten adults in the US (Brevoort et al., 2015), across the distributions of age and wealth. These data do not observe children. Adults without credit reports (so-called "credit invisibles") are unobserved and can be inferred to have zero debt of the type reported to CRAs). Credit invisibles disproportionatel includes some racial and ethnic minorities, younger consumers, and unbanked consumers. This form of selection is different from surveys where non-response can bias who responds (Dutz et al., 2022), and is becoming an increasingly important issue to address due to declining response rates. It is also different from other sources of household financial data, such as that gathered from individual firms, where only consumers who use a product or consent to share data are observed (e.g., Baker, 2018; Baker and Kueng, 2022). In some non-US countries, checking accounts, mobile phones, and utilities appear on credit reports. We expect such developments have further increased coverage of these data, though evidence is lacking on this.

Second and third, **data granularity and coverage**. As these data are built from individual accounts, they are highly granular. This enables researchers great flexibility in their approaches to classifying or aggregating different types of accounts. An obvious limitation of these data's coverage is that they only cover the liabilities side of a household's balance sheet, not their assets (although some inferences for autos and houses can be made). For these liabilities, they include some contract terms, and some others can be estimated; however, they do not contain the full contractual features one may ideally desire. We also do not observe individuals' income, only estimates of it, and have partial coverage of their consumption. While these data include most of a consumers' liabilities, there are some important gaps. Some subprime loans not typically furnished to CRAs (e.g., some subprime auto loans and payday loans), most unpaid medical, utility, or business bills, and most missed rent payments. 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. 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 consumers more broadly are difficult to find as there are few comprehensive sources for this information.

Fourth, **data accuracy**. Lenders use these data in their decisions and have good incentives to accurately report data. This is because laws require consumer credit reporting data to be furnished accurately, and mis-reporting their customer's accounts may adversely affect their own business. The Federal Trade Commission (FTC) has conducted a series of reports reviewing credit files errors and estimated in 2012 that 5% of consumers' credit files contained errors that meaningfully adversely affected their credit access (Federal Trade Commission, 2012). Also see Avery et al. (2004); Staten and Cate (2005); Smith et al. (2013); Hynes (2017) for more on such errors. 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 suggesting a role for regulation. While there are errors, credit reporting data line-up well with other sources (Brown et al., 2015), and do not suffer from the mis-measurement of credit card balances known to occur in survey data (Zinman, 2009; Beshears et al., 2018).

Fifth and finally, these are a **panel dataset** and so researchers can follow individual consumers over time. We observe consumers' locations and have high coverage across geographic regions. Consumers who move remain in these data, unless they move outside the sampling frame (e.g., abroad). Depending on their dataset's sampling, researchers can not only observe individual consumers, but follow adults in "households." "Households" can both be by geography (e.g., living in the same residence), and/or by shared financial activities (e.g., sharing a joint account). This enables researchers to study both intra-household and inter-generational behaviors. As mentioned in the main paper, death can be difficult to measure and complicate the definition of a household in these data. This is a problem not encountered in survey data.

So overall we consider consumer credit reporting data an extremely useful dataset for researchers. It complements other data sources which have their own strengths and weaknesses. The ideal dataset does not exist, however, with developments such as increased merging credit reporting data to other sources we can get closer to the researcher's ideal.

E.2 Details on Measuring Delinquency

Section 3.4.3 of the main paper describes ways to measure delinquency. This appendix provides some additional details on how to measure delinquency, for users of these data.

A researcher needs to be aware of the difference between the measure of delinquency reported by a lender and the measure calculated from credit record data. When an account is charged off from a lender's portfolio and transfers to a collection, the account is excluded from the calculation of delinquency rate reported directly by the lender. However, the same account will continue to be reported to the CRAs for a varying duration of time and be included in the delinquency rate calculated based on credit report data.

Researchers may want to use the flow of new accounts that become delinquent, as used by the Federal Reserve Bank of New York in its quarterly reports. These show the number or percent of balances of accounts that transition from being non-delinquent to being delinquent in a given time period (e.g., quarter). Delinquency measures based on the flow to delinquency typically do not face the problems faced by stock-based measures, since they directly capture the transition from being current to becoming delinquent. They are robust to the varying charge-off rules by lenders and loan types and how long those accounts remain on credit reports. As a disadvantage, however, flow-based delinquency rates generally require a more granular level of data, such as panel data at the account level, and, if using more severe statuses, still ultimately depend on these being accurately reported. See Haughwout et al. (2019) for a discussion of two approaches of measuring delinquencies, and see H.2.11 for some code examples to construct these measures.

The threshold for measuring delinquency can vary depending on a researcher's needs. While in the main paper we suggest 30+ or 90+ days past due as a typical threshold, other thresholds may be relevant. For example, 60+ days may be useful as a little over a quarter of accounts reported with a delinquency of 31–59 days are typically reported 60–89 days past due; whereas close to half of accounts reported with a 60–89-day delinquency are typically next reported with a 90–119-day delinquency.

As loans become increasingly severely delinquent, they may be reported in increasingly different ways by different lenders. One reason for this is because the debt may be transferred from the lender to a debt collector who has different reporting approaches.

The timing to charge-off an account could vary across loan types and also by lender. For example, credit cards typically charge off after 180 days past due and in comparison, federal Direct student loans are never charged off. Some lenders report all of their severely delinquent accounts with zero outstanding balances, or do not update the delinquency status to the charge-off stage, and therefore a researcher may want to use the outstanding balance before a loan becomes severely delinquent to reflect how much debt is being charged-off. Account-level data may include a variable for the amount charged-off, however, reporting of this variable appears to be inconsistent across lenders. See Guttman-Kenney and Shahidinejad (2024) for an example estimating charge-offs.

When calculating delinquency a researcher may need to account for how accommodations for the COVID-19 pandemic affected reporting. During the period March 2020 to August 2023 (120 days after the end of the National Emergency for COVID-19 on April 10 2023), amendments to the FCRA by the CARES Act required certain non-paying accounts to be recorded as having an accommodation added; in some analyses these accounts would most appropriately be interpreted as delinquent. When taking this approach, researchers can recode any remark codes on tradelines for deferred payments, forbearance, or being affected by natural disaster, and also recode open credit cards with positive statement balances where they have zero scheduled payments due. See Cherry et al. (2021); Dinerstein et al. (2024) for studies of COVID-19 accommodations and Guttman-Kenney (2024) for a study of natural disaster flags.

A neat feature of tradeline-level credit reporting data is that one cross-sectional archive of data contains a variable containing an 84-month array showing monthly historical delinquencies for the past seven years. However, these past delinquencies are most reliably observed for accounts that remain open, rather than having been closed or charged off. See Gross et al. (2021) and and H.2.14 for examples. This feature is especially useful when studying data from the early 2000s when lenders often did not furnish data every month. If a lender only furnishes quarterly, the payment status variable will not be updated with new information each month. However, when the array is updated each quarter, it will not only show information on the last month in a quarter, but the two prior months as well. These arrays are best used for measuring up to two years of history as some accounts' historic delinquency statuses will be removed from these arrays over time, and will also be updated (e.g., disputes) meaning that the array accurately records delinquency but not how delinquency was historically recorded on a consumers' report. As with other variables, if a lender stops updating the array, as may occur with severely delinquent or closed accounts, it becomes out-of-date.

E.3 Details on Measuring New Account Openings

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. In the case of mortgages one may want to try to distinguish between new purchase originations and mortgage refinances. Equation 1 calculates the value of new auto loans (a_t) using information on outstanding auto loan balances (b_t). This calculates the difference in auto loan balances (b_t) and, when this difference is above a threshold κ, this increase is classified as a new auto loan. This measure is zero otherwise. Restrict to balances where the consumer is up-to-date on payments. If using such an approach we recommend sensitivity analysis for how large an increase in loan balances is required to classify a new purchase. See Agarwal et al. (2023a) for an example of such an approach setting setting $\kappa = \$2,000$ (and testing sensitivities between $\$2,000$ and $\$5,000$). Consumer-level aggregated data may also contain CRA-created variables for the number of new accounts opened. For researchers without access to tradeline data, using the aggregated number of new accounts originated is sufficient for most purposes.

$$a_t = \begin{cases} b_t - b_{t-1} & \text{if } b_t - b_{t-1} > \kappa \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The above approaches can be improved using more granular tradeline data (e.g., Bhutta and Keys, 2016; Gross et al., 2020), to ensure the timing and amount of loan originations are more precisely measured. Because there may be a lag between when a loan is originated and when a loan first appears on a credit report, 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. This measure can be computed by researchers who have lower-than-monthly frequency data of tradeline data (e.g., annual or quarterly), because one cross-section of tradeline data includes origination details for all of a consumer's accounts (opened and closed) over the last ten years.

Newly opened auto loans can be used as a proxy for auto purchases. Researchers use this as a measure of consumption, such as Benmelech et al. (2017) and Di Maggio et al.

(2017). In credit reporting data we observe autos purchased on finance ("auto loans")—over 80% of new auto purchases have auto loans (Benmelech et al., 2017). Some subprime auto loan providers do not appear in credit reports, and therefore credit report measures will not include some auto purchases by this segment (Low et al., 2021). Benmelech et al. (2017) and Di Maggio et al. (2017) verify the accuracy of this consumption measure. They show auto loans originations in credit reports match up to external data and also track total sales (with and without loan financing). It is less common for used autos to purchased on finance than it is for new autos. We estimate, during 2022 and 2023, approximately 55% of used cars and 65% of total (new and used) cars, are purchased on finance, with these shares stable over time.

Researchers with tradeline-level data can study the contract terms of these new accounts (e.g., origination amount, monthly scheduled payment, and scheduled loan duration). Section 3.5.4 of the main paper shows how researchers with tradeline-data can estimate interest rates, or purchase estimates from the CRAs. Researchers may be interested in calculating interest rates at loan origination (i.e., how much they would pay if they exactly follow the terms of the loan), using the first month of data observed (or first few months as required for mortgages), or over a loan's duration (i.e., how much they actually paid), using multiple months of data.

E.4 Details on Measuring Credit Card Spending

Credit cards are broadly used by US consumers with high coverage across geography and credit scores. The amount of spending on credit cards therefore makes them well-suited as a measure of consumption. However, when using credit card spending as a measure of consumption, it is an important caveat to note that this will not include all of a consumer's consumption: it excludes consumption via debit cards, bank transfers, checks, cash, and payroll deductions. Approximately 30% of 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). Researchers will often use these measures by comparing them to a control group.

When calculating credit card spending, we generally recommend combining (general-purpose) credit cards with (private-label) retail credit cards (which can only be used at one or a small group of merchants). Retail cards are a much smaller market (Consumer Financial Protection Bureau, 2021) but are useful to include as they cover different socio-economic groups.

The object of interest—'credit card spending' ($S_{c,t} = \sum_{C_i} s_{i,c,t}$)—is the total value of new purchases by a consumer (i), across their spending on each of their credit cards (c), at time t. We show four ways to attempt to measure this. These increase in complexity and data requirements. The first two measures can be calculated at the consumer-level, or at the individual credit card account level and then aggregated up to the consumer-level. The last two measures require calculating at the individual credit card account level and then aggregating up to the consumer-level. Calculating from the account-level helps to ensure that the measure of spending produced is are not being driven by reporting practices over time on individual credit card accounts.

The simplest but least accurate measure of credit card spending is shown in $s_{c,t}^{simple}$ in Equation 2. This measures spending by the credit card statement balance ($b_{c,t}$). This is a likely inaccurate measure of spending as it includes spending from previous periods that was revolved as debt. It also includes financing charges (the sum of interest and fees) and excludes spending repaid before the statement balance is issued. If using this, we would recommend defining it as credit card statement balances, a useful but different object, and not consider it a consumption measure.

$$s_{c,t}^{simple} = b_{c,t} \tag{2}$$

A more accurate measure of credit card spending ($s_{c,t}^{GNW}$), as used in Gross et al. (2020), is the *change* in credit card statement balance. This is shown by Equation 3. This measure removes some double counting of revolved debt. However, changes in statement balances are the net of the change in new spending less the change in payments and change in financing charges. This means, for example, a credit cardholder whose new spending is unchanged but reduces their payments may, by this measure, appear to spend more even though their spending is unchanged. Guttman-Kenney and Shahidinejad (2024) shows this is a biased measure of spending. It is preferable to calculate this at the tradeline level as doing so enables the researcher to account for changes in tradeline reporting which may erroneously affect aggregates. This measure can also be calculated from consumer-level aggregates data—including with non-consecutive periods though doing so will further reduce this measure's accuracy.

$$s_{c,t}^{GNW} = \Delta b_{c,t} = b_{c,t} - b_{c,t-1} \tag{3}$$

Our third measure of credit card spending ($s_{c,t}^{GN}$) is shown in Equation 4, as used in Ganong and Noel (2020), and is the first of our measures that removes revolving debt. This measure takes the changes in statement balances and adds payments ($p_{c,t}$). If the measure produces a negative number, it is bounded at zero. Making this adjustment both subtracts revolving debt appropriately and also includes spending that is repaid before the statement balance is issued, but this apporach contains some error as it includes financing charges. This measure also relies on the researcher being able to observe the actual payment amounts variable at the tradeline level. However, from 2015 to, at least, 2023 this actual payment amounts variable is only reported for a highly selected subset of credit card lenders and this subset excludes the six largest lenders (Guttman-Kenney and Shahidinejad, 2024). If using this measure, researchers need to restrict to only study the cards of furnishers who consistently report the actual payment amounts (e.g., Ganong and Noel (2020) exclude furnishers where over 90% of card months are zero or missing). In the future reporting of this variable may increase. We therefore recommend that researchers who want to use this measure should confirm the reporting coverage of the actual payment amounts variable for the time period they are planning to study before using or purchasing data.

$$s^{GN}_{c,t} = \begin{cases} b_{c,t} - b_{c,t-1} + p_{c,t} & \text{if } \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Our final measure of credit card spending ($s^{GKS}_{c,t}$) is shown in Equation 5. This measure is introduced in Guttman-Kenney and Shahidinejad (2024). This adapts $s^{GN}_{c,t}$ to remove estimated financing charges ($f_{c,t}$). All the caveats on the coverage of $p_{c,t}$ also apply to this measure. Financing charges are estimated following Guttman-Kenney and Shahidinejad (2024).

$$s^{GKS} = \begin{cases} b_{c,t} - b_{c,t-1} + p_{c,t} - f_{c,t} & \text{if } \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

It is also possible to produce estimates of spending using other methodologies. Researchers may create their own predictive models. Researchers may also purchase measures of spending calculated by the CRAs. A challenge of these CRA-created measures is researchers typically will not be told the algorithm used to create them. As CRA-created measures are commercially-sold products, CRAs can be sensitive to publishing quality assurance. Without such assurance it is difficult for readers to evaluate the accuracy of such CRA-created measures. Ultimately, unless both statement balances and actual payment amounts are consistently observed in the underlying tradeline data, models created by researchers or CRAs will struggle to accurately measure spending.

F Details on Credit Scoring

F.1 History of Credit Scoring

This section provides a brief history of credit scoring. The history of credit scoring is intertwined with the history of credit reporting, so interested readers wanting additional details should also consult Miller (2003); Barron and Staten (2003); Hunt (2005); Lauer (2017). Thomas (2009) and Thomas et al. (2017) provide a more detailed introduction to credit scoring, including its history.

The early history of credit scoring in the US is synonymous with FICO, the Fair Isaac Corporation. FICO was founded in 1956 by Bill Fair and Earl Isaac, and, by 1958, FICO produced their first credit scoring method to sell to lenders.[15] The passage of ECOA in 1974 helped stimulate the use of FICO's scoring method, as algorithmic approaches to underwriting aligned with ECOA's definition of a "credit scoring system" in which "all applicants [are treated] objectively...thus avoid problems of disparate treatment."[16] The use of FICO's scoring method grew in the 1970s and 1980s as CRAs and most national lenders started adopting FICO's products (Federal Reserve Board, 2007), facilitated by growth in information technology and credit card lending. It was not until 1981 that FICO created a credit score based on CRA, and in 1989 the first version of the what became the FICO score become available. By 1991, all three CRAs were using FICO scores.

