

Credit Report Accuracy and Access to Credit

Robert B. Avery, Paul S. Calem, and Glenn B. Canner, of the Board's Division of Research and Statistics, prepared this article. Shannon C. Mok provided research assistance.

Information that credit-reporting agencies maintain on consumers' credit-related experiences plays a central role in U.S. credit markets. Creditors consider such data a primary factor when they monitor the credit circumstances of current customers and evaluate the creditworthiness of prospective borrowers. Analysts widely agree that the data enable domestic consumer credit markets to function more efficiently and at lower cost than would otherwise be possible.

Despite the great benefits of the current system, however, some analysts have raised concerns about the accuracy, completeness, timeliness, and consistency of consumer credit records and about the effects of data limitations on the availability and cost of credit. These concerns have grown as creditors have begun to rely more on "credit history scores" (statistical characterizations of an individual's creditworthiness based exclusively on credit record information) and less on labor-intensive reviews of the detailed information in credit reports. Moreover, decision-makers in areas unrelated to consumer credit, including employment screening and underwriting of property and casualty insurance, increasingly depend on credit records, as studies have shown that such records have predictive value.

A previous article in this publication examined in detail the credit records of a large, nationally representative sample of individuals as of June 30, 1999.¹ That analysis revealed the breadth and depth of the information in credit records. It also found, however, that key aspects of the data may be ambiguous, duplicative, or incomplete and that such limitations have the potential to harm or to benefit consumers.

Although the earlier analysis contributed to the debate about the quality of the information in credit records, it did not attempt to quantify the effects of data limitations on consumers' access to credit. To

date, publicly available information about the extent of data quality problems has been limited, as has research on the effects of those problems.² The lack of information has inhibited discussion of the problems and of the appropriate ways to address them.

The main reason for the lack of information is that conducting research on the effects of data limitations on access to credit is complicated. Two factors account for the complexity. First, the effects vary depending on the overall composition of the affected individual's credit record. For example, a minor error in a credit record is likely to have little or no effect on access to credit for an individual with many reported account histories, but the same error may have a significant effect on access to credit for someone with only a few reported account histories. Second, assessments of the effects of data limitations require detailed knowledge of the model used to evaluate an individual's credit history and of the credit-risk factors that compose the model. Because information about credit-scoring models and their factors is ordinarily proprietary, it is difficult to obtain.

In this article, we expand on the available research by presenting an analysis that tackles these complexities and quantifies the effects of credit record limitations on the access to credit.³ The analysis considers the credit records of a nationally representative sample of individuals, drawn as of June 30, 2003, that incorporates improvements in the reporting system over the past few years and, consequently, better reflects today's circumstances. We examine the possible effects of data limitations on consumers by estimating the changes in consumers' credit history scores that would result from "correcting" data problems in their credit records. We also investigate

2. General Accounting Office (2003), *Consumer Credit: Limited Information Exists on Extent of Credit Report Errors and Their Implications for Consumers*, report prepared for the Senate Committee on Banking, Housing, and Urban Affairs, GAO-03-1036T, July 31, pp. 1–18. In 2004, the General Accounting Office became the Government Accountability Office.

3. This analysis builds on recent research that attempted to quantify the effects of credit record limitations on the access to credit. See Robert B. Avery, Paul S. Calem, and Glenn B. Canner (2003), "Credit Reporting and the Practical Implications of Inaccurate or Missing Information in Underwriting Decisions," paper presented at "Building Assets, Building Credit: A Symposium on Improving Financial Services in Low-Income Communities," Joint Center for Housing Studies, Harvard University, November 18–19.

1. Robert B. Avery, Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner (2003), "An Overview of Consumer Data and Credit Reporting," *Federal Reserve Bulletin*, vol. 89 (February), pp. 47–73.

whether different patterns emerge when individuals in the sample are grouped by strength of credit history (credit history score range), depth of credit history (number of credit accounts in a credit record), and selected demographic characteristics (age, relative income of census tract of residence, and percentage of minorities in census tract of residence). Such segmentation allows us to determine whether the effects of data limitations differ for various subgroups of the population.

CONSUMER CREDIT REPORTS

A consumer credit report is the organized presentation of information about an individual's credit record that a credit-reporting agency communicates to those requesting information about the credit history of an individual. It includes information on an individual's experiences with credit, leases, non-credit-related bills, collection agency actions, monetary-related public records, and inquiries about the individual's credit history. Credit reports, along with credit history scores derived from the records of credit-reporting agencies, have long been considered one of the primary factors in credit evaluations and loan pricing decisions. They are also widely used to select individuals to contact for prescreened credit solicitations. More recently, credit reports and credit history scores have often been used in identifying potential customers for property and casualty insurance and in underwriting and pricing such insurance.⁴

The three national credit-reporting agencies—Equifax, Experian, and Trans Union—seek to collect comprehensive information on all lending to individuals in the United States, and as a consequence, the information that each agency maintains is vast. Each one has records on perhaps as many as 1.5 billion credit accounts held by approximately 210 million individuals.⁵ Together, these agencies generate more than 1 billion credit reports each year, providing the vast majority of the reports for creditors, employers, and insurers. One study found that con-

sumers receive only about 16 million of the credit reports distributed each year.⁶

Credit-reporting agencies collect information from “reporters”—creditors, governmental entities, collection agencies, and third-party intermediaries. They generally collect data every month, and they typically update their credit records within one to seven days after receiving new information. According to industry sources, each agency receives more than 2 billion items of information each month. To facilitate the collection process and to reduce reporting costs, the agencies have implemented procedures to have data submitted in a standard format, the so-called Metro format.⁷ Data may be submitted through various media, including CD-ROM and electronic data transfer. Reporters submit information voluntarily: No state or federal law requires them to report data to the agencies or to use a particular format for their reporting. As a result, the completeness and frequency of reporting can vary.

Using Credit Records to Evaluate Creditworthiness

In developing credit history scores, builders of credit-scoring models consider a wide variety of summary factors drawn from credit records. In most cases, the factors are constructed by combining information from different items within an individual's credit record. These factors compose the key elements of credit models used to generate credit history scores. Although hundreds of factors may be created from credit records, those used in credit-scoring models are the ones proven statistically to be the most valid predictors of future credit performance. The factors and the weights assigned to each one can vary across evaluators and their different models, but the factors generally fall into four broad areas: payment history, consumer indebtedness, length of credit history, and the acquisition of new credit.⁸

4. For purposes of insurance, the scores are typically referred to as insurance scores.

5. John A. Ford (2003), chief privacy officer of Equifax, Inc., in *Fair Credit Reporting Act: How It Functions for Consumers and the Economy*, hearing before the Subcommittee on Financial Institutions and Consumer Credit of the House Committee on Financial Services, House Hearing 108-33, 108 Cong. 2 Sess. (Washington: Government Printing Office), June 4. Also see Consumer Data Industry Association (formerly Associated Credit Bureaus), “About CDIA,” www.cdiaonline.org.

6. Loretta Nott and Angle A. Welborn (2003), *A Consumer's Access to a Free Credit Report: A Legal and Economic Analysis*, report to the Congress by the Congressional Research Service, September 16, pp. 1–14.

7. Currently, reporters may submit data in the Metro I or Metro II format. As of 2005, the Metro II format will be required for all submissions.

8. For a more detailed discussion of factors considered in credit evaluation, including the relative weights assigned to different factors, see the description on the website of Fair Isaac Corporation, www.myfico.com. Also see Robert B. Avery, Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner (1996), “Credit Risk, Credit Scoring, and the Performance of Home Mortgages,” *Federal Reserve Bulletin*, vol. 82 (July), pp. 621–48.

Payment History

The most important factors considered in credit evaluation are those that relate to an individual's history of repaying loans and any evidence of non-credit-related collections or money-related public actions. Credit evaluators consider whether an individual has a history of repaying balances on credit accounts in a timely fashion. The analysis takes into account not only the frequency of any repayment problems but also their severity (lateness), date of occurrence (newness), and dollar magnitude. Evaluators assess repayment performance on the full range of accounts that an individual holds, distinguishing accounts by type (such as revolving, installment, or mortgage) and by source (such as banking institution, finance company, or retailer). In general, an individual with serious deficiencies in repayment performance, such as a credit account that is currently delinquent, will find qualifying for new credit difficult, may face higher interest rates for the credit received, or may be limited in further borrowing on existing revolving accounts.

Consumer Indebtedness

When evaluating credit, creditors consider the type and amount of debt an individual has and the rate of credit utilization. For revolving accounts, the rate of credit utilization is measured as the proportion of available credit in use (outstanding balance divided by the maximum amount the individual is authorized to borrow, referred to as the credit limit). For installment and mortgage accounts, credit utilization is generally measured as the proportion of the original loan amount that is unpaid. High rates of credit utilization are generally viewed as an additional risk factor in credit evaluations, as they may indicate that an individual has tapped all available credit to deal with a financial setback, such as a loss of income.

Length of Credit History

Credit evaluators consider the length of a person's credit history because it provides information about how long the individual has been involved in credit markets and about whether he or she has obtained credit recently. The age of the account is relevant to an evaluation of credit quality because the longer the account has been open, the more information it con-

veys about an individual's willingness and ability to make payments as scheduled. New accounts may convey little information other than that a consumer has had a recent need for additional credit and has been approved for credit.

Acquisition of New Credit

Whether a person is seeking new credit provides information about the credit risk posed by the individual. The number of new accounts the individual has recently established and the number of attempts to obtain additional loans, as conveyed by records of recent creditor inquiries (requests for credit reports), all provide a picture of the individual's recent credit profile.⁹ Attempts to open a relatively large number of new accounts may signal that a person risks becoming overextended.

Calculating a Credit History Score

Statistical modelers working with data from credit-reporting agencies construct credit history scores using selected factors of the types described above. Modelers divide each factor into ranges and assign each range a point count. The score for an individual is the sum of these points over all factors considered in the model. Typically, the points and the factors used in the model are derived from a statistical analysis of the relationship between the factors at an initial date and the credit performance over a subsequent period.

Role of the Fair Credit Reporting Act

Although participation by reporters in the credit-reporting process is voluntary, reporters are subject to rules and regulations spelled out in the Fair Credit Reporting Act (FCRA). The FCRA regulates access to credit information and prescribes how the agencies are to maintain each credit report they hold.¹⁰ Under the FCRA, only persons with a permissible pur-

9. Inquiries made to create a mailing list for sending prescreened solicitations or to monitor existing account relationships are omitted from the credit reports. Also omitted are individuals' requests for copies of their own reports.

10. For a discussion of how the FCRA governs and encourages accurate credit reporting, see Michael Staten and Fred Cate (2003), "Does the Fair Credit Reporting Act Promote Accurate Credit Reporting?" paper presented at "Building Assets, Building Credit: A Symposium on Improving Financial Services in Low-Income Communities," Joint Center for Housing Studies, Harvard University, November 18–19.

Provisions of the Fair and Accurate Credit Transactions Act of 2003

The Fair and Accurate Credit Transactions Act of 2003 amended the Fair Credit Reporting Act in several ways. The amendments, known collectively as the FACT Act, seek to (1) improve the use of credit information and give consumers greater access to such information, (2) prevent identity theft and facilitate credit history restitution, (3) enhance the accuracy of consumer report information, (4) limit the sharing and use of medical information in the financial system, and (5) improve financial literacy and education.

The amendments that address the use and availability of credit information provide the following consumer rights and protections:

- **The right to obtain a free copy of a consumer report.** A consumer may request a free credit report once a year from each of the national credit-reporting agencies, and each agency must establish a toll-free telephone number to receive the requests. A consumer may also obtain a credit history score and related information from each agency for a “fair and reasonable” fee. For a given credit history score, related information includes the range of possible scores under the model used to produce the score, a list of the key factors (not to exceed four) that adversely affected the score, the date the score was established, and the name of the entity that provided the score.

