Exploring the Use of Anonymized Consumer Credit Information to Estimate Economic Conditions: An Application of Big Data

Chet Wiermanski*

Stephanie M. Wilshusen**

November 2015

Summary: *The emergence of high-frequency administrative data and other big data offers an opportunity for improvements to economic forecasting models. This paper considers the potential advantages and limitations of using information contained in anonymized consumer credit reports for improving estimates of current and future economic conditions for various geographic areas and demographic markets. Aggregate consumer credit information is found to be correlated with macroeconomic variables such as gross domestic product, retail sales, and employment and can serve as leading indicators such that lagged values of consumer credit variables can improve the accuracy of forecasts of these macro variables.*

Keywords: consumer credit information, administrative data, big data, real-time data, nowcasting, forecasting
JEL Classification Numbers: D12, D14, C53, C55

* Aether Analytics, LLC and visiting scholar, Payment Cards Center, Federal Reserve Bank of Philadelphia. Nothing in the text should be construed as an endorsement of any organization or its products or services.
**Payment Cards Center, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106. Corresponding author's e-mail: stephanie.wilshusen@phil.frb.org. The authors thank Dina Anderson and Jeff Morrison. The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. No statements here should be treated as legal advice. This paper is available free of charge at www.philadelphiafed.org/consumer-credit-and-payments/payment-cards-center/publications/discussion-papers/.

1. Introduction

Accurate, reliable, and relevant estimates of national, state, and local economic conditions are important to policymakers, businesses, and investors because this information provides a comprehensive and consistent perspective of the path of the U.S. economy. This perspective is useful for analyzing trends and identifying factors affecting economic growth and productivity. While there have been improvements and advancements in the reporting, collection, and treatment of national accounts data typically used in economic forecasting models, the emergence of high-frequency administrative data and other big data offers an opportunity for further improvements to these models.[1, 2] The health and stability of many sectors of the U.S. economy depend on consumer sentiment and expenditures.[3] Therefore, understanding the relationship between various aspects of the U.S. economy and consumer behavior may improve the estimation of current and future economic conditions.[4]

A promising source of high-frequency consumer behavior information is the data contained in consumer credit reports. These data provide a unique view of consumer credit activity within a given period that, in aggregate, may accurately and reliably reflect various aspects of economic activity and stability within local geographies. This paper considers potential advantages and limitations of using anonymized information contained in consumer credit reports

[1] For the purposes of our discussion, *administrative data* refers to the information that is generated by businesses and government agencies for "program administration, regulatory, or law enforcement purposes." (U.S. Office of Management and Budget (OMB), 2015).

[2] Diebold (2012) traces the origin of the term *big data* and describes in Diebold (2003) the "Big Data phenomenon" as "the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage technology." Volume, variety, and velocity were noted in Laney (2001) as big data's notable characteristics. Similarly, Podesta et al. (2014) define big data as "the ability to gather large volumes of data, often from multiple sources, and with it, produce new kinds of observations, measurements and predictions." See Varian (2014) for a discussion of the econometric tools and techniques for managing and analyzing big data.

[3] Nearly 70 percent of the U.S. economy, as measured by gross domestic product (GDP), is generated by consumer expenditures.

[4] For example, these could be measures of employment, GDP, retail sales, and housing.

for improving estimates of current and future economic conditions for various geographic areas and demographic markets. Aggregate consumer credit information is found to be correlated with macroeconomic variables such as gross domestic product (GDP), retail sales, and employment and can serve as leading indicators. In other words, lagged values of consumer credit variables have the potential to improve the accuracy of forecasts of macro variables.

The paper begins in Section 2 with a discussion of the development and availability of big data, including high-frequency administrative data — with examples of its use in macroeconomic forecasting — and introduces a particular example of these data: aggregate consumer credit information. Section 3 describes the construction of the information contained in consumer credit reports, improvements in the comprehensiveness of these data, and policy considerations in the collection and use of these data. Section 4 discusses some of the limitations associated with the data and techniques currently used in economic forecasting, while Section 5 discusses several advantageous characteristics of credit bureau data for the purposes of estimating economic activity and forecasting. Section 6 outlines some of the constraints associated with credit bureau data including its accessibility. Section 7 discusses the dynamic nature of credit use and payment behavior. Illustrations of the relationships between credit bureau data and macroeconomic variables are described in Section 8. Section 9 concludes and offers suggestions for further research.

2. Background

2.1 Big Data to Improve Economic Measurement and Forecasting

Policymakers and researchers have described the uncertainty attendant with economic measurement and forecasts. To address this uncertainty, they have suggested that new types of less aggregated and higher frequency information are needed to provide a more complete picture of economic conditions and the heterogeneity of conditions across households and geography.[5]

[5] See, for example, Bernanke (2012) and Ng and Wright (2013).

Recent studies have called for and explored the use of big data, including high-frequency administrative data, in economic policy and research. Einav and Levin (2014) attribute the increased availability of information about economic activity over the past couple of decades in part to the growth of the Internet and the data it captures as individuals search, shop, and consume goods, services, and information. In addition to the Internet, the electronic collection and linking of behavior, transactions, and demographics at the level of the individual is possible through other technologies in both the private and public sectors. The authors note several features of big data that distinguish it from other types of information that previously have been used in economic research, including its immediate, (near) real-time availability, its large scale, and its ability to capture previously unobserved activity.

For many years, businesses have used big data in the creation, marketing, and pricing of products and services. More recently, economic researchers have used these large-scale, high-frequency data sets to examine a number of public policy questions related to educational outcomes, allocation of resources in political campaigns, operation of health systems, and innovations in commerce and consumer lending. Given the substantial contribution of household expenditures to GDP and the importance of understanding this activity to explain economic output and employment, Hall (2014) calls for the improvement of existing aggregated data sources on household expenditures, asset holdings, and debt to account for the heterogeneity among households. He suggests that this could be accomplished by collecting data at greater frequency and from panels of households as well as by linking administrative data collected by government entities and the records of financial institutions.[6]

[6] Hall (2014) describes federal sources of aggregate data about the "financial position and expenditures" for all U.S. households as well as sources for household-level information. He notes that while the Federal Reserve Board's Survey of Consumer Finances (SCF) collects detailed information about a sample of families' assets and liabilities, that collection is limited to only every three years. Historically, the SCF has tracked respondents over time for only special purposes, such as studying the effects of the Great Recession.