Standardized credit scores became popular also in part because they enabled different loans to be compared and aggregated for securitization. In 1995, Freddie Mac recommended the use of FICO scores for all new mortgage applications, with Fannie Mae following shortly after (Lauer, 2017). By 2024, the regulatory requirements for lenders to use FICO led to Community Home Lenders of America to state that the "combination of FICO's extremely high market share, and the fact that Washington agencies require lenders to use this company's product, means that FICO has unilateral, solid-gold market power, the type rarely seen in any US industry short of highly regulated utilities, whereby rates are set by public-utility boards or commissions".[17]

In response to the rise of FICO, the three major CRAs created a competing joint venture: VantageScore. VantageScore was designed to apply an identical model across all the CRAs, so that the only reason an individual's score would differ across the CRAs would be due to differences in the underlying data on that person at each CRA. VantageScore

[15] https://time.com/3961676/history-credit-scores/

[16] See https://www.netinterest.co/p/monetising-an-algorithm and https://www.consumerfinance.gov/rules-policy/regulations/1002/interp-2/#2-p-Interp-2.

[17] https://www.thebignewsletter.com/p/inside-fico-and-the-credit-bureau

was also designed to extend greater coverage across the population, reducing the fraction of Americans where credit scores such as FICO cannot be calculated due to insufficient data. However, the greater coverage of VantageScore to thin-file consumers also means that the score is based on less information for these consumers.

The introduction of VantageScore interfered with FICO's monopoly and provided an outside option for lenders to use in their negotiations with FICO. In 2021, for example, Synchrony Bank switched from FICO to VantageScore.[18] FHFA also announced a new requirement to use both FICO and VantagScore for GSE mortgage securitization.[19] Anecdotally, VantageScore's entry into the market helped to keep prices down, although more research to establish whether this is the case, and if so how much by, would be valuable given the size of the economic markets served. Exactly which lenders use FICO and VantageScore and for what purposes is not well-established beyond marketing materials provided by individual lenders.

F.2 Proprietary Credit Scores

Sophisticated lenders typically create their own proprietary in-house credit scoring models. These proprietary are typically trained and tested on lenders' own datasets, which may include data not included in credit reports. Proprietary scores may use other credit scores, such as FICO or VantageScore, as one of their data inputs. While some of these in-house scores may share similarities with FICO and VantageScore, they are generally heterogeneous across lenders. Proprietary scores target outcomes specifically designed for their business needs, which may differ from FICO or VantageScore. Scores can be associated with deposit accounts, fraud detection, small businesses, alternative financial services, and internal scores used for account management by financial institutions based on private information on their own customers.

Einav et al. (2013) provide an example of how a large auto finance company's adoption of an automated in-house credit scoring increases their profitability, through improved screening and targeting. They also find the use of credit scoring crowded out "soft" information previously relied upon by dealership, a finding that relates to a broader organizational economics literature (Stein, 2002; Berger et al., 2005).

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., Bhutta et al., 2015; Agarwal et al., 2018; Gathergood

[18] https://fintechtakes.com/articles/2024-01-12/fico-score/
[19] https://www.fhfa.gov/policy/credit-scores

et al., 2019a; Argyle et al., 2023). If the researcher uses a different scoring model than the one used by the lender, those cutoffs will not align. However, Laufer and Paciorek (2022) have an innovative example for mapping a credit score used for lending decisions (FICO) to another credit score (Equifax Risk Score). Linking credit applications from data sources other than credit reports can be useful. For example, Bhutta et al. (2015) and Gathergood et al. (2019a) merge both successful and unsuccessful payday loan applications with credit scores used in lending decisions, enabling a regression discontinuity design to study the effects of payday loans on consumers' finances.

F.3 What Information Is Not In Credit Scores?

A credit scoring model is only as good as the data fed into it—-so factors such as errors, fraud and identity theft, incomplete coverage, and reporting lags that affect the credit reporting system also affect credit scores. Credit scores, such as FICO and VantageScore, only use as inputs information contained in credit reporting data, summarized in Table 1 of the main paper, with more detail in Online Appendix D.

What information is not in traditional credit scores, such as FICO and VantageScore? Credit scores are not based on income, education, or occupation. Similarly, credit scores do not have information on liquid and illiquid assets, other than through the existence of secured loans to finance durable purchases such as mortgages and auto loans. Likewise, information not present on traditional credit reports, such as usage of payday loans, subprime auto loans, and other alternative financial services and new or marginal forms of credit such as buy-now-pay-later / point of sale, and marketplace loans are also excluded from traditional credit scoring models. Information on deposit accounts, bank overdrafts, and related financial activity are also not reported to CRAs, and thus excluded from scoring models.

However, credit scores can be correlated with information that is not a data input of credit scores. Chatterjee et al. (2023) provide a theory of credit scores, where credit scores are "in part, the market's assessment of a person's unobservable type, which here we take to be patience." (Meier and Sprenger, 2012) show time preferences predict FICO credit scores, with more patient consumers having significantly higher credit scores. Arya et al. (2013) finds higher credit scores are correlated with lower impulsivity, greater patience, and greater trustworthiness but are not correlated with risk preferences.

F.4 Uses of Credit Scores

As credit scores have become more widely adopted, they are also used for a broader array of purposes including account management of existing portfolios and as a screening tool in non-credit markets such as rental, telecommunications, and insurance markets. Consumers also use them to learn about their own creditworthiness and to build and monitor their credit, and creditors and third-party providers give access to consumer credit scores as a way to build consumer loyalty and serve as a platform for advertisements.

The cost to one lender of gathering a FICO score for an applicant can be up to $60 in 2024, quadrupling over the prior two years.[20]

The types of information used in credit scores generate important and sometimes counter-intuitive economic implications for consumers. For example, because consumers are penalized for new hard credit inquiries, consumers experience a short-term decline in credit scores when applying for credit. Although scoring models allow consumers to make several credit applications within a short span of time without additional penalty (e.g., 14 to 45 days, depending on the specific product and score version used), consumers may still be penalized for search behavior in practice. Thus, the details of how the most common credit scoring models are constructed may generate frictions and can have important implications for consumer search and price dispersion. See Woodward and Hall (2012); Stango and Zinman (2016); Alexandrov and Koulayev (2018), and Argyle et al. (2023) for evidence of price dispersion and lack of search in consumer credit markets.

Consumers are sometimes said to need credit in order to build credit (e.g., Kovrijnykh et al., 2023). Because credit cards are the most common and often the first major form of credit used by consumers in the United States, the importance of credit cards in building credit may drive consumers to use credit products even without a liquidity need. More broadly, there is potential for "credit history hysteresis" that makes disadvantage persistent, via credit scores, for historically disadvantaged groups.

F.5 Different Types Of Credit Scores

Credit scores visible to consumers can be different from those used by lenders (Consumer Financial Protection Bureau, 2012a). Since the 2010s, consumers have increasing options to access "educational" scores to monitor and improve their own credit scores, offered by firms such as banks, credit card companies, and third-party platforms such as CreditKarma (Consumer Financial Protection Bureau, 2012a). Since 2005 all US consumers have been able to request a free credit report from each CRA each year (Kumar, 2022) as

[20] https://www.thebignewsletter.com/p/inside-fico-and-the-credit-bureau

a result of the Fair and Accurate Credit Transactions Act (FACTA) of 2003.[21] FACTA also required CRAs to provide credit scores directly to consumers for a reasonable fee.

FICO and VantageScore have many versions.[22] Both scoring models are updated over time and span different uses such as account management versus account origination; predicting any default versus default on a given new tradeline; predicting default for a given population of borrowers; and predicting default on a given type of trade such as credit cards versus auto loans. See Bergemann et al. (2018) for a theoretical explanation for different demands from data buyers for scores.

The outcome variable being predicted may also differ across different versions. The 24-month default rate predicted by a traditional model, may be different or calculated differently in different versions. For example, it could cover new versus all accounts, all trades versus specific types (e.g., auto, credit card), or other outcomes.

Over time, CRAs and other providers of credit scores update their scoring models to reflect changes in available data and in how the predictiveness of different data points changes over time. For example, among other, things, FICO 9 decreased the weight given to unpaid medical collections and assigned no weight to paid collections, while FICO 10 increased the weight placed on credit utilization. Newer scoring model versions (such as FICO 10T and VantageScore 4.0) include updates such as using trended data (when available) to incorporate information on changes in balances over time. Typically these models are unchanged within the same version number, but there are occasionally exceptions, such as when VantageScore updated their VantageScore 3.0 and 4.0 models to change the treatment of accounts reported in forbearance early in the pandemic.[23]

Any given model, such as FICO 9, may also produce different results when calculated based on the data from each of the three consumer reporting agencies. These differences can arise because each CRA includes slightly different data for each individual in the population based on its unique data-collection process and the network of furnishers that report to that CRA. It can also arise because each CRA can have different approaches to cleaning or aggregating data.

[21] https://www.pbs.org/wgbh/pages/frontline/shows/credit/more/scores.html

[22] https://www.capitalone.com/learn-grow/money-management/when-did-credit-scores-start/

[23] https://web.archive.org/web/20200602033654/https:/www.vantagescore.com/news-story/340/vantagescore-credit-scores-and-covid-19-pandemic

G Details on Accessing Credit Reporting Data

G.1 Established Consumer Credit Reporting Panels

Table 4 of the main paper lists the established US consumer credit reporting panels we are aware of. In this appendix subsection we provide additional details about these datasets. Researchers using such data should expect the CRAs to review outputs prior to them being released but should consult the access terms.

The most established dataset is the **Federal Reserve Bank of New York's Consumer Credit Panel**, using data from Equifax (FRBNY-CCP / Equifax). These data are accessible to researchers across the Federal Reserve system. Researchers outside the Federal Reserve system can co-author on research projects that use these data but generally are not able to access the underlying data (unless they have an employee status, such as with an internship). Lee and Van der Klaauw (2010) provides a comprehensive introduction to these data, with additional data dictionary and frequently asked questions documents online. These data are an anonymized 5% sample of consumers with a credit report in the US, based on the last two digits of SSN, and for these consumers, they also observe all consumers with the same address. These are quarterly data from Q1 1999 to the present. Each quarter, the Federal Reserve Bank of New York, releases a report on trends in these data. They also release public summary data at the national and state level.[24] A large body of research is released through the Federal Reserve blogs and working paper series, see (Haughwout et al., 2024) for an overview. Some key examples of research include: Albanesi et al. (2022),Athreya et al. (2019),Bhutta (2014),Bhutta and Keys (2016),Brown et al. (2016),Chakrabarti and Pattison (2019),Davis et al. (2021),Foote et al. (2021), and Mazumder and Miller (2016).

The **University of Chicago Booth School of Business**'s Consumer Credit Panel, housed at the Kilts Center for Marketing, uses data from TransUnion. Researchers outside the University of Chicago system can co-author on research projects that use these data but are not be able to access the underlying data. These data are an anonymized 10% sample of consumers with a credit report in the US. Data are monthly from July 2000 to the present. Data include the tradeline file. See footnote for the latest details on these data, terms of access, and links to papers using these data.[25] Examples of research using these data include: Gathergood et al. (2019b); Kluender et al. (2021); Blattner et al. (2022); Mian and Sufi (2022); Jansen et al. (2023); Keys et al. (2023); Yannelis and Zhang (2023); Cookson et al. (2024); Dinerstein et al. (2024); Granja and Nagel (2024); Guttman-Kenney (2024);

[24] https://www.newyorkfed.org/microeconomics/hhdc/background.html
[25] https://www.chicagobooth.edu/research/kilts/research-data/transunion

Guttman-Kenney et al. (2022); Guttman-Kenney and Shahidinejad (2024); Shahidinejad (2024). See Keys et al. (2023) and Dinerstein et al. (2024) for examples with public code. For researchers with access to this panel, Booth has is a non-public internal depository that contains data cleaning code.

The **Consumer Financial Protection Bureau**'s Consumer Credit Information Panel (CFPB-CCIP) uses data from one of the three nationwide consumer reporting agencies. This panel is only accessible to CFPB staff, although researchers outside the CFPB can co-author with CFPB staff but they are not able to directly access the data. The CFPB-CCIP is a 1:50 sample of deidentified credit records based on an internal ID beginning in 2002. Since 2014, the data are monthly. The data are tradeline-level, include information on coborrowers, alternative data (including payday and high-cost installment loans, and rental payments), and have been used as a sampling frame for multiple surveys (see, for example, Consumer Financial Protection Bureau (2017) and Fulford and Shupe (2021b)). The CFPB credit record data have been used for a variety of CFPB reports (Brevoort and Kambara (2017), Conkling and Gibbs, 2019, and Brennecke et al., 2021)[26] and independent research (e.g., Brevoort et al., 2015, Romeo and Sandler, 2021, Nelson, 2022, and Fulford and Nagypál, 2023).

The **University of California**'s Consumer Credit Panel (UC-CCP), at the California Policy Lab, uses data from one of the three consumer credit reporting agencies. These data are accessible to researchers affiliated with the University of California or the California Policy Lab. Un-affiliated researchers can co-author with affiliated researchers, but un-affiliated researchers are not be able to access the underlying data. As of 2024, researcher data access costs $6,929 per project, and linking data costs $12,981, with grants available to researchers. These data are an anonymized 2% sample of consumers with a credit report in the US. It also contains a 100% sample of Californians (with a credit report) who lived in Califonia at any point between 2004 and 2019, or who move to California after 2019. They also observe consumers who share an address or a tradeline with consumers in these samples. Data are quarterly from July 2004 to the present, with monthly data in 2020. Data include the tradeline file. A unique feature of these data is the ability to link them with administrative datasets, including those facilitated by the California Policy Lab. See footnote for the latest details on these data, terms of access, and links to papers using these data.[27] Examples of research using these data include: Flamang and Kancherla (2023); Liebersohn and Rothstein (2024); Papich (2023); Pinto and Steinbaum

[26]See https://www.consumerfinance.gov/data-research/research-hub/ for more examples

[27]https://www.capolicylab.org/data-resources/university-of-california-consumer-credit-pa

(2023).

University of Illinois at Urbana-Champaign Gies College of Business's Consumer and Small Business Credit Panel (GCCP), uses data from Experian. Researchers outside the University of Illinois at Urbana-Champaign can co-author on research projects that use these data but are not be able to access the underlying data. These are an anonymized 1% sample of consumers with a credit report in the US. Consumer data are annual from 2004 to the present, however, the data are the trended data product that enable a history of monthly tradeline data to be observed from each annual cross-section. These also contain alternative credit records ("Clarity"), containing detail of alternative sources of credit, such as payday loans, from 2012 onwards. A unique feature of this panel is they observe small business credit reports, annually from 2009 to 2022, and these are linked to entrepreneurs' consumer credit reports. Examples of research using these data include: Fonseca (2023); Howard and Shao (2023); Fonseca and Wang (2024); Fonseca and Liu (2024); Correia et al. (2024).

The Ohio State University's Consumer Credit Panel uses data from Experian. These data are an anonymized 1% sample of consumers with a credit report in the US. It also contains a 100% sample of Ohioans (with a credit report). Data are quarterly from 2017 to the present. See Moulton et al. (2023) for an example of research using these data.

The consumer credit panel at **Rice University Jones Graduate School of Business** uses data from Experian. These are an anonymized 1% sample of consumers with a credit report in the US, sampled based on the last two digits of SSN. Data are annual from 2004 to the present. Examples of research using these data include: Berger et al. (2018); Butler et al. (2023b,a); Mayer (2024); Xu (2023).

The **Georgia Institute of Technology Scheller College of Business**'s Consumer Credit Panel uses data from Equifax. These are an anonymized 1% sample of consumers with a credit report in the US, sampled based on the last two digits of SSN. Data are semi-annual from 2005 to 2008, and monthly thereafter to the present. Data include the tradeline file. Examples of research using these data include Chava et al. (2023a,b), and Zhang (2023b).

