Federal Reserve Bank of New York

October 2015
Volume 21 Number 1

Economic Policy Review

1 Challenges in Identifying Interbank Loans
Olivier Armantier and Adam Copeland

19 Do We Know What We Owe? Consumer Debt as Reported by Borrowers and Lenders
Meta Brown, Andrew Haughwout, Donghoon Lee, and Wilbert van der Klaauw

45 The Great Recession's Impact on School District Finances in New York State
Rajashri Chakrabarti, Max Livingston, and Elizabeth Setren

Federal Reserve Bank of New York research publications—the *Economic Policy Review, Current Issues in Economics and Finance, Staff Reports, Research Update,* and the annual publications catalogue—are available at no charge from the Research Group's website.

www.newyorkfed.org/research

Follow us on Twitter: *@NYFedResearch*

Federal Reserve Bank of New York Economic Policy Review

October 2015
Volume 21 Number 1

Contents

Articles

1 Challenges in Identifying Interbank Loans

Olivier Armantier and Adam Copeland

Although interbank lending markets play a key role in the financial system, the lack of disaggregated data often makes the analysis of these markets difficult. To address this problem, recent academic papers focusing on unsecured loans of central bank reserves have employed an algorithm in an effort to identify individual transactions that are federal funds loans. The accuracy of the algorithm, however, is not known. The authors of this study conduct a formal test with U.S. data and find that the rate of false positives produced by one of these algorithms is on average 81 percent; the rate of false negatives is 23 percent. These results raise concerns about the information content of the algorithm's output.

19 Do We Know What We Owe? Consumer Debt as Reported by Borrowers and Lenders

Meta Brown, Andrew Haughwout, Donghoon Lee, and Wilbert van der Klaauw

Household surveys are the source of some of the most widely studied data on consumer balance sheets, with the Survey of Consumer Finances (SCF) generally cited as the leading source of wealth data for the United States. At the same time, recent research questions survey respondents' propensity and ability to report debt characteristics accurately. This study compares household debt as reported by borrowers to the SCF with household debt as reported by lenders to Equifax using the new FRBNY Consumer Credit Panel (CCP). The borrower and lender debt distributions are compared by year, age of household head, household size, and region of the country, in total and across five standard debt categories. The authors' central finding is that the SCF and CCP debt patterns are strikingly similar. There are, however, two

noteworthy exceptions: the aggregate credit card debt implied by SCF borrowers' reports is estimated to be 37 to 40 percent lower than that implied by CCP lenders' reports, and the aggregate student debt implied by the SCF is roughly 25 percent lower than that implied by the CCP. In contrast to the credit card debt mismatch, bankruptcy history is reported comparably in the borrower and lender sources, indicating that not all stigmatized consumer behaviors are underreported.

45 The Great Recession's Impact on School District Finances in New York State

Rajashri Chakrabarti, Max Livingston, and Elizabeth Setren

A slowly emerging literature explores the effects of the Great Recession on different parts of the economy; however, very little research examines the impact of the Great Recession (or any other recession) on schools. Given the fundamental role of education in human capital formation and growth, understanding the effect of recessions on schools is essential. This article contributes to filling this gap. Exploiting detailed panel data on a multitude of school finance indicators and a trend shift analysis, it examines how the Great Recession affected school finances in New York State. While it finds no evidence of effects on either total funding or expenditures, both funding and expenditures experienced important compositional changes. There is strong evidence of substitution of funds on the funding side: the infusion of funds with the federal stimulus occurred simultaneously with statistically and economically significant cuts in state and local financing, especially the former. On the expenditure side, instructional expenditure was maintained, while several noninstructional categories such as transportation, student activities, and utilities suffered. Important heterogeneities in experience are also observed by poverty level, metropolitan area, and urban status (urban, suburban, or rural). Affluent districts were hurt the most, while analysis by metro area reveals that the New York City metropolitan area, especially Nassau and Suffolk counties, sustained the largest reductions in most expenditure categories. The findings of this study promise to enhance our understanding of how recessions affect schools and the role policy can play in mitigating the consequences.

Olivier Armantier and Adam Copeland

Challenges in Identifying Interbank Loans

- Empirical analyses of the federal funds market often use the so-called "Furfine algorithm" to identify activity in the market at the most disaggregated level—individual loans between two specific banks.
- However, a formal test of the accuracy of the algorithm in identifying fed funds transactions shows that the algorithm may be ill-suited to this task.
- Given access to the identifiers used by two large banks to denote fed funds payments, the authors are able to compare a set of payments known to be fed funds transactions with the set of payments pegged as such by the algorithm.
- The authors find that for the 2007-11 period, an average of 81 percent of all pairs of payments identified by the algorithm are not, in fact, fed funds transactions conducted by the two banks, while an average of 23 percent of the banks' actual fed funds transactions are overlooked by the algorithm.

1. Introduction

The U.S. federal funds (fed funds) market is an interbank market for unsecured, mostly overnight loans of reserves held by banks at Federal Reserve Banks. It is an over-the-counter market where banks arrange trades either on their own on a bilateral basis or through brokers. Historically, the fed funds market has been a key financial market with major macroeconomic and monetary policy implications. In particular, the average fed funds market rate, known as the effective fed funds rate, has substantial influence on the terms at which commercial banks lend to businesses and individuals. Furthermore, the Federal Reserve implements monetary policy by creating conditions under which fed funds trade around a specific target or within a target range set by the Federal Open Market Committee (FOMC).[1]

The traditional source of data on the fed funds market is based on fed funds trades reported by the major fed funds brokers to the Federal Reserve Bank of New York (FRBNY). Using these data, various market-level interest rate statistics are calculated and published daily by the FRBNY. These

[1] Although other forms of short-term interbank lending may be informally referred to as "fed funds," we are solely concerned in this article with loans of reserves between eligible counterparties as officially defined as fed funds ▸

Olivier Armantier is an assistant vice president and Adam Copeland a research officer at the Federal Reserve Bank of New York.

Correspondence: adam.copeland@ny.frb.org

Note: For its analysis of interbank lending markets in the conduct of monetary policy, the Federal Reserve Bank of New York relies on different sources of data, not on an algorithm's output. Consequently, our results have no bearing on the Federal Reserve Bank of New York's operational understanding of interbank lending markets and its calculation of market-level measures, including the effective federal funds rate.

The authors thank Gara Afonso, Marco Cipriani, Todd Keister, Anna Kovner, Antoine Martin, Jamie McAndrews, David Skeie, and James Vickery for their help on this project, as well as Isaac Davis, Sha Lu, and Michael Walker for valuable research assistance. They also thank seminar participants at the Federal Reserve Bank of New York. The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

statistics—in particular, the effective fed funds rate—are used widely by policymakers, financial market participants, and researchers in academia.

An alternative source of data, used exclusively to conduct academic research, is inferred from algorithms based on the original work of Furfine (1999). Although there are now different versions of the original algorithm, they all seem to rely on the same principles. A number of recent empirical papers use a version of the Furfine algorithm's output to make important contributions. These papers assume, but do not formally test, the accuracy of the output of their algorithms. As we explain in more detail below, the main purpose of this article is to formally test these Furfine-based algorithms. Importantly, the results presented here do not extend to the traditional source of data collected by the FRBNY from the fed funds brokers. In particular, the results have no bearing on the ability of the Markets Group of the FRBNY to understand the fed funds market and to accurately calculate market-level measures, including the effective fed funds rate.

In this study, we focus on the revised Furfine algorithm used by the Research Group of the FRBNY. This algorithm exploits the fact that privately traded fed funds transactions are often settled over the Fedwire® Funds Service (Fedwire), the large-value real-time payments settlement system operated by the Federal Reserve.[2] As further explained in section 2, the algorithm searches all payments sent over Fedwire to identify the pairs of payments that look like fed funds loans. Specifically, the algorithm tries to identify first a "sent" payment from bank A to bank B on a given date for an amount that could reasonably constitute a loan principal, and then a "return" payment from bank B to bank A on the following day for an amount that could reasonably constitute the principal plus interest payment.

If the algorithm correctly identifies fed funds transactions with sufficient accuracy, then its output could be useful to academic economists in studying the fed funds market. Indeed, it would provide data at the lowest level of aggregation (that is, individual transactions between specific pairs of banks) that could help shed light on the underpinnings of the U.S. fed funds market. The algorithm's output is especially attractive when trying to explain the behavior of the market during the 2008-09 financial crisis, as well as the specific role played by individual banks. Indeed, there has been a surge in the number of papers that use the algorithm's output (we found eleven papers written in the past two years; all are listed in the References section of this article).[3]

An important question remains, however: To what extent does the algorithm identify individual fed funds transactions? Indeed, nothing guarantees that a pair of payments between two banks labeled by the algorithm as a fed funds transaction is indeed a fed funds loan between those two banks. In 2009, we started to test the algorithm's output. In this article, we report the outcome of a formal test assessing the ability of the algorithm to identify individual overnight fed funds transactions.

The basic methodology underlying the test, discussed more fully in section 3, may be summarized as follows. From the flow of payments a bank receives over Fedwire, its back office needs to be able to identify those corresponding to the fed funds transactions initiated by the front office. While back offices use a variety of strategies, at least two banks require their fed funds counterparties to incorporate a unique identifier into the message portion of the Fedwire payment. These two institutions, which are among the biggest banks and account for a large fraction of transactions in the fed funds market, gave us access to their unique identifier. As a result, we can flag every fed funds payment these two banks receive through Fedwire on a given day. To assess the quality of the algorithm, we can then compare the set of payments constructed with the unique identifiers to the set of transactions identified for these two banks by the algorithm. Our identification method rests on the hypothesis that the unique identifiers provided by the two banks are included in every fed funds transaction they settle over Fedwire. At the end of section 3.1, we present evidence supporting this hypothesis.

The outcome of the test is discouraging: In the first quarter of 2007, we estimate that 64 percent of all pairs of payments identified by the algorithm are not fed funds transactions conducted by the two banks (type I error), while 24 percent of the

Footnote 1 (continued)
by the Board of Governors of the Federal Reserve System in Regulation D (see http://www.federalreserve.gov/bankinforeg/reglisting.htm#D). See the FedPoint document at http://www. newyorkfed.org/aboutthefed/fedpoint/fed15.html for a concise definition of fed funds. Examples of papers considering similar definitions of fed funds are Hamilton (1996, 1997), Demiralp, Preslopsky, and Whitesell (2006), Afonso, Kovner, and Schoar (2011), and Afonso and Lagos (2012a, 2012b).

[2] Fed funds transactions can be settled over Fedwire, possibly settled over CHIPS (Clearing House Interbank Payments System, another high-value payments settlement system), or conducted on a bank's books. However, on the basis of conversations with industry participants, Bartolini, Hilton, and McAndrews (2008) report that fed funds loans settle almost exclusively over Fedwire as opposed to other payment services. Still, to the best of our knowledge, the exact extent to which Fed funds are primarily settled over Fedwire has not been established formally.