2.2 Availability of Administrative Data

Businesses generate large-scale, high-frequency data as they manage their products and services. Financial transactions are a key example. Payments technology company First Data processes almost half of all U.S.-based consumer payments. First Data uses the information created while processing these payments to provide insight into U.S. consumer spending patterns in the form of its SpendTrend product. Using actual consumer spending data across 4 million merchant locations, SpendTrend offers a view of aggregated transaction and dollar volume growth by industry, region, and payment type. A more recent example of harnessing proprietary business data is the creation of the JPMorgan Chase Institute. Launched in May 2015, the institute's mission is to study the transaction-level data of Chase's customers across all of its business lines to provide insights into economic activity and to inform public policy.

Large-scale, high-frequency data also can be found in the public sector as government agencies administer their programs. The 2016 federal budget emphasizes the importance of creating a framework to improve access to the administrative data collected by federal agencies that provide insight into economic conditions and inform public policy.[7] It also outlines several advantages of using administrative data including larger sample sizes, longer periods of collection, lower costs of collection, and greater accuracy compared with self-reported survey data.

Card, Chetty, Feldstein, and Saez (2011) note difficulties with survey data, including the inability to track individuals over time, attrition, underreporting, and increasing nonresponse rates. The authors call for researchers to have direct access to anonymized administrative microdata. They offer suggestions for the development of research partnerships between U.S. administrative agencies and external researchers and suggest extending such a model to private institutions.

[7] For a discussion of efforts to expand access to federal administrative data, see U.S. OMB, 2015.

To address the protocols, privacy issues, and restrictions placed on access to and use of large-scale administrative data sets maintained by government agencies, a series of open data initiatives instituted by the federal government directs federal agencies to make available to the public more of its administrative information and establishes guidelines for the stewardship of these data.[8] In addition to soliciting feedback from the public about which data they would like to access, this open data policy requires federal agencies to collect information and create data in a manner that facilitates processing and promotes dissemination using machine-readable and open formats while protecting personal and confidential information. In support of this effort to provide researchers and policymakers with access to their administrative data, agencies including the Federal Reserve and the U.S. Department of Commerce have created chief data officer positions to facilitate the development of standards for data collection and distribution.

Making administrative data available to social scientists for economic research to inform public policy is the focus of a project conducted by the Council of Professional Associations of Federal Statistics (COPAFS), "Increasing Researcher Access to Federal Administrative Data." Project findings and recommendations are expected to be released after the project's conclusion in December 2015.[9]

2.3 Use of High-Frequency and Administrative Data to Forecast Current and Future Economic Conditions and Economic Activity

One application of high-frequency data is *nowcasting*. Nowcasting is a forecasting method that uses the staggered nature of large sets of information to update and improve forecasts of

[8] See Executive Order 13642 (May 9, 2013) *Making Open and Machine Readable the New Default for Government Information*, Open Data Policy — Managing Information as an Asset (OMB Memorandum M-13-13); and *Guidance for Providing and Using Administrative Data for Statistical Purposes* (OMB Memorandum M-14-06).

[9] COPAFS organized a session at the 2015 Annual Meeting of the American Economic Association titled "Increasing Researcher Access to Federal Administrative Data: Rewards, Risk, and Demand." The presentations by panelists John N. Freidman (Brown University and National Bureau of Economic Research), Amy Finkelstein (Massachusetts Institute of Technology and J-PAL North America), Ron Jarmin (U.S. Census Bureau), and Katherine R. Smith (COPAFS) are available at www.copafs.org/projects.aspx.

macroeconomic variables for the current period.[10] Nowcasting exploits the correlations between other sources of data and the macro variable of interest (most often, GDP or inflation). These component or correlated data series are released earlier or more frequently than the macroeconomic variable being forecast and, therefore, can provide signals about its value. The models used in nowcasting are designed to rigorously exploit the variable frequency of the data to improve the timeliness and accuracy of forecasts of current economic conditions.[11]

For example, Galbraith and Tkacz (2007) use electronic transaction information in the form of Canadian debit card data as real-time indicators of economic activity. Since debit cards are the preferred means of payment in Canada, debit card purchase data offer a high-frequency and timely measure of a significant portion of household expenditures and therefore GDP. The authors find that these data can help to lower GDP and consumption forecast errors and improve predictions by reducing the magnitude of revisions to these key national account variables. Galbraith and Tkacz (2013) document the marginal contribution of monthly payments data (debit, checking, and credit card transactions) to GDP nowcasts over time. They find that the inclusion in models of payment data, particularly debit card activity, can improve the accuracy of nowcasts. The contribution varies over time with the amount of available information.

Another example of high-frequency administrative data is the information contained in consumer credit reports. The timeliness and broad coverage of information in consumer credit repositories has led thousands of lenders and insurance companies to use these data when underwriting new credit and insurance applications and managing portfolio performance. While credit bureau data have been used for years in loan portfolio analysis to determine which economic factors influence portfolio performance, to quantify the impact of industry trends on

[10] For a discussion of data and methodologies used in nowcasting, see, for example, Bańbura, Giannone, and Reichlin (2011) and Modugno (2013).

[11] Nowcasts produced by the Federal Reserve Banks of Philadelphia and Atlanta are publicly available. See the Aruoba-Diebold-Scotti Business Conditions Index and the GDPNow model forecasts, available at www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/ and www.frbatlanta.org/cqer/researchcq/gdpnow.aspx, respectively.

portfolio behavior, and to understand changes in consumers' credit use and performance, this paper adds to the literature by exploring the use of anonymized, consumer credit information to estimate macroeconomic conditions. [12]

To understand actual choices under uncertainty, it is important to understand economic decision-making in the context of the information available at the time of the decision (see Box 1: "Real-Time Data Set for Macroeconomists"). The information contained in consumer credit reports available to researchers can provide that context. The historical information contained in the credit bureau data used by researchers is not subject to revision; it is the information that existed at the time the data were created. That is generally not the case for the data available to market participants.

Box 1. Real-Time Data Set for Macroeconomists

An element of information processing in a big data environment is the unique role of *real-time* data. Policy decisions based on forecasts that are too reliant on data that have been substantially revised over time are subject to an inconsistency — they presume more information than historically existed at the time people made their decisions. Real-time data offer insights about the actual choices under uncertainty by providing the information set available to consumers, businesses, and policymakers at the time they made their decisions.

Several years ago, researchers at the Federal Reserve Bank of Philadelphia (FRBP) decided to investigate the practical significance of using real-time rather than revised data in macroeconomic forecasting (Croushore and Stark, 2001). To do that, they needed to construct real-time data sets — vintages of data as they appeared during a particular calendar date. In the 1990s, the FRBP started to build a large U.S. real-time data set — the Real-Time Data Set for Macroeconomists.[a] Using the real-time and revised data, economists could assess the degree to which relying on revised data might lead to forecasts that perform less well in real time than after the fact. Croushore (2008) discusses advances researchers have made toward understanding the large and systematic impact of data revisions on macroeconomic analysis and the resulting development of real-time data sets.