The **Urban Institute**'s consumer credit panel uses data from one of the three consumer reporting agencies. These are an anonymized 2% sample of consumers with a credit report in the US, sampled based on the last two digits of SSN. Data are annual from 2010 to the present. The Urban Institute releases public geographic data on its data catalog, on an ad hoc basis.[28] For examples of research using these data see Wei et al. (2016) and Braga

[28] For examples, see Financial Health and Wealth Dashboard 2022: `https://datacatalog.urban.org/dataset/financial-health-and-wealth-dashboard-2022` and its Debt in America: An Interactive Map: `https://apps.urban.org/features/debt-interactive-map/?type=overall&variable=totcoll`

et al. (2019).

Researchers have linked credit report data to public Home Mortgage Disclosure Act (HMDA) data themselves using data on mortgage characteristics (e.g., loan amount, loan origination date, geography, birth date)—see Bayer et al. (2016), Bartlett et al. (2022), Bhutta and Hizmo (2021), and Shahidinejad (2024) for examples—and researchers with access to more granular confidential HMDA can potentially do more precise merges (e.g., Bhutta and Canner, 2013). Recent richer mortgage datasets, such as the expanded HMDA data and the National Mortgage Database, enhance the value of linking credit reports to these.[29] A final benefit of these linked data is that they enable researchers using mortgage datasets to evaluate selection into their dataset compared to the more complete population of mortgages in credit reports.

In addition to the above panels, we are aware researchers at some other institutions purchase an off-the-shelf product, the Equifax Analytic Dataset. This contains monthly data (including the tradeline file) from 2005 to the present, for a 10% sample of US consumers with credit reports. For examples using these data see Cherry et al. (2021) and Piskorski and Seru (2021). We understand Experian and TransUnion can construct similar products.

G.2 Constructing Credit Panels

Researchers can encounter and construct credit reporting data in a variety of forms, including samples based on individuals or loans drawn directly from a CRA's database, as well as samples constructed via a match to a preexisting data source. In this section we briefly provide guidance on how to construct different types of data panels, how to merge credit data with other data sources, and how to run surveys off of credit data panels, with special attention to maintaining confidentiality and to where issues may arise if researchers do not account for the nature and structure of the credit record data. The specific requirements of a data agreement may vary, but the CRAs typically prohibit reidentification of consumers and require a right to review research before public circulation to ensure that researchers are properly using the data.

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.[30] As discussed in the main text, two of the most common ways to draw and main-

[29] The FHFA/CFPB's National Mortgage Database is based on Experian credit reporting data on mortgages linked to various administrative databases from Fannie Mae, Freddie Mac, Federal Housing Administration, the U.S. Department of Veterans Affairs, and the Rural Housing Authority. https://www.fhfa.gov/PolicyProgramsResearch/Programs/Pages/National-Mortgage-Database.aspx

[30] Alternatively, some researchers have relied on static panels, which follow a given set of birth cohorts

tain a nationally representative sample are to select the sample based on the last few digits of the social security numbers on the credit records or the internal ID assigned by the CRA. These result in similar but not identical panels.[31] The social security number method will miss records that do not have an SSN or similar ID number (Lee and Van der Klaauw, 2010). The internal ID method will include more fragment files, which researchers need to account for when constructing consumer-level measures as discussed below. Because credit records are regularly merged and split, using an external ID method like SSN will provide a representative sample of people (with that ID) while the internal ID method will provide a representative sample of records.[32] All these 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).

Cookson et al. (2024) find that approximately eighty percent of "consumers" without missing birth dates who appear in tradeline data at any point over 2000 to 2023 have SSNs. Examination of "consumers" who appear in credit reports but who lack SSNs suggests they are likely fragmented records—they typically have younger ages and credit reports that do not persist over time, suggesting that these "consumer" records were later consolidated with another credit record.

National estimates of various measures of consumer credit align well when comparing datasets using these two different approaches, but there can be larger differences in some areas, such as with third-party collections. Brown et al. (2015) find aggregate debt estimates from credit reporting data to line up well with estimates from the Survey of Consumer Finances (SCF). However, when distinguishing by loan type, they find considerable under-reporting of credit card debt in the SCF, a finding consistent with evidence presented by Zinman (2009) based on a comparison of credit card debt in the SCF with aggregate credit card debt estimates from the G.19 and call reports.

The CRAs suppress a small subset of records for use by researchers to comply with laws and internal guidelines, like excluding records for those under age 18. CRAs typically apply other filters to their data relevant for business purposes, such as only including tradelines with a recently reported update or excluding records considered "inactive"

of individuals or loan origination vintages. Representative static panels can be drawn using the same sampling approaches as applied for representative dynamic panels.

[31]The credit panels we know to be in existence at the time of writing were created using both sampling approaches.

[32]Other methods of drawing a panel are less common because they offer fewer advantages. For example, CRAs can also draw a sample by assigning random numbers to all records. Maintaining a dynamic representative sample can be difficult with this approach because credit record files are regularly merged or split as CRAs receive more information.

accounts. Researchers often have different purposes than other users of credit record data, and they may wish to confirm with the CRA if any filters have been applied and how they are defined. They may want to adjust these filters to their needs; for example, some researchers may want to exclude or include inquiry-only files.

Credit record panels almost always only include anonymized IDs for consumers (and possibly for 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), and each CRA has requirements for the types of flags they will provide and the minimum number of furnishers covered by such flags.

Researchers constructing panels may also want to consider the geographic unit to cover. Some panels such as the University of California's Consumer Credit Panel, include consumers in California and track if they move to other states. The CRAs include some areas other than US States (and DC), like US territories (e.g., Puerto Rico) and US Armed Forces bases, which may be of interest to some researchers, but other researchers may also want to exclude these. If a consumer moves overseas, their location and non-US debts are unobserved.

G.2.1 Household-Level Analysis

In constructing a panel, the population of records may include just a primary sample of records or may also include records for borrowers who have some sort of association with the primary sample. For example, both the FRBNY-CCP and the UC-CCP samples include credit records of individuals living at the same address, while the UC-CCP and CFPB-CCIP include credit records of associated borrowers, defined as borrowers who share a credit account with a primary sample borrower (joint, cosigned, or authorized user) even if they are not at the same address. These types of linkages permit computation of household-level debt aggregates, comparable to household-level debt measures from the SCF. To calculate aggregate individual and household-level statistics based on such expanded population samples requires applying appropriate sampling weights to avoid double-counting debts held by multiple people (see section H.2.3 and Lee and Van der Klaauw (2010)).

Constructing households or "decision making units" based on shared addresses or credit accounts can present problems. For example, some records have "generalized" addresses where only the main street address is captured for multi-unit dwellings without unit number, such as those living in a mobile home park, a college dorm, or military barrack. In those cases, the "household" constructed around the primary sample member contains both the valid household members and their neighbors and leads to the creation of unrealistically large (because they are actually multi-unit) households. Researchers can attempt to validate these cases by considering other information such as shared accounts, ages, and geographic history. In the other direction, relying on shared tradelines to construct "households" may miss household members who do not share credit accounts. Borrowers may also share accounts with people who are not part of the household and live elsewhere, but, again, researchers can rely on other information in the data (such as geography and age) to help address these cases.

An additional concern with drawing representative samples of households relates to the continued inclusion of records of deceased persons, as previously discussed. If a deceased person is sampled as a primary sample member and then a "household" is inferred based on all other individuals currently living at the deceased's former address, then the computed sampling weights can be invalid and this can produce biases in derived household-level aggregate statistics.

G.2.2 Data Frequency and Aggregation

If researchers are interested in studying credit reports *as the information appears to lenders* (e.g., to study how lenders respond to credit information), then reporting lags may not cause an issue. However, if researchers are interested in other aspects that require consistent timing (e.g., following an individual's credit accounts and debt over time), then they will need to create a time series incorporating information on the timing of furnishing updates to help remove noise in the data and reflect the timing of debt balances and performance.

While credit record data are typically updated monthly, researchers might also consider whether their project could potentially use less frequent data extracts. As previously noted, CRAs typically store their data as archives, or snapshots in time, so the various data elements can be measured at different times. But many of these data elements do not change over time or change infrequently. Some measures, such as the payment history of an account, include up to seven years of monthly history. As a result, researchers may be able to save money (or acquire more data) by obtaining credit record data at a lower frequency. For example, many ad hoc panels are constructed at an annual or biannual level (Butler et al., 2023a and Mezza and Sommer, 2016).

Researchers should also be aware that accounts in dispute are suppressed by CRAs during the investigation process, so researchers may need to contend with missing observations (e.g., fill in using preceding month if the account reappears with a reference to a prior dispute).

H Code

H.1 General Practical Guidance

Credit reporting data are large datasets. Here we provide useful practical advice for researchers working with these large datasets.

- First, reduce the size of variables you are working with. A researcher's dataset may contain anonymized identifiers (e.g., for individual consumers, tradeline accounts, or furnishers) created by the CRA that are very long alphanumeric strings. Creating a lookup file mapping these to short numeric versions, and using these more concise identifiers can substantially reduce the size of these datasets a researcher is loading and working with.

- Second, reduce the number of variables you are working with. Credit reporting data contain large number of variables. For example, the CRA aggregated datasets often contain hundreds, or sometimes thousands of variables. Often only a handful of these are used by researchers so only load these (or save a subset of these data). For the tradeline data, some variables may be long strings, such as the 84-month array variable, which may be dropped if not using.

- Third, save the raw data in an efficient format for load it. If you have access to a very large credit reporting dataset (e.g., tradeline-level, large sample of consumers) it may be efficient to save the raw data in parquet files. These can often be queried quicker than csv, or other formats for initial processing.

H.2 Code for Common Tasks

Here we describe several common tasks researchers perform with credit reporting data and offer some overarching guidance to approaching these tasks, code snippets, or references to existing code repositories from published papers. Unless otherwise specified, these code examples rely on tradeline-level data.

We do not include code for loading and cleaning data. We instead refer readers to the following examples of public code depositories that contain complete code from loading raw data to conducting analysis. Ganong and Noel (2020) and Keys et al. (2023) use tradeline-level TransUnion data. Bhutta and Keys (2016) and Laufer and Paciorek (2022) use FRBNY-CCP Equifax data. Gross et al. (2020) use tradeline-level CFPB-CCP data. Beshears et al. (2022) use consumer-level Experian data. Beraja et al. (2019) and Berger et al. (2021) use CRISM data.

H.2.1 Variable Names

In the code we provide, we have standardized the variable names and noted whether the unit of observation is at the tradeline- or consumer-level. The exact variable names will differ across credit reporting datasets and, therefore, will differ in public code depositories accompanying papers. A public data dictionary for Equifax Analytic Dataset is available at the time of writing.[33]

[33] https://aws.amazon.com/marketplace/pp/prodview-vgmxklm42lhmq#dataSets

Table H1: Variable Names

Variable Name	Variable Description	Variable Unit of Observation
date	Archive furnishing date	Consumer/Tradeline
personid	Anonymous consumer identifier	Consumer
hhid	Anonymous household identifier	Consumer
state	State of consumer's primary residence	Consumer
address	Anonymous address identifier	Consumer
loanid	Anonymous account identifier	Tradeline
servicerid	Anonymous identifier for a tradeline's servicer	Tradeline
ecoa	Sole, joint, or other user of tradeline	Tradeline
account_type	Tradeline account type	Tradeline
balance	Outstanding balance	Tradeline
status	Delinquency status or manner of payment (MOP)	Tradeline
schpayment	Scheduled payment amount	Tradeline
actpayment	Actual payment amount	Tradeline
payment_history	Array of status history	Tradeline
balance_date	Date corresponding to balance	Tradeline
open_date	Account opening date	Tradeline
origination_amount	Account origination amount	Tradeline
terms_frequency	Account frequency	Tradeline
terms_duration	Account term duration	Tradeline
special_comment_code	Narrative codes from Metro 2	Tradeline
inquiry_date	Date of inquiry	Inquiry

H.2.2 Joint Account Adjustment

Joint account information is typically contained in an ECOA variable. One can create a ″wgt″ variable of 0.5 if joint accounts, and 1 otherwise, and use it as a weight for calculating average or aggregate statistics.

This Stata example is for data that do not include authorized users.

```
gen wgt = 1
replace wgt = .5 if inlist(ECOA, "C", "M", "J", "S")
* C = comaker, M = maker, J = joint, S = shared

* aggregation example
table state [iw=wgt], stat(sum balance) stat(mean balance)
```

This SQL code example is more general and applies to data which include authorized users.

```
-- Using Metro 2 codes:
CASE
    -- individual:
    WHEN ecoa_code IN ('1') THEN 1
    -- authorized user, deceased:
    WHEN ecoa_code IN ('3','X') THEN 0
    -- joint contractual liability, co-maker, maker:
    WHEN ecoa_code IN ('2','5','6','7') THEN 0.5
END AS wgt
```

H.2.3 Household Weights

Based on the FRBNY-CCP, primary samples are selected based on certain combinations of social security numbers representing 5% of population with social security numbers, and all those sharing the same addresses are selected as household samples. As a result, 1-person household member is selected with 5% probability, and 2-person household is selected with $1 - 0.95^2$, and households with N members are selected with probability $1 - 0.95^N$. As N approaches to infinity, the probability of the households being selected approaches 1.

This example uses Stata code and a 5% sampling rate.

```
egen N_hh = sum(1), by(hhid)
```

```
* number of household members in hhid
gen hh_wgt = 1 - 0.95^N_hh

* example of aggregation
table state [iw=hh_wgt], stat(sum balance) stat(mean balance)
```

H.2.4 Population Counts with Credit Reports

It is estimated that roughly 1 in 10 US adults do not have a credit report (Brevoort et al., 2015). This share is calculated by estimating the number of adults *with* a credit report, dividing by Census/ACS population counts, and subtracting from 1. While seemingly straightforward, measuring this share requires taking a stance on which credit records are fragment files. By definition, a fragment file is a non-unique credit record for a given individual (see section B), so fragment files should be excluded when computing the count of individuals with a credit report. Common approaches to removing fragment files include removing records that do not persist for at least four years (or some other threshold), inquiry-only credit records, collection-only credit records, public-record-only credit records, and/or credit records with missing consumer age or birthyear. The choice of which approach is used to drop fragment records can change by 10 million or more the implied count of consumers without a credit record (see Appendix A Table 1 in Brevoort et al. (2015)). An example of code to calculate these ratios can be found in 03_count_with_trade.py and 03_sumstats.do in the replication package to Keys et al. (2023).[34]

H.2.5 Mobility

Mobility can be measured by a change in a consumer's address in credit reporting data. Often researchers will not observe the exact address of a consumer in their credit reporting data. The replication package of Keys et al. (2023) identifies and analyzes movers across coarsened geographies such as commuting zones (02_mover.py, main function defined in lines 128 to 178).[35]

If a researcher observes a consumer's (anonymized) address, they can use the following Stata code that examines the address history of a person and flags it as a move only

[34] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/G85KDR

[35] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/G85KDR

when the address is new in the entire history of the person.

```
sort personid address time
by personid address: gen move= (_n==1)
* first appearance of new address is flagged as a move.
```

Another example of similar Stata code for measuring mobility at different frequencies can be found in 2_clean_efx_moves.do in the replication package to Abel and Fuster (2021).[36]

Readers measuring geographic mobility in credit report data should also review the caveats discussed in Section 3.7 of the main text.

It is sometimes important for researchers to exclude real-estate investors who hold multiple properties. These individuals' primary residence often cannot be well-established in credit reports. This makes it challenging to assess moves, and also whether new mortgages are for new properties or for refinancing. The replication package of Bhutta and Keys (2016) contains an example for identifying these individuals (see lines 100 to 109 of dofiles/datasets/create_extraction_dataset.do).[37] In their paper, "A borrower is classified as an investor if (i) he has three or more mortgage accounts; (ii) he has exactly two closed-end mortgages where the smaller loan is at least one-third the size of the larger; or (iii) he has two or more HELOCs, or a HELOC with a line size that is more than 50 percent the size of his only closed-end loan: 3+ accounts; 2 closed end accounts of "similar" size; large HELOC relative to closed; or two HELOCs."