[3] The following papers use a version of the Furfine algorithm to varying degrees (although the main results may not depend on the algorithm's output): Ashcraft and Bleakley (2006), Ashcraft and Duffie (2007), Atalay and Bech (2010), Acharya and Skeie (2011), Ashcraft, McAndrews, and Skeie (2011), Bech et al. (2011), Afonso, Kovner, and Schoar (2011, 2013), Afonso and Lagos (2012a, 2012b), and Armantier et al. (2011). We are not implying that these authors did anything improper. Specific concerns about the algorithm only emerged recently. Furthermore, some of these papers explicitly discuss the potential problems with the algorithm.

fed funds transactions actually conducted by the two banks are not identified by the algorithm (type II error). This negative result seems to be robust with respect to the time period considered. If we go forward to the first quarter of 2011, the type I error is estimated to be 93 percent, while the type II error is estimated to be 17 percent. Although our results may not extend to every bank, we argue that they apply to the majority of the algorithm's output for at least two reasons. First, the two banks that provided their unique identifier are either senders or receivers for about three-tenths of all pairs of payments output by the algorithm over the 2007-11 period. Second, if we assume that the estimates of type I and type II errors generalize to other large banks with similar Fedwire activity, then our estimates apply to almost half of all pairs of transactions output by the algorithm. Consequently, we conclude that there is substantial doubt about the ability of the algorithm to produce transaction-level measures that characterize accurately and comprehensively the fed funds market.

The algorithm has an additional, perhaps insurmountable, problem: Even if it could correctly find every fed funds transaction, there is no guarantee that it correctly identifies the ultimate originator and beneficiary of a payment. Indeed, while Fedwire data list which bank is sending the payment over Fedwire, it is not at all clear whether that bank or one of that bank's correspondents is originating the payment. Similarly, the algorithm cannot guarantee the identity of the ultimate beneficiary of the payment. Although we are unaware of the exact extent of this problem, conversations with market participants suggest that having cash accounts at other (typically large) banks is not uncommon.[4] Not being able to identify with certainty the true counterparties of a Fedwire payment poses a fundamental challenge to constructing transaction-level or even bank-level estimates of fed funds activity.[5]

These negative results cast doubt on the robustness of empirical work that uses the output of Furfine-based algorithms at the transaction level. Our findings strongly suggest that, going forward, a better understanding of the federal funds market at a disaggregate level depends upon finding data, or improving (and validating) the Furfine algorithm, rather than using the current algorithm's output. Alternatively, researchers may want to forgo the lure of disaggregate measures of fed funds activity and use the transaction-based market-level statistics published by the FRBNY, which are based on data from fed funds brokers (for example, see Hamilton [1996, 1997]).

While our work focuses on the revised Furfine algorithm used by the Research Group of the FRBNY, slightly different versions of this algorithm are used by researchers outside the Bank. Demiralp, Preslopsky, and Whitesell (2006) and Bech, Klee, and Stebunovs (2012) use the same proprietary Fedwire data and a similar algorithm to create measures of overnight fed funds activity. We therefore expect their algorithm's output to suffer from the same problems we highlight in this study. Beyond fed funds, researchers have used algorithms based on Furfine (1999) to construct estimates of unsecured interbank lending. For instance, Kuo, Skeie, and Vickery (2012) have expanded the algorithm to identify loans with maturities longer than overnight. In addition, similar algorithms have been applied to Canadian and European data to identify overnight loans.[6] In particular, using data from TARGET2 (a large-value payments system for European banks), Arciero et al. (2013) conduct a test suggesting that their algorithm produces substantial type I errors but virtually no type II errors.

The test we conducted only demonstrates the inability of the algorithm to identify correctly individual overnight fed funds transactions conducted by two specific banks. Although we believe our results extend more generally, it is possible that the algorithm performs better for some specific types of banks. It is also possible that when the output of the algorithm is aggregated to the bank-to-bank level (that is, all transactions conducted between two banks), to the bank level (that is, all transactions conducted by a bank), or to the market level, it produces useful summary statistics to analyze the fed funds market. In the conclusion, we argue that our negative test results apply at the transaction, bank-to-bank, and bank levels, and we identify conditions under which the algorithm could be considered to produce accurate statistics at the market level. Finally, we discuss in the conclusion the possibility that, beyond fed funds, the algorithm output captures more general overnight interbank loans. Although we provide some evidence to support this hypothesis, we ultimately conclude that the algorithm cannot systematically recognize that a given pair of payments corresponds to an overnight interbank loan between two specific banks. In any case, the hypothesis that the algorithm's output captures overnight interbank loans would need to be formally tested in order to be validated. Until then, researchers and policymakers should be reluctant to use the algorithm's output as a proxy for interbank lending.

The remainder of the article is structured as follows. In section 2, we describe the algorithm and discuss its potential

[4] For example, foreign banks often have nostro accounts at domestic banks.

[5] Several of the papers mentioned in footnote 3 discuss this issue (such as Ashcraft and Bleakley [2006] and Afonso, Kovner, and Schoar [2011, 2013]). See also Furfine (1999).

[6] Hendry and Kamhi (2009), Allen et al. (2012), and Allen, Kastl, and Hortacsu (2012) make use of a similar algorithm applied to Canadian payments data. Millard and Polenghi (2004) and Acharya and Merrouche (2011) make use of a similar algorithm applied to U.K. payments data.

problems as a tool for identifying individual fed funds transactions. In section 3, we present the methodology underlying our test and report the outcome of the test. We conclude in section 4 with a discussion of our results' implications.

2. The Algorithm

2.1 Background

Fedwire is a real-time gross settlement system operated by the Federal Reserve. It enables depository institutions and other financial institutions to make large-value payments that are immediate and final.[7] To initiate a transfer through Fedwire, a participant must populate a number of fields in an electronic form specifying in particular the identity of the sending and receiving parties and the amount sent.

While data from Fedwire are not publicly available, some researchers within the Federal Reserve System have access to the transaction-level payments data. As part of this group of researchers, we can observe the universe of payments sent over Fedwire on any given day. However, we are only allowed to observe a subset of the message fields. Specifically, we observe the American Bankers Association (ABA) number of the sending and receiving banks, the amount sent, the time the payment was sent and received, a payment type code, and a payment business code. These last two fields give the bank sending the payment the opportunity to characterize the nature of the payment. Unfortunately, there are no industry-wide standards regarding the use of the payment type and business code fields. Consequently, the content of these two fields is not sufficient to determine unambiguously the nature of the payment sent.

To infer overnight fed funds transactions settled over Fedwire, Furfine (1999) proposed an algorithm that has been slightly adapted over the years by researchers at the FRBNY and the Federal Reserve Board. The current algorithm used by the FRBNY to produce some of its reports follows these general steps:

1. Transfers from or to a settlement institution (that is, CHIPS, CLS, or the Depository Trust Company) are dropped because loans to or from these institutions are not considered fed funds loans as defined by Regulation D.
2. On a given business day t, the algorithm considers every pair of banks $\{i,j\}$. Then, it constructs the set of possible send payments X_{ijt} consisting of all the transfers x_{ijt} from bank i to bank j on day t that are both greater than or equal to \$1 million and in increments of \$100,000. Each payment x_{ijt} in X_{ijt} is therefore considered to constitute the principal on a possible fed funds loan from bank i to bank j on day t.
3. For each payment x_{ijt} in the set X_{ijt}, the algorithm now constructs the set $Y(x_{ijt})$ of possible return payments the next business day (t+1). Specifically, every payment y_{jit+1} from bank j to bank i on day t+1 is evaluated to determine whether it could represent the principal x_{ijt} plus a plausible interest payment. To make this determination, the algorithm calculates the (annualized) interest rate implied by the pair of payments x_{ijt} and y_{jit+1}.[8] This implied interest rate is then compared with the range $[\underline{i}, \bar{i}]$, where $\underline{i}$ (respectively, $\bar{i}$) is the minimum (respectively, maximum) fed funds rate published by the FRBNY at date t minus (respectively, plus) 50 basis points.[9] If the implied interest rate is within the range $[\underline{i}, \bar{i}]$, then y_{jit+1} is included in the set $Y(x_{ijt})$ of possible return payments for x_{ijt}. Otherwise, y_{jit+1} is not considered a possible return payment for x_{ijt}.
4. Next, the algorithm determines the most likely return payment for each payment x_{ijt} in X_{ijt}. Three scenarios are possible. First, if there are no candidate return payments (that is, $Y(x_{ijt}) = \varnothing$), then x_{ijt} is not considered part of an overnight loan. Second, if there is a unique matching return payment (that is, $Y(x_{ijt})$ is a singleton), then x_{ijt} and the unique y_{jit+1} in $Y(x_{ijt})$ are linked and said to be an overnight loan. Third, if there are multiple candidate return payments (that is, $\dim[Y(x_{ijt})] > 1$), then the algorithm first computes the median interest rate implied by all the candidate payments in $Y(x_{ijt})$. The algorithm then chooses the return leg of the overnight loan with an implied interest rate that is closest to the median rate from above.[10] If linked to a send payment x_{ijt}, a return payment y_{jit+1} is then removed from consideration as a candidate match for all remaining send payments x'_{ijt} in X_{ijt}.[11]

[7] See Armantier, Arnold, and McAndrews (2008) for further details on Fedwire operations.

[8] This interest rate is equal to $((y_{jit+1}-x_{ijt})/x_{ijt})*(360/n)$, where n is the number of calendar days between business day t and t+1, while 360 is used to annualize an overnight loan, per convention in the fed funds market.

[9] Every day, the FRBNY conducts a survey of the four largest fed funds brokers. As mentioned in the introduction, the FRBNY uses this source of data to publish the mean, standard deviation, minimum, and maximum interest rates of brokered fed funds transactions for the prior day.

Currently, the minimum bound on an interest rate is the maximum of 0.9 basis point and the minimum fed funds rate reported by the FRBNY (using the data collected from fed funds brokers) minus 50 basis points. In the past, the minimum bound was the maximum of 1/32 and the minimum fed funds rate reported by the FRBNY minus 50 basis points. The absolute lower bound was pushed down from 1/32 to 0.009 percent because the extremely low nominal rates in recent times made interest rates below 1/32 plausible.

[10] In the case of ties, the algorithm chooses a return leg randomly among those with an implied interest rate closest to the median rate from above.

[11] The algorithm's output may differ depending upon the ordering of the x_{ijt} in the set X_{ijt}, because a matched return payment y_{jit+1} is removed from consideration, without replacement, as a candidate match for all remaining send payments x'_{ijt}. We have not yet studied how changes in the ordering of payments affect the algorithm's output.

5. Finally, the algorithm determines whether the overnight loans identified should be considered fed funds or Eurodollars. If the send leg on the pair of transactions has been given a "CTR" business code, then the pair of transactions is deemed an overnight Eurodollars loan.[12] Otherwise, the pair of transactions is classified as an overnight fed funds loan.[13]

At the end of these steps, the algorithm's final output consists of a series of paired Fedwire payments labeled as fed funds loans. To get a sense of the amount of filtering done by the algorithm, the algorithm identified slightly more than 0.7 percent of the 493,000 Fedwire payments sent on an average day in the first quarter of 2011 as being a leg of a fed funds loan.

If the algorithm is perfectly accurate, then the pairs of payments identified should capture the entire population of individual overnight fed funds loans settled over Fedwire that day. The algorithm therefore produces data at the most granular level, that is, individual loans between two specific banks. From each pair of payments, several characteristics may be inferred, such as the loan's interest rate, duration, or time of repayments. While the algorithm's output has a variety of uses, FRBNY researchers have used it to calculate summary statistics that describe features of the fed funds markets (for example, average rates and volumes) at the bank-to-bank, bank, and market levels.