- **The right to be told when, as a result of negative information in a credit report, a creditor has offered a consumer credit on terms that are materially less favorable than those offered to most other consumers.** At the time of notification, the creditor must provide a statement that explains the consumer’s right to obtain a free credit report from a credit-reporting agency and that provides contact information for obtaining the report (as of this writing, the rules for implementing this provision were not yet final).

- **Protection against faulty reporting of credit record data.** Federal supervisors of financial institutions must establish and maintain guidelines regarding the accuracy and integrity of the information that data reporters submit to credit-reporting agencies. In certain circumstances, a data reporter must reinvestigate a dispute involving the information it reported.

pose for obtaining a credit report—for example, to facilitate a credit transaction, to screen prospective employees, or to underwrite property and casualty insurance involving a consumer—may have access to this credit information. The FCRA prohibits a

from furnishing information credit-

11. About 85 percent of the credit reports that consumers receive each year are associated with adverse actions. See Nott and Welborn, *A Consumer’s Access to a Free Credit Report*, p. 10.

12. For example, if a reporter submits a file that includes a much larger or a much smaller number of records than have historically been received, then the agency will flag the file for review. Similarly, if an unexpectedly large or an unexpectedly small percentage of the data items have a given characteristic (for example, the number of accounts sixty or more days late exceeds a designated threshold), then the agency will also flag the data for review.

potential for error. For example, because data reporting is voluntary and because the ability of the agencies to enforce certain standards is limited, the agencies have had to devise techniques for recognizing that sometimes data items reported with the same identifying information, such as the same name, may actually be associated with different individuals. Similarly, a social security number may be missing from or may be reported incorrectly in reported information on an individual. In such cases, the likelihood of associating the reported item with the wrong person increases significantly.

Although the agencies' data are extensive, they are incomplete in two respects. First, not all information on credit accounts held by individuals is reported to the agencies. Some small retailers and mortgage and finance companies do not report to the agencies, and individuals, employers, insurance companies, and foreign entities typically do not report loans they extend. Also, information on student loans is not always reported. Second, some accounts that are reported contain incomplete or out-of-date information. Sometimes creditors do not report or update information on the credit accounts of consumers who consistently make their required payments as scheduled or on the accounts of those who have been seriously delinquent in their payments, particularly accounts with no change in status. Similarly, credit limits established on revolving accounts, such as credit cards, are not always reported or updated. Moreover, creditors may not notify the agencies when an account has been closed, transferred, or assigned a new payment status. For example, sometimes creditors fail to report delinquent payments that are fewer than thirty or sixty days past due, and they report changes in payment status only when a more serious payment problem arises. Each of these possibilities contributes to problems of data completeness and integrity, and each has the potential to compromise the evaluation of an individual's creditworthiness.

Another problem that may compromise credit evaluations concerns the timeliness of the data. The information reported on credit accounts reflects each account's payment status and outstanding balance as of a date shortly before the information is forwarded to the agencies. Thus, the information is sensitive to the date on which the information is forwarded. For example, a credit account reported the day after a creditor has posted a payment to the account will show a smaller balance than will the same account reported the day before the posting. Similarly, the payment status reflected in a credit report is sensitive to timing; the record on an account may indicate no

late payment problems on a given day but may show a delinquency if reported to the agency one or two days later.

Besides the accuracy, completeness, and timeliness of information in a given credit record, the consistency of information about an individual across agencies is an issue of concern. The information may differ from agency to agency for several reasons. First, the rules governing the processing of reported information differ across agencies. For example, each agency has its own rules for determining whether identifying information is sufficient to link reported information to a single individual. The inability to link reported information accurately in all cases can be an important source of data quality concerns because it results in the creation of "fragmentary files"—that is, multiple and therefore incomplete credit reports for the same individual—and sometimes in the assignment of the wrong credit records to an individual. Fragmentary files often result because consumers use different addresses or names (for example, after a marriage or a divorce), in some cases fraudulently, to obtain credit or other services. Each agency also has its own rules governing the treatment of out-of-date information, such as accounts last reported to have a positive balance. Second, the agencies receive and post information at different times. Third, a given reporter may provide information to one or two of the agencies but not to all three. Finally, changes made to disputed information may be reflected only in the credit records of the agency that received the disputed claim.

Although the agencies endeavor to maintain high-quality data and accurate files, the degree to which consumer credit reports are accurate, complete, timely, or consistent across agencies is in dispute. Moreover, analysts disagree on the extent to which data errors and omissions affect credit history scores. A recent analysis by the General Accounting Office (GAO) cites information drawn from the relatively few studies that have attempted to address data accuracy and importance.¹³ Specifically, the GAO cites a 2002 joint study by the Consumer Federation of America and the National Credit Reporting Association that found evidence that the information included in the credit reports of any given individual can differ widely across agencies.¹⁴ This study also found that credit history scores based on data from the agencies can vary substantially regardless of whether the individual has a generally good or a generally bad credit

13. General Accounting Office, *Consumer Credit*.

14. Consumer Federation of America and National Credit Reporting Association (2002), *Credit Score Accuracy and Implications for Consumers*, December 17, www.consumerfed.org.

history. As a consequence, the study concluded, “millions of consumers are at risk of being penalized by inaccurate credit report information and inaccurate credit scores.”¹⁵

The GAO report also discusses research on errors and omissions that occur within the credit files of a single agency. The report highlights different perspectives on the data quality issue. For example, one investigation by a consumer organization estimated that up to 79 percent of credit reports may contain some type of error and that about 25 percent of all consumer credit reports may contain errors that can result in the denial of access to credit.¹⁶ A study by Arthur Andersen and Company reviewing the outcomes for individuals who were denied credit and then disputed information in their credit reports concluded, however, that only a small proportion of the individuals were denied credit because of inaccurate information in their credit reports.¹⁷

THE FEDERAL RESERVE SAMPLE OF CREDIT RECORDS

The Federal Reserve Board obtained from one of the three national credit-reporting agencies the credit records (excluding any identifying personal or creditor information) of a nationally representative random sample of 301,000 individuals as of June 30, 2003.¹⁸ The sample data omitted home addresses but

included census tracts, states, and counties of residence. We used this geographic information with census 2000 files—which provide population characteristics, such as income, race, and ethnicity, by census tract of residence—to analyze the credit record data.

Four general types of credit-related information appear in credit records, including those in the Federal Reserve sample: (1) detailed information from creditors (and some other entities such as utility companies) on credit accounts—that is, current and past loans, leases, and non-credit-related bills; (2) information reported by collection agencies on actions associated with credit accounts and non-credit-related bills, such as unpaid medical or utility bills; (3) information purchased from third parties about monetary-related public records, such as records of bankruptcy, foreclosure, tax liens (local, state, or federal), lawsuits, garnishments, and other civil judgments; and (4) information about inquiries from creditors regarding an individual’s credit record.

Credit accounts constitute the bulk of the information in the typical individual’s credit record, and thus they compose the bulk of the information that the agencies maintain. Credit account records contain a wide range of details about each account, including the date that an account was established; the type of account, such as revolving, installment, or mortgage; the current balance owed; the highest balance owed; credit limits if applicable; and payment performance information, such as the extent to which payments are or have been in arrears for accounts in default.

A basic element of agency data is information on the open or closed status of each account. An account is considered open if a credit relationship is ongoing and closed if the consumer can no longer use the account. Another important element of account information is the date on which the information was most recently reported. The date is critical in determining whether the information on the account in the credit agency files is current or stale (unreported for some time and therefore potentially in need of updating).

Significantly less-detailed information is available on collection agency accounts, public records, and creditor inquiries about a consumer’s credit history. Generally, only the amount of the collection or public record claim, the name of the creditor, and the date last reported are available. For creditor inquiries, information is even more limited and includes just the type of inquirer and the date of the inquiry. The

15. Consumer Federation of America and National Credit Reporting Association, *Credit Score Accuracy and Implications for Consumers*. The study found that the difference between the high and the low credit history scores for an individual across the three agencies averaged 41 points (on a scale of 300 to 850) and that about 4 percent of individuals had score differences of 100 points or more.

16. Alison Cassady and Edmund Mierzwinski (2004), *Mistakes Do Happen: A Look at Errors in Consumer Credit Reports*, National Association of State Public Interest Research Groups, June, www.uspirg.org. Also see Jon Golinger and Edmund Mierzwinski (1998), *Mistakes Do Happen: Credit Report Errors Mean Consumers Lose*, U.S. Public Interest Research Group, March, www.uspirg.org.

17. Consumer Data Industry Association (1998), press release, March 12, www.cdiaonline.org. Also see Robert M. Hunt (2002), “The Development and Regulation of Consumer Credit Reporting in America,” Working Paper No. 02-21 (Philadelphia: Federal Reserve Bank of Philadelphia, November). The study found that 8 percent of the consumers who were denied credit requested copies of their credit reports. Of these consumers, 25 percent found and disputed errors. Of those consumers who found errors, about 12 percent (3 percent of those who requested credit reports) eventually received credit because of favorable dispute resolutions.

18. Agency files include personal identifying information that enables the agencies to distinguish among individuals and construct a full record of each individual’s credit-related activities. The records received by the Federal Reserve excluded the personal identifying information that agency files contain—the consumer’s name, current and previous addresses, and social security number—as well as other personal information that credit files sometimes contain—telephone

numbers, name of spouse, number of dependents, income, and employment information. Under the terms of the contract with the credit-reporting agency, the data received by the Federal Reserve cannot be released to the public.

1. Individuals with credit-reporting agency records, by type of information in credit record, as of June 30, 2003

Type of information in credit record	Number	Share of sample (percent)
Sample size	301,536	100.0
Credit account	259,211	86.0
Collection agency account	109,964	36.5
Public record	36,742	12.2
Creditor inquiry ¹	188,616	62.6
None of the above	15	*
MEMO		
Credit account only	63,501	21.1
Collection agency account only	34,978	11.6
Public record only	53	*
Creditor inquiry only ¹	31	*
Credit account and		
Collection agency account	67,747	22.5
Public record	34,715	11.5
Creditor inquiry ¹	182,553	60.5

NOTE. In this and subsequent tables, components may not sum to totals because of rounding.

1. Item includes only inquiries made within two years of the date the sample was drawn.

* Less than 0.5 percent.

agencies generally retain inquiry information for twenty-four months.

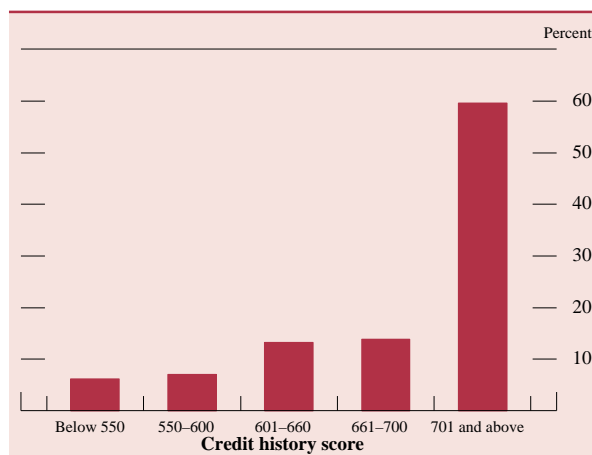
In aggregate, the Federal Reserve sample contained information on about 3.7 million credit accounts, more than 318,000 collection-related actions, roughly 65,000 monetary-related public record actions, and about 913,000 creditor inquiries. Not every individual had information of each type. In the sample, approximately 260,000, or 86 percent, of the individuals had records of credit accounts as of the date the sample was drawn (table 1).¹⁹ Although a large portion of individuals had items indicating collection agency accounts, public record actions, or creditor inquiries, only a very small share (well less than 1 percent) of the individuals with credit records had only public record items or only records of creditor inquiries. However, for about 12 percent of the individuals, the only items in their credit records were collection actions.