Mian and Sufi (2022) construct three measures of housing speculators, which they show to be highly correlated. They write:

- "First, a mortgage origination is classified as being taken out by a speculator if the individual taking out the mortgage in question also takes out another distinct first-lien purchase mortgage in a 2-year period around the origination in question. We refer to this as the "multiple houses" categorization of a speculator.

- "Second, a given first-lien purchase origination is classified as being taken out by a speculator if the first-lien purchase mortgage is subsequently closed within a year, and there is no associated refinancing for the individual in the six months after the purchase mortgage is closed. We refer to such an individual as a "short-term" trader, where we are making the assumption that the closed mortgage reflects a sale.

[36] https://www.openicpsr.org/openicpsr/project/116461/version/V1/view
[37] https://www.openicpsr.org/openicpsr/project/116153/version/V1/view

- "Third, a given first-lien purchase origination is classified as being taken out as a speculator if the individual taking out the mortgage already has at least two existing first-lien mortgages on his balance sheet at the time of the new origination. We refer to such an individual as a "2+ mortgage" speculator."

H.2.6 Merging CCP Mortgages with Other Mortgage Datasets

With a wide array of non-CCP mortgage datasets available, researchers sometimes merge CCP and non-CCP mortgage data at the loan level. A challenge is that details such as loan balance or loan date may vary slightly between datasets depending on when the loan was recorded or which aspects of the loan transaction were recorded in each dataset. To address this, a common approach is to use a "fuzzy" merge that allows close, but not exact, matches, and then to select the closest match based on various criteria available in both datasets. While particulars will vary depending on the two datasets in question, an example of this strategy is in lines 47–74 of empirics/CRISMcleaning/re4_match_efx_lps.do from the replication package to Beraja et al. (2019).[38] The replication package of Berger et al. (2021) use a similar, slightly updated approach, see empirical_code/CRISM_Data_Processing/4_match_efx_lps.do, lines 52–164.[39]. Similarly, see 4_match_efx_lps.do in the replication package of Abel and Fuster (2021).[40]

H.2.7 Mortgage Purchases and Refinances

We first identify a new origination by observing the first observation of a loan. A new origination is either a purchase origination or a refinance origination. We flag refinance originations first and code the rest as purchase originations. To be classified as a refinance origination, we check 1) if there is a preceding prepaid/closed mortgage loan not too long before, such as within 12 months, and 2) if the address of the person at the time of the origination is a new one for the person in the entire history.

```
* stata code for FRBNY CCP

drop if balance == 0
* to drop the trailing zero balances after pay off
```

[38] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GETNJK
[39] https://www.openicpsr.org/openicpsr/project/134161/version/V1/view
[40] https://www.openicpsr.org/openicpsr/project/116461/version/V1/view

```
* create new address indicator
sort personid address date
by personid address: gen new_address==1 if _n==1

* create indicators for first and last observations of a loan

sort personid loanid date
by personid loanid: g startofloan=1 if _n==1
by personid loanid: g endofloan=1 if _n==_N


* now flag refinance
sort personid date
by personid: gen purpose="refinance" if startofloan==1
& endofloan[_n-1]==1 & date<date[_n-1]+12 & new_address==0

* impose condition of refinance such that there should be
a preceding loan within 12 month before the startofloan,
and the address should be a new one in the history of the person

replace purpose="purchase" if purpose=="" & startofloan==1
```

If using the above approach, it is important to note that loans are typically censored at the beginning of sample periods, so a researcher should modify the purchase/refinance indicators that existed from the beginning of sample periods.

The code below is adapted from Mian and Sufi (2022) for their tightest definition of mortgage refinancing. We thank the authors for allowing us to publish this code. If using the code below, please cite the source as Mian and Sufi (2022). In this code, numbmor is a consumer-level variable recording the number of mortgages outstanding, and censgeocode is the census tract of the consumer. These two variables are measured at different time periods denoted by the suffix (l12, l6, l3, l1, f1, f3, f6, f12) where l12 is twelve months prior to the date of mortgage origination and f12 is 12 months after the date of mortgage origination.

```
gen refi=1 if numbmor\_l1==numbmor\_f6
& censgeocode\_l1==censgeocode\_f6
replace refi=1 if numbmor\_l1==numbmor\_f3
& censgeocode\_l1==censgeocode\_f3
replace refi=1 if numbmor\_l1>numbmor\_f6
& censgeocode\_l1==censgeocode\_f6
replace refi=1 if numbmor\_l1==0 & numbmor\_f6==1
& censgeocode\_l1==censgeocode\_f6
replace refi=. if censgeocode\_l1==.
```

The replication package of Beraja et al. (2019) contains code (empirics/CRISMcleaning/5_link_new_lps_loans.do) for identifying refinanced loans in CRISM data.[41] They write in their Online Appendix:

- "We thus use the following rules to identify refinances. We start by looking for any loan in the Equifax data set that has an open date within 4 months of the McDash loan's termination date. We find at least one such loan for about 81% of the voluntary terminations in 2008 and 2009. We classify these new loans as a refinance if either:
  - The loan also appears in McDash and is tagged as a refinance in the purpose-type variable (61% of the McDash-matched loans).
  - The loan also appears in McDash and is tagged as an "Unknown" or "Other" purpose type, and has the same property zip code as the original loan.
  - The loan appears only in Equifax but the borrower's Equifax address does not change in the 6 months following the termination of the original loan."

A similar, slightly updated approach is taken in Berger et al. (2021)'s replication package code (empirical_Data_Processing/5_link_new_lps_loans.do).[42] They write in their online appendix:

- "As in our primary analysis, we begin with all remaining outstanding fixed rate first liens in the McDash which are voluntarily paid off. We then look for any loan in the Equifax data set that has an open date within 4 months of the McDash loan's termination date. We classify these new loans as a refinance if either:

[41] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GETNJK

[42] https://www.openicpsr.org/openicpsr/project/134161/version/V1/view

- “The loan also appears in McDash and is tagged as a refinance in the purpose-type variable.
- “The loan also appears in McDash and is tagged as an ”Unknown” or ”Other” purpose type, and has the same property 5 digit zip code (where available, or 3-digit zip code and MSA-div where not) as the original loan.
- “The loan appears only in Equifax but the borrower’s Equifax address does not change in the 6 months following the termination of the original loan.”

H.2.8 Identifying First and Second Liens Mortgages

For mortgage accounts, identifying first and second liens usually come from account type and narrative codes. HELOCs have account type of ”R”, revolving, along within mortgages, and first and second lien installment mortgages can be derived from narrative codes. Mortgage security types such as Fannie, Freddie, FHA, VA can be classified as first mortgages, and accounts with narrative codes of ”second mortgage,” ”home equity loan,” and ”home improvement loan” can be classified as second liens. Among those that are not classified before, one can use the origination amount as a proxy for first lien vs. second lien. New York Fed uses a threshold of $40,000 to draw a distinction.

```
variables: narrative = narrative code
account_type: "R" if revolving, "I" if installment

gen mortgage_type ="heloc" if account_type =="R"
* "R" = revolving accounts

replace mortgage_type="second lien" if account_type=="I"
& inlist(narrative, "home equity loan", "home improvement loan",
"second mortgage")

replace mortgage_type="first lien" if account_type=="I"
& (narrative=="Fannie Mae" | narrative=="Freddie Mac"
| narrative=="FHA" | narrative =="VA" )
```

Alternatively, Mian and Sufi (2022) use a threshold of less than 30% of CLTV to designate which loans are a second lien, as in the following block of Stata code which takes as input a dataset of (only) mortgage tradelines.[43]

```
egen currtotal=sum(balance), by(personid date)
gen frac=balance/currtotal
gen second_lien=(frac<=0.3)
```

For an example of code identifying second lien balances based on CRISM matched data see 2_second_lien_balances.do in the replication package to Abel and Fuster (2021).[44]

The replication package of Beraja et al. (2019) contains code (empirics/CRISMcleaning/5_piggybackseconds.do) for identifying piggyback second liens.[45] They define a piggyback second lien as one that: "(1) Has the same open month in Equifax within three months of the matched loan's Equifax open month. (2) Has an origination balance of less than 125% of the LPS loan's origination balance if it's a CES or HELOC, OR (3) Has an origination balance of less than 25% of the LPS loan's origination balance if it's a first mortgage." A similar approach is taken in Berger et al. (2021)'s replication package code (empirical_Data_Processing/5_piggybackseconds.do).[46] They define a piggyback second lien as one that: "(1) Has the same open month in Equifax within three months of the matched loan's Equifax open month, (2) Has an origination balance of less than 125% of the LPS loan's origination balance (if it's a CES or HELOC), OR (3) Has an origination balance of less than 25% of the LPS loan's origination balance (if it's a first mortgage)."

H.2.9 Mortgage Cash Out Refinance

Cash out amount from mortgage refinance is measured by taking the difference between the mortgage loan origination amount, for mortgages tagged as refinanced (as using code in prior sections), to the outstanding mortgage balance in the prior month. The Stata code below provides a simple example for calculating this, however, we refer readers to replication packages for more refined approaches.

```
sort personid date
```

[43] We are grateful to Amir Sufi for sharing this code, which we have adapted to follow the naming convention in Table H1.

[44] https://www.openicpsr.org/openicpsr/project/116461/version/V1/view

[45] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GETNJK

[46] https://www.openicpsr.org/openicpsr/project/134161/version/V1/view

```
by personid: gen cashout = origination_amount - balance[_n-1]
if purpose=="refinance"
* purpose as defined in code in earlier section
```

An example of similar Stata code for measuring equity extraction can be found in create_extraction_dataset.do (especially lines 111 to 113) in the replication package to Bhutta and Keys (2016).[47]

The replication package of Beraja et al. (2019) contains code (empirics/CRISMcleaning/6_cashout_panel.do) for identifying cash-out refinancing in CRISM data.[48] A similar, slightly updated approach is taken in Berger et al. (2021)'s replication package code (empirical_Data_Processing/6_cashout_panel.do).[49] Both of these papers contain excellent Online Appendix documentation describing their methods in detail.

H.2.10 Cost of Borrowing

While interest rates are not directly reported in credit bureau data, the costs of borrowing can be estimated by researchers from tradeline-level data, or purchased from products the credit reporting agencies sell. For fixed-rate installment loans, such as auto loans and unsecured personal loans, once a researcher observes the principal origination amount ($origination\_amount$), origination term duration ($terms\_duration$), and scheduled monthly payment amount ($schpayment$), they can calculate the interest rate (i) at origination using a root-solver shown in Equation 6. For cases where $origination\_amount \geq schpayment \times terms\_duration$, loans are assumed to have zero-percent interest rates. Researchers may also wish to top-code unreasonably high interest rates. See (Yannelis and Zhang, 2023) for an example of this methodology.

If estimating interest rates, the interest rate's n-th digit is unlikely to be of use and may provide false precision. We therefore recommend rounding the interest rate to two-to-four decimal places.

$$schpayment = \frac{origination\_amount \times i}{1-(1+i)^{-terms\_duration}} \quad \text{if} \quad origination\_amount < schpayment \times terms\_duration \tag{6}$$

The cost of borrowing on mortgages at origination can be calculated using a similar methodology, as developed by Shahidinejad (2024). Equation 7 shows the root-solver

[47] https://www.openicpsr.org/openicpsr/project/116153/version/V1/view

[48] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GETNJK

[49] https://www.openicpsr.org/openicpsr/project/134161/version/V1/view

equation to calculate the interest rate at origination (i), using data from the first few months (e.g., months two to seven) of the loan. This uses the outstanding balance ($balance$) at two points in time j and k, and origination term duration ($terms\_duration$). Importantly, this methodology accounts for the fact that the scheduled payment variable observed in credit reporting data typically include taxes, insurance, or homeowner association (HOA) fees, as well as the interest and principal payment, which would mean estimates of borrowing costs produced using Equation 6 would often be biased. As Shahidinejad (2024) is a working paper, the exact methodology to estimate borrowing costs may develop over time, and we encourage readers to examine updated version of this paper to locate code.

$$\frac{balance_j}{balance_k} = \frac{(1+i)^{terms\_duration-j} - 1}{(1+i)^{terms\_duration-j} - (1+i)^{k-j}} \tag{7}$$

Across auto loans, mortgages, and other installment loans, if a researcher is interested in the realized effective interest rate, to capture costs changing post-origination, researchers can use these same methodologies using multiple observations after origination. For one example, see Conkling and Gibbs (2019).

For computational reasons, it may be more efficient to calculate rates by searching over a grid restricted to reasonable range (e.g., 0% to 50%) and decimal place increments, instead of using a root-solver. This is especially likely to apply if researchers are calculating realized interest rates where substantially more data is required compared to calculating interest rates at origination.

Without using a root-solver, we also provide Stata code below to estimate realized effective interest rates. This will not be accurate for many mortgages, for the reasons discussed above as the scheduled payment variable includes non-interest and non-principal components. In addition, it also requires an assumption that the scheduled payment is equal the actual payments made. This is often not the case as consumers often prepay loans.

```
sort loanid date
by loanid: gen cost =  schpayment[_n-1] - (balance[_n-1] - balance)
by loanid: gen rate = cost / balance[_n-1]
```

Separately from installment loans, Guttman-Kenney and Shahidinejad (2024) develop a novel methodology for estimating financing charges on credit cards. This methodology's key insight is that researchers can use an institutional detail of how credit card minimum payment formulae, to infer borrowing costs. Credit cards' minimum payment follow a deterministic formula where the minimum payment is the maximum of (i) $\$\mu$ amount (e.g., \$10, \$25), and (ii) $\theta\%$ statement balance, plus interest and fees. This formula structure means that if a researcher can work out a credit card's μ and θ, a researcher can then work out, for each card, each month the minimum payment before financing charges (sum of interest and fees), and compare this to the observed minimum payment in credit reporting data (the scheduled payment amount), to estimate the costs of borrowing. The terms μ and θ can be deduced by researchers using tradeline data to examine the relationships between statement balances and scheduled payment amounts. Researchers can more accurately deduce these terms with more granular credit reporting data available (e.g., actual payment amounts, anonymized furnisher identifiers). As Guttman-Kenney and Shahidinejad (2024) is a working paper, the exact methodology may develop over time, and we encourage readers to examine updated version of this paper to locate code.

H.2.11 Flow Delinquency

The flow of new accounts into delinquency can be measured by comparing the number of accounts in stages of delinquency over time. Delinquency information is contained in the "status" variable.

```
sort personid account\_type date
by personid account\_type: gen flow_delinquent = delinquent[_n-1]==0
& status !="1" if _n>1
```

H.2.12 Linking Tradelines Across Transfers

Users may have CRA-supplied tradeline identifers, but these IDs typically change when the tradeline is transferred between servicers or when a credit card replaces a lost/stolen card. Usually, the account opening dates carry over enabling account linking. The Stata code below creates a new variable loanid2 that links loanid over transfers. Account transfers are distinguished from mortgage refinances since the latter is a new origination and with new account opening date.

```
gen loanid2=loanid
sort personid origination_date date
by personid origination_date: replace loanid2 = loanid2[1]
```

Note: The data can be more complicated than this simple example. There may be multiple loans with the same origination date in each date t, in which case a researcher may use additional information such as ”credit limit” or ”origination amount” to make the loans more unique to separate among multiple such loans. For example,

```
gen loanid2 = loanid
sort personid origination_date origination_amount date
by personid origination_date origination_amount:
replace loanid2 = loanid2[1]
```

Below we provide some more general guidance to link the same tradeline over time for a variety of credit products (assuming the origination date for the trade does not change), but users may wish to do something slightly different for specific products or contexts:

- For installment loans: group based on loan product, person identifier, open date, loan amount
  - For student loans still in deferment, users may need to allow the loan amount to change within the first 6-9 months after origination (many of these loans have a second disbursement the semester after origination).
  - Installment loans with the same new ID in the same time period and the same repeated balance dates (and/or different, non-zero dollar balance amounts) are likely separate loans that should have different IDs. This is especially common with student loans where a borrower may have multiple loans open on the same day.
- For revolving accounts: group based on loan product, person identifier, open date
  - This ignores credit limit because limits can change over time.
  - Revolving accounts with the same new ID in the same time period but with different credit limits are likely different accounts.
  - Revolving accounts with the same new ID in the same time period but and the same credit limits but different balances, scheduled payment, or actual payment amounts are likely difference accounts.