The Real-Time Data Set for Macroeconomists captures historical revisions to macroeconomic observations as they are released over time by government statistical agencies. It includes more than 3 million data points in 13,000 vintages on 90 macroeconomic variables, including real GDP and its components, personal income and its components, inflation measures, labor market metrics, corporate profits, and a variety of financial variables.[b] The information contained in this data set allows economists to assess current macroeconomic conditions and feeds into economic models that produce forecasts on which policy decisions are based.

[a] These data are available on the FRBP's website at www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/data-files/.

[b] Each data series vintage contains the entire time series available at a given point in time.

[12] For examples of research on the traditional uses of credit bureau data, see, for example, Barron, Elliehausen, and Staten (2000) and Morrison (2008 and 2010).

3. Construction of, Recent Improvements to, and Policy Issues Associated with Consumer Credit Information

Consumer credit information is compiled from individual consumer credit reports and maintained by each of the three leading national consumer credit reporting agencies (CRAs: Equifax, Experian, and TransUnion). Reflected for each account found on a consumer's credit report is an indicator describing the nature of the debt and the date the account was opened. In addition to a diverse set of historical account repayment indicators, information in a consumer's credit report contains a fairly comprehensive picture of the amount of outstanding debt held by a consumer and each consumer's capacity to take on additional debt quickly, as reflected by credit limits on open revolving accounts. The credit report also contains public record items such as judgments, bankruptcies, and state and federal tax liens. Outstanding obligations placed with collection agencies may be reported by third-party debt collectors to credit reporting agencies.

A variety of entities voluntarily provides information about consumers' borrowing and payment behavior on mortgage, installment, and revolving debt.[13] These organizations include creditors; government agencies; collection agencies; and in some instances, third-party intermediaries that collect data from some utility companies, telecommunication service providers, and property rental agencies.

In addition to information about how a consumer has managed his or her credit obligations, a consumer's credit report also contains personally identifying information referred to as the header section of the file and includes the consumer's name, date of birth, Social Security number, current and previous addresses, and employment history. A consumer's credit report also contains a record of all entities that have accessed the report for permissible purposes.[14] Information about an individual's race, color, religion, national origin, education,

[13] See Hunt (2002) for a discussion of the economics of information sharing by consumer credit bureaus.

[14] The Fair Credit Reporting Act (15 U.S.C. §1681 *et. seq*.) specifies the types of businesses to which and under what circumstances CRAs may provide consumer credit report information.

marital status, and health are not included in a consumer's credit report. The data provided to credit reporting agencies by approximately 30,000 data furnishers are generally updated monthly, and each year more than 1 billion consumer credit reports are provided to consumers and lenders.[15]

3.1 Construction of Anonymized Consumer Credit Bureau Information

Anonymized consumer credit report information is available from the three leading national consumer CRAs. Although each CRA deploys different proprietary processes and makes this information available under different user restrictions, the essence of this information is very similar at each CRA. The source data used to create the consumer samples are drawn from each CRA's consumer credit reporting database used for consumer credit underwriting.

When compiling anonymized consumer credit behavior information, the major CRAs do not rely solely upon a statistical sample. Instead, the entire universe of more than 200 million individual consumer credit reports may be evaluated. Each CRA makes anonymous data available for different applications. When the application relies upon a statistical sampling approach, the sampling techniques used may also vary. One CRA might use a random sample, while another might select consumers with two or three specific numbers found within their Social Security numbers. Each sampling technique may have its own advantages and disadvantages.[16] Because the samples drawn from each CRA are fairly consistent over time, these anonymized consumer credit databases provide a relatively good representation of the consumers residing within each CRA's consumer credit reporting database. Depending on the CRA, anonymized consumer credit

[15] See Federal Trade Commission (2015) and the Consumer Data Industry Association's website at www.cdiaonline.org.

[16] For example, when using either an Nth random sample or specific Social Security number approach, consumers with joint credit obligations may have a joint account included in the anonymous credit database twice. This will represent credit performance for mortgages, auto loans, and bank cards more heavily than department store or personal loans, which more often are opened for an individual. Using a specific sampled Social Security number introduces a bias to include only those individuals with a Social Security number reported to the CRA. This may underrepresent non-U.S. citizens and recent immigrants.

information for these applications is gathered on the same date on either a monthly or quarterly basis, typically spanning more than a decade.

Each consumer selected within each census or sample is then processed against the particular CRA's propriety data hygiene process for inclusion into the final sample.[17] This process typically involves excluding deceased consumers and a review of the contents of each consumer's credit report to ensure that a minimum amount of information is resident and that the information has been updated within a specific period of time. This is an attempt to ensure that "file fragments" and "stale" credit files, which may contain duplicate or obsolete credit information, are not included within the final sample.[18]

The consumer credit reports are then anonymized and summarized using the CRA's proprietary credit characteristics to produce a variety of credit statistics describing the consumer's acquisition, use, and repayment of credit obligations. Credit risk scores, income estimates, and other modeled credit behaviors also may be calculated for consumers within the sample. The credit characteristics used to describe the consumer's credit behavior are generally the same as those evaluated and used to develop various credit performance models. The anonymized credit information derived from each consumer credit repository reflects credit conditions and trends within relatively small geographic areas (state, metropolitan statistical area (MSA), county, census tract, and zip code) and among consumer groups segmented according to demographic or credit profiles.

3.2 Recent Improvements in Consumer Credit Bureau Data

Over time, consumer CRAs have expanded the comprehensiveness of the information within their respective repositories. Perhaps one of the most important events over the past two decades

[17] This process refers to practices used to maintain data consistency.

[18] File fragments can arise when information on credit applications, such as name, address, or Social Security number, is not consistent, not provided, or incorrect across data furnishers resulting in incomplete or improper linking of information on consumers' credit records. See Wardrip and Hunt (2013) for a discussion of time trends in the percent of file fragments in credit bureau data.

affecting the quality and inclusiveness of consumer credit report information occurred when Fannie Mae and Freddie Mac required lenders possessing guaranteed residential mortgages to report their portfolios to the national CRAs.[19] The reporting of additional tens of millions residential mortgages, which began in the mid-1990s, filled a critical missing component of a consumer's credit profile. The addition of personally identifying information such as the name, Social Security number, and address associated with a consumer responsible for repaying mortgage obligations was also a catalyst for creating more comprehensive credit reports. The personally identifying information reported by mortgage lenders fueled the development of proprietary data compilation programs by CRAs to assist with consolidating multiple consumer credit records (file fragments) into a single, more comprehensive credit report.