2.2 Potential Problems

The algorithm described above produces pairs of payments that are labeled overnight fed funds loans. Here, we describe the potential mistakes the algorithm may make that would generate false positives and false negatives.[14]

False positives are pairs of payments that are incorrectly categorized as fed funds activity between the two specific banks sending and receiving the payments over Fedwire. Beyond the obvious case of two completely random payments incorrectly paired by the algorithm, we can suggest four general reasons why the algorithm could generate false positives.

First, the pair of transactions could be a fed funds loan, but not between the two banks sending the payments over Fedwire. As noted in the introduction, the algorithm cannot distinguish between a bank sending or receiving a payment on its own behalf and a bank doing so on behalf of a correspondent. In such a case, the algorithm would have identified a legitimate fed funds loan, but attributed it to the incorrect bank(s). This type of misassignment of counterparties will not affect aggregate market-level analysis, but it may bias estimates of fed funds activity at the transaction, bank-to-bank, or bank level. While we know this type of correspondent banking activity does occur, we do not know how often it occurs and how large a share of total fed funds activity it represents.

Second, the pair of transactions could be an overnight unsecured loan different from a fed funds transaction as defined under Regulation D. Observe that these types of loans may not exclusively capture interbank lending. In particular, the algorithm could pick up loans conducted on behalf of wealth-management funds, hedge funds, or even firms outside the financial sector.

Third, the pair of payments could be related to a collateralized loan. For the vast majority of collateralized loans, the cash portion is not sent over Fedwire. There is potentially a concern, however, with tri-party repo transactions.[15] While the cash portion of these repo transactions typically moves around on the books of the clearing banks, there are cases when the cash portion of a tri-party repo transaction is sent and returned between the cash investor and the clearing bank over Fedwire. This payment activity could be picked up in the algorithm and incorrectly labeled as a fed funds transaction.

Fourth, the algorithm could identify a legitimate fed funds loan, but incorrectly link one of the two payments related to that transaction. Such an error may occur when the algorithm finds multiple candidates for one of the legs of the transaction. Instead of picking the payment corresponding to the actual fed funds transaction, the algorithm incorrectly selects an unrelated but similar payment. In most cases, this mismatch might not severely bias the most important characteristics of the fed funds transaction (that is, interest rate, amount), but it could affect other characteristics, such as the timing of transactions.

False negatives are actual overnight fed funds loans settled through Fedwire that are not identified by the algorithm. The constraints embedded in the algorithm could produce such errors in at least two ways: First, the algorithm requires the principal amount of fed funds loans to be greater than or equal to $1 million and in increments of $100,000. Actual fed funds activity in which the principal is less than $1 million or is not in an increment of $100,000 will be missed by the algorithm. Second, if there is considerable variability in the fed funds rates

[12] "CTR" stands for customer transfer, and is meant to designate that the beneficiary of the payment is not a bank.

[13] The motivation for using the CTR business code to differentiate fed funds loans from Eurodollars loans is based on internal work at the FRBNY. The classification, however, may include errors because the use of the CTR code by banks is neither mandatory nor an explicit industry standard.

[14] Some of the potential mistakes listed in this section have been previously discussed in, for example, Furfine (1999).

[15] See Copeland, Martin, and Walker (2010) for a description of the tri-party repo market.

across banks, the plus or minus 50 basis point range around the minimum and maximum fed funds rate published by the FRBNY might rule out actual fed funds activity.

In addition to the systematic problems that may arise with the algorithm, idiosyncratic difficulties exist. A bank, for example, may return the principal and interest associated with a fed funds loan in two separate payments. Likewise, fails can occur when a bank, perhaps because of operational difficulties, does not return the principal and interest the next day. According to a handful of industry participants, these events rarely occur. When they do occur, however, the algorithm will not identify the underlying fed funds activity. Finally, the objective of the algorithm is to identify fed funds activity settled through Fedwire. As a result, the algorithm cannot provide any information about fed funds loans settled outside Fedwire—for example, over other payment systems or on a bank's books.

3. Testing the Quality of the Algorithm

3.1 The Test's Methodology

From the perspective of a given bank, each of its fed funds transactions consists of two legs: a "send leg," in which the money flows from the bank to its counterparty, and a "receive leg," in which the money flows from the counterparty to the bank. When a bank sells fed funds, the send leg precedes the receive leg; when the bank purchases fed funds from a counterparty, the receive leg precedes the send leg. The perspective of the bank's counterparty is the mirror image—that is, the send leg for a bank that sells fed funds is the receive leg for the counterparty that purchases the fed funds.

Every day, banks may send and receive a large number of payments over Fedwire (more than 150,000 in some cases), a tiny portion of which correspond to fed funds transactions (typically less than 0.1 percent). Because banks must keep track in real time of every fed funds transaction they conduct, they have to be able to flag automatically a fed funds transaction from within the flow of Fedwire payments they receive. To do so, large banks typically require their fed funds counterparties to incorporate an identifier into the message portion of the Fedwire payment. Two of these banks voluntarily gave us access to their unique identifiers. Using these identifiers, we can locate the receive leg of every fed funds transaction the two banks have conducted by searching for the unique identifier within the message fields of all Fedwire payments they receive.[16] Unfortunately, we do not have access to the unique identifiers for the two banks' counterparties (except, of course, when these two banks interact with each other). Thus, we can identify only the receive legs but not the send legs of the fed funds transactions conducted by the two banks. Consequently, we do not know for sure the true interest rate associated with a receive leg of a fed funds transaction, because it takes both legs to infer unambiguously the interest rate of a fed funds loan. Although this limitation has no impact on our estimates of type I and type II errors, we will need to keep it in mind when studying the interest rates produced by the algorithm.

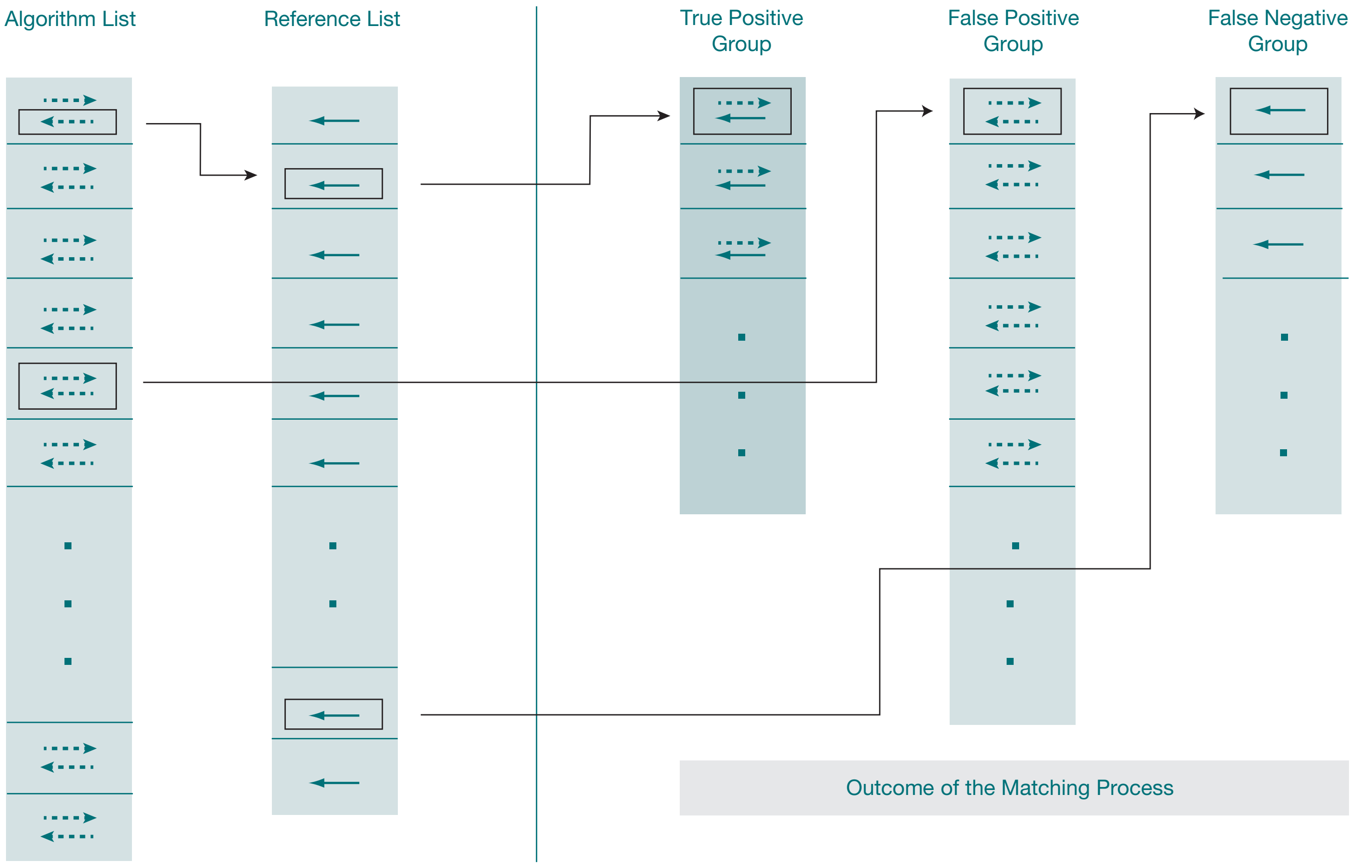
Our goal is to establish how well the algorithm identifies overnight fed funds transactions conducted by the two banks over Fedwire. To do so, we consider all possible pairs of payments $\{x_{ijt}, y_{jit+1}\}$ on consecutive business days between bank i and bank j, where bank i or j is one of the two banks for which we have a unique identifier. The null hypothesis is that $\{x_{ijt}, y_{jit+1}\}$ is not a fed funds loan, while the alternative hypothesis is that $\{x_{ijt}, y_{jit+1}\}$ is a fed funds loan. The algorithm can be seen as a test of the null hypothesis because it provides a method to decide which $\{x_{ijt}, y_{jit+1}\}$ should or should not be considered a fed funds loan. Because the presence of the unique identifier flags unambiguously which receive legs are, and which receive legs are not, part of a fed funds loan for our two banks, we can estimate when the algorithm incorrectly rejects the null hypothesis (type I error) and when the algorithm incorrectly accepts the null hypothesis (type II error). The method we use to construct these estimates consists of three steps (see the exhibit).

First, we run the algorithm for the two banks for every business day within a quarter. This gives us a list of paired payments, each consisting of a send leg and a receive leg. We call this the "algorithm list." Second, we construct another list of payments (the "reference list") by searching for the unique identifier over all the Fedwire payments the two banks received on every business day within the quarter. This reference list therefore consists of receive legs identifying all fed funds payments the banks received that quarter. Third, we compare the algorithm and reference lists, searching for matches. Specifically, we verify whether each of the receive legs in the reference list can be found in the algorithm list.

As illustrated in the exhibit, this matching process produces three different groups. The "true positive group" consists of every pair of payments in the algorithm list with a match in the reference list. The "false positive group" consists

[16] To be clear, the unique identifier is included in the receive leg of every fed-funds-related transaction conducted by the two banks, regardless of whether the two banks purchased or sold fed funds in that transaction.