Credit History Scores in the Sample

The credit-reporting agency provided credit history scores for about 250,000, or 83 percent, of the individuals in the sample. The agency used its propri-

19. The credit account information was provided by 92,000 reporters, 23,000 of which had reported within three months of the date the sample was drawn.

1. Distribution of individuals, by credit history score



NOTE. Data are from a Federal Reserve sample drawn as of June 30, 2003. The distribution is composed of individuals in the sample who had been assigned credit history scores. Authors have adjusted the scores, which are proprietary, to match the distribution of the more familiar FICO credit history scores, developed by Fair Isaac Corporation.

etary credit-risk-scoring model as of the date the sample was drawn to generate the scores (one for each individual), which it constructed from selected factors of the type described previously. The proprietary credit-risk score is like other commonly used consumer credit history scores in that larger values indicate greater creditworthiness. The agency did not assign scores to anyone who did not have a credit account. A small proportion of individuals without scores did have credit accounts, but most of these individuals were not legally responsible for any debt owed.

To facilitate this discussion, we have adjusted the proprietary credit-risk scores assigned to individuals in the Federal Reserve sample to match the distribution of the more familiar FICO credit history scores, for which information is publicly available.²⁰ Among the individuals in our sample who had scores, about 60 percent had adjusted scores of 701 or above (chart 1). Individuals with FICO scores in this range are relatively good credit risks. According to Fair Isaac Corporation, less than 5 percent of such con-

20. For a national distribution of FICO scores, see www.myfico.com/myfico/creditcentral/scoringworks.asp. All three agencies use versions of the FICO score, which is generated from software developed by the Fair Isaac Corporation. Each agency gives the score a different name. Equifax calls it the Beacon score; Experian, the Experian/Fair Isaac Risk score; and Trans Union, the Empirica score. In developing the scores, Fair Isaac used the same methods at each agency but estimated the FICO model differently at each one, using separate samples. Thus, just as the information about an individual can differ across the three companies, so can the FICO model.

sumers are likely to become seriously delinquent on any debt payment over the next two years.²¹ In contrast, about 13 percent of individuals in our sample had adjusted scores at or below 600. According to Fair Isaac, more than half of these consumers are likely to become seriously delinquent on a loan over the next two years.

Because credit history scores can be used to measure credit risk, creditors use them, along with other measures of creditworthiness, such as collateral, income, and employment information, to determine whether to extend credit and, if so, on what terms. Credit history scores are closely aligned with the interest rates offered on loans—that is, higher scores are associated with lower interest rates. For example, as of August 30, 2004, the national average interest rate for a thirty-year fixed-rate conventional mortgage for an individual with a FICO score of 720 or more was 5.75 percent, whereas the average interest rate for someone with a score below 560 was 9.29 percent.²²

Assessing the Effects of Data Limitations

The analysis to assess the potential effects of data limitations on an individual's access to credit involves two steps: identifying data problems in an individual's credit record and simulating the effects of "correcting" each problem on the availability or price of credit as represented by the change in the individual's credit history score. To conduct this exercise, one must know (1) the factors used to construct the score, (2) the points assigned to these factors in deriving an individual's score, and (3) the process used to create the underlying factors from the original credit records.

The Federal Reserve's sample includes all the information that would be necessary to construct any credit history score and its underlying factors from the original credit records. However, the details of the credit-reporting agency's credit-scoring model, including the factors and point scales used in the model, are proprietary and were not made available to the Federal Reserve. Nevertheless, we were able to approximate the model by using three types of infor-

mation: (1) the proprietary credit-risk score assigned to each individual in our sample; (2) a large set of credit factors for each individual—a subset of which was known to comprise the factors used in the proprietary credit-scoring model; and (3) detailed account-level information in each individual's credit record. We used the first two items to construct an approximation of the proprietary credit-scoring model, employing regression techniques to estimate the points to assign to each factor. We used the second and third items to "reverse-engineer" the credit factors included in our version of the credit-scoring model. This information enabled us to recalculate how the factors—and ultimately the credit history scores—would change if alterations were made to the underlying credit records so that we could simulate the effects of correcting a data problem or omission.

Because of the numerous potential factors and specifications that could have been used to construct the proprietary credit-risk score, our version of the credit-scoring model undoubtedly differs from the actual proprietary model. However, we were able to identify almost exactly the process used to construct the factors in the actual model from the underlying credit records. Moreover, the approximated and actual model scores corresponded quite closely. Thus, we believe that our approximation of the scoring process provides a reasonable estimate of the potential effects of a change in a credit record item on an individual's credit history score.

Other model builders consider different credit-risk factors in creating their scoring models, assign different points to the factors, and employ different rules for constructing the factors. As a consequence, even if we had identified the proprietary model exactly, the results of our analysis would not necessarily have been the same as those implied by other models. Nevertheless, our results should be viewed as indicative of the implications of data quality issues for scoring models in general and as applicable in many, if not all, respects.

DATA QUALITY ISSUES

As noted earlier, a previous article in this publication examined in detail the credit records of a sample of individuals as of June 30, 1999, and found that key aspects of the data were ambiguous, duplicative, or incomplete. The article highlighted four areas of concern: (1) The current status of "stale" accounts, which show positive balances (amounts owed that are greater than zero) but are not currently reported, is ambiguous; (2) some creditors fail to report

21. The term "seriously delinquent" means falling behind on a loan payment ninety days or more, defaulting on a loan, or filing for bankruptcy.

22. See www.myfico.com. Loan rate includes 1 discount percentage point and is based on a loan amount of \$150,000 for a single-family, owner-occupied property and on an 80 percent loan-to-value ratio. As the data on the web site show, interest rates vary little by credit history score for individuals with scores above 700.

credit account information, including nonderogatory accounts (accounts whose payments are being made as scheduled) or minor delinquencies (accounts 30 to 119 days in arrears); (3) credit limits are sometimes unreported; and (4) the reporting of data on collection agency accounts and public records may be inconsistent or may contain redundancies, and some of the items regarding creditor inquiries are often missing. Our simulations, discussed below, address these areas of concern.

Ambiguous Status of Stale Accounts

A primary concern about data quality involves stale accounts. About 29 percent of all accounts in the sample showed positive balances at their most recent reporting, but the report date was more than three months before the sample was drawn. These accounts fell into one of three categories based on their status when last reported: major derogatory (accounts that are 120 days or more in arrears and involve a payment plan, repossession, charge-off, collection action, bankruptcy, or foreclosure), minor delinquency, or paid as agreed. Of all stale accounts with a positive balance at last report, about 15 percent were reported to be major derogatories, 3 percent were minor delinquencies, and 82 percent were paid as agreed.

Analysis of the credit records in the sample suggests that many of these stale accounts, particularly those involving mortgages and installment loans, were likely to have been closed or transferred but were not reported as such. Many were reported by creditors that were no longer reporting data to the agency about any individuals when the sample was drawn, and thus information on these accounts was unlikely to be up to date. The significant fraction of positive-balance stale accounts that were likely closed or transferred implies that some consumers will show higher current balances and a larger number of open accounts than they actually hold.

Because the current status of stale accounts is often unclear, users of consumer credit reports must obtain additional information or make assumptions about the status. In credit-scoring models, such assumptions are inherent in “stale-account rules” that credit modelers typically apply when they calculate an individual’s credit history score. A stale-account rule defines the period for which reporting is considered current and thus identifies stale accounts. The rule also dictates how accounts identified as stale should be treated. In most cases, the rule treats them as closed accounts with zero balances.

To some extent, rules that consider stale accounts closed and paid off may mitigate concerns about stale account information. Another possible mitigating factor is that consumers who review their credit reports for mistakes are likely to catch stale-account errors and to have them corrected. Nevertheless, stale-account rules and consumer action can only partially correct the problem of noncurrent information in credit account records. For example, a rule that is conservative in identifying stale accounts may permit noncurrent information to be used over an extended period, whereas an overly aggressive rule may nullify information that is still current.

Failure to Report Credit Account Information

Some reporters provide incomplete performance information on their accounts, and others fail to report any information about some credit accounts. For example, in the Federal Reserve sample, 2.7 percent of the large creditors reported only credit accounts with payment problems.²³ The failure to report accounts in good standing likely affected the credit evaluations of consumers with such accounts. The way in which credit evaluations are affected depends on the circumstances of an account. For consumers with a low utilization of nonreported accounts, the failure to report may worsen their credit evaluations. For consumers with a high utilization of nonreported accounts, however, the failure to report may result in better credit evaluations than are warranted.

In addition, some creditors report minor delinquent accounts as performing satisfactorily until the accounts become seriously delinquent. Almost 6 percent of the large creditors in the Federal Reserve sample followed this practice. Because the credit histories for consumers who fall behind in their payments to such lenders appear somewhat better in the credit records than they actually are, these consumers may benefit from such underreporting.

Finally, some lenders withhold account information. For example, in 2003, Sallie Mae, the nation’s largest provider of student loans, decided to withhold information on its accounts from two of the three credit-reporting agencies. Clearly, while this policy was in effect, the failure to report information harmed some consumers and benefited others depending on

23. Some lenders, particularly those that specialize in lending to higher-risk individuals (referred to here as subprime lenders), choose to withhold positive performance information about their customers for competitive advantage.

whether the withheld information was favorable or unfavorable.

Unreported Credit Limits

A key factor that credit evaluators consider when they assess the creditworthiness of an individual is credit utilization. If a creditor fails to report a credit limit for an account, credit evaluators must either ignore utilization or use a substitute measure such as the highest-balance level—that is, the largest amount ever owed on the account. Substituting the highest-balance level for the credit limit generally results in a higher estimate of credit utilization because the highest-balance amount is typically lower than the credit limit; the higher estimate leads, in turn, to a higher perceived level of credit risk for affected consumers.

For the June 30, 1999, sample of individuals, proper utilization rates could not be calculated (the highest-balance levels had to be used) for about one-third of the open revolving accounts because the creditors had not reported the credit limits. At that time, about 70 percent of the consumers in the sample had missing credit limits on one or more of their revolving accounts. Circumstances have improved substantially since then because public and private efforts to encourage the reporting of credit limits have resulted in more-consistent reporting. Nevertheless, in the sample drawn as of June 30, 2003, credit limits were missing for about 14 percent of revolving accounts, and the omissions affected about 46 percent of the consumers in the sample. Thus, although the incidence of missing credit limits has fallen substantially, it remains an important data quality issue.

Problems with Collection Agency Accounts, Public Records, and Creditor Inquiries

Data on collection agency accounts, public records, and creditor inquiries are a source of inconsistency, redundancy, and missing information in credit records.

Collection Agency Accounts

Evidence suggests that collection agencies handle claims in an inconsistent manner. Most notably, some collection agencies may report only larger collection amounts to credit-reporting agencies, whereas others

may report claims of any size.²⁴ Inconsistent reporting does not imply inaccuracy of the information that does get reported, but it does imply some arbitrariness in the way individuals with collections are treated. Those whose collection items happen to be reported to the credit-reporting agency will have lower credit history scores than will those whose collection items go unreported. This situation raises the question as to the extent and effect of such arbitrary differences in treatment, particularly for small collection amounts. In addition, anecdotes abound about consumers who have had difficulty resolving disputes over collection items or who have had trouble removing erroneous items from their credit records.

Another potentially important data quality issue for collection agency accounts is duplication of accounts within collection agency records. Duplications can occur, for example, when a collection company transfers a claim to another collection company. Duplications can also occur when a debt in collection is satisfied but the paid collection is recorded as a separate line item by the collection agency. Analysis of the collection agency accounts in the latest Federal Reserve sample suggests that about 5 percent of collection items are likely duplications resulting from such transfers or payouts.