More complete credit reports combined with the increased use of credit in the mid-1990s in the form of general purpose bank-issued credit cards, installment loans, and residential mortgages and faster computer processing capabilities enabled CRAs to produce timely and reliable credit metrics extracted from their respective databases regularly.

However, as outlined in the section that follows, there are operational and regulatory costs that may constrain credit bureaus' efforts to capture more information. Consumer CRAs must weigh the costs associated with less reliable data in their decision to improve the comprehensiveness of their data. This is not an unconstrained decision because laws and regulations influence the tradeoff between, and the weights assigned to, inclusiveness and precision.

3.3 Policy Issues Associated with Credit Bureau Data

Information maintained and provided by CRAs is subject to state and federal laws limiting the use and distribution of personal data. The two federal laws that directly address the use of consumer credit reports and anonymized information derived from data provided to CRAs are the

[19] See Avery, Bostic, Calem, and Canner (1996) for a discussion of this development.

Fair Credit Reporting Act (FCRA), as amended by the Fair and Accurate Credit Transactions Act, and the Gramm–Leach–Bliley Act (GLBA).

The FCRA was enacted in 1970 and regulates access to and maintenance of personal information to protect the security and confidentiality of these data. It also restricts the collection and use of personal information to permissible purposes such as making credit granting and insurance underwriting decisions.[20] Because the data provided by the CRAs for research purposes are anonymized, the information is not classified as personal information. Under this interpretation, CRAs have been providing anonymized credit report information for more than two decades to lenders and third parties, including federal government agencies, for the development of credit scoring systems, and for research and analysis of consumer credit behavior.

Enacted in 1999, the GLBA restricts the sharing and disclosure of nonpublic personal information that individuals provide to financial institutions. These restrictions apply to *financial institutions* as defined by the GLBA or to entities that receive nonpublic personal information from such financial institutions. The sharing and use of nonpublic personal information that a nonaffiliated third party receives from a financial institution are limited under a "reuse and redisclosure" provision of the GLBA. In addition, under the GLBA, financial institutions are subject to standards established by financial regulators to maintain the security and confidentiality of the information contained in consumers' records.[21] CRAs formulated guidelines under the reuse and redisclosure provision to comply with their release and subsequent use of anonymized consumer credit records. Broadly speaking, these guidelines require that the use of anonymized consumer credit information be associated with analysis to formulate strategies to better manage the risk within portfolios and consumer segments.

[20] See U.S. Government Accountability Office (2013).

[21] See U.S. Government Accountability Office (2013).

4. Limitations Associated with Current Data Sources and Forecasting Techniques

Government agencies responsible for reporting economic conditions routinely evaluate new data sources to improve their methodologies and their estimates to continue to provide a reliable and relevant picture of the economy. Those involved in gathering, estimating, and reporting economic conditions are acutely aware of the limitations associated with some of their data sources and the manner in which information is gathered and derived from the surveys administered. The discussion in the following sections describes a few of the limitations identified by these providers.

4.1 Manner in Which Data Are Gathered

The use of random samples limits the degree to which information can be reliably estimated and reported for smaller geographic areas and demographic subpopulations. The size of random samples used to compile economic conditions varies widely, from several thousand to hundreds of thousands of observations. Although the samples surveyed typically are stratified, the reporting of most economic variables are limited to the national, regional, or state level, and, in some instances, a few major MSAs. Economic activity typically is reported for a major geographic area (the nation or a state) as a whole, without describing economic conditions for any specific demographic group of consumers.

This limitation exists because the information gathered is first aggregated by the data provider, without providing detailed information as to where units of interest were purchased or consumed. For example, sales of certain consumer goods or services reported by businesses to government agencies do not include where these consumer goods or services were purchased or consumed or, in the case of surveys administered to consumers, may not include enough observations to accurately or reliably estimate consumption for smaller geographies (for example, county or MSA) or among demographic subpopulations.

Although the reporting of economic activity for smaller geographies and demographic subpopulations can be obtained with special targeted samples, usually either concentrated within

a few smaller geographic areas or among demographic subpopulations of interest, this information typically is not reported on a timely or regular basis or may only be available for more densely populated areas. When economic activity is reported for a specific demographic subpopulation sampled, the economic activity may be projected for the entire demographic subpopulation studied, without reporting differences across geography.

4.2 Nonsampling Errors Within Economic Estimates Derived from Sample Surveys

Most estimates regarding the economic health and stability of a geographic area are based on voluntary responses from surveys administered to randomly selected consumers and businesses. Some of the factors attributed to nonsampling error within an estimate include the inability to obtain information about all units in the sample, differences in the interpretation of survey questions, mistakes in coding or keying the data obtained, and other errors of collection, response, coverage, and processing. Although the agencies responsible for preparing economic estimates exercise many precautionary steps during the collection, processing, and tabulation phase of survey data to minimize nonsampling error, little information on the potential biases caused by nonsampling error is available.[22]

5. Advantages of Credit Bureau Information

The inclusion of anonymized consumer credit information may potentially assist with providing more accurate, reliable, and relevant estimates of economic activity. Some of the potential advantages of adding anonymized consumer credit information to current approaches used to estimate certain economic conditions include reducing initial estimate variance associated with scheduled time series releases; enhancing existing economic data releases by providing estimates of economic activity within smaller geographic areas such as counties and MSAs; reporting the economic health and stability among demographic subpopulations including age, gender, income, and ethnic background; and the creation of new consumer-centric estimates of current and future

[22] Manski (2015) discusses the measurement of different types of nonsampling error and ways in which to classify and communicate this uncertainty when government agencies release economic statistics.

economic activity highly correlated with the acquisition, amount, use, and repayment of consumer credit. However, data sets that may be used for research with this level of demographic specificity will require matching credit bureau information with other data sets (see Section 5.2).

5.1 Unit of Observation

Depending upon the application, anonymized consumer credit information is organized at various levels of observation. Currently, consumer credit information is available at three different levels: account, consumer, and household. Because the vast majority of consumers maintains more than one credit obligation, one approach is to report credit behavior at the account level, which is consistent with the manner in which various federal regulatory bodies report credit activity sourced directly from financial institutions.