The Test's Methodology



Notes: The "algorithm list" consists of all pairs of payments identified by the algorithm as federal funds loans. The "reference list" consists of all Fedwire payments with the unique identifier. The "true positive group" consists of every pair of payments in the algorithm list with a match in the reference list. The "false positive group" consists of every pair of payments in the algorithm list without a match in the reference list. The "false negative group" consists of every receive leg in the reference list without a match in the algorithm list. A dashed line indicates a send or a receive leg of a federal funds transaction identified by the algorithm. A solid line indicates a receive leg of a fed funds transaction with the unique identifier.

of every pair in the algorithm list without a match in the reference list. Finally, the "false negative group" consists of the receive legs in the reference list without a match in the algorithm list. The size of the false positive group relative to the size of the algorithm list gives us an estimate of the algorithm's type I error for the two banks. Similarly, the size of the false negative group relative to the size of the reference list gives us an estimate of the type II error for the two banks.[17]

[17] Technically, the type I error rate is the probability of the receiving leg not being part of a fed funds loan conditional on the algorithm labeling the receiving leg as part of a fed funds loan. The type II error rate is the probability of the algorithm not labeling the receiving leg as part of a fed funds loan conditional on the receiving leg being part of a fed funds loan.

This methodology, in fact, provides only a lower bound on the extent of type I errors for at least two reasons. First, we can test whether the algorithm correctly identifies the receive leg of a fed funds transaction but, because of the possibility of correspondent banking, we cannot confirm that the bank that sent the Fedwire payment is indeed the counterparty in the fed funds transaction. Second, a pair of payments is in the true positive group if it possesses the receive leg of an actual fed funds transaction. This does not imply, however, that the algorithm correctly identified the send leg of that fed funds transaction. As mentioned earlier, our methodology does not allow us to test this hypothesis. The consequences of such mismatches, however, should not be expected to be too severe. Although the mismatches may seriously affect some characteristics of the fed funds transactions (for example, the exact duration of the loan), in general they should not substantially affect the more important characteristics (that is, the amount loaned and the interest rate inferred). Indeed, by construction, the algorithm can only match an incorrect send leg to the receive leg of an actual fed funds transaction if the amount of this incorrect send leg is similar to the amount of a true send leg. As a result, we expect the interest rates inferred for the pairs of payments in the true positive group to be reasonably accurate.

In contrast, our methodology provides an upper bound on the extent of type II errors. Indeed, the two banks under consideration ask their counterparties to include the unique identifier for payments corresponding to any fed funds transactions, which include overnight as well as term fed funds transactions. As a result, some of the fed funds payments in the false negative group may not correspond to overnight loans, and our test's methodology may therefore exaggerate the extent of type II errors. Although we cannot quantify precisely the extent of this problem, conversations with fed funds traders at each of the two banks suggest that the number of term fed funds transactions they conduct is relatively small.

To conclude this section, we want to acknowledge that the validity of our test hinges on the fact that the unique identifiers provided by the two banks are included in every fed funds transaction they settle over Fedwire. Note that the validity of the unique identifiers has been confirmed at various points in time by different members of the two banks in question. Further, we were able to find independent evidence from a third, unrelated bank. Indeed, this third bank confirmed that a necessary condition to remain a fed funds counterparty to the two banks on which we base our test is that every fed funds payment sent over Fedwire must include the unique identifiers.[18]

3.2 Type I and Type II Errors

The results reported in Table 1 are discouraging. In the first quarter of 2007, the type I error produced by the algorithm is estimated to be 64 percent (18,633/29,077). While much lower, the estimated type II error, at 24 percent

Table 1
Estimates of Type I and Type II Errors for 2007:Q1

Algorithm List		Reference List
29,077		13,655
False Positive Group	True Positive Group	False Negative Group
18,633	10,444	3,211
Type I error: 64 percent		Type II error: 24 percent

Source: Authors' calculations, based on Fedwire data.

Note: The type I error is equal to the false positive group divided by the algorithm list; the type II error is equal to the false negative group divided by the reference list.

(3,211/13,655), is not inconsequential. To measure how well the algorithm performed through the recent financial crisis, we estimated the type I and type II errors for these two banks for the first quarters of each year between 2007 and 2011 (see Table 2).[19] The type I error is estimated to be higher as we go forward in time, reaching 93 percent in the first quarters of 2010 and 2011. Conversely, the type II error is estimated to be lower as we go forward in time, slightly declining to 17 percent in the first quarter of 2011.[20] On average, the type I error is estimated to be 81.4 percent from 2007 to 2011 and the average type II error is estimated to be 23.0 percent.

As noted earlier, type I errors may be the result of several factors (for example, the algorithm matches two completely unrelated payments or identifies a loan other than an overnight fed funds transaction). Although we are unable to trace back the source of these type I errors, we conjecture that correspondent banking, whereby the algorithm incorrectly assigns to our two banks fed funds transactions conducted on behalf of some of their clients, plays a major role.

In contrast, we can quantify some of the reasons behind type II errors. While we focus on the first quarter of 2007 for this analysis, similar results were found in the first quarter of 2011. First, the algorithm classifies some pairs of transactions as Eurodollars when they are in fact fed funds. Our results suggest that this occurs relatively frequently. In particular, out of the 3,211 fed funds transactions not recognized by the algorithm in the first quarter of 2007, 1,455, or 45 percent, had

[18] Ideally, we would have liked to double-check the validity of the hypothesis by comparing the transactions carrying unique identifiers with another source of data on fed funds transactions. However, we are not aware of such an alternative source. In particular, the data reported by depository institutions in the Federal Financial Institutions Examination Council's Consolidated Reports of Condition and Income and by bank holding companies in the Federal Reserve's FR Y-9C forms do not isolate fed funds transactions as defined by Regulation D. Instead, these filings report a broader measure of purchases and sales of unsecured funds among financial institutions.

[19] Because of technical limitations, the furthest back we can go to test the algorithm is 2007.

[20] We do not know why there are opposing trends in our estimates of the type I and type II errors. The total number of payments sent and received by these two banks over Fedwire is roughly flat over this time period. Further, the number of payments exceeding $1 million sent and received by these two banks over Fedwire is also roughly flat, except for a decline of 20 percent from the first quarter of 2008 to the first quarter of 2009.

Table 2
Estimates of Type I and Type II Errors over Time
Percent

	2007:Q1	2008:Q1	2009:Q1	2010:Q1	2011:Q1	Average
Type I	64	72	85	93	93	81.4
Type II	24	28	27	19	17	23.0

Source: Authors' calculations, based on Fedwire data.

been discarded by the algorithm as being Eurodollars.[21] Second, by construction, the algorithm ignores fed funds loans where the principal is less than $1 million. In the first quarter of 2007, there were 170 such small fed funds transactions, accounting for 5 percent of the 3,211 false negatives. These small overlooked fed funds transactions, however, account for only 0.07 percent of the false negatives in terms of dollar value. Third, the algorithm could have faced multiple-candidate receive legs and did not choose the correct receive leg with the identifier. This only happened in the case of 128 of the 3,211 false negatives (4 percent). Fourth and finally, even if a payment is above $1 million, the algorithm may not find a potential match because, for example, it is a term loan, or the negotiated interest rate is outside the range specified by the algorithm. For the first quarter of 2007, 1,458, or 45 percent, of the transactions fall into this category.

3.3 Is the Output of the Algorithm Biased?

Given the high rates of type I and type II errors, it would appear that the algorithm's transaction-level output is ill-suited to study the fed funds market, and more generally to conduct research. Nevertheless, it is possible that the algorithm's errors may be considered white noise, in which case the algorithm's output would be unbiased. Unfortunately, we find evidence that the algorithm does produce biased outputs along at least three dimensions: the set of counterparties, the distribution of amounts loaned, and the distribution of interest rates. Once again, we focus on the first quarter of 2007 for this analysis, but find that the algorithm produces similar biases in the first quarter of 2011.

We first examine the set of counterparties for both fed funds sold and fed funds purchased by the two banks in the first quarter of 2007.[22] For each of the two banks, we compare the top ten counterparties, as ranked by the number of transactions, for the reference and algorithm lists.[23] For both banks, only three of the top ten counterparties in the algorithm list also appear in the top ten counterparties in the reference list. When ranking counterparties by the total value of their transactions, for both banks we find that five of the top ten counterparties in the algorithm list also appear in the equivalent top ten counterparties in the reference list. This comparison illustrates the algorithm's poor performance in correctly identifying the most important counterparties of the two banks.

We now turn to quantities. In the reference list, we observe the amount of the receive legs of the fed funds loans conducted by the two banks. From the algorithm list, we construct a comparable set of amounts by extracting the receive leg from each pair of payments linked by the algorithm. As illustrated in Chart 1, the distributions of amounts differ across these two sets of payments. Specifically, the amounts in the reference list tend to be smaller than those in the algorithm list. In particular, the mean and median amounts in the reference list are $18.1 million and $72.5 million, as compared with $50 million and $143.8 million in the algorithm list. Using the Mann-Whitney U test, we can reject at the 1 percent significance level the null hypothesis that the distributions of amounts across both samples are equal (the Z-score is -54.8). We therefore find statistical evidence that the algorithm output is biased with respect to the amounts of fed funds loans. Similar biases are identified when we consider separately the amount of fed funds sold and the amount of fed funds purchased by the two banks (see Appendix Charts A1 and A2).

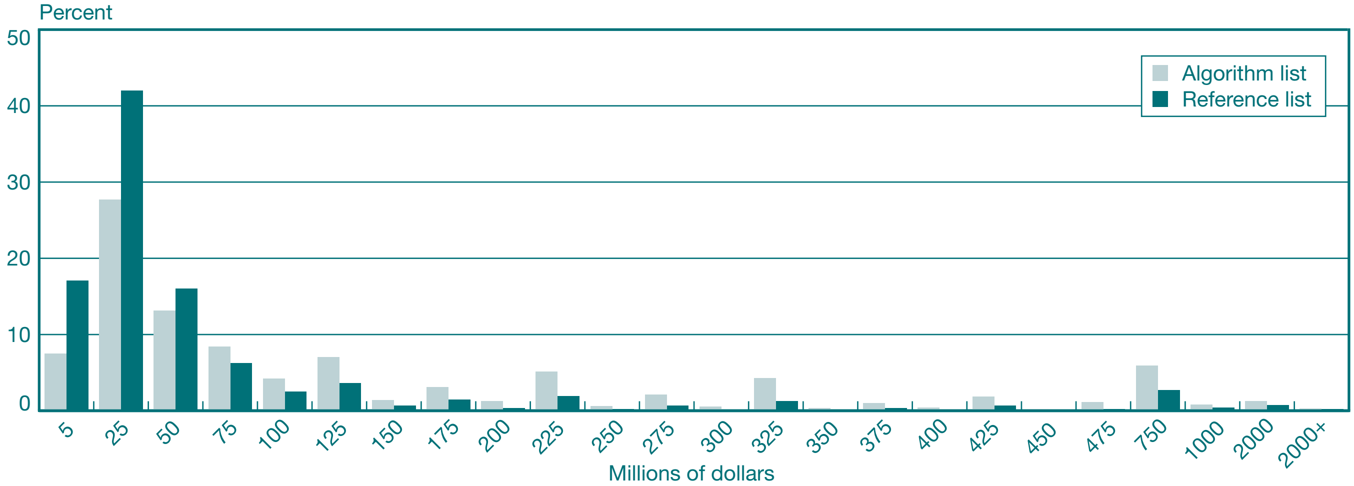
Finally, we consider interest rates. To compute the interest rate for a transaction in the reference list, we need to pair the receive leg with its send leg. As the latter is unobserved, the pairing can only be approximated. For the comparisons conducted below, we focus on the set of true positives in the first quarter of 2007, that is, the 10,444 send legs in the algorithm list that can be matched to a receive leg in the reference list. We can then compare the inferred interest rates from this set of transactions to the inferred interest rates in the algorithm list. In Charts 2 and 3, we plot the interest rate distributions

[21] In the first quarter of 2007, 32,647 pairs of payments were classified as Eurodollars instead of fed funds because the send leg had been given a "CTR" business code (see step 5 of the algorithm in section 2.1). We find that out of these 32,647 pairs of payments, only 1,455, or 4.5 percent, were in fact fed funds transactions. Our results therefore support the presumption that the "CTR" business code is an effective (albeit imperfect) way to distinguish Eurodollar from fed funds loans.