Credit evaluators also have some concern about the appropriateness of using medical collection items in credit evaluations because these items (1) are relatively more likely to be in dispute, (2) are inconsistently reported, (3) may be of questionable value in predicting future payment performance, or (4) raise issues of rights to privacy and fair treatment of the disabled or ill. The last concern recently received special attention with the inclusion of provisions in the FACT Act that address medical-related collections. One provision requires the credit-reporting agencies to restrict information that identifies the provider or the nature of medical services, products, or devices unless the agencies have a consumer's affirmative consent. In the future, the agencies may be able to meet this requirement by using a code, with the name of the creditor suppressed, to distinguish medical-related collections from other collections. Because the coding system is prospective, however, even if implemented today, years may pass before all the collection items in the agency files have this code. In the interim, if the name of the creditor is suppressed, distinguishing medical collection items

24. One indication of the inconsistent reporting of collection items is the wide dispersion across states in the ratio of small collection items to all collection agency accounts. The percentage ranges from 30 percent to 60 percent.

will depend on the ability of the credit-reporting agencies to mechanically code historical data. If such coding is done imperfectly, it may adversely affect consumers who deal with creditors that want to discount collection items involving medical incidents. (As of September 2004, at least one of the agencies had developed a system that suppresses the name of the creditor and uses a code to distinguish medical-related collections.)

Public Records

Public records suffer from similar consistency and duplication problems that affect collection items. In particular, a single episode can result in one or more public record items depending on how it is recorded. For example, tax liens can be recorded on a consolidated basis or treated as separate items. Similarly, amendments to a public record filing, such as a bankruptcy or a foreclosure, can be treated as updates, which result in no change in the number of items, or as new filings.

In addition, evidence suggests that the credit-reporting agencies inconsistently gather information on lawsuits that the courts have not yet acted on, in part because some agency officials believe that the mere filing of a lawsuit does not necessarily relate to future credit performance. For the most part, such lawsuits are missing from the public records. However, for idiosyncratic reasons, some lawsuits have been reported in nonrandom ways. Specifically, 80 percent of the lawsuits in the Federal Reserve sample came from only two states, an indication that residents of these states may be at a disadvantage in credit evaluations.

About one-fourth of non-bankruptcy-related public records reflect dismissals. In such cases, the courts seem to have determined that the individuals are not legally liable. Such information may be of questionable value for credit evaluations.

Creditor Inquiries

Although credit evaluators use information on creditor inquiries to predict future loan performance, the value of this information is limited in an important way. Ideally, credit evaluators would use such information to distinguish the consumers who are seeking multiple loans to greatly expand their borrowing from the consumers who are shopping for the best terms for a single loan. However, the information that evaluators need to make this distinction—that is, a

code that identifies the type of credit sought from the inquiring lender—is generally not available in inquiry records (it is missing from 99 percent of the inquiries in the Federal Reserve sample). Consequently, credit evaluators must use less reliable rules, potentially harming consumers who are simply shopping for a single loan by failing to distinguish them sufficiently from consumers who are seeking an excessive amount of credit.

DESIGN OF THE SIMULATIONS

We designed a series of simulations to estimate the potential effects of the data quality issues identified in the preceding section. Each simulation identified a set of “data problems” or potential problems, applied a plausible “correction” to each problem, and used an approximation of the proprietary credit-risk model to evaluate the effect of the correction on the credit history scores of individuals who had the problem in their credit records.²⁵ We estimated how many consumers each data problem affected; and for those who were affected, we estimated how many would see a decrease or an increase in their scores and by how much when the problem was corrected.

Selecting Factors in the Approximated Model

The first step in setting up the simulations was selecting the factors to be used in the approximated credit-scoring model. The approximated model used seventy-three factors, including the number of credit accounts of different types and the various characterizations of payment history patterns, such as the number of accounts with all payments made on time, in various stages of delinquency, or with major derogatory status. Also included were measures of outstanding balances, credit limits on revolving accounts, ages of credit accounts, variables derived from collection agency accounts and public records, and account inquiry information. Our discussions with credit evaluators suggested that most credit history models are based on a smaller number of factors than were included here. However, most of the “additional” variables in our model were decompositions or interactions that involved more general factors and were unlikely to lead to significant distortions in our representations of the effects of data quality issues.

25. We use the terms “data problem” and “correction” in their broadest sense. For example, “data problem” may mean an actual problem or only a potential problem. Similarly, “correction” may mean a solution to a problem or simply a “best guess” at a solution.

2. Share of individuals with selected factors used in credit evaluation, distributed by type of account

Percent except as noted

Factor used in credit evaluation	Factor used in credit evaluation	Type of account			
		Revolving	Installment	Mortgage	Total
<i>Number of credit</i>	<i>Number of credit accounts</i>				
No account	30 days past due in past				
1	12 months				
2	0	n.a.	n.a.	n.a.	75
3-5	1	n.a.	n.a.	n.a.	13
6-8	2	n.a.	n.a.	n.a.	5
9 or more	3 or more	n.a.	n.a.	n.a.	7
Total	Total	n.a.	n.a.	n.a.	100
<i>Number of open cr</i>	<i>Number of credit accounts</i>				
<i>accounts paid as</i>	60 days past due in past				
0	12 months				
1	0	n.a.	n.a.	n.a.	82
2	1	n.a.	n.a.	n.a.	10
3-5	2	n.a.	n.a.	n.a.	4
6-8	3 or more	n.a.	n.a.	n.a.	4
9 or more	Total	n.a.	n.a.	n.a.	100
Total					
<i>Number of credit</i>	<i>Number of credit accounts</i>				
<i>opened in most-recent</i>	90 days past due in past				
<i>12 months</i> ¹	12 months				
0	0	n.a.	n.a.	n.a.	86
1	1	n.a.	n.a.	n.a.	8
2 or more	2	n.a.	n.a.	n.a.	3
Total	3 or more	n.a.	n.a.	n.a.	3
	Total	n.a.	n.a.	n.a.	100
<i>Years since most-r</i>	<i>Number of credit accounts</i>				
<i>credit account opened</i>	more than 90 days past due				
0	0	n.a.	n.a.	n.a.	68
Less than 1	1	n.a.	n.a.	n.a.	11
1-2	2	n.a.	n.a.	n.a.	6
3-4	3 or more	n.a.	n.a.	n.a.	15
5 or more	Total	n.a.	n.a.	n.a.	100
Total					
<i>Age of oldest credit</i>	<i>Worst delinquency ever on</i>				
<i>(years)</i> ²	<i>credit account (number of</i>				
No oldest account	<i>days delinquent)</i>				
Less than 1	0	n.a.	n.a.	n.a.	51
1-4	30	n.a.	n.a.	n.a.	12
5-9	60	n.a.	n.a.	n.a.	5
10 or more	90	n.a.	n.a.	n.a.	2
Total	120	n.a.	n.a.	n.a.	4
	More than 120	n.a.	n.a.	n.a.	26
	Total	n.a.	n.a.	n.a.	100
<i>Amount owed on</i>	<i>Balance owed on collection</i>				
<i>nonmortgage credit</i>	<i>accounts (dollars)</i>				
<i>(dollars)</i>	No collection account or				
0	zero balance owed	73
1-499	1-99	2
500-999	100-499	9
1,000-4,999	500-999	5
5,000-9,999	1,000 or more	11
10,000 or more	Total	100
Total					
<i>Utilization rate for</i>	<i>Number of public records</i>				
<i>accounts (percent)</i>	0	86
No account or not	1	9
calculable	2 or more	5
0	Total	100
1-24					
25-49	<i>Number of creditor inquiries</i>				
50 or more	<i>in past 6 months</i>				
Total	0	55
	1	20
<i>Share of individuals</i>	2	11
<i>credit accounts never</i>	3	6
<i>delinquent</i>	4 or more	8
0	Total	100
1-20					
21-60					
61-90					
91 or more					
Total					

NOTE. Data include only individuals with at least one credit account (of any type) and a credit history score.

1. Data for revolving accounts include only bank-issued credit cards.
2. Data for installment accounts include only bank-issued installment loans.
3. Utilization rate is the proportion of available credit in use (outstanding balance divided by the credit limit—that is, the maximum amount an individual

is authorized to borrow). The rate cannot be calculated in all cases because of unreported information on credit limit, highest balance, or outstanding balance.

- ... Not applicable.
- n.a. Not available.

We report many of the factors used in our model and show the distribution of individuals in the sample across each factor (table 2). For example, more than 60 percent of the individuals in the sample who had a record of a credit account had information on nine or more accounts, and more than half the individuals had opened at least one new account within twelve months of the date the sample was drawn. The patterns show that payment performance varies greatly among individuals: Although about two-thirds of individuals had never been more than ninety days past due on a credit account, 15 percent had been this late on three or more accounts. In addition, nearly 15 percent had a record of at least one bankruptcy, tax lien, or other monetary-related public action, each of which weighs heavily in credit evaluations.

Estimating the Approximated Model

To estimate our approximation of the proprietary credit-scoring model, we used standard statistical regression techniques to fit the actual proprietary credit-risk score against the selected credit factors for the individuals in the sample data. Although credit modelers typically break factors into ranges, because we did not know the break points that had been selected, we approximated the process with linear splines.²⁶ For the estimation, the sample included only individuals with proprietary credit-risk scores who had not filed for bankruptcy. Our simulations were also restricted to this sample.²⁷

We estimated the regression equation separately for three subpopulations. The first group consisted of individuals with one or more major derogatory credit accounts in their credit records. Both the second and third groups consisted of individuals who had no major derogatory accounts, but individuals in the second group had no more than two credit accounts whereas those in the third group had more than two credit accounts. We conducted the analysis in this way because allowing the estimated coefficients to

26. The use of linear approximations rather than ranges is likely to mean that our simulations implied more small but consistent changes in credit history scores when factors were altered than would the “true” model, which divides consumers into two groups: those whose scores did not change because they stayed within the same range and those whose scores changed more substantially because they moved to a different range.

27. Although individuals who had filed for bankruptcy or did not have a proprietary credit-risk score were excluded from our analysis, these individuals may also have been affected by data quality problems. However, because they had not been scored or they had filed for bankruptcy, they were likely subject to a different type of credit review process, one that may have provided greater opportunities for the loan underwriter to identify and address data quality problems.

differ across population subgroups provided a noticeably better fit. The approach was also consistent with the common industry practice of using different “scorecards” for different subpopulations. The R^2 (a statistic characterizing how well a model fits the data) for each of the three subpopulation regressions was about 0.85, and the combined R^2 for the full population was about 0.94.

Proprietary considerations constrain our ability to report details of the regression equation specification or the coefficient estimates. However, a few variables in the estimated credit-scoring model were statistically insignificant and sometimes exhibited an unexpected relationship to the credit history score. As a consequence, as will be seen below, simulations of the effects of changes in an individual’s credit record led in a few instances to anomalous outcomes in the sense that some scores moved in unexpected directions when changes in the individual’s credit record were simulated.

Conducting the Simulations

As noted, the simulations identified problems in the data and applied hypothetical corrections to them. Only in the case of missing credit limits, however, could we identify the problem unambiguously. In other cases—specifically, stale accounts and the data quality issues associated with collections, public records, and inquiries—we could determine only that the information was *likely* inaccurate, incomplete, or of questionable value.²⁸ Finally, in other situations, a data problem was unobservable, such as when accounts were unreported or inconsistently reported. In these situations, we could identify only the potential effect on credit history scores of correcting the problem but not the proportion of people affected.

We conducted fifteen simulations: three that addressed issues related to stale credit accounts, four that pertained to nonreported credit account information, and eight that addressed data quality issues for collection agency accounts, public record items, and creditor inquiries.