Reporting at the consumer and household levels provides insight as to how consumers manage and maintain their personal credit portfolios. Vital statistics about the number of accounts and amount of credit available and an individual's or household's ability to repay those credit obligations offer a different perspective from the manner in which federal regulatory report information is gathered from consumer surveys. Currently, for subscribers to its data services, one or more CRAs provide consumer credit report information at the level of individual consumers, households, or accounts. It is important to note that all three CRAs have the ability to produce anonymous account-, consumer-, and household-level credit information.

5.2 Characteristics and Segmentation Capabilities

Because each CRA archives detailed credit information for each consumer over time, each CRA can produce historical longitudinal panel and random credit snapshots based upon new or custom credit variables that were not calculated or contemplated when the consumers were originally selected for a sample. In addition, because names and addresses for consumers in these data sets are available, the CRAs have developed a process to append demographic and other behavioral information from third parties, which enables users to segment credit information by consumer

categories of interest to researchers. After appending the relevant variables, the CRA then anonymizes the data set before making it available to the researcher.

5.3 Timeliness

Because credit information contained within the three national consumer CRAs' databases is frequently updated, typically on a monthly basis, anonymized aggregate consumer credit samples can quickly reflect underlying trends and changes in credit use as they evolve. This may reduce the level of noise associated with aging census or industry statistics used to weight survey responses.

5.4 Insight into Local Credit Conditions

Each of the national CRAs is able to produce representative random samples of consumer credit activity for specific geographies. Accounting for geographic variation in consumer credit variables could augment data obtained under existing methods to refine estimates of economic activity, potentially providing more reliable initial estimates.

For example, large, random samples consisting of tens or hundreds of millions of consumers could quickly be drawn from the national CRAs to augment estimates of retail sales activity for housing and automobiles, which may allow for accurate and reliable economic estimates at the county or MSA level. It also may reduce the variance between initial and subsequent data revision. Additionally, by understanding the relationship between the magnitude and direction in which consumers use and repay various credit obligations and economic activity, it is possible that proxies for certain economic conditions, such as the percent change in GDP, employment, and disposable income, may be reliably projected for smaller geographic areas on a regular basis.

Historically, consumer credit markets were much more localized. That made it possible to use regulatory data (Call Reports) to construct local measures of credit conditions based on the location of the headquarters of financial institutions. As consumer credit markets became more national in scope, this approach became less and less viable. Today, by virtue of access to credit

bureau data, it is possible to view the geography of lending based on where the consumer lives. This contribution makes consumer credit information useful for estimating local macroeconomic conditions.

6. Credit Bureau Data Constraints

6.1 Research Data Sets Using Credit Bureau Data Typically Are Not Free

Access to aggregate consumer credit information typically is governed by an agreement between a CRA and researcher that outlines the conditions of its use and the cost of obtaining the data. In other words, these are proprietary data sets that are sold or licensed for a fee. But there are exceptions. For example, a recent project jointly funded and managed by the Federal Housing Finance Agency and the Consumer Financial Protection Bureau (CFPB) marries high-frequency administrative data with credit bureau data to create a new, nationally representative loan-level mortgage database that will be made available to researchers at no cost to the user. The National Mortgage Database will contain anonymized, full credit bureau file information for all borrowers associated with a national sample of active mortgage loans. To create a comprehensive picture of each mortgage, the consumer credit information on borrowers' mortgages and payment histories will be matched with other data files, including information about the mortgage product and terms and property characteristics. This project may serve as a model for future efforts to make more granular consumer credit data available to the public.

6.2 Credit Bureau Data Do Not Measure All Purchase Activity

Credit bureau data do not reflect transactions drawing on consumers' demand deposit accounts or consumers' use of cash. This point is illustrated in the Federal Reserve System's triennial payments study. The payments study is a survey of networks, processors, and banks about noncash payments in the U.S. The study has been conducted every three years since 2003 by researchers at the Board of Governors and the Retail Payments Office of the Federal Reserve System. Results from the most recent edition reveal that payments have become increasingly card-based over time. Credit and debit card transactions now comprise the majority of payment

transactions, with debit card transactions representing 38 percent of noncash payments in 2012, a greater share of the number of transactions than any other payment type. While the study's measures of average transaction value and share of the total value of transactions made using credit cards are greater than that for debit cards, the noncash forms of payment preferred by consumers in the study (that is, debit and prepaid cards) are not captured in the credit bureau information maintained by the three major CRAs. However, recent trends in e-commerce and mobile commerce, including consumers' increasing use of smartphones to make payments and merchant adoption of mobile card readers, suggest that more card-based purchase activity, if made by credit card, will be reflected in consumers' credit reports.

6.3 Accuracy of Information Reported to Bureaus by Lenders

While each of the national CRAs follows a series of data integrity checks for accuracy of the information received from data furnishers, there is considerable debate about the frequency with which lenders report inaccurate information to CRAs and the number of instances in which consumer CRAs include incorrect account performance information in a consumer's credit report. Under the FCRA, consumers can dispute and have inaccurate information removed from their reports.[23] Two studies of the accuracy of consumer credit reports found that approximately one-fifth to one-quarter of consumer credit reports contained potential inaccuracies.[24] Many of these potential inaccuracies include information found in the "header" section of a consumer's credit report, which contains personal identification and other personal information, the consumer's credit section, and the section containing public record information. Among consumers who disputed inaccuracies found in the credit section of their reports and achieved resolution of the

[23] In 2011, the national credit reporting agencies received 8 million consumer contacts disputing the accuracy of one or more items on their credit reports, representing 1.3 percent to 3.9 percent of credit active consumers. (Consumer Financial Protection Bureau, 2012).

[24] Federal Trade Commission (2012) and Turner, Varghese, and Walker (2011). For follow-up studies, see Federal Trade Commission (2015) and Turner, Varghese, Walker, and Chaudhuri (2013).

disputed items on their credit files, 1 to 2 percent saw an increase of 25 or more points in their credit scores.[25]

The existence of inaccuracies in the account information contained within the national CRAs' databases does affect some consumers and creates some inefficiencies in the credit reporting system. However, there is little reason to believe that those errors would induce a systematic bias in the correlations between credit bureau variables and macroeconomic variables. The information content of this rich but admittedly imperfect data can be used as input into forecasts of consumption and economic activity more generally.

To address questions about the accuracy of account information in the CRAs' databases, a standard electronic data reporting file format — Metro 2 — was introduced in 1997 by the credit reporting industry. Lenders are encouraged to use the Metro 2 format when furnishing consumer credit information to CRAs to address regulatory compliance concerns by improving the consistency, timeliness, and matching of information in consumer credit records. Standardization of lenders' reporting of the information associated with consumer credit obligations was designed to reduce differences in the manner in which account performance information is reported to any of the national CRAs.