[22] Recall that neither the algorithm nor the unique identifiers for the two banks allow us to identify with certainty the fed funds counterparty of the banks. So instead of comparing counterparties, we may actually be comparing the correspondent banks of the true counterparties.

[23] According to the reference list, the top ten counterparties for each of the two banks account for, very roughly, two-tenths of the total number of fed funds transactions conducted by the two banks and one-half of their total value of fed funds activity.

Chart 1
Comparison of Transaction Amounts across the Algorithm and Reference Lists



Source: Authors' calculations, based on Fedwire data.

Notes: The comparison is conducted for 2007:Q1. For the algorithm list, amounts plotted are those in the receive leg of the paired payment transactions. The horizontal axis label is the amount bin's larger end point, except for "2000+," which denotes the bin with all payments greater than $2,000 million.

for the fed funds sold and purchased by the two banks. Our findings are similar to those for our analysis of amounts: the distributions of rates produced by the algorithm differ from the distributions of rates of the true positives. In particular, the median rates of fed funds sold and purchased are, respectively, 537 and 519 basis points for the true positives, while the median rates of fed funds sold and purchased are, respectively, 525 and 523 basis points for the algorithm list.[24] Using a Mann-Whitney U test, we can reject at the 1 percent significance level the null hypothesis that these distributions of rates are equal (the Z-score is -26.3 for fed funds sold and -33.3 for fed funds purchased). Because the algorithm is biased downward for fed funds sold and upward for fed funds purchased, these biases partially offset each other when the interest rates of fed funds sold and purchased by the two banks are combined. Nevertheless, even when fed funds sold and purchased are combined, there remain significant differences between the distribution of interest rates inferred from the algorithm and the distribution of interest rates from true positives (Appendix Chart A3). Hence, we find that the interest rates produced by the algorithm are statistically biased for fed funds sold and fed funds purchased—by 12 and 4 basis points, respectively. To gauge the economic magnitude of these biases, we note that over the same time period, the average spread between the overnight Libor rate (for U.S. dollars) and the one-month (six-month) Libor rate was 1.5 (5.7) basis points.

These three comparisons provide statistical evidence of significant bias in the set of counterparties as well as the distributions of transaction amounts and interest rates inferred from the algorithm's output for our two banks. In other words, the algorithm's errors are not just white noise. Rather, the main characteristics of the pairs of payments produced by the algorithm seem to exhibit systematic biases. Further, the nature of these biases is such that they do not subside when the algorithm's output is aggregated to the bank-to-bank level, or at the bank level. Finally, the algorithm's errors and biases remain essentially unchanged when its implementation is slightly modified (for example, by relaxing the minimum $1 million loan amount or widening the range of possible interest rates).

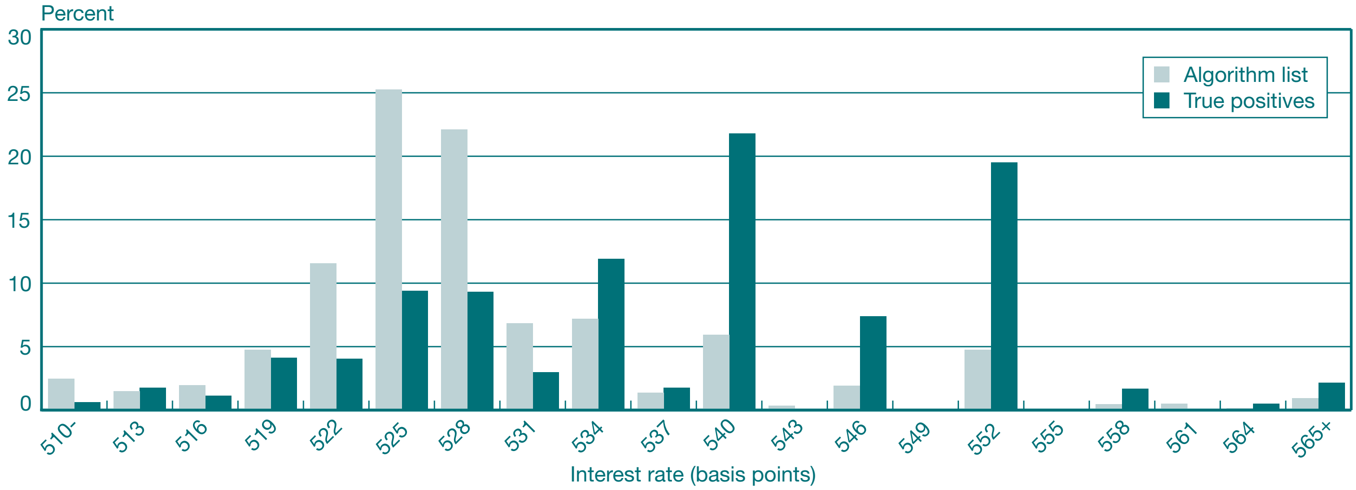
4. Discussion

Because the federal funds market has been one of the key financial markets in the United States, it has attracted considerable attention from researchers, especially after the 2008-09 financial crisis. Empirical analyses of this market have typically relied on transactions inferred by an algorithm

[24] In the first quarter of 2007, the target fed funds rate was 525 basis points.

Chart 2

Comparison of Interest Rates for Federal Funds Sold

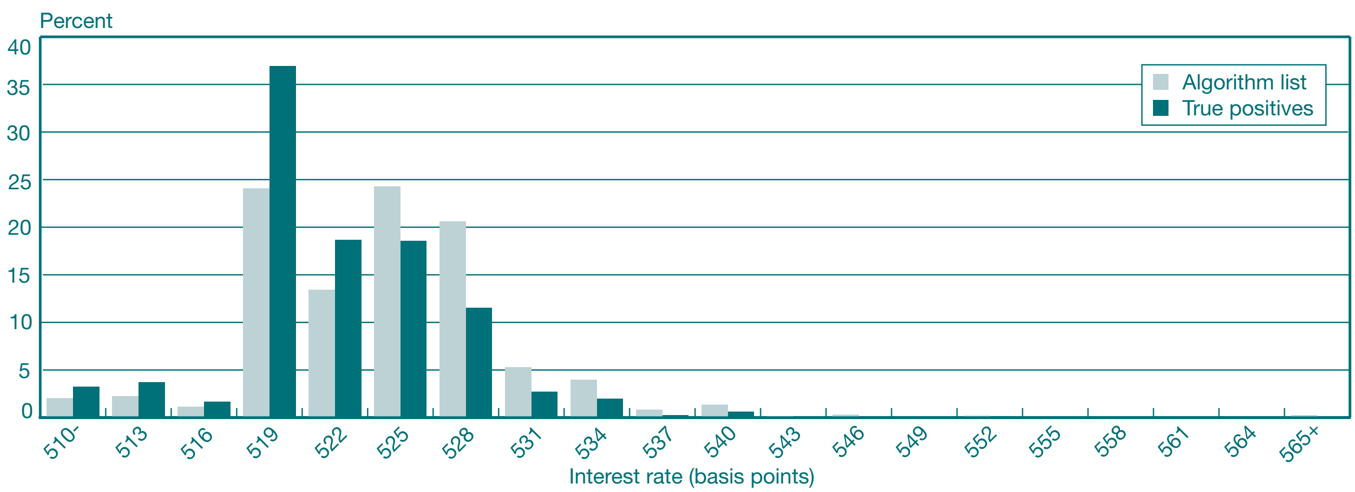


Source: Authors' calculations, based on Fedwire data.

Notes: The comparison is conducted for 2007:Q1. Note that the federal funds rate targeted by the Federal Open Market Committee in this quarter was 525 basis points. The horizontal axis label is the rate bin's larger end point, except for "565+," which denotes the bin with all interest rates greater than 565 basis points.

Chart 3

Comparison of Interest Rates for Federal Funds Purchased



Source: Authors' calculations, based on Fedwire data.

Notes: The comparison is conducted for 2007:Q1. Note that the federal funds rate targeted by the Federal Open Market Committee in this quarter was 525 basis points. The horizontal axis label is the rate bin's larger end point, except for "565+," which denotes the bin with all interest rates greater than 565 basis points.

comparable to the one used by the Federal Reserve Bank of New York. There is no guarantee, however, that this algorithm correctly identifies individual fed funds transactions.

In this article, we reported on a test aimed at assessing the transaction-level quality of the algorithm. For two large banks, among the more active in the fed funds market, we find the type I and type II errors to be large, averaging 81 percent and 23 percent, respectively, from 2007 to 2011. Further, we find evidence suggesting that these large errors cannot be considered white noise. Rather, they introduce significant biases in the computed rate and volume of fed funds activity, as well as in the set of counterparties. To be sure, we want to acknowledge that our study has possible limitations. In particular, our test applies only to fed funds as defined under Regulation D, and is based only on two banks. Despite these limitations, however, we argue below that our results have important implications.

4.1 How General Are Our Results?

The two institutions on which our test is based are large banks and so are not representative of all participants in the fed funds market. Hence, there is a possibility that the results of our test do not generalize to other fed funds participants. We provide two reasons, however, why we believe our results do, in fact, apply quite broadly. First, the two banks that provided their unique identifiers are either senders or receivers for a sizable share of all pairs of transactions that are output by the algorithm. Over the 2007-11 period, the two banks were involved, on average, with 29.4 percent of the algorithm's output. Our results, then, directly relate to a large fraction of the algorithm's output. Second, we believe it is reasonable to assume that our results are applicable to other large banks with similar Fedwire activity. We define large banks as those that receive or send over 800,000 payments a quarter (in a typical quarter, only nine or ten banks met this criterion). Assuming that our type I and type II errors generalize to these banks implies that, on average, 44.3 percent of the algorithm's output is affected (see Table 3).

The algorithm, however, may perform better for smaller banks. Indeed, these banks send fewer payments over Fedwire, and these payments may reflect fewer types of transactions. As a result, it might be easier for the algorithm to recognize fed funds transactions initiated by smaller banks. Although we cannot test it formally at this point, this hypothesis finds some support in the fact that there is separate preliminary evidence that the algorithm may perform well for some government-sponsored enterprises. If one can establish that

Table 3

Percent of Algorithm's Output to Which the Type I and Type II Error Estimates Apply

	2007:Q1	2008:Q1	2009:Q1	2010:Q1	2011:Q1	Average
Two banks	29.0	25.0	28.0	31.4	33.6	29.4
Large banks	39.5	37.4	40.7	49.4	54.7	44.3

Source: Authors' calculations, based on Fedwire data.