Stale Accounts Last Reported as Paid as Agreed or as Minor Delinquencies

Recognizing the prevalence of stale accounts in credit records, most credit-scoring modelers apply stale-

28. In the case of stale accounts, the information was clearly outdated. In the case of inquiries, the information was incomplete in that we could not determine whether the inquiries were associated with shopping for a single loan.

account rules to such accounts when they develop credit evaluation models. For credit accounts that have never been in major derogatory status (paid-as-agreed accounts or accounts with only minor delinquencies recorded), the rules typically retain the historic information on payment performance but dictate that certain accounts that have gone unreported for an extended period no longer have balances outstanding. Any balances shown at last report for these accounts are reset to zero.

In reverse-engineering the factors used in this analysis, we discovered that the credit-reporting agency had imposed a one-year stale-account rule when it created most factors related to paid-as-agreed accounts and to accounts with only minor delinquencies. Our simulation examined the effects on these accounts of a more-aggressive stale-account rule, one that redefined stale accounts on the basis of a three-month period for current reporting.²⁹

Stale Accounts with Major Derogatories

Some stale accounts were last reported in major derogatory status. Here the payment status was more likely to have remained the same since the last report than it was in the case of stale accounts that were paid as agreed or showed only minor delinquencies at last report. Many seriously delinquent accounts can remain in that state for an extended period with no change in status (and thus the account information need not be updated). However, in several situations, the reported account status is likely to be no longer accurate, such as when a consumer has taken out a new mortgage after the date on which the stale major derogatory was last reported. Generally, a mortgage lender will not extend a new loan until a consumer pays off (or otherwise addresses) all major derogatories. Another situation in which the reported account status is likely to be inaccurate is when the account creditor no longer reports about any individuals. In this case, the account has probably been paid off or transferred.

We evaluated the effect of non-updating of credit account information in these situations by treating as paid off all stale major derogatories for which (1) the consumer had taken out a new mortgage after the date on which the major derogatory was last reported

or (2) the creditor for the derogatory account had not reported information on any consumer within three months of the date on which the sample was drawn. The credit-reporting agency had imposed a one-year stale-account rule when it created factors related to major derogatory accounts. The rule implied that paying off a major derogatory account that had not been reported within a year generally would have no effect on an individual's credit history score. Thus, we again restricted our analysis of the effect of stale accounts to those that had last been reported three to twelve months before the date on which the sample was drawn.

Failure of Some Subprime Creditors to Report Accounts

As a potential source of data inaccuracy, the failure of some subprime creditors (lenders that specialize in loans for high-risk individuals) to report accounts differs from the others studied here in that non-reporting is by definition unobservable. Consequently, the task for researchers is conceptually more difficult, and simulations cannot address the incidence of such nonreporting. To simulate the potential effect of such creditor behavior, we chose a random, never-delinquent mortgage, installment, or revolving account at a subprime lender for each individual with such an account and rescored the individual as if the account had not been reported. We defined subprime lenders as those that were reporting credit accounts as of the date the sample was drawn and for which more than one-half of their customers in the sample had credit history scores in the high-risk range (a score below 600).

Failure of the Largest Student Loan Creditor to Report Any Accounts

As noted above, in 2003 Sallie Mae stopped reporting information on its accounts to two of the three largest credit-reporting agencies. Moreover, Sallie Mae asked that the agencies suppress all historic information on the accounts it had previously reported. By the time the Federal Reserve sample was drawn, Sallie Mae had reversed its initial decision. Our sample omits information that would allow us to identify Sallie Mae specifically. Thus, to approximate the potential effect of Sallie Mae's original decision, we deleted information on the loans of random student-loan lenders—representing approximately the same number of student loans that Sallie

29. Analysis of the patterns of verification showed that the vast majority of open accounts were verified by the reporter every month or two. Thus, in choosing a three-month rule, we simulated the effect of a maximally aggressive stale-account rule on the likely inaccuracy associated with the account information. We had no obvious way of simulating the effect of lengthening the time period.

Mae stopped reporting—from the credit records in the Federal Reserve sample, and we rescored the affected individuals.

Failure of Some Creditors to Report Minor Delinquencies

Our review of the sample indicates that a small percentage of lenders fail to report that paid-as-agreed accounts have become minor delinquencies. Rather, the lenders report the accounts as paid as agreed until the accounts become major derogatories. To simulate the potential effects of unreported minor delinquencies, for each individual we randomly selected a currently reported account that was not in major derogatory status, was associated with a lender that did report minor delinquencies for each individual, and had been thirty or sixty days delinquent at least once. We assigned “paying as agreed” performance status to each thirty- and sixty-day delinquency in the selected account’s performance record. This adjustment replicates what the credit record would show for a lender that reported thirty- and sixty-day minor delinquencies to be paid as agreed.

Failure of Some Creditors to Report Credit Limits on Revolving Accounts

As noted, about 14 percent of revolving credit accounts were reported without information about credit limits, affecting roughly 46 percent of the individuals in the Federal Reserve sample. Therefore, credit evaluators must use other means to derive credit utilization rates for these individuals. The most common approach (and the one that model developers customarily use for credit-risk factors) is to substitute the highest balance for the missing credit limit; the typical result is higher calculated utilization rates than if the credit limits had been reported.

We simulated the effects of the nonreporting of credit limits on individuals by creating an estimated credit limit for each revolving account without a reported limit. Because information on the true credit limit in these cases was missing, the simulation in effect compared our method of calculating credit utilization rates with that of the credit-reporting agency. The primary difference between the two estimation procedures is that our approach is statistically unbiased, whereas the agency’s method, which relies on the highest-balance amount, tends to be biased upward. That is, our estimates reflect the “best guess” for the missing credit limit based on other information in the individual’s credit record. Specifi-

cally, we used samples of accounts of individuals with reported credit limits to estimate a regression model that predicted the credit limits for revolving accounts with missing limits.³⁰

Duplications in Collection Agency Accounts

A review of the sample credit records suggests that some collection agency accounts may be duplicated. Duplication can occur because of changed account numbers or transfers of accounts from one collection agency to another. To address the potential effects of this problem, we conducted simulations that consolidated likely duplicated collection account records into single items. We identified simulated duplicates in two ways. One procedure was to match the collection amount and the identity of the creditor when one account was reported paid and the other unpaid. The second procedure was to identify likely account transfers that were not reported as such to the credit-reporting agencies.

Additional duplicate collection agency accounts likely exist in the data but are difficult to identify. For example, accounts that match on collection amount and identity of the original creditor but that are reported by a single agency with reporting dates that are close in time may be duplicates, but they may just as likely result from repeated missed payments of the same amount. Accounts that match on identity of the original creditor and are spaced apart in time but do not match on amount could indicate a new report filed after a partial payment was received, in which case they would involve duplication. Alternatively, they could reflect separate incidents of missed payments with the same creditor.

Inconsistent Reporting of Small Collection Agency Accounts

Analysis of collection accounts reveals that many are for very small amounts that may be inconsistently reported. Recognizing this possibility, some credit evaluators choose to exclude small collection accounts from credit evaluations. To test the effect of inconsistently reported small collection items on

30. Independent factors used in the estimation included outstanding balance and highest-balance level, the age and type of account, the type of lender, balances and limits on other accounts, and payment performance information. The resulting distribution of estimated credit limits and utilization for accounts with imputed limits was virtually identical to the distribution of accounts with reported limits within the population, an indication that missing limits are primarily a function of the lender and are almost always unrelated to the characteristics of the account.

credit history scores, we removed all collection records involving items under \$100 from the credit records.

Medical Collection Items

Some credit evaluators report that they remove collection accounts related to medical services from credit evaluations because such accounts often involve disputes with insurance companies over liability for the accounts or because the accounts may not indicate future performance on loans. Unfortunately, evaluators must use manual overrides based on the creditors' identities to remove medical collection accounts because the credit record data lack a code identifying claims associated with medical services. The absence of a code means that this process cannot be used in automated calculations of credit history scores. To test the potential effect of including medical collection items in the calculation of credit history scores, we developed a medical collection code based on an inspection of the creditor name, and we used the code to identify medical collection accounts to drop from the credit history score calculation (as noted earlier, as of this writing, at least one of the agencies had developed such a code, potentially reducing the relevance of this simulation).

Potentially Misassigned Collection Agency Accounts

Most (72 percent) of the individuals in the sample with a non-credit-related collection agency account also had a credit-related major derogatory. About 45 percent of those individuals with information reported by a single collection agency had no credit-related major derogatories. In contrast, only about 15 percent of those with information reported by more than one collection agency had no credit-related major derogatories. These patterns suggest that misassigned collection agency accounts may be more common among those with information reported by a single collection agency. We simulated the effects of correcting such misassigned collections by dropping the collection accounts of individuals who had information reported by one collection agency but had no credit-related major derogatories.

Duplications in Public Records

As with our analysis of collection agency accounts, our review of the sample public record reports

suggests that some records may be duplicated. To address the potential effects of this problem, we conducted simulations that removed likely duplicates of public record items. We identified duplicates by matches on the recording date, amount owed, and creditor. In many instances, the duplicates involved the original filing of a judgment or lien, which was followed by a record of a paid judgment or lien with all information identical to that in the first record. In other instances, second or third filings may have ended up as duplicates with the same (or almost identical) information.

Inconsistent Reporting of Lawsuits and Dismissed Items in Public Records

As noted earlier, our analysis of credit record files in the Federal Reserve sample suggests that lawsuits are inconsistently included in the credit-reporting agency files. An additional issue concerns the inclusion in the public records of dismissed liens, judgments, or suits, which may be of questionable value for predicting credit performance. To simulate the potential effects of including these items in the calculation of credit history scores, we removed all lawsuits and dismissals from the credit records of individuals with such items.

Failure to Consolidate Multiple Inquiries for the Same Loan

Analysts have cautioned that simple counts of inquiries in scoring models may unfairly penalize consumers who shop for credit. However, the information needed to help distinguish consumers shopping to obtain a single loan from those seeking to obtain multiple loans is generally not available in credit records because of incomplete reporting of the type of inquiry.

To simulate the potential magnitude of the effect of incomplete reporting of the type of inquiry, we conducted two experiments. First, we identified all individuals in the sample who had taken out a mortgage or an auto loan in the two years before the sample was drawn. For each loan type, we consolidated into a single inquiry the multiple inquiries that had occurred in the two-month period preceding the date on which the loan was opened (if any non-auto or non-mortgage loans were also taken out during this period, we did not consolidate any inquiries). The second simulation was somewhat broader. We divided all inquiries into three groups based on the type of inquirer as a proxy for the likelihood that

the consumer was shopping for a single loan or potentially “bulking up on credit.” The first group represented inquiries that were unlikely to be credit-related, including inquiries from insurance companies, utilities, and collection agencies. The second group involved inquiries likely related to the purchase of a single large item, such as inquiries from auto companies or real estate firms. We put all other inquiries in the third group. Inquiries from the first group were dropped in the simulation because they did not appear to be credit related. For the second group, we consolidated all inquiries within a two-week period into a single inquiry. Only inquiries from the third group were left unchanged.