- In all cases: if the CRA-supplied tradeline identifier match for conflicting IDs based on the above groupings, reassign the new ID so it matches. For example, if tradelines A and C share a new ID based on the above and tradelines A and B share a CRA-supplied tradeline identifier, tradelines A, B, and C should all have the same new ID.

H.2.13 Identifying Direct Student Loans versus FFELP Student Loans

If the student servicer IDs are available in the data, a researcher can assign student loans servicers between Direct Loans and Federal Family Education Loans (FFELP) among federal student loans using the COVID-19-related administrative student loan forbearance that only affected Direct and federally-held FFELP loans. FFELP loans are private student loans guaranteed by federal government and those that were still held by commercial lenders were not subject to the same policy. This Stata code assumes that each servicer is serving exclusive direct loans or FFELP loans only; if a servicer ID services both of them under one ID, it will create a problem.

```
egen test1 = sum((payment ==0)*(t=="June 2020")*(balance>0)),
by(servicerid)
* number of accounts with payment 0 in June 2020

egen test2 = sum((t=="June 2020")*(balance>0)) , by(servicerid)
* number of accounts with positive balance in June 2020

g test3 = test1 / test2
* share of payment=0 accounts among those with positive balance
in June 2020 by servicers

g direct = (test3>0.99)
* If zero payment share is 99%, it's a direct loan servicer.
One needs to to check the distribution of test3 to determine
the right threshold.
```

Note: this will separate Direct loan servicers from FFELP / private loans servicers, but it will not distinguish between FFELP servicers and private student loan servicers. To distinguish between the FFELP and private student servicers, users needs to check other information such as timing of originations (FFELP loans stopped originating new loans in 2010), origination amounts, joint account behavior, credit scores, etc. Some servicers might service both types of loans under the same servicer ID, in which case the distinction at the servicer level is not possible.

Additionally, the code for Dinerstein et al. (2024) presents a slightly different approach in 02_IdentifyLoanType.py and 05_AlternativeClassification.py.[50]

H.2.14 Delinquencies from payment history

When researchers want to construct the monthly payment status history for a tradeline without monthly data, they can pull this information from the payment history if it is included in their data. This variable is typically a grid showing of monthly payment history with the most recent month in the leftmost position based on Metro2 guidance. The payment history takes values corresponding to 30-day increments in delinquency and has additional values for other statuses, such as collections and charge-offs.

To align the payment history information correctly, users must also use the balance date for the same data archive. See below for an example in Stata transforming the data into a long dataset with the last 12 months of payment history.

```
forvalues i = 0/11 {
    gen payment_status`i' = substr(payment_history,`i'+1,1)
}
reshape long payment_status, i(tradeline_id) j(payment_month)
replace payment_month = mofd(balance_date) - payment_month
format payment_month %tmMonCCYY
```

For another example calculating delinquencies in the any month in the prior quarter, see starting at line 561 in process_bk_sample.do in the code repository for Gross et al. (2020).[51]

H.2.15 Missing Loans

Occasionally loans disappear from credit record data for several months due to reporting issues or transfers between furnishers. Sometimes these gaps are long and sizeable, such as with student loans from December 2011 to June 2012 and in 2023. To impute missing

[50] See https://www.openicpsr.org/openicpsr/project/193167/version/V1/view
[51] See https://www.openicpsr.org/openicpsr/project/115211/version/V1/view

loan balances for the former period with consumer-level data, Beshears et al. (2022) use linear interpolation starting at line 1148 in ae_sample_compile_all.do.[52]

Note that, especially when accounting for balances of new originations, users may want to use seasonal patterns from other years rather than a linear trend. To impute missing delinquency statuses during these periods, users can refer to the payment grid after the tradeline reappears as described in H.2.14.

H.2.16 Forbearances

Accounts are placed into forbearance (or deferment) when borrowers are not required to make payments. This is most common for student loans, but it can also happen with mortgages and other types of credit, especially during natural disasters or other major events such as the COVID-19 pandemic. There is a special comment code ("CP") and loan term frequency ("D") that identify many of these forbearances. Sometimes furnishers do not report these deferment codes and users must infer the forbearance when a tradeline with a positive balance also has a $0 scheduled payment, however, when doing so they need to check whether a furnisher is reporting the scheduled payment variable across its portfolio, otherwise it may erroneously interpret a $0 as a deferment when it is not.

```
gen forbearance = (balance_amount > 0 & payment_amount == 0)
replace forbearance = 1 if special_comment_code == "CP"
    | terms_frequency == "D"
```

While deferments and forbearances are different things for federal student loans, they are not reliably reported over time and across furnishers. As a result, it is difficult to separately categorize deferments and forbearances in the credit record data.

We also refer interested readers to Cherry et al. (2021); Dinerstein et al. (2024) as these studies of COVID-19 accommodations include deferments and forbearance.

H.2.17 Aggregating Balances Across Loan Types

Consumer-level aggregated CRA data include numerous – sometimes hundreds – of variables reporting the sum of a consumer's outstanding debt across all loans in various categories. When using these aggregates to create other aggregate measures, care should be taken that the variables used have applied comparable filters to which underlying tradelines are included: only loans or also non-loan tradelines such as medical collections; all

[52] https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fjofi.13069&file=jofi13069-sup-0002-ReplicationCode.zip

loans or only currently open ones, as loans are sometimes closed but have a nonzero balance; all loans or only "verified" ones, which have been recently reported by a furnisher (e.g., in the last 90 days). Care should also be taken not to add to variables where one is a superset of the other. For example, a measure of a consumer's total "installment" debt may already include student loans, mortgage loans, and auto loans (each of which is sometimes referred to as an installment loan, because the consumer pays in periodic installments).

For an example of such aggregation exercises, see lines 1230–1349 of ae_sample_compile_all.do from the replication package for Beshears et al. (2022).[53] Also see the replication package code 03_ConstructPanel.py in Dinerstein et al. (2024).[54]

H.2.18 Consumption: Automobile Purchases and Credit Card Spending

Two types of consumption that can be usefully studied in credit report data are automobile purchases and credit card spending. These are further described in Section 3.6.1 and 3.6.2. Helpful examples of code to compute these two consumption measures can be found in 11_make_tu_creditcard_file.R and 12_make_tu_auto_file.R in the replication package to Ganong and Noel (2020).[55]

H.2.19 Determining Inquiry Success

When users have data with non-aggregated inquiries that include the date of the inquiry, they can match those inquiries to new tradeline openings to calculate inquiry success (or credit tightness). Users must first define the time period after an inquiry happens that a new account must open by. Common search windows are 7 days for credit cards, 14 days for auto loans, 120 days for mortgages, and 30 days for all other types of credit.

Users should keep in mind that they are unlikely to observe all inquiries that might be part of a consumer's search window (except for mortgages prior to 2024) because the inquiries corresponding to an application may have gone to a CRA other than the researcher's data source. To partially account for this, it is helpful to collapse down to the consumer-search window level. Below we provide an example in SQL.

```
-- Define relevant search window by credit product type
CASE WHEN loan_product = 'credit_card' THEN 7
```

[53] https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fjofi.13069&file=jofi13069-sup-0002-ReplicationCode.zip

[54] https://www.openicpsr.org/openicpsr/project/193167/version/V1/view

[55] https://www.openicpsr.org/openicpsr/project/118401/version/V1/view

```
     WHEN loan_product = 'auto_loan' THEN 14
     WHEN loan_product = 'mortgage' THEN 120
     ELSE 30
END AS search_window
-- Join inqs w/trades based on inquiry dates and opening dates
matched AS (
SELECT * FROM inquiries AS a
    LEFT JOIN trades AS b
    WHERE a.person_id = b.person_id
    AND datediff(day,a.inquiry_date,b.open_date) <= a.search_window)
-- Collapse to person-time period level (time_period)
SELECT *,
    CASE WHEN MAX(open_date) IS NOT NULL THEN 1
        ELSE 0
    END AS inquiry_success
    FROM matched
    GROUP BY person_id, loan_product, time_period
```

I References

Aaronson, D., Agarwal, S., and French, E. (2012). The spending and debt response to minimum wage hikes. *American Economic Review*, 102(7):3111–3139.

Abel, J. and Fuster, A. (2021). How do mortgage refinances affect debt, default, and spending? evidence from harp. *American Economic Journal: Macroeconomics*, 13(2):254–91.

Acemoglu, D., Makhdoumi, A., Malekian, A., and Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, 14(4):218–256.

Acquisti, A., Taylor, C., and Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, 54(2):442–492.

Adams, P., Guttman-Kenney, B., Hayes, L., Hunt, S., Laibson, D., and Stewart, N. (2022). Do nudges reduce borrowing and consumer confusion in the credit card market? *Economica*, 89(S1):S178–S199.

Adelino, M., McCartney, W. B., and Schoar, A. (2020). The role of government and private institutions in credit cycles in the us mortgage market. *NBER Working Paper No. 27499*.

Agarwal, S., Amromin, G., Chomsisengphet, S., Landvoigt, T., Piskorski, T., Seru, A., and Yao, V. (2023a). Mortgage refinancing, consumer spending, and competition: Evidence from the home affordable refinancing program. *Review of Economic Studies*, 90(2):499–537.

Agarwal, S., Chomsisengphet, S., and Liu, C. (2010). The importance of adverse selection in the credit card market: Evidence from randomized trials of credit card solicitations. *Journal of Money, Credit and Banking*, 42(4):743–754.

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

Agarwal, S., Kigabo, T., Minoiu, C., Presbitero, A. F., and Silva, A. F. (2023b). Serving the underserved: microcredit as a pathway to commercial banks. *Review of Economics and Statistics*, 105(4):780–797.

Agarwal, S., Mikhed, V., and Scholnick, B. (2020). Peers' income and financial distress: Evidence from lottery winners and neighboring bankruptcies. *The Review of Financial Studies*, 33(1):433–472.

Albanesi, S., DeGiorgi, G., and Nosal, J. (2022). Credit growth and the financial crisis: A new narrative. *Journal of Monetary Economics*, 132:118–139.

Albanesi, S. and Nosal, J. (2018). Insolvency after the 2005 bankruptcy reform. *NBER Working Paper No. 24934*.

Albanesi, S. and Vamossy, D. F. (2019). Predicting consumer default: A deep learning approach. *NBER Working Paper No. 26165*.

Alexandrov, A., Brown, A., and Jain, S. (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*.

Alexandrov, A. and Koulayev, S. (2018). No shopping in the us mortgage market: Direct and strategic effects of providing information. *Consumer Financial Protection Bureau Office of Research Working Paper No 2017-01*.

Allen, J., Clark, R., Li, S., and Vincent, N. (2022). Debt-relief programs and money left on the table: Evidence from canada's response to covid-19. *Canadian Journal of Economics*, 55(S1):9–53.

Andre, J., Blavin, F., Braga, B., and Gangopadhyaya, A. (2023). The impact of hospital closures on medical debt in collections: Analysis using consumer credit bureau data. *IZA Discussion Paper No. 16448*.

Andre, J., Braga, B., Martinchek, K., and McKernan, S.-M. (2024). The effects of the state utility shut-off moratoria on credit delinquencies during the covid-19 pandemic. *Journal of Economics and Business*, 129:106170.

Argyle, B., Iverson, B., Nadauld, T. D., and Palmer, C. (2021). Personal bankruptcy and the accumulation of shadow debt. *NBER Working Paper No. 28901*.

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

Argys, L. M., Friedson, A. I., Pitts, M. M., and Tello-Trillo, D. S. (2020). Losing public health insurance: Tenncare reform and personal financial distress. *Journal of Public Economics*, 187:104202.

Arya, S., Eckel, C., and Wichman, C. (2013). Anatomy of the credit score. *Journal of Economic Behavior & Organization*, 95:175–185.

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

Avery, R. B., Bilinski, M. F., Bucks, B. K., Chai, C., Critchfield, T., Keith, I. H., Mohamed, I. E., Pafenberg, F. W., Patrabansh, S., Schultz, J. D., and Wood, C. E. (2017). A profile of 2013 mortgage borrowers: Statistics from the national survey of mortgage originations. *Technical Report No. 16-01*.

Avery, R. B., Bostic, R. W., Calem, P. S., and Canner, G. B. (1996). Credit risk, credit scoring, and the performance of home mortgages. *Federal Reserve Bulletin No. 82*.

Avery, R. B. and Brevoort, K. P. (2015). The subprime crisis: Is government housing policy to blame? *Review of Economics and Statistics*, 97(2):352–363.

Avery, R. B., Brevoort, K. P., and Canner, G. B. (2009). Credit scoring and its effects on the availability and affordability of credit. *Journal of Consumer Affairs*, 43(3):516–537.

Avery, R. B., Calem, P. S., and Canner, G. B. (2004). Credit report accuracy and access to credit. *Federal Reserve Bulletin*, 90:297.

Avery, R. B., Calem, P. S., Canner, G. B., and Bostic, R. W. (2003). An overview of consumer data and credit reporting. *Federal Reserve Bulletin No. 89*.

Babina, T., Bahaj, S., Buchak, G., De Marco, F., Foulis, A., Gornall, W., Mazzola, F., and Yu, T. (2024). Customer data access and fintech entry: Early evidence from open banking. *Stanford University Graduate School of Business Research Paper*.

Bach, H., Campa, P., De Giorgi, G., Nosal, J., and Pietrobon, D. (2023). Born to be (sub) prime: An exploratory analysis. *AEA Papers and Proceedings*, 113:166–171.

Baker, S. R. (2018). Debt and the response to household income shocks: Validation and application of linked financial account data. *Journal of Political Economy*, 126(4):1504–1557.

Baker, S. R. and Kueng, L. (2022). Household financial transaction data. *Annual Review of Economics*, 14:47–67.

Ballance, J., Clifford, R., and Shoag, D. (2020). "no more credit score": Employer credit check bans and signal substitution. *Labour Economics*, 63:101769.

Balyuk, T. (2023). Fintech lending and bank credit access for consumers. *Management Science*, 69(1):555–575.

Bank, T., Segev, N., and Shaton, M. (2023). Relationship banking and credit scores: Evidence from a natural experiment. *Working Paper*.

Barron, J. M., Elliehausen, G., and Staten, M. E. (2000). Monitoring the household sector with aggregate credit bureau data: A new source of data provides a firmer foundation for credit analysis and decisions. *Business Economics*, pages 63–76.

Barron, J. M. and Staten, M. (2003). The value of comprehensive credit reports: Lessons from the us experience. *Credit reporting systems and the international economy*, 8:273–310.

Bartik, A. W. and Nelson, S. T. (2022). Deleting a signal: Evidence from pre-employment credit checks. *Review of Economics and Statistics*, Forthcoming.

Bartlett, R., Morse, A., Stanton, R., and Wallace, N. (2022). Consumer-lending discrimination in the fintech era. *Journal of Financial Economics*, 143(1):30–56.

Batty, M., Gibbs, C., and Ippolito, B. (2022). Health insurance, medical debt, and financial well-being. *Health Economics*, 31(5):689–728.

Bayer, P., Ferreira, F., and Ross, S. L. (2016). The vulnerability of minority homeowners in the housing boom and bust. *American Economic Journal: Economic Policy*, 8(1):1–27.

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

Ben-David, I., Johnson, M., Lee, J., and Yao, V. (2022). Fintech lending with lowtech pricing. *Working Paper*.

Benetton, M., Kudlyak, M., and Mondragon, J. (2022). Dynastic home equity. *Working Paper*.

Benjamin, C., Hartley, D., Keys, B. J., and Ng, X. (2022). Credit when you need it. *Working Paper*.

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

Beraja, M., Fuster, A., Hurst, E., and Vavra, J. (2019). Regional heterogeneity and the refinancing channel of monetary policy. *The Quarterly Journal of Economics*, 134(1):109–183.

Berg, T., Burg, V., Gombović, A., and Puri, M. (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies*, 33(7):2845–2897.