Notes: "Two banks" are the two institutions on which our tests are based. "Large banks" are those that sent and received more than 800,000 payments in the relevant quarter. The same nine banks met this criterion every quarter in the table, including the two banks at the center of our analysis. A tenth bank met this criterion in the first quarters of 2007, 2008, and 2011, although the identity of this tenth bank is not the same across the three quarters.

the algorithm is only inaccurate for a few large banks, then a possible remedy could be to exclude these banks from any empirical analysis. Still, we see at least three problems with this approach. First, ignoring at least a third of all transactions output by the algorithm would prevent any comprehensive analysis of the fed funds market. Second, one would have to show that excluding banks in a nonrandom way does not introduce biases in the algorithm output. Third, this approach would not only exclude the fed funds transactions conducted by these large banks, but also those involving their smaller clients as part of correspondent banking. As a result, excluding a few large banks may not permit an accurate analysis of the fed funds transactions conducted by smaller banks.

4.2 Does Aggregating the Algorithm's Output Make It More Precise?

Our test suggests that the algorithm is unlikely to identify individual fed funds transactions correctly. However, if aggregated to the bank-to-bank level, the bank level, or the market level, could the algorithm's output be useful to study the fed funds market? In part because the algorithm cannot identify the ultimate originator or beneficiary of a fed funds transaction, we do not think that the algorithm can provide, in general, meaningful measures at the bank-to-bank or the bank level. In particular, the algorithm will attribute 1) more transactions to large banks that serve as intermediaries, and 2) fewer transactions to small banks using correspondent banks. Because small and large banks may transact fed funds at different rates, the average rate identified by the algorithm for those banks may be biased.

Our analysis provides little evidence that the algorithm may or may not provide accurate market-level measures of fed funds activities. Nevertheless, we note that correspondent banking may possibly be the major source of type I errors in our test. In other words, the algorithm may correctly identify fed funds transactions but attribute them to the wrong originator or beneficiary. If this is the case, then the algorithm would produce unbiased market-level data on the distribution of rates and volumes of fed funds. The algorithm's output would then be a useful complement to the data obtained through brokers by the FRBNY, because it would cover fed funds transactions arranged both through brokers and privately between banks. To confirm this hypothesis, however, further work is necessary to test whether the algorithm's type I errors are almost exclusively produced by correspondent banking.

4.3 Does the Algorithm's Output Capture More General Interbank Overnight Loans?

While the available evidence points to the algorithm's output being imprecise measures of fed funds activity at the transaction and bank levels, the algorithm may still be of value if it captures a broader type of overnight funding. This would follow if most of the false positives identified in our test were indeed loans, but simply not fed funds loans (for example, if they were loans to financial institutions other than banks). This hypothesis finds support in the fact that 89 percent of the transactions paired by the algorithm in first quarter of 2007 are found to have inferred interest rates that, once rounded, can be considered to be in whole basis points or 32nds of an interest rate.[25] Discussions with market participants suggest that overnight unsecured loans are typically traded in these discrete amounts, suggesting that the pairing of transactions by the algorithm is not random.[26]

We note, however, that even if the algorithm correctly identifies loans, it may not accurately identify interbank loans. This would be the case in particular if loans are placed on behalf of bank clients that are outside the banking system or even the financial sector. Furthermore, even in the case of an interbank loan, the algorithm cannot guarantee the identity of the originator and the beneficiary because of the possibility of correspondent banking. More generally, the hypothesis that the algorithm's output captures overnight interbank loans would need to be formally tested in order to be validated. Until then, we believe that the algorithm's output should not be used as a proxy for interbank lending.

In conclusion, our results raise serious concerns about the appropriateness of using the algorithm's output to study the fed funds market. As a consequence, it raises questions about the validity of empirical results previously obtained using the algorithm's output. Finally, our analysis underscores the need to validate formally, prior to any analysis, that the indirect inferences produced by an algorithm are accurate.

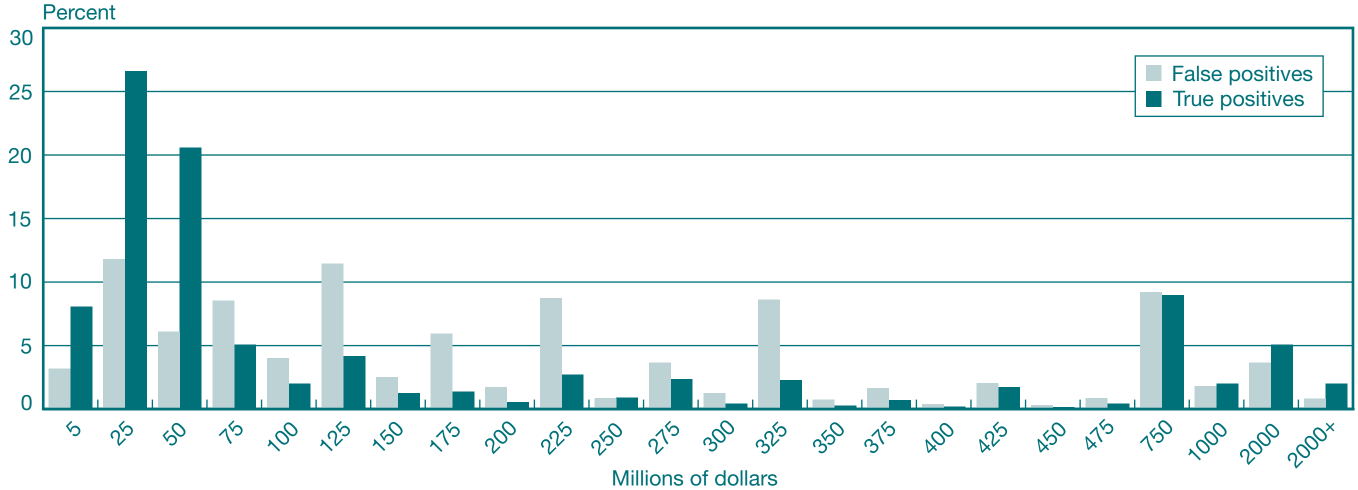
[25] The dollar amount a bank can send to another bank over Fedwire is constrained to be rounded to the nearest cent. Because of rounding, the interest rate agreed upon by the banks when agreeing to a trade may differ from the interest rate we compute from the payment flows. Hence, when checking whether an implied interest rate is in whole basis points, we account for rounding. We do this by computing the implied interest rate when the principal and interest payment amount is increased by one cent and then when the amount is decreased by one cent. If these two inferred interest rates straddle an interest rate in whole basis points or 32nds of an interest rate, then we say that the algorithm's implied interest rate is consistent with a loan with an interest rate in whole basis points or 32nds of an interest rate.

[26] Substantiating these claims by market participants, we found that the interest rates of brokered fed funds trades between February 11, 2002, and September 24, 2004, provided by BGC Brokers, were all in whole basis points or 32nds of an interest rate. See Bartolini, Hilton, and McAndrews (2008) for details on these data.

Appendix

Chart A1

Comparison of Transaction Amounts of Federal Funds Sold

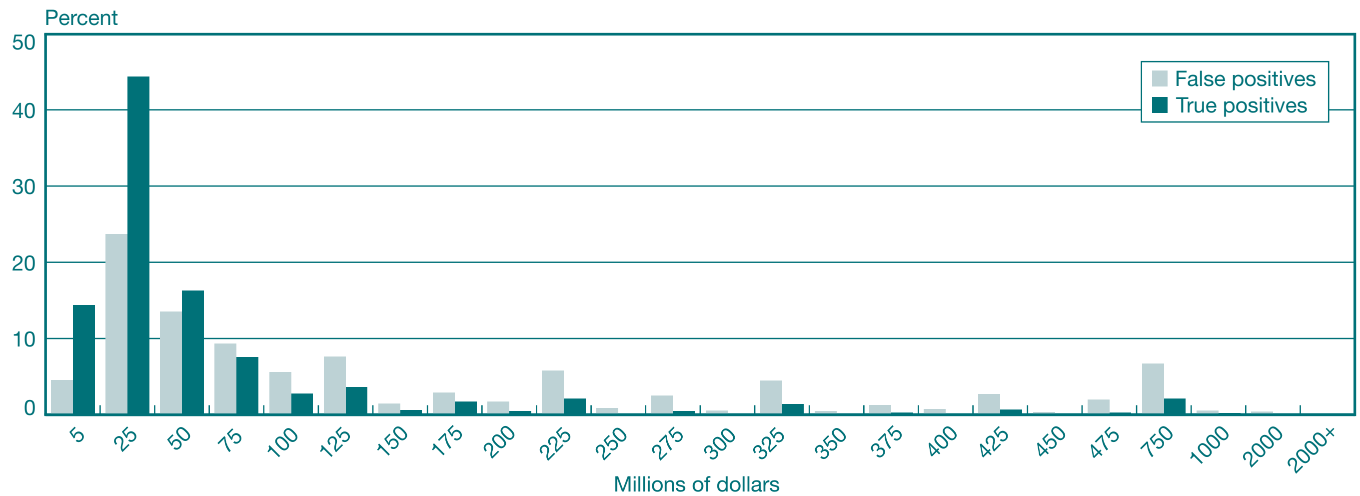


Source: Authors' calculations, based on Fedwire data.

Notes: The comparison is conducted for 2007:Q1. The principal amount of the federal funds sale is graphed. The horizontal axis label is the amount bin's larger end point, except for "2000+," which denotes the bin with all payments greater than $2,000 million. A Mann-Whitney U test rejects the null hypothesis that the distribution of amounts across false positives and true positives is equal at the 1 percent significance level (the Z-score is -15.0).

Chart A2

Comparison of Transaction Amounts of Federal Funds Purchased

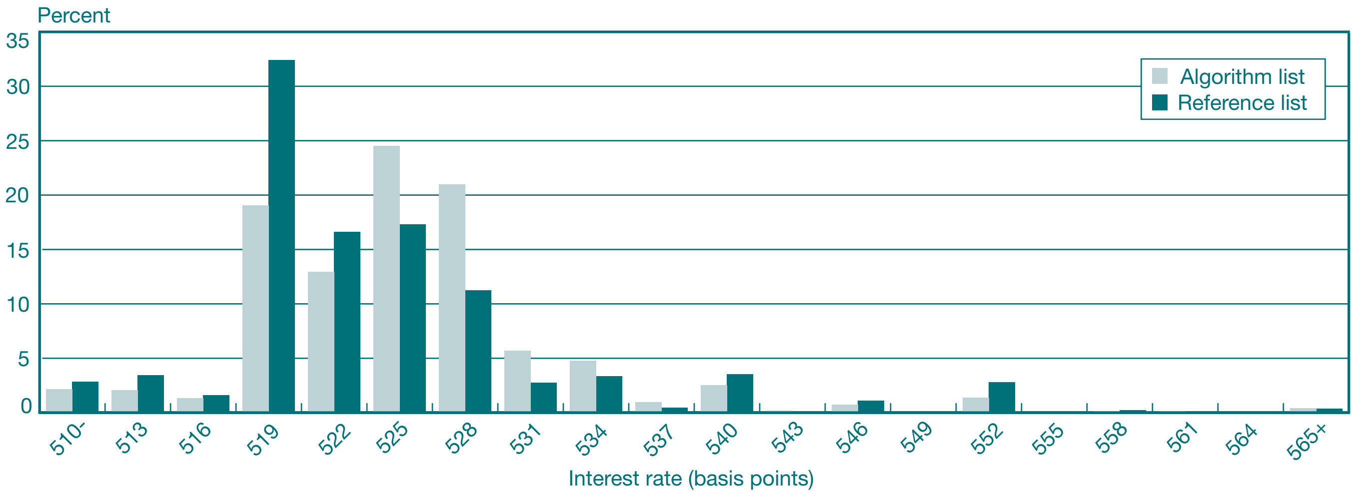


Source: Authors' calculations, based on Fedwire data.