Analyzing the Populations of Interest

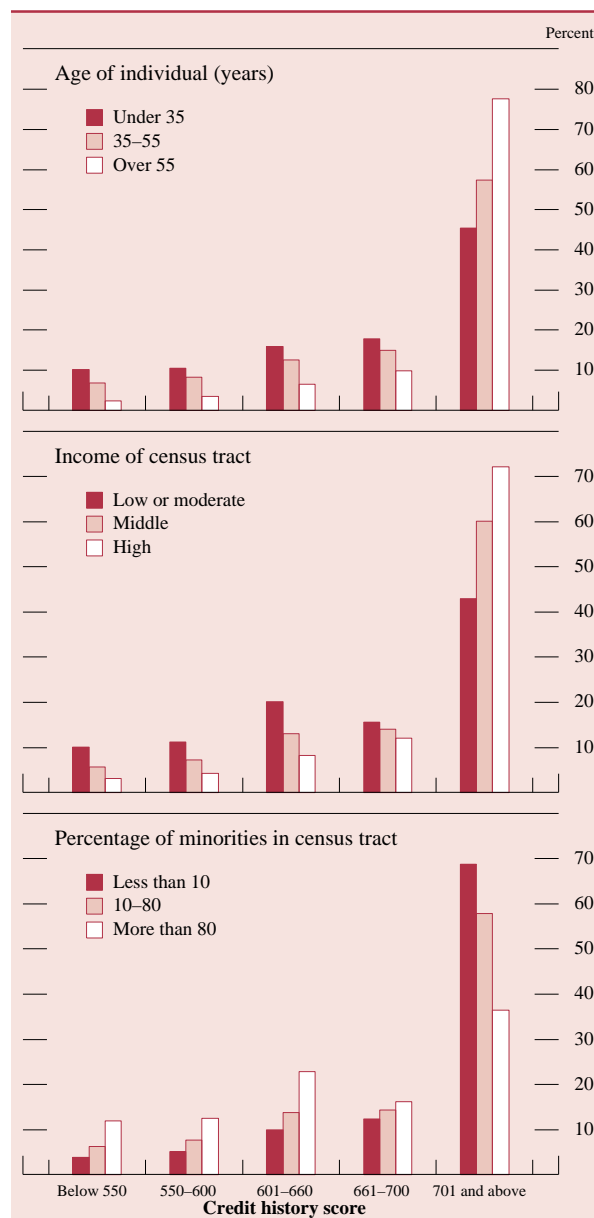
Each of the data quality issues that we focus on may have different implications for different individuals depending on the individuals’ credit characteristics. For example, the effect of a missing credit limit will be different for individuals who have many open revolving accounts than for those who have few. Therefore, we also examined the effect of these data quality issues for three subpopulations of interest. Because data quality problems are less likely to affect the access to credit of individuals with relatively high credit history scores, we divided the analysis population (the same one used to estimate the approximated model) into categories based on credit history score. We also categorized the analysis population by depth of credit file and by selected demographic characteristics.

For the analysis by credit history scores, we sorted individuals into one of three risk groups based on their proprietary credit-risk scores. The first group included individuals whose scores were 661 or above (74 percent of the sample population), the second group included individuals with scores between 600 and 660 (13 percent of the sample), and the third group included individuals whose scores were below 600 (13 percent of the sample).³¹

31. Individuals with credit scores above 660 have scores sufficiently high that they are likely to qualify for the lowest interest rates available on loans, and individuals with scores below 600 have scores sufficiently low that they are likely to be denied credit or to pay substantially higher rates than those charged to better-qualified borrowers. Individuals in the middle category have scores that place them at the margin.

The credit history score ranges used here are not immutable; in practice, the bounds of these ranges vary somewhat by loan product and by the appetite for risk of individual market participants. Moreover, credit history is only one factor considered in credit underwriting, although an important one, and so a low credit history score may be offset by, for example, a low debt-to-income ratio, a significant down payment, collateral, or potential for strong future earnings.

2. Distribution of individuals, by credit history score and by selected demographic characteristics



NOTE. See note to chart 1. Income categories are defined as follows: *low or moderate*, less than 80 percent of the median family income of individual’s metropolitan statistical area (MSA) or of nonmetropolitan portion of individual’s state; *middle*, 80–119 percent of the median family income of individual’s MSA or of nonmetropolitan portion of individual’s state; *high*, 120 percent or more of the median family income of individual’s MSA or of nonmetropolitan portion of individual’s state.

For the analysis by depth of credit file, we sorted individuals with records of credit accounts into two groups based on the number of credit accounts in their credit records. One group consisted of individuals with “thin files”—that is, files with fewer than four credit accounts. The second group consisted of all other individuals. Individuals with thin files, who

accounted for about 19 percent of the sample, are an important segment of the population to examine because their credit history scores may exhibit relatively greater sensitivity to data problems. A data problem affecting a particular account may be more likely to have a substantial effect on the credit history score of an individual with a thin file because of a lack of information from other accounts that could dilute the effect of the problem.

For the other analyses, we investigated whether different patterns emerge when individuals are grouped by age, relative income of census tract of residence, and percentage of minorities in census tract of residence. Such segmentation allows us to determine whether issues of data accuracy and completeness likely affect various subgroups of the population in different ways. For example, residents of higher-income census tracts may, on average, have more revolving accounts than residents of lower-income areas and therefore may face a greater probability of encountering a missing credit limit. We report the distribution of proprietary credit-risk scores

for these various subgroups (chart 2).³² In general, younger individuals, those who live in lower-income areas, and those who live in areas with high minority populations have lower scores.

RESULTS

First, we report the proportion of individuals who are affected by a simulated change in (correction to) the credit records—that is, the proportion subject to the data quality issue in question (table 3). Second, we report the proportion among those affected by the simulated change in credit records for which the net effect on approximated credit history scores was zero. Third, we report the proportions of individuals among those affected by the simulated change for which approximated credit history scores changed

32. Scores in chart 2 are somewhat higher than scores for individuals in the simulation samples, which exclude individuals who have had bankruptcies.

3. Estimated effects of data “corrections” on the credit history scores of individuals, by data problem corrected

Percent except as noted

Data problem corrected	Individuals affected	Distribution of individuals affected					Total	Memo	
		Effect on credit history score						Mean change in points	
		No change	Decrease		Increase			Individuals with decrease in score	Individuals with increase in score
			1–9 points	10 or more points	1–9 points	10 or more points			
<i>Involving credit accounts</i>									
Failure to close a									
Paid-as-agreed account	12.9	10.9	27.0	8.1	48.7	5.2	100.0	-8.1	4.4
Minor delinquent account	1.3	4.5	20.0	17.8	43.1	14.5	100.0	-12.6	8.6
Major derogatory account	4.7	82.3	9.2	.3	8.2	.0	100.0	-1.9	1.2
Failure of a subprime lender to report									
a paid-as-agreed account	n.c.	28.5	41.0	8.9	17.9	3.7	100.0	-6.0	6.2
Failure of largest student loan									
creditor to report	3.5	16.1	45.0	13.1	21.5	4.4	100.0	-7.0	7.5
Failure to report a									
Minor delinquency	n.c.	15.1	39.3	20.8	22.4	2.4	100.0	-11.0	4.0
Credit limit	33.0	31.7	1.7	.0	53.3	13.3	100.0	-1.4	6.1
<i>Involving collection agency accounts</i>									
Failure to eliminate duplicate									
collection agency accounts	1.2	6.8	1.1	.0	67.4	24.7	100.0	-1.4	8.5
Reporting of									
Collection agency accounts									
under \$100	11.1	41.2	7.0	1.2	41.7	8.9	100.0	-4.3	5.8
Medical collection accounts	15.5	11.8	5.4	1.5	49.6	31.6	100.0	-5.9	11.2
Potentially misassigned collection									
accounts	8.2	12.8	9.0	3.4	42.8	31.9	100.0	-6.9	13.4
<i>Involving public records</i>									
Reporting of duplicate public records	.4	38.6	1.9	.0	59.4	.1	100.0	-1.9	1.3
Inclusion of lawsuits and dismissals	1.1	18.5	3.8	1.0	53.1	23.6	100.0	-5.9	9.1
<i>Involving creditor inquiries</i>									
Failure to consolidate									
Multiple inquiries for auto and									
mortgage loans	3.7	16.8	8.3	.5	73.8	.7	100.0	-2.9	2.3
Other multiple inquiries	14.6	5.2	4.9	.1	85.4	4.4	100.0	-2.3	4.2

NOTE. The table reports the effect of “correcting” a data problem. Individuals whose scores increase because of a correction would be better off if the problem were corrected.

n.c. Not calculable.

materially—that is, increased or decreased 10 or more points. These calculations provide insight into the proportion of consumers who may or may not face a change in credit terms (either a higher or a lower interest rate) or who may be unable to gain access to credit because of the particular data problem. Also, to provide another basis for determining how much variation in credit history scores may occur when simulated corrections are made to individuals' credit records, we present the overall mean change in credit history scores for the individuals who were materially affected. Because the hypothesized correction may increase or decrease an individual's credit history score depending on the nature of the problem and the composition of the individual's credit record, the mean change for individuals with a decrease in score and the mean change for those with an increase in score are shown separately.

For each simulation, the overall effect of a simulated change on an individual can be either positive or negative. Some of the effect is undoubtedly due to imprecision in our approximation of the credit-scoring model or to consumers' being shifted from one "scorecard" to another. However, we believe the results mainly reflect the complexity of interactions among the various factors that produce a credit history score. For example, a failure to report a paid-as-agreed account as closed can help individuals with few active and paid-as-agreed credit accounts but can hurt individuals with a substantial number of accounts that have high balances and utilization rates.

Effects of Stale Accounts

The first group of simulations presented in the table involved hypothetical corrections to selected credit account records. The first three of these pertained to the use of a more aggressive stale-account rule that designated accounts as stale after three months of nonreporting and treated the accounts as being closed and having a zero balance. Several conclusions emerged from these simulations. On the one hand, a significant proportion of consumers appeared to be subject to stale credit account issues. Almost one-fifth of the individuals in the Federal Reserve sample had at least one stale credit account as defined by the assumptions of the first three simulations. Further, 21 percent of the individuals with stale major derogatories (percentage not shown in table) had at least one account that met the conditions of the third simulation and thus had potentially been paid off.

On the other hand, the application of the more aggressive stale-account rule appeared to have only

a modest effect on the credit history scores of these individuals. Our simulations suggest that more than 80 percent of the individuals with stale major derogatories would have shown no change in score if they had paid off the account the month after the date on which the lender last reported it and the lender had reported the payoff to the credit-reporting agency. The effect of paying off accounts was somewhat larger for paid-as-agreed accounts and for those with minor delinquencies, but even here most consumers showed changes of fewer than 10 points. One likely explanation for the relatively minor effect of the corrections on individuals is the large number of credit accounts in the typical consumer's file. For example, consumers with a stale paid-as-agreed account had, on average, almost sixteen credit accounts, and 90 percent of these consumers had at least five.

Many of the credit-risk factors reflect "extreme" values such as the age of the oldest account or the number of months since the most-recent delinquency. These factors will change as the result of a correction only if the affected account is the "marginal" account—for example, the oldest or the most recently delinquent. Moreover, although factors reflecting sums, such as total balances, will be sensitive to changes in any account, the effect of the change will be reduced if many other accounts contribute to the factor. Another explanation for the relatively minor effects of the corrections for stale accounts probably lies in the rules used to calculate the factors employed by credit modelers. For example, modelers appear to place little weight on outstanding balances for major derogatory accounts, perhaps recognizing the inconsistency in the reporting of account payoffs. Thus, when payoffs are recorded, the effect on scores is minimal.

Effects of Unreported Credit Account Information

We conducted an additional four simulations for data problems in credit accounts. The simulations addressed the nonreporting of certain categories of accounts (paid-as-agreed accounts of a subprime lender and accounts of the largest student loan creditor) and of certain types of information (minor delinquencies and credit limits).