Bergemann, D. and Bonatti, A. (2019). Markets for information: An introduction. *Annual Review of Economics*, 11:85–107.

Bergemann, D., Bonatti, A., and Smolin, A. (2018). The design and price of information. *American economic review*, 108(1):1–48.

Berger, A. N., Miller, N. H., Petersen, M. A., Rajan, R. G., and Stein, J. C. (2005). Does function follow organizational form? evidence from the lending practices of large and small banks. *Journal of Financial economics*, 76(2):237–269.

Berger, D., Milbradt, K., Tourre, F., and Vavra, J. (2021). Mortgage prepayment and path-dependent effects of monetary policy. *American Economic Review*, 111(9):2829–78.

Berger, E., Butler, A. W., and Mayer, E. J. (2018). Credit where credit is due: Drivers of subprime credit. *Working Paper*.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., and Zinman, J. (2010). What's advertising content worth? evidence from a consumer credit marketing field experiment. *The Quarterly Journal of Economics*, 125(1):263–306.

Beshears, J., Blakstad, M., Choi, J. J., Firth, C., Gathergood, J., Laibson, D., Notley, R., Sheth, J. D., Sandbrook, W., and Stewart, N. (2024). Does pension automatic enrollment increase debt? evidence from a large-scale natural experiment. *NBER Working Paper No. 32100*.

Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2018). Behavioral household finance. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 1, pages 177–276. Elsevier.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., and Skimmyhorn, W. L. (2022). Borrowing to save? the impact of automatic enrollment on debt. *The Journal of Finance*, 77(1):403–447.

Bhole, M. (2017). Why do federal loans crowd out the private market? evidence from graduate plus loans. *Working Paper*.

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

Bhutta, N. (2015). The ins and outs of mortgage debt during the housing boom and bust. *Journal of Monetary Economics*, 76:284–298.

Bhutta, N. and Canner, G. B. (2013). Mortgage market conditions and borrower outcomes: Evidence from the 2012 hmda data and matched hmda-credit record data. *Federal Reserve Bulletin*, 99(4).

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

Bhutta, N., Goldin, J., and Homonoff, T. (2016). Consumer borrowing after payday loan bans. *The Journal of Law and Economics*, 59(1):225–259.

Bhutta, N. and Hizmo, A. (2021). Do minorities pay more for mortgages? *The Review of Financial Studies*, 34(2):763–789.

Bhutta, N., Hizmo, A., and Ringo, D. (2022). How much does racial bias affect mortgage lending? evidence from human and algorithmic credit decisions. *Finance and Economics Discussion Series No. 2022-067*.

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

Bhutta, N., Skiba, P. M., and Tobacman, J. (2015). Payday loan choices and consequences. *Journal of Money, Credit and Banking*, 47(2-3):223–260.

Billings, S. B., Gallagher, E. A., and Ricketts, L. (2022). Let the rich be flooded: the distribution of financial aid and distress after hurricane harvey. *Journal of Financial Economics*, 146:797–819.

Björkegren, D. and Grissen, D. (2018). The potential of digital credit to bank the poor. *AEA Papers and Proceedings*, 108:68–71.

Björkegren, D. and Grissen, D. (2020). Behavior revealed in mobile phone usage predicts credit repayment. *The World Bank Economic Review*, 34(3):618–634.

Black, S. E., Denning, J. T., Dettling, L. J., Goodman, S., and Turner, L. J. (2020). Taking it to the limit: Effects of increased student loan availability on attainment, earnings, and financial well-being. *NBER Working Paper No. 27658*.

Blascak, N., Cheney, J. S., Hunt, R. M., Mikhed, V., Ritter, D., and Vogan, M. (2019). Financial consequences of identity theft: Evidence from consumer credit bureau records. *Federal Reserve Bank of Philadelphia Working Paper No. 19-02*.

Blascak, N. and Mikhed, V. (2023). Health insurance and young adult financial distress. *Journal of Policy Analysis and Management*, 42(2):393–423.

Blattner, L., Hartwig, J., and Nelson, S. (2022). Information design in consumer credit markets. *Working Paper*.

Blattner, L. and Nelson, S. (2022). How costly is noise? data and disparities in consumer credit. *Working Paper*.

Blattner, L., Nelson, S., and Spiess, J. (2021). Unpacking the black box: Regulating algorithmic decisions. *Working Paper*.

Bleemer, Z., Brown, M., Lee, D., Strair, K., and Van der Klaauw, W. (2021). Echoes of rising tuition in students' borrowing, educational attainment, and homeownership in post-recession america. *Journal of Urban Economics*, 122(1):1–24.

Bleemer, Z., Brown, M., Lee, D., Van der Klaauw, W., et al. (2017). Tuition, jobs, or housing: What's keeping millennials at home. *Federal Reserve Bank of New York Staff Reports*, 700.

Bleemer, Z. and van der Klaauw, W. (2019). Long-run net distributionary effects of federal disaster insurance: the case of hurricane katrina. *Journal of Urban Economics*, 110:70–88.

Board of Governors of the Federal Reserve System (2023). Economic well-being of us households in 2022.

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

Bond, P., Elul, R., Garyn-Tal, S., and Musto, D. K. (2017). Does junior inherit? refinancing and the blocking power of second mortgages. *The Review of Financial Studies*, 30(1):211–244.

Bono, T., Croxson, K., and Giles, A. (2021). Algorithmic fairness in credit scoring. *Oxford Review of Economic Policy*, 37(3):585–617.

Bornstein, G. and Indarte, S. (2022). The impact of social insurance on household debt. *Working Paper*.

Bos, M., Breza, E., and Liberman, A. (2018). The labor market effects of credit market information. *The Review of Financial Studies*, 31(6):2005–2037.

Bouckaert, J. and Degryse, H. (2004). Softening competition by inducing switching in credit markets. *Journal of Industrial Economics*, 52(1):27–52.

Braga, B., McKernan, S.-M., and Hassani, H. (2019). Delinquent debt decisions and their consequences over time. *Urban Institute Research Brief*.

Braxton, J. C., Chikhale, N., Herkenhoff, K. F., and Phillips, G. M. (2024a). Intergenerational mobility and credit. *NBER Working Paper No. 32031*.

Braxton, J. C., Herkenhoff, K. F., and Phillips, G. M. (2024b). Can the unemployed borrow? implications for public insurance. *Journal of Political Economy*, Forthcoming.

Brennecke, C. (2016). Information acquisition in antebellum us credit markets: Evidence from nineteenth-century credit reports. *FDIC Center for Financial Research Paper No. 2016-04*.

Brennecke, C., Siravyan, G., and Witzen, B. H. (2021). Commercial credit on consumer credit reports. *Consumer Financial Protection Bureau Office of Research Reports Series No. 21-7*.

Brevoort, K., Grodzicki, D., and Hackmann, M. B. (2020). The credit consequences of unpaid medical bills. *Journal of Public Economics*, 187:104203.

Brevoort, K. P. (2011). Credit card redlining revisited. *Review of Economics and Statistics*, 93(2):714–724.

Brevoort, K. P., Avery, R. B., and Canner, G. B. (2013). Credit where none is due? authorized-user account status and piggybacking credit. *Journal of Consumer Affairs*, 47(3):518–547.

Brevoort, K. P. and Cooper, C. R. (2013). Foreclosure's wake: The credit experiences of individuals following foreclosure. *Real Estate Economics*, 41(4):747–792.

Brevoort, K. P., Grimm, P., and Kambara, M. (2015). Credit invisibles. *Bureau of Consumer Financial Protection Data Point Series No. 15-1*.

Brevoort, K. P. and Kambara, M. (2017). Becoming credit visible. *Bureau of Consumer Financial Protection Data Point Series No. 17-1*.

Brown, A. and Wilson, E. (2023). Data point: Consumer credit and the removal of medical collections from credit reports. *Consumer Financial Protection Bureau Report, 26 April 2023*.

Brown, J. R., Cookson, J. A., and Heimer, R. Z. (2019). Growing up without finance. *Journal of Financial Economics*, 134(3):591–616.

Brown, M., Grigsby, J., Van Der Klaauw, W., Wen, J., and Zafar, B. (2016). Financial education and the debt behavior of the young. *The Review of Financial Studies*, 29(9):2490–2522.

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

Brown, M., Jappelli, T., and Pagano, M. (2009). Information sharing and credit: Firm-level evidence from transition countries. *Journal of Financial Intermediation*, 18(2):151–172.

Brown, M. and Zehnder, C. (2007). Credit reporting, relationship banking, and loan repayment. *Journal of Money, Credit and Banking*, 39(8):1883–1918.

Bruhn, J. M., Greenberg, K., Gudgeon, M., Rose, E. K., and Shem-Tov, Y. (2023). The effects of combat deployments on veterans' outcomes. *NBER Working Paper No. 30622*.

Bruhn, M., Farazi, S., and Kanz, M. (2013). Bank competition, concentration, and credit reporting. *World Bank Policy Research Working Paper No. 6442*.

Buchak, G. (2024). Financing the gig economy. *Journal of Finance*, 79(1):219–256.

Burke, J., Jamison, J., Karlan, D., Mihaly, K., and Zinman, J. (2023). Credit building or credit crumbling? a credit builder loan's effects on consumer behavior and market efficiency in the united states. *The Review of Financial Studies*, 36(4):1585–1620.

Burlando, A., Kuhn, M. A., and Prina, S. (2024). The role of credit reports in digital lending: a case study from mexico. *Oxford Review of Economic Policy*, 40(1):104–117.

Butcher, K. and Munoz, A. P. (2017). Using credit reporting agency data to assess the link between the community reinvestment act and consumer credit outcomes. *Cityscape*, 19(2).

Butler, A. W., Demirci, I., Gurun, U. G., and Tellez, Y. (2022). Consumption smoothing: Financial stress and food insecurity. *Working Paper*.

Butler, A. W., Mayer, E. J., and Weston, J. P. (2023a). Racial disparities in the auto loan market. *The Review of Financial Studies*, 36(1):1–41.

Butler, A. W., Spyridopoulos, I., Tellez, Y., and Xu, B. (2023b). Financial breakups. *Working Paper*.

Caldwell, S., Nelson, S., and Waldinger, D. (2023). Tax refund uncertainty: Evidence and welfare implications. *American Economic Journal: Applied Economics*, 15(2):352–376.

Campbell, J. Y. (2006). Household finance. *The journal of finance*, 61(4):1553–1604.

Carter, S. P. and Skimmyhorn, W. (2017). Much ado about nothing? new evidence on the effects of payday lending on military members. *Review of Economics and Statistics*, 99(4):606–621.

Castellanos, S. G., Hernández, D. J., Mahajan, A., Alcaraz Prous, E., and Seira, E. (2022). Contract terms, employment shocks, and default in credit cards. *Working Paper*.

Chakrabarti, R., Fos, V., Liberman, A., and Yannelis, C. (2023). Tuition, debt, and human capital. *Review of Financial Studies*, 36(4):1667–1702.

Chakrabarti, R. and Pattison, N. (2019). Auto credit and the 2005 bankruptcy reform: The impact of eliminating cramdowns. *The Review of Financial Studies*, 32(12):4734–4766.

Chan, S., Haughwout, A., Hayashi, A., and van der Klaauw, W. (2016). Determinants of mortgage default and consumer credit use: The effects of foreclosure laws and foreclosure delays. *Journal of Money, Credit and Banking*, 48(2/3):393–413.

Chan, T., Hamdi, N., Hui, X., and Jiang, Z. (2022). The value of verified employment data for consumer lending: Evidence from equifax. *Marketing Science*, 41(4):795–814.

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

Chatterjee, S., Corbae, D., Dempsey, K. P., and Ríos-Rull, J.-V. (2023). A quantitative theory of the credit score. *Econometrica*, 91(5):1803–1840.

Chava, S., Ganduri, R., Paradkar, N., and Zeng, L. (2023a). Shocked by bank funding shocks: Evidence from consumer credit cards. *Review of Financial Studies*, 36(10):3906–3952.

Chava, S., Ganduri, R., Paradkar, N., and Zhang, Y. (2021). Impact of marketplace lending on consumers' future borrowing capacities and borrowing outcomes. *Journal of Financial Economics*, 142(3):1186–1208.

Chava, S., Tookes, H., and Zhang, Y. (2023b). Leaving them hanging: Student loan forbearance, distressed borrowers, and their lenders. *Working Paper*.

Cheng, I.-H., Severino, F., and Townsend, R. R. (2021). How do consumers fare when dealing with debt collectors? evidence from out-of-court settlements. *The Review of Financial Studies*, 34(4):1617–1660.

Chernousov, M. M., Flagg, J. N., Hannon, S. M., Lewis, V. L., Stephens, S. M., and Volz, A. H. (2024). A note on revolving credit estimates. *Board of Governors of the Federal Reserve System's FEDS Notes, June 14*.

Cherry, S., Jiang, E. X., Matvos, G., Piskorski, T., and Seru, A. (2021). Government and private household debt relief during covid-19. *Brookings Papers On Economic Activity*, Fall Edition.

Chioda, L., Gertler, P., Higgins, S., and Medina, P. (2024). Fintech lending to borrowers with no credit history. *Working Paper*.

Clarkberg, J. and Kambara, M. (2018). Public records. *Consumer Financial Protection Bureau Office of Research Reports Series No. 18-1*.

Cochran, K. T., Stegman, M., and Foos, C. (2021). Utility, telecommunications, and rental data in underwriting credit. *Urban Institute, December*.

Conkling, T. S. and Gibbs, C. (2019). Borrower experiences on income-driven repayment. *Consumer Financial Protection Bureau Data Point Series No. 19-5*.

Consumer Financial Protection Bureau (2012a). Analysis of differences between consumer-and creditor-purchased credit scores. *Consumer Financial Protection Bureau Report (September)*.

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

Consumer Financial Protection Bureau (2014). Report on the use of remittance histories in credit scoring. *Consumer Financial Protection Bureau Report (July)*.

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 (2021). The consumer credit market. *Consumer Financial Protection Bureau Report (September)*.

Consumer Financial Protection Bureau (2023). Market snapshot: An update on third-party debt collections tradelines reporting. *Consumer Financial Protection Bureau Report (February)*.

Conway, J., Glazer, J., and Plosser, M. (2023). Does the community reinvestment act improve consumers' access to credit? *Federal Reserve Bank of New York Staff Report No. 1048*.

Cookson, J. A., Gallagher, E., and Mulder, P. (2023). Money to burn: Wildfire insurance via social networks. *Working Paper*.

Cookson, J. A., Gilje, E. P., and Heimer, R. Z. (2022). Shale shocked: Cash windfalls and household debt repayment. *Journal of Financial Economics*, 146(3):905–931.

Cookson, J. A., Guttman-Kenney, B., and Mullins, W. (2024). Access to credit & credit invisibility. *Unpublished Work-in-Progress*.

Cooper, D., Luengo-Prado, M. J., and Parker, J. A. (2020). The local aggregate effects of minimum wage increases. *Journal of Money, Credit and Banking*, 52(1):5–35.

Corbae, D. and Glover, A. (2018). Employer credit checks: Poverty traps versus matching efficiency. *NBER Working Paper No. 25005*.

Correia, F., Han, P., and Wang, J. (2024). The online payday loan premium. *Working Paper*.

Cortés, K. R., Glover, A., and Tasci, M. (2022). The unintended consequences of employer credit check bans for labor markets. *Review of Economics and Statistics*, 104(5):997–1009.

Cubides, E. and O'Brien, S. (2023). 2023 findings from the diary of consumer payment choice. *The Federal Reserve Bank of San Francisco, Fed Notes May 5 2023*.

Davis, M. A., Gregory, J., Hartley, D. A., and Tan, K. T. (2021). Neighborhood effects and housing vouchers. *Quantitative Economics*, 12(4):1307–1346.

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

De Janvry, A., McIntosh, C., and Sadoulet, E. (2010). The supply-and demand-side impacts of credit market information. *Journal of Development Economics*, 93(2):173–188.