Notes: The comparison is conducted for 2007:Q1. The principal amount of the federal funds purchased is graphed. The horizontal axis label is the amount bin's larger end point, except for "2000+," which denotes the bin with all payments greater than $2,000 million. A Mann-Whitney U test rejects the null hypothesis that the distribution of amounts across false positives and true positives is equal at the 1 percent significance level (the Z-score is -53.0).

Chart A3

Comparison of Interest Rates across the Algorithm and Reference Lists When Federal Funds Sold and Federal Funds Purchased Are Combined



Source: Authors' calculations, based on Fedwire data.

Notes: The comparison is conducted for 2007:Q1. For the reference list, interest rates were inferred for only those transactions in the set of true positives. Note that the federal funds rate targeted by the Federal Open Market Committee in this quarter was 525 basis points. The horizontal axis label is the rate bin's larger end point, except for "565+," which denotes the bin with all interest rates greater than 565 basis points.

References

Acharya, V., and O. Merrouche. 2011. "Precautionary Hoarding of Liquidity and Inter-Bank Markets: Evidence from the Sub-Prime Crisis." Review of Finance 17, no. 1: 107-60.

Acharya, V., and D. Skeie. 2011. "A Model of Liquidity Hoarding and Term Premia in Inter-Bank Markets." Journal of Monetary Economics 58, no. 5: 436-47.

Afonso, G., A. Kovner, and A. Schoar. 2011. "Stressed, not Frozen: The Federal Funds Market in the Financial Crisis." Journal of Finance 66, no. 4: 1109-39.

———. 2013. "The Importance of Trading Relationships in the Interbank Lending Market." Federal Reserve Bank of New York Staff Reports, no. 620, May.

Afonso, G., and R. Lagos. 2012a. "Trade Dynamics in the Market for Federal Funds." Federal Reserve Bank of New York Staff Reports, no. 549, February.

———. 2012b. "An Empirical Study of Trade Dynamics in the Fed Funds Market." Federal Reserve Bank of New York Staff Reports, no. 550, February.

Allen, J., J. Chapman, F. Echenique, and M. Shum. 2012. "Efficiency and Bargaining Power in the Interbank Loan Market." Bank of Canada Working Paper no. 2012-29.

Allen, J., J. Kastl, and A. Hortacsu. 2012. "Crisis Management: Analyzing Default Risk and Liquidity Demand during Financial Stress." Unpublished paper.

Arciero, L., R. Heijmans, R. Heuver, M. Massarenti, C. Picillo, and F. Vacirca. 2013. "How to Measure the Unsecured Money Market? The Eurosystem's Implementation and Validation Using TARGET2 Data." De Nederlandsche Bank Working Paper no. 369.

Armantier, O., E. Ghysels, A. Sarkar, and J. Shrader. 2011. "Stigma in Financial Markets: Evidence from Liquidity Auctions and Discount Window Borrowing during the Crisis." Federal Reserve Bank of New York Staff Reports, no. 483, January.

Armantier, O., J. McAndrews, and J. Shrader. 2008. "Changes in the Timing Distribution of Fedwire Funds Transfers." Federal Reserve Bank of New York Economic Policy Review 14, no. 2, September.

Ashcraft, A., and H. Bleakley. 2006. "On the Market Discipline of Informationally Opaque Firms: Evidence from Bank Borrowers in the Federal Funds Market." Federal Reserve Bank of New York Staff Reports, no. 257, August.

Ashcraft, A., and D. Duffie. 2007. "Systemic Illiquidity in the Federal Funds Market." American Economic Review: Papers and Proceedings 97, no. 2: 221-5.

Ashcraft, A., J. McAndrews, and D. Skeie. 2011. "Precautionary Reserves and the Interbank Market." Journal of Money, Credit, and Banking 43, no. 7: 311-48.

Atalay, E., and M. Bech. 2010. "The Topology of the Federal Funds Market." Physica A: Statistical Mechanics and Its Applications 389, no. 22.

Bartolini, L., S. Hilton, and J. McAndrews. 2008. "Settlement Delays in the Money Market." Federal Reserve Bank of New York Staff Reports, no. 319, March.

Bech, M. L., C. T. Bergstrom, R. J. Garratt, and M. Rosvall. 2011."Mapping Change in the Federal Funds Market." Federal Reserve Bank of New York Staff Reports, no. 507, August.

Bech, M., E. Klee, and V. Stebunovs. 2012. "Arbitrage, Liquidity, and Exit: The Repo and Federal Funds Markets before, during, and Emerging from the Financial Crisis." Board of Governors of the Federal Reserve System Finance and Economics Discussion Series, no. 2012-21.

Copeland, A., A. Martin, and M. Walker. 2010. "The Tri-Party Repo Market before the 2010 Reforms." Federal Reserve Bank of New York Staff Reports, no. 477, November.

Demiralp, S., B. Preslopsky, and W. Whitesell. 2006. "Overnight Interbank Loan Markets." Journal of Economics and Business 58, no. 1: 67-83.

Furfine. C. 1999. "The Microstructure of the Federal Funds Market." Financial Markets, Institutions, and Instruments 8, no. 5: 24-44.

Hamilton, J. 1996. "The Daily Market for Federal Funds." Journal of Political Economy 104, no. 1: 26-56.

———. 1997. "Measuring the Liquidity Effect." American Economic Review 87, no. 1: 80-97.

Hendry, S., and N. Kamhi. 2009. "Uncollateralized Overnight Loans Settled in LVTS." Applied Finance Economics 19, no. 1: 869-80.

Kuo, D., D. Skeie, and J. Vickery. 2012. "A Comparison of Libor to Other Measures of Bank Borrowing Costs." Unpublished paper, Federal Reserve Bank of New York.

Millard, S., and M. Polenghi. 2004. "The Relationship between the Overnight Interbank Unsecured Loan Market and the CHAPS Sterling System." Bank of England Quarterly Bulletin, Spring: 42-7.

Meta Brown, Andrew Haughwout, Donghoon Lee, and Wilbert van der Klaauw

Do We Know What We Owe? Consumer Debt as Reported by Borrowers and Lenders

- Economists' understanding of the finances of U.S. consumers is based heavily on survey data, and on the Survey of Consumer Finances (SCF) in particular. However, recent research calls into question survey respondents' willingness and ability to report their debts accurately.
- This study compares U.S. household debt as reported by borrowers to the SCF with debt reported by lenders to Equifax using the FRBNY Consumer Credit Panel (CCP). Debt levels, distributions, and trends are compared by loan type, both in aggregate form and for age, region, and household-size subsamples.
- Our most striking finding is that, overall and in most disaggregated debt categories, debt levels reported in the SCF and CCP are quite similar. Even bankruptcy measures correspond well.
- The exceptions lie in the unsecured debts. Under our most inclusive assumptions, SCF-implied aggregate credit card debt is 37 percent lower than that implied by the CCP, and SCF-implied aggregate student debt is 25 percent lower.

1. Introduction

The state of scientific knowledge regarding U.S. consumers' affluence and relationship to financial markets is based in many ways on survey data, and, in particular, on the Survey of Consumer Finances (SCF) published by the Board of Governors of the Federal Reserve System. For example, an extensive and influential line of research establishes the prevalence and importance of consumer liquidity constraints in the United States using SCF debt and related data.[1] Much of our understanding of U.S. wealth inequality over recent decades derives from analysis of SCF net worth figures.[2] Recent papers use SCF debt data to address a wide variety of topics relating to consumer balance sheets, such as the use of debt by low-income, unemployed, and bankrupt households.[3]

[1] This research includes Fissel and Jappelli (1990), Jappelli (1990), Cox and Jappelli (1993), Jappelli, Pischke, and Souleles (1998), Johnson and Li (2010), and others.

[2] See, for example, Wolff (1992), Davies and Shorrocks (1999), Keister (2000), Gokhale et al. (2001), Castañeda, Díaz-Giménez, and Ríos-Rull (2003), De Nardi (2004), and Cagetti and De Nardi (2008). Note that net worth calculations using the SCF rely on households' debt reports.

[3] See Cagetti and De Nardi (2006), Bucks and Pence (2008), Iacoviello (2008), Sullivan (2008), Scholz and Seshadri (2009), Han and Li (2011), and Kiyotaki, Michaelides, and Nokolov (2011), among others.

Meta Brown is a senior economist, Andrew Haughwout a senior vice president, Donghoon Lee a research officer, and Wilbert van der Klaauw a senior vice president at the Federal Reserve Bank of New York.

meta.brown@ny.frb.org; andrew.haughwout@ny.frb.org; donghoon.lee@ny.frb.org; wilbert.vanderklaauw@ny.frb.org

The authors thank Sonia Gilbukh, John Grigsby, Maricar Mabutas, and David Yun for excellent research assistance. They are grateful to Neil Bhutta, Brian Bucks, Joanne Hsu, Dean Karlan, Arthur Kennickell, Annamaria Lusardi, Kevin Moore, Karen Pence, Robert Pollak, John Sabelhaus, John Karl Scholz, Hui Shan, Basit Zafar, Eastern and Midwest Economic Association meeting participants, seminar participants at the Federal Reserve Bank of New York and the Board of Governors of the Federal Reserve System, and particularly the editor and referee for helpful comments. The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

However, other recent findings bring into question survey respondents' propensity and ability to report debts accurately. Lusardi and Tufano (2009) pose simple questions to U.S. survey respondents on the functioning of debt contracts. They report discouraging findings: "Debt literacy is low: only about one-third of the population seems to comprehend interest compounding or the workings of credit cards." Karlan and Zinman (2008) find that, among first-time borrowers from a leading South African "cash loan" firm, 50 percent fail to report their high-interest loans in a subsequent survey. Most pertinent to the question at hand is Zinman (2009), who compares the aggregate credit card debt levels implied by the SCF for 1989-2004 to aggregate credit card debt levels from the lender-reported Consumer Credit-G.19 data provided by the Federal Reserve Board of Governors. Zinman finds an undercounting of credit card debt in the SCF relative to the G.19 data of roughly 50 percent, and a divergence of the survey and the G.19 measures over the period.