The accuracy of information received by CRAs is a priority for the CFPB. In 2012, the CFPB published its larger participant rule permitting it to supervise companies with annual receipts of more than $7 million from consumer reporting, as defined in the rule. In its supervision of large banks, the CFPB began examining the processes institutions use to assure accuracy when furnishing information to the national CRAs and when responding to consumer disputes about information contained in their credit reports.

In addition to oversight by the CFPB, the CRAs are also subject to investigation and enforcement action by state attorneys general. For example, in March 2015, the New York State

[25] Federal Trade Commission (2012).

Office of the Attorney General announced a settlement agreement with the three national CRAs after an investigation into the CRAs' practices regarding the accuracy of the information they maintain and their methods of investigating consumer disputes of information in consumers' credit reports. Under the terms of the settlement, the CRAs agreed to phase in over a three-year period several reforms to improve the dispute resolution process. Among the reforms is an agreement to employ specially trained employees to review consumers' supporting documentation for disputes related to mixed files, fraud, or identity theft; the addition to the CRAs' websites of a prominently labeled hyperlink to AnnualCreditReport.com; the provision by the CRAs of an additional free credit report for consumers whose credit reports change as a result of initiating a dispute; and the creation by the CRAs of a National Credit Reporting Working Group to develop practices to enhance the monitoring and data accuracy of companies that furnish data to the CRAs. [26]

6.4 Voluntary Reporting of Information

The consumer credit reporting business model in the U.S. is based upon voluntary data contribution, as are a majority of the credit reporting systems throughout the world. Data furnishers can elect to contribute information from any or all of their various credit portfolios. All three national CRAs encourage full reporting of all information as detailed with the Metro 2 reporting format and maintain rigorous proprietary minimum data reporting standards before adding information into their consumer reporting database. Lenders electing not to contribute certain critical tradeline information will either not have the consumer's entire tradeline added to the consumer credit reporting database or will be subject to each CRA's proprietary "give-to-get" data use policy.[27]

[26] For complete details of the settlement agreement, see Attorney General of the State of New York (2015).

[27] An example is a recent data field added to a consumer credit file involving the tradeline payment amount, which reflects the actual amount paid by the consumer toward his or her current balance. Some lenders believe this information gives too much competitive insight into their customer base and do not report information for this field.

With rare exception, most lenders contribute information for all of their collateralized and uncollateralized portfolios, including credit card, mortgage, auto, and personal loan information to at least one of the three national CRAs. Some local lenders, collection agencies, and most notably public utility companies (gas and electric companies) have elected to report all critical tradeline components for all portfolios to at least one national CRA.[28]

Because reporting consumer credit information to a CRA is voluntary, one or two CRAs may lack a local lender's tradeline information. Still, given the very high market share enjoyed by national and regional lenders in most areas of consumer credit, incorporating information from consumer tradelines in estimates of current and future economic conditions is worthy of exploration.

7. Dynamic Nature of Credit Use

Consumers' use of credit is dynamic, which is one reason it can be useful for estimating economic conditions. To illustrate this point, consider that the order in which consumers prefer to pay their monthly credit obligations is not static, particularly in times of financial distress. This preference is often referred to as *consumers' payment hierarchy*. Historically, mortgages and auto loans were the first monthly bills consumers paid, with credit card debt considered a lower priority. However, as described below, this hierarchy can change. Changes in payment hierarchy may be an important signal of changes in the macroeconomy, and thus may be a useful indicator for the purposes of forecasting. At the same time, the fact that payment hierarchy can change introduces certain complications to the forecasting process.[29]

Guitart and Becker (2014) describe research from TransUnion that revealed a shift in consumers' traditional bill payment preferences beginning in 2008 such that "the propensity to go

[28] Conversations with market participants suggest the discrepancies in reporting and coverage across the national CRAs have fallen over time.

[29] One concern is reverse causation. Consumers may change their repayment behavior in response to changes in current macroeconomic conditions or expectations of future conditions. This is a common problem in much empirical and forecasting work, and there are a variety of techniques for addressing it.

delinquent on a mortgage exceeded the propensity to go delinquent on a credit card." The authors cite several contributing factors to this shift, with particular emphasis on falling home prices from their peak in 2006 and rising unemployment as a result of the recession. For many consumers, their reaction to the lost equity in their homes as home values depreciated was to redirect their resources to financial obligations that would preserve liquidity. In addition, persistent and long-lasting unemployment as a result of the recession increased consumers' reliance on more liquid forms of credit. The disparity in the timing of the consequences from default on a credit card or auto loan (account closure or vehicle repossession after 90 days of delinquency) versus nonpayment of a mortgage (depending on state laws, the foreclosure process can take years) reinforced consumers' focus on liquidity preservation and equity protection through repayment of nonmortgage obligations.[30] Earlier research from TransUnion found that for borrowers (especially subprime borrowers) in those states with larger state-level home-value depreciation and higher state-level unemployment rates, this shift in the traditional payment hierarchy toward repayment of the credit card in place of the mortgage was more pronounced.[31]

In the wake of the economic recovery, Guitart and Becker (2014) revisit TransUnion's examination of consumers' payment hierarchy by tracking consumers' payment behavior during periods of marked home price appreciation and depreciation. Using a longitudinal sample covering the period January 2008 to December 2012, the authors study the payment behavior of 48 cohorts with approximately 20 million consumers in each cohort. Based on 12-month rolling consumer group cohorts, the payment hierarchy was defined by the spread in 30 or more days past due delinquency rate between mortgages and bank cards for each cohort. The authors identify a shift in early 2008 from the traditional payment hierarchy that was driven by falling

[30] Additional evidence is found in Calem, Jagtiani, and Lang (2015), who document that longer foreclosure timelines led to improved household performance on nonmortgage consumer credit both during and after the foreclosure process. It is unclear if the affected households' consumer credit payment performance will deteriorate once their balance sheets reflect housing expenses.

[31] See Komos, Reardon, Wise, and Becker (2012).

housing prices and rising unemployment. They also observe a return to the traditional payment hierarchy beginning in 2012 that was strongly related to appreciation in housing prices. These findings suggest that consumers' payment preferences are related to changes in home prices during both downturns and upturns in the business cycle. Put more simply, consumers' decisions about debt payments are correlated with local and national economic conditions. To the extent those decisions are also based on expectations of future macroeconomic conditions, information about repayment behavior may be useful in improving the accuracy of forecasts.