We could not determine the incidence of subprime creditors' failure to report paid-as-agreed credit accounts. By our estimates, Sallie Mae's failure to report loans affected less than 4 percent of individuals. Nonreporting of these types of accounts appeared

4. Estimated effects of data “corrections” on the credit history scores of individuals, by data problem corrected, for selected credit history score ranges

Percent except as noted

Data problem corrected	Individuals affected	Distribution of individuals affected					Total	Memo	
		Effect on credit history score						Mean change in points	
		No change	Decrease		Increase			Individuals with decrease in score	Individuals with increase in score
		1-9 points	10 or more points	1-9 points	10 or more points				
Individuals with credit history scores above 660									
<i>Involving credit accounts</i>									
Failure to close a									
Paid-as-agreed account	13.6	11.3	22.0	4.7	55.8	6.2	100.0	-6.1	4.5
Minor delinquent account	.2	3.1	19.2	52.9	21.7	3.1	100.0	-20.2	5.0
Major derogatory account	1.2	89.1	6.1	.2	4.6	.0	100.0	-1.8	1.0
Failure of a subprime lender to report a paid-as-agreed account									
Failure of largest student loan creditor to report	n.c.	45.5	30.1	2.8	20.3	1.3	100.0	-4.3	3.0
Failure to report a									
Minor delinquency	n.c.	19.6	45.7	20.0	14.2	.6	100.0	-9.3	3.0
Credit limit	35.8	34.8	1.4	.0	54.1	9.7	100.0	-1.1	5.1
<i>Involving collection agency accounts</i>									
Failure to eliminate duplicate collection agency accounts									
Reporting of	.1	11.7	.4	.0	81.4	6.6	100.0	-1.0	4.6
Collection agency accounts under \$100									
Medical collection accounts	3.6	21.8	9.3	2.7	42.8	23.4	100.0	-5.8	10.6
Potentially misassigned collection accounts	6.5	5.2	8.0	2.9	35.7	48.3	100.0	-6.8	16.6
Potentially misassigned collection accounts	5.4	4.7	11.0	4.4	31.4	48.6	100.0	-1.6	6.4
<i>Involving public records</i>									
Reporting of duplicate public records	.2	39.1	2.3	.0	58.6	.0	100.0	-1.0	1.1
Inclusion of lawsuits and dismissals	.7	19.2	5.0	1.7	45.5	28.7	100.0	-7.0	10.8
<i>Involving creditor inquiries</i>									
Failure to consolidate									
Multiple inquiries for auto and mortgage loans	3.4	10.9	3.8	.0	84.7	.7	100.0	-1.6	2.3
Other multiple inquiries	12.2	3.1	1.4	.0	94.0	1.5	100.0	-1.4	3.6
Individuals with credit history scores between 600 and 660									
<i>Involving credit accounts</i>									
Failure to close a									
Paid-as-agreed account	12.1	11.0	49.4	13.0	25.4	1.3	100.0	-6.4	3.3
Minor delinquent account	2.6	4.0	27.2	22.6	41.7	4.6	100.0	-11.9	4.9
Major derogatory account	10.2	87.9	6.4	.1	5.7	.0	100.0	-1.7	1.3
Failure of a subprime lender to report a paid-as-agreed account									
Failure of largest student loan creditor to report	n.c.	22.2	48.6	6.4	19.4	3.5	100.0	-4.2	4.9
Failure to report a									
Minor delinquency	3.8	8.1	33.7	17.6	33.0	7.6	100.0	-9.5	6.0
Credit limit	n.c.	11.0	33.1	21.5	31.2	3.2	100.0	-11.7	3.7
Credit limit	28.4	14.4	2.3	.0	57.2	26.1	100.0	-1.8	7.8
<i>Involving collection agency accounts</i>									
Failure to eliminate duplicate collection agency accounts									
Reporting of	3.0	8.6	.8	.0	80.7	9.9	100.0	-1.0	5.3
Collection agency accounts under \$100									
Medical collection accounts	28.1	43.6	5.7	1.2	42.7	6.9	100.0	-5.1	4.4
Potentially misassigned collection accounts	38.8	11.1	4.4	1.7	56.5	26.4	100.0	-7.2	9.2
Potentially misassigned collection accounts	11.8	18.1	9.7	6.9	48.1	17.2	100.0	-9.8	9.6
<i>Involving public records</i>									
Reporting of duplicate public records	.7	44.3	1.0	.0	54.7	.0	100.0	-1.0	1.1
Inclusion of lawsuits and dismissals	2.1	20.8	2.2	.2	62.2	14.7	100.0	-4.2	6.4
<i>Involving creditor inquiries</i>									
Failure to consolidate									
Multiple inquiries for auto and mortgage loans	5.0	32.7	15.1	.1	51.6	.6	100.0	-1.9	2.0
Other multiple inquiries	17.0	10.0	7.8	.0	80.9	1.3	100.0	-1.5	3.9

4.—Continued

Data problem corrected	Individuals affected	Distribution of individuals affected					Total	Memo	
		Effect on credit history score						Mean change in points	
		No change	Decrease		Increase			Individuals with decrease in score	Individuals with increase in score
			1–9 points	10 or more points	1–9 points	10 or more points			
Individuals with credit history scores below 600									
<i>Involving credit accounts</i>									
Failure to close a									
Paid-as-agreed account	9.1	7.0	46.0	35.8	10.4	.8	100.0	-16.8	3.3
Minor delinquent account	7.1	5.0	17.3	9.9	47.4	20.4	100.0	-9.7	9.8
Major derogatory account	22.9	77.3	11.7	.4	10.6	.1	100.0	-2.0	1.3
Failure of a subprime lender to report a paid-as-agreed account									
Failure of largest student loan creditor to report	n.c.	7.2	52.4	19.8	13.1	7.6	100.0	-8.1	12.4
Failure to report a									
Minor delinquency	n.c.	5.8	26.0	22.7	38.5	7.0	100.0	-17.0	5.3
Credit limit	19.3	19.9	4.2	.1	37.7	38.1	100.0	-1.9	13.1
<i>Involving collection agency accounts</i>									
Failure to eliminate duplicate collection agency accounts									
Reporting of	6.8	5.2	1.4	.0	58.9	34.6	100.0	-1.5	10.6
Collection agency accounts under \$100	43.2	50.7	6.7	.3	40.4	2.0	100.0	-2.4	2.6
Medical collection accounts	51.6	18.0	4.1	.2	55.9	21.7	100.0	-2.7	8.0
Potentially misassigned collection accounts	23.5	22.6	5.7	.1	57.8	13.9	100.0	-1.6	6.4
<i>Involving public records</i>									
Reporting of duplicate public records									
Inclusion of lawsuits and dismissals	1.0	34.1	1.8	.0	63.7	.4	100.0	-3.8	1.8
Inclusion of lawsuits and dismissals	2.6	15.1	3.1	.3	59.8	21.7	100.0	-2.5	8.5
<i>Involving creditor inquiries</i>									
Failure to consolidate									
Multiple inquiries for auto and mortgage loans	4.3	27.7	23.5	3.7	44.4	.8	100.0	-4.6	2.2
Other multiple inquiries	28.2	8.5	13.1	.6	62.9	14.9	100.0	-2.9	6.5

NOTE. See note to table 3.

n.c. Not calculable.

to have only a modest effect on the credit history scores of affected individuals. For example, the simulation results indicate that if nonreporting by a subprime lender or by Sallie Mae had been corrected, in each case less than 5 percent of affected individuals would have gained 10 percentage points or more in their credit history scores. Moreover, such nonreporting may help or hurt the individuals. For example, the simulations suggest that, on average, consumers were helped by Sallie Mae's not reporting their loans, a somewhat surprising result. Fifty-eight percent of affected individuals would have experienced decreases in their credit history scores if the accounts had been reported. However, the median number of credit accounts for individuals with a corrected student loan account was twenty-two, a figure well above average for all individuals. Thus, the positive effects on credit history scores of reducing indi-

viduals' outstanding balances by not reporting their student loans may have outweighed the negative effects of eliminating one additional paid-as-agreed account.

We also could not determine the proportion of individuals affected by creditors' suppression of minor delinquencies; however, we could estimate the impact of the suppression on affected individuals. The simulation suggests that when suppression occurs, it is likely to improve the credit history scores of many affected individuals by a significant amount.

Effects of Unreported Credit Limits

Nonreporting of credit limits affects a substantial number of individuals (33 percent of the individuals

in the simulations), but the effect tends to be small. The likely reason for this result is that affected individuals tend to have a large number of credit accounts in their credit records (eighteen on average), while the frequency of accounts missing limits is low. Thus, accounts with missing limits tend to have a small effect on the overall utilization rates of individuals.

Unlike the results in most of the other simulations, the effects of missing credit limits were predominantly in one direction—most affected individuals' scores would have likely been higher if missing credit limits had been reported. This finding suggests that the rule that credit modelers typically adopt for addressing missing limits—use of the reported highest-balance amount—is likely biased. To further test this conjecture, we examined credit accounts for which the credit limit was reported and compared the actual limit with the estimated limit that credit modelers would have applied if the limit had not been reported. On average, the rule that the credit-reporting agencies used when they constructed utilization rates would imply a credit limit of less than one-half the actual limit. The rule would imply a lower credit limit than the actual limit in about 90 percent of the cases. In contrast, our rule, as noted earlier, was statistically unbiased.

Effects of Problems with Collection Agency Accounts, Public Records, and Creditor Inquiries

Results for eight simulations involving collection agency accounts, public records, and creditor inquiries were varied.

Collection Agency Accounts

The proportion of individuals affected by potential data problems or inconsistencies in reporting by collection agencies ranged from 16 percent for reporting of medical collection items to only about 1 percent for duplication of collection items, although, as noted, our ability to detect such duplications was limited. However, the effect of corrections on affected individuals tended to be large, particularly in comparison with simulated problems in credit accounts, and was generally associated with increases in credit history scores. For example, for three of the four collection account simulations, one-fourth or more of the affected individuals showed increases of 10 points or more in their scores. These results illustrate that

collection accounts weigh heavily in the scoring model and that most individuals have relatively few such accounts and thus are affected more significantly when a problem occurs in any given account.

Public Records

Both simulations that addressed potential data problems or inconsistencies in public records indicated that the proportion of individuals affected was small (1 percent or less). However, the effects of the corrections differed significantly between the two simulations. In the simulation involving duplicate public record items, less than 1 percent of affected individuals experienced increases in their credit history scores of 10 points or more, whereas in the simulation involving lawsuits and dismissals, nearly one-fourth of affected individuals did so. This dichotomy reflects an important distinction between duplicate public records and lawsuits and dismissals. Whereas removing a lawsuit or a dismissal may completely eliminate adverse public record items from an individual's credit record, eliminating a duplicate record cannot do so.

Creditor Inquiries

The simulation that consolidated inquiries related to auto and mortgage loans affected only 4 percent of individuals in the sample; the broader consolidation simulation affected about 15 percent of individuals. In both cases, the size of the effect was modest and almost always resulted in a higher score. Only a small percentage of individuals experienced increases in their scores of more than 10 points.