Debbaut, P., Ghent, A., and Kudlyak, M. (2016). The card act and young borrowers: The effects and the affected. *Journal of Money, Credit and Banking*, 48(7):1495–1513.

Dell'Ariccia, G. and Marquez, R. (2006). Lending booms and lending standards. *Journal of Finance*, 61(5):2511–2546.

Demyanyk, Y., Hryshko, D., Luengo-Prado, M. J., and Sørensen, B. E. (2017). Moving to a job: The role of home equity, debt, and access to credit. *American Economic Journal: Macroeconomics*, 9(2):149–81.

Demyanyk, Y., Loutskina, E., and Murphy, D. (2019). Fiscal stimulus and consumer debt. *Review of Economics and Statistics*, 101(4):728–741.

Desai, C. A. and Elliehausen, G. (2017). The effect of state bans of payday lending on consumer credit delinquencies. *The Quarterly Review of Economics and Finance*, 64:94–107.

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

DeWaard, J., Johnson, J., and Whitaker, S. (2019). Internal migration in the united states: A comprehensive comparative assessment of the consumer credit panel. *Demographic Research*, 41:953.

DeWaard, J., Johnson, J. E., and Whitaker, S. D. (2020). Out-migration from and return migration to puerto rico after hurricane maria: Evidence from the consumer credit panel. *Population and Environment*, 42(1):28–42.

Di Maggio, M., Kalda, A., and Yao, V. (2023). Second chance: Life with less student debt. *Journal of Finance*, Forthcoming.

Di Maggio, M., Kermani, A., Keys, B. J., Piskorski, T., Ramcharan, R., Seru, A., and Yao, V. (2017). Interest rate pass-through: Mortgage rates, household consumption, and voluntary deleveraging. *American Economic Review*, 107(11):3550–88.

Di Maggio, M., Kermani, A., and Palmer, C. J. (2020). How quantitative easing works: Evidence on the refinancing channel. *The Review of Economic Studies*, 87(3):1498–1528.

Di Maggio, M., Kermani, A., Ramcharan, R., Yao, V., and Yu, E. (2022). The pass-through of uncertainty shocks to households. *Journal of Financial Economics*, 145(1):85–104.

Di Maggio, M., Ma, A. T., and Williams, E. (2024). In the red: Overdrafts, payday lending and the underbanked. *Journal of Finance*, Forthcoming.

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

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The Review of Economic Studies*, 51(3):393–414.

Dinerstein, M., Yannelis, C., and Chen, C.-T. (2024). Debt moratoria: Evidence from student loan forbearance. *American Economic Review: Insights*, 6(2):196–213.

Djankov, S., McLiesh, C., and Shleifer, A. (2007). Private credit in 129 countries. *Journal of Financial Economics*, 84(2):299–329.

Djeundje, V. B., Crook, J., Calabrese, R., and Hamid, M. (2021). Enhancing credit scoring with alternative data. *Expert Systems with Applications*, 163:113766.

Dobbie, W., Goldsmith-Pinkham, P., Mahoney, N., and Song, J. (2020). Bad credit, no problem? credit and labor market consequences of bad credit reports. *The Journal of Finance*, 75(5):2377–2419.

Dobbie, W., Goldsmith-Pinkham, P., and Yang, C. S. (2017). Consumer bankruptcy and financial health. *Review of Economics and Statistics*, 99(5):853–869.

Dobkin, C., Finkelstein, A., Kluender, R., and Notowidigdo, M. J. (2018). The economic consequences of hospital admissions. *American Economic Review*, 108(2):308–52.

Doblas-Madrid, A. and Minetti, R. (2013). Sharing information in the credit market: Contract-level evidence from us firms. *Journal of Financial Economics*, 109(1):198–223.

Dokko, J., Li, G., and Hayes, J. (2015). Credit scores and committed relationships. *FEDS Working Paper No. 2015-081*.

Dooley, J. and Gallagher, E. (2024). Blood money: Selling plasma to avoid high-interest loans. *Review of Financial Studies*, Forthcoming.

Dupor, B., Karabarbounis, M., Kudlyak, M., and Mehkari, M. (2022). Regional consumption responses and the aggregate fiscal multiplier. *Federal Reserve Bank of San Francisco Working Paper No. 2018-04*.

Durbin, E., Li, G., Low, D., and Ricks, J. (2021). Characteristics of mortgage borrowers during the covid-19 pandemic. *Consumer Financial Protection Bureau Office of Research Reports Series No. 21-3*.

Dutz, D., Huitfeldt, I., Lacouture, S., Mogstad, M., Torgovitsky, A., and van Dijk, W. (2022). Selection in surveys. *NBER Working Paper No. 29549*.

Einav, L., Jenkins, M., and Levin, J. (2013). The impact of credit scoring on consumer lending. *The RAND Journal of Economics*, 44(2):249–274.

Einav, L. and Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210):1243089.

Elul, R., Payne, A., and Tilson, S. (2023). Owner-occupancy fraud and mortgage performance. *Real Estate Economics*, 51(5):1137–1177.

Evans, D. S. and Schmalensee, R. (2004). *Paying with plastic: the digital revolution in buying and borrowing*. Mit Press.

Fedaseyeu, V. (2020). Debt collection agencies and the supply of consumer credit. *Journal of Financial Economics*, 138(1):193–221.

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, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., and Group, O. H. S. (2012). The oregon health insurance experiment: evidence from the first year. *The Quarterly Journal of Economics*, 127(3):1057–1106.

FinRegLab, Blattner, L., and Spiess, J. (2022). Machine learning explainability & fairness: Insights from consumer lending. *FinRegLab Empirical White Paper*.

Fiorin, S., Hall, J., and Kanz, M. (2023). How do borrowers respond to a debt moratorium? experimental evidence from consumer loans in india. *Working Paper*.

Flamang, N. and Kancherla, S. (2023). Unemployment insurance as a financial stabilizer: Evidence from large benefit expansions. *Working Paper*.

Foley, C. F., Hurtado, A., Liberman, A., and Sepulveda, A. (2022). The effects of information on credit market competition: Evidence from credit cards. *Working Paper*.

Fonseca, J. (2023). Less mainstream credit, more payday borrowing? evidence from debt collection restrictions. *Journal of Finance*, 78(1):63–103.

Fonseca, J. and Liu, L. (2024). Mortgage lock-in, mobility, and labor reallocation. *Working Paper*.

Fonseca, J. and Wang, J. (2024). How much do small businesses rely on personal credit? *Working Paper*.

Foote, C. L., Loewenstein, L., and Willen, P. S. (2021). Cross-sectional patterns of mortgage debt during the housing boom: evidence and implications. *The Review of Economic Studies*, 88(1):229–259.

Fos, V., Hamdi, N., Kalda, A., and Nickerson, J. (2024). Gig-labor: Trading safety nets for steering wheels. *Journal of Financial Economics*, Forthcoming.

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

Frisancho, V. (2023). Spillover effects of financial education: The impact of school-based programs on parents. *Journal of Financial Literacy and Wellbeing*, pages 1–16.

Fritsch, N. and Heimer, R. Z. (2020). Intergenerational homeownership and mortgage distress. *Federal Reserve Bank of Cleveland Economic Commentary No. 2020–12*.

Fu, G., Sun, M., and Xu, Q. (2020). An alternative credit scoring system in china's consumer lending market: A system based on digital footprint data. *Working Paper*.

Fulford, S. and Nagypál, E. (2023). The equilibrium effect of information in consumer credit markets: Public records and credit redistribution. *Working Paper*.

Fulford, S. and Nagypál, É. (2023). Using the courts for private debt collection: How wage garnishment laws affect civil judgments and access to credit. *Consumer Financial Protection Bureau Office of Research Working Paper No. 23-02*.

Fulford, S. and Schuh, S. (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, S. and Schuh, S. D. (2017). Credit card utilization and consumption over the life cycle and business cycle. *Working Paper*.

Fulford, S. and Shupe, C. (2021a). Consumer finances during the pandemic: Insights from the making ends meet survey. *Consumer Financial Protection Bureau Office of Research Reports Series No. 21-16*.

Fulford, S. and Shupe, C. (2021b). 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, S. L. (2015). How important is variability in consumer credit limits? *Journal of Monetary Economics*, 72:42–63.

Fuster, A., Guttman-Kenney, B., and Haughwout, A. (2018). Tracking and stress-testing us household leverage. *Economic Policy Review*, 1(24):35–63.

Fuster, A., Plosser, M., Schnabl, P., and Vickery, J. (2019). The role of technology in mortgage lending. *The Review of Financial Studies*, 32(5):1854–1899.

Galenianos, M. and Gavazza, A. (2022). Regulatory interventions in consumer financial markets: The case of credit cards. *Journal of the European Economic Association*, 20:1897–1932.

Gallagher, E. A., Billings, S. B., and Ricketts, L. R. (2023a). Human capital investment after the storm. *The Review of Financial Studies*, 36(7):2651–2684.

Gallagher, J. and Hartley, D. (2017). Household finance after a natural disaster: The case of hurricane katrina. *American Economic Journal: Economic Policy*, 9(3):199–228.

Gallagher, J., Hartley, D., and Rohlin, S. (2023b). Weathering an unexpected financial shock: The role of disaster assistance on household finance and business survival. *Journal of the Association of Environmental and Resource Economists*, 10:283–567.

Gallego, D. and Meisenzahl, R. (2022). Internal migration patterns after the 2008 financial crisis: Evidence from a credit panel. *Chicago Fed Insights*.

Ganong, P. and Noel, P. (2020). Liquidity versus wealth in household debt obligations: Evidence from housing policy in the great recession. *American Economic Review*, 110(10):3100–3138.

Gao, J., Yi, H. L., and Zhang, D. H. (2024). Algorithmic underwriting in high risk mortgage markets. *Working Paper*.

Garcia, D. I. (2022). Second-home buying and the housing boom and bust. *Real Estate Economics*, 50(1):33–58.

Garmaise, M. J. and Natividad, G. (2017). Consumer default, credit reporting, and borrowing constraints. *The Journal of Finance*, 72(5):2331–2368.

Gathergood, J., Guttman-Kenney, B., and Hunt, S. (2019a). How do payday loans affect borrowers? evidence from the uk market. *The Review of Financial Studies*, 32(2):496–523.

Gathergood, J., Mahoney, N., Stewart, N., and Weber, J. (2019b). How do americans repay their debt? the balance-matching heuristic. *Economics Bulletin*, 39(2):1458–1466.

Gehrig, T. and Stenbacka, R. (2007). Information sharing and lending market competition with switching costs and poaching. *European Economic Review*, 51(1):77–99.

George, T., Newberger, R., and O'Dell, M. (2019). The geography of subprime credit. *ProfitWise News and Views, Federal Reserve Bank of Chicago*, 6.

Ghent, A. C. and Kudlyak, M. (2016). Intergenerational linkages in household credit. *Federal Reserve Bank of San Francisco Working Paper No. 2016-31*.

Ghosh, P. and Vats, N. (2023). Safety nets, credit, and investment: Evidence from a guaranteed income program. *Working Paper*.

Gibbs, C. (2023). Office of research blog: Initial fresh start program changes followed by increased credit scores for affected student loan borrowers.

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

Goldsmith-Pinkham, P., Pinkovskiy, M., and Wallace, J. (2021). The great equalizer: Medicare and the geography of consumer financial strain. *Working Paper*.

Gomes, F., Haliassos, M., and Ramadorai, T. (2021). Household finance. *Journal of Economic Literature*, 59(3):919–1000.

Gopalan, R., Hamilton, B. H., Kalda, A., and Sovich, D. (2021a). Home equity and labor income: The role of constrained mobility. *The Review of Financial Studies*, 34(10):4619–4662.

Gopalan, R., Hamilton, B. H., Kalda, A., and Sovich, D. (2021b). State minimum wages, employment, and wage spillovers: Evidence from administrative payroll data. *Journal of Labor Economics*, 39(3):673–707.

Gopalan, R., Hamilton, B. H., Sabat, J., and Sovich, D. (2023). Aversion to student debt? evidence from low-wage workers. *Journal of Finance*, 79(2):1249–1295.

Granja, J. and Nagel, F. (2024). Current expected credit losses and consumer loans. *Working Paper*.

Gresenz, C. R., Mitchell, J. M., Rodriguez, B., Turner, R. S., and van der Klaauw, W. (2024). The financial consequences of undiagnosed memory disorders. *Federal Reserve Bank of New York Staff Report No. 1106*.

Grodzicki, D. and Koulayev, S. (2021). Sustained credit card borrowing. *Journal of Consumer Affairs*, 55(2):622–653.

Gross, T., Kluender, R., Liu, F., Notowidigdo, M. J., and Wang, J. (2021). The economic consequences of bankruptcy reform. *American Economic Review*, 111(7):2309–41.

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

Guiso, L. and Sodini, P. (2013). Household finance: An emerging field. In *Handbook of the Economics of Finance*, volume 2, pages 1397–1532. Elsevier.

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

Guttman-Kenney, B. (2024). Disaster flags: Credit reporting relief from natural disasters. *Working Paper*.

Guttman-Kenney, B., Adams, P., Hunt, S., Laibson, D., Stewart, N., and Leary, J. (2023). The semblance of success in nudging consumers to pay down credit card debt. *NBER Working Paper No. 31926*.

Guttman-Kenney, B., Kluender, R., Mahoney, N., Wong, F., Xia, X., and Yin, W. (2022). Trends in medical debt during the covid-19 pandemic. *JAMA Health Forum*, 3(5):e221031.

Guttman-Kenney, B. and Shahidinejad, A. (2024). Unraveling information sharing in consumer credit markets. *Working Paper*.

Hahm, J.-H. and Lee, S. (2011). Economic effects of positive credit information sharing: the case of korea. *Applied Economics*, 43(30):4879–4890.

Hall, J. P. (2024). Credit cards and retail firms: Historical evidence from the u.s. *Working Paper*.

Hamdi, N., Kalda, A., and Sovich, D. (2024). The labor market consequences of student loan forbearance. *Working Paper*.

Hampole, M. (2024). Financial frictions and human capital investments. *Working Paper*.

Han, S., Keys, B. J., and Li, G. (2018). Unsecured credit supply, credit cycles, and regulation. *The Review of Financial Studies*, 31(3):1184–1217.

Hand, D. J. and Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A*, 160(3):523–541.

Haughwout, A., Lee, D., Mangrum, D., Scally, J., and van der Klaauw, W. (2023). The great pandemic mortgage refinance boom. *Liberty Street Economics Blog May 15, Federal Reserve Bank of New York*.

Haughwout, A., Lee, D., Scally, J., and van der Klaauw, W. (2019). Just released: Mind the gap in delinquency rates. *Liberty Street Economics*.

Haughwout, A., Lee, D., Tracy, J. S., and Van der Klaauw, W. (2011). Real estate investors, the leverage cycle, and the housing market crisis. *Federal Reserve Bank of New York Staff Report No. 514*.

Haughwout, A. and van der Klaauw, W. (2015). Do we know what we owe? consumer debt as reported by borrowers and lenders. *Economic Policy Review*, 1(21):19–44.

Haughwout, A. F., Lee, D., Mangrum, D., Scally, J., and Van der Klaauw, W. (2024). The new york fed consumer credit panel: A foundational cmd data set. *Liberty Street Economics Blog April 17, Federal Reserve Bank of New York*.

Hauswald, R. and Marquez, R. (2006). Competition and strategic information acquisition in credit markets. *Review of Financial Studies*, 19(3):967–1000.

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

Herbst, D. (2022). The impact of income-driven repayment on student borrower outcomes. *American Economic Journal: Applied Economics*, 15(1):1–25.

Herkenhoff, K., Phillips, G., and Cohen-Cole, E. (2023). How credit constraints impact job finding rates, sorting & aggregate output. *Review of Economic Studies*, Forthcoming.

Herkenhoff, K., Phillips, G. M., and Cohen-Cole, E. (2021). The impact of consumer credit access on self-employment and entrepreneurship. *Journal of Financial Economics*, 141(1):345–371.

Herman, L., Kaplan, J., and Mueller, A. (2020). Payment amount furnishing & consumer reporting. *Consumer Financial Protection Bureau Office of Research Reports Series No. 20-13*.