8. Relationships Between Credit Variables and Macroeconomic Variables

During a 2012 Payment Cards Center workshop, researchers from TransUnion presented findings from their investigation of the correlation between anonymized aggregate credit information and certain economic conditions. For this exercise, the researchers used a TransUnion database that captured representative snapshots of consumer credit activity within a given geography and specific period of time. The database contained several hundred credit characteristics calculated from each individual consumer credit file in the sample and aggregated at the MSA, county, and state level.

The researchers evaluated the relationships between macroeconomic variables and credit bureau variables and found that predictive models for economic activity can be designed using credit variables. The data structure they used was pooled, cross-section time series data covering the two most recent recessions (2000:Q1 through 2012:Q1). The cross-sectional aspect of the data allowed researchers to explore the variation in the state and MSA dimensions of the data set.

The preliminary treatment of the data was a simple correlation analysis among a variety of macroeconomic indicators and consumer credit variables measured contemporaneously as well as lagged one and two quarters. For example, over the period from 2000 to 2011, auto loan credit limits and balances were found to be highly correlated with contemporaneous retail sales as well as values lagged one and two quarters. This step determined which consumer credit variables to include in a model based on the results of the correlation exercise.

The next step taken by the researchers was the more formal bivariate Granger causality test, which identifies the information content of the time series dimension of the data. For example, for the period 2000:Q1 to 2011:Q3, this test revealed that the retail sales variable was influenced by its recent past historical values as well as by captive auto finance credit limits with a two-quarter lead. The Granger test showed the additional information from captive auto finance data to be statistically significant: There appears to be valid information content in the past values of captive auto finance average credit limits in predicting retail sales. Looking over an 11-year period that includes the two most recent recessions, auto loan credit limits and balances were highly correlated with nominal GDP in contemporaneous quarters as well as lagged one and two quarters. Disposable income was negatively correlated with 60 or more days past due delinquency on retail and finance installment accounts while positively correlated with credit limits. Strong positive correlations exist between motor vehicles and parts sales and auto loan amounts as well as between captive auto finance credit trends and measures of employment.

The relationships between macroeconomic and credit bureau variables over time showed a divergence between state-level unemployment and captive auto finance balances since the 2001 recession, but they showed a convergence as the Great Recession neared. These preliminary studies of correlation between macroeconomic variables and credit bureau variables showed that there is information content in past values of consumer credit variables.

Based on the observation that credit variables often lead macroeconomic variables, the researchers built a model of lagged (by one period) consumer credit characteristics to determine if they could form a leading indicator to predict macro variables. A collection of consumer credit variables lagged one quarter, including auto and real estate inquiries (positive correlation), new bank auto debt to total debt (positive correlation), bank card average credit limit (positive correlation), bank card delinquency rate (negative correlation), total balance of bank revolving trades in last three months (positive correlation), and severe retail delinquency (negative correlation), was used to develop predictive models for economic activity. For illustrative

purposes, single-factor predictor indices were created for each MSA in the U.S. Based on the model's prediction of macroeconomic variables using a consumer credit variable for the current quarter and one quarter out (that is, using a one-quarter lag), researchers created a map of MSA "signals" showing expansions or contractions in the macro variable of interest; that is, an increase or decrease relative to the current predicted value. Using this method for each MSA, the model showed that retail sales signals could be developed from leading credit indicators such as changes in auto inquiries. Similarly, housing signals could be created from leading credit indicators, such as changes in real estate inquiries. Employment signals were created from leading credit indicators, such as changes in mortgage inquiries.

To understand if the overall economy in an MSA is improving or worsening and to identify forecast trends, researchers could run those models over a period of time (perhaps a year) to obtain counts of upturns or downturns in these indicators. By adjusting the structure (rule logic) of this process to look at trends in the two or three most recent quarters, changes in the direction of various signals could be used to infer future economic health, or to construct diffusion indices based on the number of MSAs in contraction, in recovery, or strong recovery. The model validation done by the researchers as the economy entered and exited the recession showed that two- or three-quarter lags were best. Stability can be introduced into the rule logic by counting up the number of positive trends in the two most recent quarters, and even greater stability can be introduced by using the number of positive trends in three most recent quarters (a more stringent test). Researchers redeveloped these models once or twice per year to determine if the two or three quarters rule logic/structure remained stable.

9. Conclusion and Suggestions for Further Research

The vast majority of houses, autos, major household appliances, and general merchandise items are purchased using some form of consumer credit (mortgage, installment loan, store revolving charge card, or bank card), and a considerable percentage of small business expenditures are often financed using personal credit. The vast majority of this credit is tracked in the data collected by

the large CRAs with a national scope. The analysis of time series credit use and repayment patterns aggregated by the consumer CRAs will reflect the local-level economic conditions described by various employment, income, consumer sentiment, and retail sales dimensions as reported by federal government agencies. The addition of aggregate consumer credit information into the process by which economic conditions are estimated potentially may improve the accuracy and reliability of these statistics, possibly leading to new metrics for gauging the health and direction of economic conditions within smaller geographies, for consumer demographic categories, and on a more frequent basis.

Access to more regular and frequent barometers of local market economic conditions potentially may assist a broad spectrum of organizations, encompassing a wide variety of investment and strategic decisions. Examples of these decisions include, but certainly are not limited to, state, county, and local municipalities estimating income tax revenues, sales tax revenues, and balancing public works projects according to economic need; and commercial organizations making inventory, capital expenditures, and marketing investment decisions.

The business sector that potentially may benefit the most from enhanced economic health metrics incorporating consumer credit data includes financial institutions that provide lending and insurance products and services to consumers and small businesses. Examples of the business decisions that may use these enhanced and new economic metrics include establishing underwriting standards according to risk levels; modifying account acquisition, portfolio, and collection management strategies; and allocation of capital for lending and business operations.

The reporting lags and revisions in macroeconomic data released through many surveys, if not properly addressed, can introduce errors in forecasting. Credit bureau data generally available to researchers are not subject to such revisions. Future research using information contained in consumer credit reports should quantify the improvement in model fit and the predictive accuracy of forecasting models using these data.

References

Attorney General of the State of New York. In the Matter of the Investigation by Eric T. Schneiderman, Attorney General of the State of New York, of Experian Information Solutions, Inc.; Equifax Information Services, LLC; and TransUnion, LLC, Respondents. (March 9, 2015).

Avery, Robert B., Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner. "Credit Risk, Credit Scoring, and the Performance of Home Mortgages," *Federal Reserve Bulletin* (July 1996), pp. 621–648.