Differences across Subpopulations

Individuals with scores below 600 tended to have the highest frequency of data problems, and those with scores above 660 had the lowest incidence (table 4). Two exceptions to this pattern occurred in the simulations involving the failure to close stale paid-as-agreed accounts and the failure to report a credit limit. Here individuals in the highest score range showed the largest incidence of data problems primarily because they tended to have more credit accounts. Significant differences were also apparent in the impact of simulated corrections on affected individuals across the three groups. Generally, individuals with scores below 600 were the most likely to experience a score increase of 10 points or more in

5. Estimated effects of data “corrections” on the credit history scores of individuals with “thin” files, by data problem corrected
Percent except as noted

Data problem corrected	Individuals affected	Distribution of individuals affected					Total	Memo	
		Effect on credit history score						Mean change in points	
		No change	Decrease		Increase			Individuals with decrease in score	Individuals with increase in score
			1–9 points	10 or more points	1–9 points	10 or more points			
<i>Involving credit accounts</i>									
Failure to close a									
Paid-as-agreed account	3.2	3.6	21.7	44.1	15.8	14.9	100.0	-17.0	11.3
Minor delinquent account	.7	8.1	22.4	22.0	45.1	2.4	100.0	-16.0	3.7
Major derogatory account	2.4	88.7	5.4	.0	5.9	.0	100.0	-1.8	1.5
Failure of a subprime lender to report a paid-as-agreed account	n.c.	4.4	35.9	38.0	16.4	5.4	100.0	-12.3	6.8
Failure of largest student loan creditor to report	1.0	3.4	33.6	51.8	8.0	3.2	100.0	-20.8	6.8
Failure to report a									
Minor delinquency	n.c.	4.3	18.1	46.6	14.1	16.9	100.0	-24.9	9.8
Credit limit	9.1	18.2	1.4	.0	36.0	44.3	100.0	-1.2	13.2
<i>Involving collection agency accounts</i>									
Failure to eliminate duplicate collection agency accounts	1.9	7.4	.8	.0	82.4	9.5	100.0	-1.0	5.1
Reporting of									
Collection agency accounts under \$100	15.2	48.0	3.0	.6	35.8	12.6	100.0	-5.1	9.5
Medical collection accounts	20.9	10.6	1.7	.9	52.0	34.9	100.0	-8.7	14.7
Potentially misassigned collection accounts	8.6	16.3	4.1	3.1	32.7	43.7	100.0	-10.7	26.6
<i>Involving public records</i>									
Reporting of duplicate public records	.3	50.4	1.7	.0	47.9	.0	100.0	-1.0	1.0
Inclusion of lawsuits and dismissals	.7	22.4	1.6	.6	52.3	23.0	100.0	-6.3	13.4
<i>Involving creditor inquiries</i>									
Failure to consolidate									
Multiple inquiries for auto and mortgage loans	.9	19.1	7.2	.0	69.2	4.5	100.0	-2.1	3.4
Other multiple inquiries	9.5	4.9	3.4	.0	87.0	4.7	100.0	-1.5	4.8

NOTE. See note to table 3. A “thin” file has a record of a credit account but has fewer than four such accounts. n.c. Not calculable.

response to corrections of data problems. Collection account problems provided an exception to this pattern: Affected individuals in the credit history score range above 660 were the most likely to experience large score increases. The reason for this result is that relatively high-score individuals with collection agency accounts generally have no other major derogatory information in their credit records and thus can show significant score increases when a derogatory is corrected.

For individuals with thin files, the incidence of data quality issues involving credit accounts was generally lower than that for all individuals, but the incidence of issues involving collection agency accounts was somewhat higher (compare table 5 with table 3). The result regarding credit accounts reflects the smaller number of accounts in the credit records of individuals with thin files and, consequently, the generally lower probability that such individuals will have data quality issues. The result concerning collection agency accounts is due to the higher probability that people with thin files will have such accounts. However, in simulations involving corrections to

credit accounts, the effects on the credit history scores of individuals with thin files were either similar to or substantially larger than the effects on the scores of persons in the general population. For example, correcting a failure to close a paid-as-agreed account resulted in a decline in credit history score that was twice as large, on average, for individuals with thin files as it was for those in the population at large.

In general, older individuals and those living in higher-income and nonminority neighborhoods had the lowest incidence of data problems (table 6). The most-notable exception to this pattern was for failure to report a credit limit, which was less common among younger individuals and among individuals living in lower-income and predominantly minority neighborhoods. We do not report the changes in credit history scores of affected individuals for these decompositions of the sample because the comparisons are difficult to interpret without also accounting for differences in the incidence of thin files and in credit history scores across groups. In most cases, the effects of data quality problems were similar across groups after controlling for the differences in depth of

6. Share of individuals affected by data problems in credit records, distributed by selected demographic characteristics

Percent except as noted

Data problem	Age (years)			Income of census tract ¹			Share of minorities in census tract (percent)		
	Under 35	35–55	Over 55	Low or moderate	Middle	High	Less than 10	10–80	More than 80
<i>Involving credit accounts</i>									
Failure to close a									
Paid-as-agreed account	16.9	16.6	10.1	11.3	13.1	13.7	13.4	12.9	11.3
Minor delinquent account	2.0	1.4	.6	1.8	1.3	.8	1.0	1.3	2.1
Major derogatory account	5.5	6.2	2.9	6.8	4.7	3.1	3.2	5.1	8.0
Failure of a subprime lender to report									
a paid-as-agreed account	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
Failure of largest student loan									
creditor to report	9.0	3.2	.8	3.4	3.3	3.8	3.0	3.8	3.3
Failure to report a									
Minor delinquency	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
Credit limit	31.5	40.3	37.4	27.7	31.7	40.0	33.7	33.5	28.0
<i>Involving collection agency accounts</i>									
Failure to eliminate duplicate									
collection agency accounts	1.9	1.2	.4	2.3	1.1	.6	.7	1.4	2.7
Reporting of									
Collection agency accounts									
under \$100	15.1	11.5	5.0	17.0	11.1	6.4	8.7	11.7	16.9
Medical collection accounts	19.5	16.5	8.3	22.8	15.7	9.3	12.7	16.1	22.3
Potentially misassigned collection									
accounts	10.8	8.8	5.3	11.6	7.9	6.1	6.4	8.5	13.1
<i>Involving public records</i>									
Reporting of duplicate public records2	.5	.3	.4	.4	.3	.4	.4	.3
Inclusion of lawsuits and dismissals6	1.7	1.1	1.2	1.1	1.1	1.0	1.2	1.4
<i>Involving creditor inquiries</i>									
Failure to consolidate									
Multiple inquiries for auto and									
mortgage loans	5.4	5.3	2.1	3.3	3.7	3.9	3.8	3.7	3.3
Other multiple inquiries	19.6	17.7	10.0	15.9	14.1	14.6	12.8	15.3	17.5

NOTE. See note to table 3.

n.c. Not calculable.

1. For definition of income of census tract, see note to chart 2.

file and in credit history score. Exceptions generally involved instances in which either the youngest or the oldest age group was disproportionately affected. For example, individuals over age 55 were more likely to have increases of more than 10 points in their credit history scores when medical collections were dropped, and individuals under age 35 were more likely to have large increases in their scores when nonreporting of a credit limit was corrected.

SUMMARY AND CONCLUSIONS

Available evidence indicates that the information that credit-reporting agencies maintain on the credit-related experiences of consumers, and the credit history scoring models derived from these experiences, have substantially improved the overall quality of credit decisions while reducing the costs of such decisionmaking. The availability of these data has also greatly enhanced the process of screening prospective customers to facilitate the marketing of credit and insurance products, thereby reducing the costs of such marketing by limiting solicitations to

customers who are most likely to qualify for the products. If not for the information that the agencies maintain, consumers on the whole would receive less credit at higher prices. Moreover, the credit-reporting system has become more comprehensive over the past decade or so with notable improvements, such as the adoption of common formats for reporting information and the enhanced reporting of information on credit limits and mortgages. Recent congressional amendments to the FCRA have advanced prospects for future improvements as consumer access to credit records and credit history scores has improved.

Despite the benefits of the credit-reporting system, analysts have raised concerns about the accuracy, completeness, timeliness, and consistency of agency records and about the effects of these shortcomings on the cost and availability of credit. Clearly, for the benefits of the credit-reporting system to be realized, some reasonable degree of accuracy and completeness of credit reports is required. Moreover, the more accurate and complete the information assembled by credit-reporting agencies, the greater the potential for more efficiency in the credit-granting process and a reduction in costs to the advantage of both consumers

and creditors. Over the years, a number of studies have focused on the contents of credit records but have reached quite different conclusions about the degree to which such information is accurate and complete and about the implications of data limitations for credit availability and pricing.

This study extends earlier research and assesses the effects of data limitations and ambiguities in credit reports on the availability and pricing of credit by using a large, nationally representative sample of individuals with credit records from one of the three national credit-reporting agencies. Specifically, we estimate the proportion of individuals who are likely to be materially affected by a number of different data problems, and we quantify the likely effect of each problem on the credit history scores of individuals. Because such effects can vary across different populations, we also separately evaluate the effects on individuals in different credit-risk categories and in different groups classified by age and by income and minority population of the neighborhoods where they live. We emphasize that we use the terms “data problem” and “correction” in their broadest sense, as we do not necessarily observe actual errors and the appropriate correction is sometimes unclear.

This analysis of the effects of data problems on credit history scores indicates that the proportion of individuals affected by any single type of data problem appears to be small, with the exception of missing credit limits, which affected almost one-third of the individuals in the sample used for the simulations. Moreover, in most cases, the effect of each type of problem on the credit history scores of affected individuals was modest. Two principal reasons explain this result. First, most individuals have a large number of credit accounts, and thus problems in any given account have only a relatively small effect on the individuals’ overall credit profiles. Second, credit modelers recognize many of these data problems when they construct and weight the factors used in credit history scoring models. Therefore, correcting the problems identified here is unlikely to substantially change the risk evaluation and access to credit for the typical individual.

The analysis suggests, however, that the effects of data problems may be more substantial in some cases than in others. In particular, problems with collection accounts are much more likely to have significant effects on the credit history scores of affected individuals. Missing credit limits, simply because they occur so frequently, also represent an important data quality problem. In general, individuals with relatively low credit history scores or those with thin files are more likely to experience significant effects when

a data problem arises. The incidence of problems also varies across groups, with older individuals, those with higher credit history scores, and those living in higher-income and nonminority neighborhoods showing the lowest incidence.

Our analysis shows that predicting the effects of “correcting” errors is not straightforward. Sometimes, effects were counterintuitive. For example, our analysis suggests that about one-fourth of the individuals affected by lenders’ failure to report student loans would show increases in their credit history scores as a result. This outcome occurs in part because, somewhat surprisingly, individuals with student loans have more accounts than does the average individual. The complexity of the results is underscored by the fact that some individuals show increases and some show decreases for every simulation. In large part, this result occurs because the corrections typically affect more than one factor, moving scores in different directions. This is particularly true for problems with credit accounts, which are likely to involve multiple factors.

The research here highlights the importance of data reporters’ supplying complete information in a timely manner. How such reporting can be fully achieved in a voluntary system is unclear. The current system relies heavily on consumers to identify and dispute “incorrect” or missing items in their credit reports. One problem with this approach is that consumers have no incentive to challenge information that is favorable to them, even if it is in error. Our research indicates that even when data are incomplete or in error, they often have little or no bearing on an individual’s credit history score or access to credit.

Currently, consumers have access only to general information about the types of factors that are weighed in credit evaluation, or in the case of credit denials, the chief reasons for the adverse action. On the one hand, lack of specific information may lead some consumers to believe that virtually any data quality issue is pertinent and should be disputed, causing the credit-reporting agencies and reporters to incur unnecessary costs to correct or update files. On the other hand, consumers may be unaware of the potential importance of specific data issues, such as missing credit limits, and may not take appropriate action. Some of these problems may be addressed by consumer education, whereas others are likely to continue for the foreseeable future.

Before these results are taken as definitive estimates of the effects of data quality issues on credit availability, several important caveats must be made. First, we have investigated only some potential sources of error. Most notably, we can say nothing

about the consequences of mistakenly including account records that do not belong to an individual in the individual's file. Second, we have used only one credit-scoring model to simulate our results and have relied on our approximation to the model to quantify our results. Third, we have omitted manual reviews of credit records, which are part of many underwriting systems. Such systems identify and address many data quality issues. Finally, we have used data from only one credit-reporting agency. Creditors, particularly in the mortgage market, typically obtain data from all three national credit-reporting agencies for credit underwriting. Reconciling inconsistencies in data across the three agencies can lead to corrections of many of the data quality issues we have identified.

Moreover, we have analyzed only the potential effects on credit history scores of addressing data quality issues. We have said nothing about how such

problems could be corrected, how much the corrections might cost, or what potential gains in efficiency might result from developing models based on more complete and accurate data. If the current level of accuracy and completeness is socially inefficient, reaching the optimal level may be difficult. Credit information has aspects of a classic public good. The parties that bear the costs of correcting errors or providing more timely and complete information may not receive much benefit from the improvement in accuracy. Further remedies, such as imposing additional legal liability penalties, may, in a system of voluntary reporting, lead to unintended consequences, including less information reporting and a less efficient and effective system. Policymakers need to weigh all of these considerations when they determine whether the current credit-reporting system should be changed and, if so, what changes should be made. □