Hertzberg, A., Liberti, J. M., and Paravisini, D. (2011). Public information and coordination: evidence from a credit registry expansion. *The Journal of Finance*, 66(2):379–412.

Homonoff, T., O'Brien, R., and Sussman, A. B. (2021). Does knowing your fico score change financial behavior? evidence from a field experiment with student loan borrowers. *Review of Economics and Statistics*, 103(2):236–250.

Hossain, M., Livshits, I., and Wardius, C. (2023). Not cashing in on cashing out: An analysis of low cash-out refinance rates. *Federal Reserve Bank of Philadelphia Working Paper No. 23-04*.

Howard, G. and Shao, H. (2023). The dynamics of internal migration: A new fact and its implications. *Working Paper*.

Hu, L., Kaestner, R., Mazumder, B., Miller, S., and Wong, A. (2018). The effect of the affordable care act medicaid expansions on financial wellbeing. *Journal of Public Economics*, 163:99–112.

Hunt, R. M. (2005). A century of consumer credit reporting in america. *Federal Reserve Bank of Philadelphia Working Paper No. 05-13*.

Hynes, R. M. (2017). Maximum possible accuracy in credit reports. *Law & Contemp. Probs.*, 80:87.

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

Israel, S., Caspi, A., Belsky, D. W., Harrington, H., Hogan, S., Houts, R., Ramrakha, S., Sanders, S., Poulton, R., and Moffitt, T. E. (2014). Credit scores, cardiovascular disease risk, and human capital. *Proceedings of the National Academy of Sciences*, 111(48):17087–17092.

Jagtiani, J. and Dolson, E. (2021). Which lenders are more likely to reach out to underserved consumers: Banks versus fintechs versus other nonbanks? *FRB of Philadelphia Working Paper No. 21-17*.

Jansen, M., Nagel, F., Yannelis, C., and Zhang, A. L. (2023). Data and welfare in credit markets. *NBER Working Paper No. 30235*.

Jappelli, T. and Pagano, M. (2002). Information sharing, lending and defaults: Cross-country evidence. *Journal of Banking & Finance*, 26(10):2017–2045.

Jiang, Z. (2022). An empirical bargaining model with left-digit bias: A study on auto loan monthly payments. *Management Science*, 68(1):442–465.

Jiang, Z., Wei, Y. M., Chan, T., and Hamdi, N. (2023). Designing dealer compensation in the auto-loan market: Implications from a policy change. *Marketing Science*, 42(5):958–983.

Jiang, Z., Zhang, D. J., and Chan, T. (2021). How do bonus payments affect the demand for auto loans and their delinquency? *Journal of Marketing Research*, 58(3):476–496.

Jones, C. I. and Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–2858.

Kallberg, J. G. and Udell, G. F. (2003). The value of private sector business credit information sharing: the us case. *Journal of Banking & Finance*, 27(3):449–469.

Keys, B. J., Mahoney, N., and Yang, H. (2023). What determines consumer financial distress? place-and person-based factors. *Review of Financial Studies*, 36(1):42–69.

Keys, B. J. and Wang, J. (2019). Minimum payments and debt paydown in consumer credit cards. *Journal of Financial Economics*, 131(3):528–548.

Khandani, A. E., Kim, A. J., and Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11):2767–2787.

Kim, Y. S., Lee, D., Scharlemann, T. C., and Vickery, J. I. (2022). Intermediation frictions in debt relief: evidence from cares act forbearance. *Federal Reserve Bank of New York Staff Report No. 1035*.

Kluender, R., Mahoney, N., Wong, F., and Yin, W. (2021). Medical debt in the us, 2009-2020. *JAMA*, 326(3):250–256.

Kluender, R., Mahoney, N., Wong, F., and Yin, W. (2024). The effects of medical debt relief: Evidence from two randomized experiments. Technical report, NBER Working Paper No. 32315.

Kovrijnykh, N., Livshits, I., and Zetlin-Jones, A. (2023). Building credit histories. *Working Paper*.

Kumar, A. (2022). Know thyself: Access to own credit report and the retail mortgage market. *Working Paper*.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. W. (1997). Legal determinants of external finance. *The Journal of Finance*, 52(3):1131–1150.

Lambie-Hanson, L. and Reid, C. (2018). Stuck in subprime? examining the barriers to refinancing mortgage debt. *Housing Policy Debate*, 28(5):770–796.

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

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

Lee, D. and Van der Klaauw, W. (2010). An introduction to the frbny consumer credit panel. *Federal Reserve Bank of New York Staff Report No. 479*.

Lee, D., Van der Klaauw, W., Sabelhaus, J., and Scaly, J. (2024a). The best of both worlds: Synthesizing administrative and survey household financial data. *Unpublished Work-in-Progress*.

Lee, J. Y., Yang, J., and Anderson, E. (2024b). Who benefits from alternative data for credit scoring? evidence from peru. *Working Paper*.

Lee, S. C. and Maxted, P. (2024). Credit card borrowing in heterogeneous-agent models: Reconciling theory and data. *Working Paper*.

Li, W. and Goodman, L. (2016). Comparing credit profiles of american renters and owners. *Urban Institute Working Paper*.

Liang, A., Lu, J., and Mu, X. (2021). Algorithm design: A fairness-accuracy frontier. *Working Paper*.

Liberman, A., Neilson, C., Opazo, L., and Zimmerman, S. (2019). The equilibrium effects of information deletion: Evidence from consumer credit markets. *NBER Working Paper No. 25097*.

Liberti, J., Sturgess, J., and Sutherland, A. (2022). How voluntary information sharing systems form: Evidence from a us commercial credit bureau. *Journal of Financial Economics*, 145(3):827–849.

Liebersohn, J. and Rothstein, J. (2024). Household mobility and mortgage rate lock. *Working Paper*.

Lin, M., Prabhala, N. R., and Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management science*, 59(1):17–35.

Low, D., Clarkberg, J., and Gardner, J. (2021). Subprime auto loan outcomes by lender type. *Consumer Financial Protection Bureau Office of Research Reports Series No. 21-12*.

Mabille, P. (2023). The missing home buyers: Regional heterogeneity and credit contractions. *Review of Financial Studies*, 36(7):2756–2796.

Madeira, C. (2024). The impact of information laws on consumer credit access: evidence from chile. *Working Paper*.

Mayer, E. J. (2024). Big banks, household credit access, and intergenerational economic mobility. *Journal of Financial and Quantitative Analysis*, Forthcoming.

Mazumder, B. and Miller, S. (2016). The effects of the massachusetts health reform on household financial distress. *American Economic Journal: Economic Policy*, 8(3):284–313.

McNamara, J. (2023). Why the largest credit card companies are suppressing actual payment data on your credit report.

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

Meier, S. and Sprenger, C. D. (2012). Time discounting predicts creditworthiness. *Psychological Science*, 23(1):56–58.

Mello, S. (2023). Fines and financial wellbeing. *Working Paper*.

Mezza, A., Ringo, D., Sherlund, S., and Sommer, K. (2020). Student loans and homeownership. *Journal of Labor Economics*, 38(1):215–260.

Mezza, A. and Sommer, K. (2016). A trillion-dollar question: What predicts student loan delinquencies? *Journal of Student Financial Aid*, 46(3):3.

Mian, A. (2012). The case for a credit registry. In Brunnermeier, M. and Krishnamurthy, A., editors, *Risk topography: Systemic risk and macro modeling*, chapter 11, pages 163–172. University of Chicago Press.

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

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

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

Mian, A. and Sufi, A. (2012). The effects of fiscal stimulus: Evidence from the 2009 cash for clunkers program. *The Quarterly Journal of Economics*, 127(3):1107–1142.

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

Mian, A. and Sufi, A. (2017). Household debt and defaults from 2000 to 2010: The credit supply view. In Fennell, L. A. and Keys, B. J., editors, *Evidence and Innovation in Housing Law and Policy*, chapter 11, pages 257–288. Cambridge University Press Cambridge.

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

Mian, A., Sufi, A., and Trebbi, F. (2010). The political economy of the us mortgage default crisis. *American Economic Review*, 100(5):1967–98.

Mikhed, V. and Vogan, M. (2018). How data breaches affect consumer credit. *Journal of Banking & Finance*, 88:192–207.

Miller, M. J. (2003). *Credit Reporting Systems and the International Economy*. Mit Press.

Miller, S. and Soo, C. K. (2020). Does increasing access to formal credit reduce payday borrowing? *NBER Working Paper No. 27783*.

Miller, S. and Soo, C. K. (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, S., Wherry, L. R., and Foster, D. G. (2023). The economic consequences of being denied an abortion. *American Economic Journal: Economic Policy*, 15(1):394–437.

Mishra, P., Prabhala, N., and Rajan, R. G. (2022). The relationship dilemma: Why do banks differ in the pace at which they adopt new technology? *The Review of Financial Studies*, 35(7):3418–3466.

Mohr, J. and Kohli, D. (2024). House of stolen cards: Payment security and lending decisions. *Working Paper*.

Moulton, S., Brown, M., Haurin, D., and Loibl, C. (2023). How does the death of a partner during the covid-19 pandemic affect the economic security of the surviving older adult? evidence from credit panel and labor force participation data. *Working Paper No. WI23-13*.

Musto, D. K. (2004). What happens when information leaves a market? evidence from postbankruptcy consumers. *The Journal of Business*, 77(4):725–748.

Musto, D. K. and Souleles, N. S. (2006). A portfolio view of consumer credit. *Journal of Monetary Economics*, 53(1):59–84.

Nagel, F. (2024). Consumer bankruptcy auditss. *Working Paper*.

Nelson, S. T. (2022). Private information and price regulation in the us credit card market. *Working Paper*.

Nicholas, L. H., Langa, K. M., Bynum, J. P., and Hsu, J. W. (2021). Financial presentation of alzheimer disease and related dementias. *JAMA Internal Medicine*, 181(2):220–227.

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

Norden, L. and Weber, M. (2010). Credit line usage, checking account activity, and default risk of bank borrowers. *The Review of Financial Studies*, 23(10):3665–3699.

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

Pal, A. (2024). Time on your side: Labor market effects of foreclosure delays. *Working Paper*.

Papich, S. (2023). Effects of buy now, pay later on financial well-being. *Working Paper*.

Petersen, M. A. and Rajan, R. G. (1994). The benefits of lending relationships: Evidence from small business data. *Journal of Finance*, 49(1):3–37.

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

Petersen, M. A. and Rajan, R. G. (2002). Does distance still matter? the information revolution in small business lending. *The journal of Finance*, 57(6):2533–2570.

Pinto, S. and Steinbaum, M. (2023). The long-run impact of the great recession on student debt. *Labour Economics*, 85:102449.

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

Piskorski, T., Seru, A., and Witkin, J. (2015). Asset quality misrepresentation by financial intermediaries: Evidence from the rmbs market. *The Journal of Finance*, 70(6):2635–2678.

Ponce, A., Seira, E., and Zamarripa, G. (2017). Borrowing on the wrong credit card? evidence from mexico. *American Economic Review*, 107(4):1335–61.

Puri, M., Rocholl, J., and Steffen, S. (2017). What do a million observations have to say about loan defaults? opening the black box of relationships. *Journal of Financial Intermediation*, 31:1–15.

Raith, M. (1996). A general model of information sharing in oligopoly. *Journal of Economic Theory*, 71(1):260–288.

Ramakrishnan, R. T. and Thakor, A. V. (1984). Information reliability and a theory of financial intermediation. *The Review of Economic Studies*, 51(3):415–432.

Rishabh, K. (2024). Beyond the bureau: Interoperable payment data for loan screening and monitoring. *Working Paper*.

Robinson, J., Park, D. S., and Blumenstock, J. E. (2023). The impact of digital credit in developing economies: A review of recent evidence. *KDI School of Public Policy & Management Paper No. 23-04*.

Romeo, C. and Sandler, R. (2021). The effect of debt collection laws on access to credit. *Journal of Public Economics*, 195:104320.

Roth Tran, B. and Sheldon, T. L. (2017). Same storm, different disasters: Consumer credit access, income inequality, and natural disaster recovery. *Working Paper*.

Sandler, R. and Nathe, L. (2022). Paid and low-balance medical collections on consumer credit reports. *Consumer Financial Protection Bureau Office of Research Report No. 22-5, 27 July 2022*.

Sauers, Z. (2022). Student debt and high-skill worker location choice. *Working Paper*.

Schenone, C. (2010). Lending relationships and information rents: Do banks exploit their information advantages? *Review of Financial Studies*, 23(3):1149–1199.

Seira, E., Elizondo, A., and Laguna-Müggenburg, E. (2017). Are information disclosures effective? evidence from the credit card market. *American Economic Journal: Economic Policy*, 9(1):277–307.

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

Shupe, C., Li, G., and Fulford, S. (2023). Consumer use of buy now, pay later: Insights from the cfpb making ends meet survey. *Consumer Financial Protection Bureau Office of Research Reports Series No. 2023-1*.

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

Smith, J., Goodman, J., and Hurwitz, M. (2020). The economic impact of access to public four-year colleges. *NBER Working Paper No. 27177*.

Smith, L. D., Staten, M., Eyssell, T., Karig, M., Freeborn, B. A., and Golden, A. (2013). Accuracy of information maintained by us credit bureaus: frequency of errors and effects on consumers' credit scores. *Journal of Consumer Affairs*, 47(3):588–601.

Stango, V. and Zinman, J. (2016). Borrowing high versus borrowing higher: price dispersion and shopping behavior in the us credit card market. *The Review of Financial Studies*, 29(4):979–1006.

Staten, M. and Cate, F. H. (2005). Chapter 10 - does the fair credit reporting act promote accurate credit reporting? In Retsinas, N. P. and Belsk, E. S., editors, *Building Assets, Building Credit: Creating Wealth in Low-Income Communities*, pages 237–265. Brookings Press.

Stein, J. C. (2002). Information production and capital allocation: Decentralized versus hierarchical firms. *The journal of finance*, 57(5):1891–1921.

Sutherland, A. (2018). Does credit reporting lead to a decline in relationship lending? evidence from information sharing technology. *Journal of Accounting and Economics*, 66(1):123–141.

Thomas, L., Crook, J., and Edelman, D. (2017). *Credit scoring and its applications (second edition)*. SIAM.

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

Tobback, E. and Martens, D. (2019). Retail credit scoring using fine-grained payment data. *Journal of the Royal Statistical Society Series A*, 182(4):1227–1246.

Wei, Y., Yildirim, P., Van den Bulte, C., and Dellarocas, C. (2016). Credit scoring with social network data. *Marketing Science*, 35(2):234–258.

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

Woodward, S. E. and Hall, R. E. (2012). Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. *American Economic Review*, 102(7):3249–3276.

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.

Xie, R. and Moulton, S. (2023). Temporary debt payment relief or long-term payment restructuring? insights from credit card accommodations during the covid-19 pandemic. *Working Paper*.

Xie, S., Xie, V. W., and Zhang, X. (2023). Extreme weather and low-income household finance: Evidence from payday loans. *Working Paper*.

Xu, B. Y. (2023). The consequences of access to credit in early adulthood: Evidence from the card act. *Working Paper*.

Yannelis, C. and Zhang, A. L. (2023). Competition and selection in credit markets. *Journal of Financial Economics*, 150:103710.

Zeballos Doubinko, V. and Akana, T. (2023). How does buy now, pay later affect customers' credit? *Federal Reserve Bank of Philadelphia, Consumer Finance Institute Discussion Paper DP 23-01*.

Zhang, D. (2023a). Closing costs, refinancing, and inefficiencies in the mortgage market. Technical report, Working Paper.

Zhang, Y. (2023b). Banking on the cloud: Impact on credit card borrowers. *Working Paper*.

Zhong, M., Braga, B., McKernan, S.-M., Hayward, M., Millward, E., and Trepel, C. (2023). Impacts of covid-19-era economic policies on consumer debt in the united kingdom. *Urban Institute Working Paper*.

Zinman, J. (2009). Where is the missing credit card debt? clues and implications. *Review of Income and Wealth*, 55(2):249–265.