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin. "Nowcasting," in Michael P. Clements and David F. Hendry (eds.), *The Oxford Handbook of Economic Forecasting*. Oxford: Oxford University Press, pp. 63–90 (2011).

Barron, John M., Gregory Elliehausen, and Michael E. Staten. "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* (January 2000), pp. 63–76.

Bernanke, Ben S. "Economic Measurement," speech delivered at the 32nd General Conference of the International Association for Research in Income and Wealth, Cambridge, MA (August 6, 2012).

Calem, Paul, Julapa Jagtiani, and William W. Lang. "Foreclosure Delay and Consumer Credit Performance," Federal Reserve Bank of Philadelphia Working Paper 15-24 (July 2015).

Card, David, Raj Chetty, Martin Feldstein, and Emmanuel Saez. "Expanding Access to Administrative Data for Research in the United States," American Economic Association, Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas (2011).

Consumer Financial Protection Bureau. "Key Dimensions and Processes in the U.S. Credit Reporting System: A Review of How the Nation's Largest Credit Bureaus Manage Consumer Data" (2012).

Croushore, Dean. "Frontiers of Real-Time Data Analysis," Federal Reserve Bank of Philadelphia Working Paper 08-4 (2008).

Croushore, Dean, and Tom Stark. "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105(2001), pp. 111–130.

Diebold, Francis X. "Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting: A Discussion of the Papers by Reichlin and Watson," in M. Dewatripont, L.P. Hansen, and S. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society*. Cambridge: Cambridge University Press, pp. 115–122 (2003).

Diebold, Francis X. “A Personal Perspective on the Origin(s) and Development of “Big Data”: The Phenomenon, the Term, and the Discipline,” Second Version. Penn Institute for Economic Research Working Paper 13-003 (November 2012).

Einav, Liran, and Jonathan Levin. “The Data Revolution and Economic Analysis,” in Josh Lerner and Scott Stern (eds.), *Innovation Policy and the Economy, Volume 14*. Chicago: University of Chicago Press, pp. 1–24 (2014).

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

Federal Trade Commission. *Report to Congress Under Section 319 of the Fair and Accurate Credit Transactions Act of 2003*. (January 2015).

Galbraith, John W., and Greg Tkacz. “Electronic Transactions as High-Frequency Indicators of Economic Activity,” Bank of Canada Working Paper 2007-58 (December 2007).

Galbraith, John W., and Greg Tkacz. “Nowcasting GDP: Electronic Payments, Data Vintage and the Timing of Data Releases,” CIRANO Scientific Series 2013s-25 (August 2013).

Guitart, Antoni, and Ezra Becker. “Take Heed of Consumers’ Changing Payment Hierarchy,” *RMA Journal* (December 2013–January 2014), pp. 49–54.

Hall, Robert E. “Monitoring the Financial Condition and Expenditures of Households,” in Markus Brunnermeier and Arvind Krishnamurthy (eds.), *Risk Topography: Systemic Risk and Macro Modeling*. Chicago: University of Chicago Press, pp. 175–182 (2014).

Hunt, Robert M. “What’s in the File? The Economics and Law of Consumer Credit Bureaus,” Federal Reserve Bank of Philadelphia *Business Review* (Q2 2002), pp. 17–24.

Komos, Matthew, Sean Reardon, Charlie Wise, and Ezra Becker. “Payment Hierarchy Analysis: A Study of Changes in Consumer Payment Prioritization from 2007 through 2011,” TransUnion White Paper (2012).

Laney, Doug. “3D Data Management: Controlling Data Volume, Velocity, and Variety,” META Group Inc. *Application Delivery Strategies* (February 6, 2001).

Manski, Charles F. “Communicating Uncertainty in Official Economic Statistics: An Appraisal Fifty Years after Morgenstern,” *Journal of Economic Literature* (2015), pp. 631-653.

Modugno, Michele. “Nowcasting Inflation using High Frequency Data,” *International Journal of Forecasting*, 29 (2013), pp. 664–675.

Morrison, Jeffrey S. “Credit Decisions in a Changing Economic Environment,” *RMA Journal* (September 2008), pp. 56–60.

Morrison, Jeffrey S. “Marrying Credit Scoring and Time-Series Data,” *RMA Journal* (March 2010), pp. 60–63.

Ng, Serena, and Jonathan H. Wright. “Facts and Challenges from the Great Recession for Forecasting and Macroeconomic Modeling,” *Journal of Economic Literature* (2013), pp. 1120–1154.

Podesta, John, Penny Pritzker, Ernest J. Moniz, John Holdren, and Jeffrey Zients. “Big Data: Seizing Opportunities, Preserving Values,” Executive Office of the President (2014).

Turner, Michael A., Robin Varghese, and Patrick Walker. “U.S. Consumer Credit Reports: Measuring Accuracy and Dispute Impacts.” Durham, NC: Policy and Economic Research Council (May 5, 2011).

Turner, Michael, Robin Varghese, Patrick Walker, and Sukanya Chaudhuri. “Comparing FTC and PERC Studies on Measuring the Accuracy of U.S. Consumer Credit Reports.” Durham, NC: Policy and Economic Research Council (2013).

U.S. Government Accountability Office. Report to the Chairman, Committee on Commerce, Science, and Transportation, U.S. Senate, “Information Resellers: Consumer Privacy Framework Needs to Reflect Changes in Technology and the Marketplace,” GAO-13-663. Washington, D.C.: U.S. GAO (2013).

U.S. Office of Management and Budget. Executive Order 13642: *Making Open and Machine Readable the New Default for Government Information*, Open Data Policy — Managing Information As an Asset (OMB Memorandum M-13-13). Washington, D.C.: U.S. OMB (May 9, 2013).

U.S. Office of Management and Budget. *Guidance for Providing and Using Administrative Data for Statistical Purposes* (OMB Memorandum M-14-06). Washington, D.C.: U.S. OMB (2014).

U.S. Office of Management and Budget. “Building Evidence with Administrative Data,” FY 2016 Budget Analytical Perspectives, Chapter 7. Washington, D.C.: U.S. OMB (2015).

Varian, Hal R. “Big Data: New Tricks for Econometrics,” *Journal of Economic Perspectives* (Spring 2014), pp. 3–28.

Wardrip, Keith, and Robert M. Hunt. “Residential Migration, Entry, and Exit as Seen Through the Lens of Credit Bureau Data,” Federal Reserve Bank of Philadelphia Payment Cards Center Discussion Paper 13-04 (2013).

FEDERAL RESERVE BANK OF PHILADELPHIA

Payment Cards Center Discussion Paper Series

www.philadelphiafed.org/PCC