The quality of survey-based debt data is of clear importance for researchers. An understanding of the debt behaviors on which households can and do report accurately, and those where they may not, is of use in evaluating the existing body of survey-based inference regarding household debt practices, and also in the design of future research. Identifying which questions are best answered using survey-based debt measures depends heavily on households' reporting tendencies, including both their level of accuracy and the informativeness of any common inaccuracies.[4]

Further, information on the accuracy of household debt reporting may be relevant to understanding the nature and effectiveness of household financial decision-making. Households with limited awareness of their debt positions may both misreport debts in surveys and make less informed financial choices as a result. The possibility of intentional misreporting implies that households' exact debt awareness cannot be inferred from evidence on the match between survey and administrative debt data. However, debt awareness is arguably a necessary precondition to closely matched survey and administrative debts.[5]

This article examines the correspondence between borrower- and lender-reported debts in recent years, at a relatively disaggregated level, with the objective of shedding light on both the quality and potential uses of survey-based debt data and the nature of household financial decision-making. We employ SCF data from 2001, 2004, 2007, and 2010 on household debts for the borrowers' picture of consumer obligations. For the lenders' side, we turn to the FRBNY Consumer Credit Panel/Equifax (CCP). The CCP is a panel of individual credit data drawn from Equifax, one of the three national credit reporting agencies. These data reported by lenders and servicers are classified as "administrative data" in much of the literature. The frequency and duration of the CCP data are sufficient to match the timing and, arguably, the representativeness of the SCF data for 2001, 2004, 2007, and 2010.

We compare consumer debt aggregates as well as moments—such as the mean and variance—of the household distributions of total debt, mortgage and home equity line of credit (HELOC) debt, vehicle loans, credit card debt, student loans, and other debts in the two sources. The latter comparisons are performed by year, household head age, household size, and region of the country. Differences between the samples are tested using standard methods; the large size of the administrative CCP data set permits a high degree of precision in such tests. We also compute household delinquency and bankruptcy rates in the two samples for the four years, noting that two of those years precede and two follow the implementation of a major bankruptcy law reform in 2005.

Our most striking finding is that, overall and in the majority of disaggregated debt categories and borrower characteristics, debt levels reported in the SCF and CCP are quite similar. Mortgages, HELOCs, and vehicle loans attain similar levels and follow similar age patterns in the SCF and CCP, for example. The growth of consumer debts over time and the accelerated growth rates of housing debt are similarly evident in the two samples. Overall, the weight of the evidence indicates a high level of accuracy in the correspondence between debts in the two sources.

A second central finding, echoing Zinman (2009), is that credit card debt appears to be up to 40 percent lower in the SCF than in the CCP. Two possible explanations for this raw difference are that (1) unlike the CCP households, SCF households may not have any member with a credit report, and (2) SCF households may not report business uses of personal credit cards that nevertheless appear on households' combined credit reports. We make generous allowances for these explanations, and find that a 37-percentage-point gap in aggregate credit card debt remains.

Further, the aggregate student debt balances implied by the SCF are roughly 25 percent lower than those implied by the CCP, which, in turn, are similar to aggregates drawn

[4] For example, Bucks and Pence (2008) show that informative patterns exist in the "don't know" responses to questions on mortgage characteristics.

[5] Here we assume that very similar debt findings are produced only in the case of accurate reporting on both sides. A remaining possibility is that borrowers and lenders make similar reporting errors. Given the very different nature of the reporting activities and objectives on the two sides, we judge this a low probability event and set aside the issue for the remainder of the article.

from other student debt sources. Hence we see that, by far, the largest differences between borrowers' debt reports in the SCF and lenders' debt reports in the CCP lie in the unsecured debts. We discuss sampling differences that may contribute to the measured student loan reporting gap. Unfortunately, information available in the two sources provides less opportunity to reconcile the difference in the case of student loans than in the case of credit card debt.

Nevertheless, bankruptcy appears to be reported at similar frequencies in the SCF and the CCP (though differences in available measures of bankruptcy in the two data sets impose qualifications on this claim). We find that, among one- and two-adult households, the CCP's two-year household bankruptcy rates in 2001, 2004, 2007, and 2010 fall comfortably between the SCF's one- and three-year bankruptcy rates, and that, if anything, one- and three-year bankruptcy rates in the SCF appear to be a bit high relative to CCP two-year rates. All measures reflect the expected drop in bankruptcy following the 2005 reform.

Finally, the match between SCF and CCP debt levels on certain individual debt measures is significantly closer for households with one adult than for households with two or more adults. In particular, survey measures appear to fall further below administrative measures for larger households, especially in the case of auto and credit card debt. This suggests that survey respondents are more able to report their own debt levels than those of other household members. This insight might help to inform both the design of surveys eliciting consumer balance sheet information and the research applications of such survey data. Further, it may tell us something about the nature of household members' interactions over financial matters.

2. Previous Studies

The SCF wealth data have been vetted in a number of studies produced both by the SCF survey staff and by others. The wealth data have been shown to be accurate, based on comparison with several administrative and survey sources.[6] The debt data of the SCF have received somewhat less attention.

Bucks and Pence (2008) ask whether SCF respondents accurately report the terms of their mortgages (and their house prices). In distribution-level comparisons between the 2001 SCF and lender-reported data, they find that "most homeowners appear to report their . . . mortgage terms reasonably accurately." Borrowers with adjustable-rate mortgages, however, may not be as well informed regarding potential interest rate changes.

Zinman (2009), as mentioned, compares credit card debt figures in the SCF to the Federal Reserve Board's G.19 statistical releases on consumer debt. Zinman was the first study (of which we are aware) to demonstrate in print the gap between SCF and administrative data credit card debt findings.[7] His lower bound estimate of the undercounting of credit card debt in the SCF is 50 percent. Further, he reports an increasing gap between credit card debt estimates from the SCF and the G.19 between 1989 and 2004, and suggests that such a trend might indicate individual heterogeneity in debt reporting that would undermine standard applications of survey-based debt data. In this study, we will generate further news on the trend in credit card debt reporting and evaluate the level of heterogeneity, by broad observable characteristics, in the extent of debt counting inaccuracies.

Johnson and Li (2009) vet the Consumer Expenditure Survey (CE) debt payments and limited debt balance data against the debt payment and balance measures in the SCF, taking the latter to be accurate. They find a match of within 5 percent on vehicle and credit card debt for the 1989-2004 waves of the SCF and comparable waves of the CE. However, they find that mortgage reports in the CE are substantially below those in the SCF, which, given the strong agreement between the SCF results and lender data for mortgages demonstrated by Bucks and Pence (2008), suggests an undercounting of mortgages in the CE.

Antoniewicz (2000) compares consumer assets and liabilities in the 1989-98 SCF fieldings to the Federal Reserve Board's flow of funds statistical release. She finds similar aggregate liabilities, consumer credit, and home mortgage debt in the two sources for 1989 and 1992, and a divergence in measured consumer debt in subsequent years. By 1995, the flow of funds estimate of total consumer credit is more than $200 billion higher than the SCF estimate. This divergence aligns with the time patterns observed by Zinman in the SCF and lender-reported debt data.

By and large, the methods used by these studies involve comparing one data source's estimates of aggregate debt or moments of debt distributions with those of another, either informally or using simple test statistics. Our approach is similar. But no other study of which we are aware has access to household-level matches of SCF data to other relevant debt

[6] See, for example, Avery, Elliehausen, and Kennickell (1988), Johnson and Moore (2005), Antoniewicz (2000), Bucks and Pence (2008), and Sierminska, Michaud, and Rohwedder (2008).

[7] Informal discussion indicates that SCF staff and users were aware of some part of this difference before the publication of Zinman (2009).

data for the purpose of comparison. To our knowledge, this article represents the most recent, most granular, and broadest validation of SCF debt data available. All of this derives from the richness of the administrative data available to us for comparison, as described below.

3. Data and Comparability

3.1 Survey of Consumer Finances

The Federal Reserve Board's Survey of Consumer Finances is a triennial survey of U.S. households, focusing primarily on household assets and liabilities. The survey was first fielded in 1983, and the present study covers the 2001, 2004, 2007, and 2010 surveys. The sample size of each survey was roughly constant through 2007, at about 4,500 households;[8] in 2010, it rose to 6,492 households. The survey includes both a geographically based representative sample of households and an over-sample of wealthy households. All results for the SCF reported here are weighted to be representative of the population of U.S. households, using the Kennickell-Woodburn consistent weights provided by the survey.[9] Further, we rely on the survey's multiple imputation methods where relevant data are missing.[10] Bucks et al. (2009) provide a detailed description of the 2001, 2004, and 2007 data. Bricker et al. (2012) detail the 2007 and 2010 data.

It may aid the reader's interpretation of observed similarities and differences between the survey and administrative debt data to include a sketch of the survey process that produces the consumer-side debt measures. The SCF measures are the product of a richly designed and meticulously managed interview of relevant household members by a well-trained interviewer. Interviews may occur in person or via phone. In 2007, an unweighted 55.3 percent of interviews were conducted in person and the balance over the phone. In 2010, 70.4 percent of the interviews were conducted in person and the balance by phone. Of the 6,492 interviews in 2010, 185 were conducted in Spanish.

Interviewers are instructed to encourage respondents to rely on documentation to obtain the details necessary to answer the highly specific battery of financial questions being fielded. They are also instructed to encourage the use of interview cards for keeping notes relevant to the sequence of questions. Specifically, interviewers are required to read each of the following statements to respondents at the start of the interview: "Feel free to consult any knowledgeable person or use any records and notes at any time during this interview. And please ask questions when anything is not clear;" and, "As we go through the interview, I will ask you to write a few things on this card to help keep us on track." Further, the SCF provides variables indicating whether respondents referenced documentation during the course of the interview, and if so, what type, along with how credible the interviewer found the responses as a whole.

3.2 FRBNY Consumer Credit Panel

The FRBNY Consumer Credit Panel is based on data supplied to the Federal Reserve Bank of New York by Equifax, one of the three national credit reporting agencies. The CCP comprises a 5 percent random sample of U.S. individuals with credit files and all of the household members of those 5 percent.[11] In all, the data set includes files on more than 15 percent of the population, or approximately 40 million individuals. We observe information from the credit reports for those individuals each quarter for the past sixteen years, and the data continue to be updated every quarter.

The sampling procedure generates a random sample of U.S. credit report holders, and ensures that the panel is dynamically updated in each quarter to reflect new entrants into credit markets. In addition, the data provider matches each primary individual's mailing address to all records in the data in order to capture information about other members of the primary individual's household. These individuals are also added to the sample. This procedure enables us to track individuals and households consistently over time, thus allowing us to study richer dynamics of consumer debt and related policy issues at both the individual and household levels.

The credit report data include residential location at the census block level and the individual's year of birth. The data also contain detailed information on each individual home-secured loan, including origination date and balance, current balance, scheduled payment, and current repayment status. In addition to information on debts secured by

[8] In 2001, the survey included 4,442 households, in 2004, 4,522 households, and in 2007, 4,422 households.

[9] We use the revised Kennickell-Woodburn consistent weights for the more recent data.

[10] Kennickell (1991, 1998) describes the imputation methods used in the SCF.

[11] See Avery et al. (2003) for a detailed discussion of the contents, sources, and quality of credit report data. See Lee and van der Klaauw (2010) for a discussion of contents and sampling design of the FRBNY Consumer Credit Panel.

residential real estate, the data set includes information on individuals' and households' other loans, such as credit cards and auto loans. The data include the following:

- total number of each type of account (for example, the total number of bank-issued credit cards),
- credit limit on each type of account (for example, the combined credit limit on all credit cards),[12] and
- total balance on each type of account in each status (for example, the total auto loan balance that is current, thirty days delinquent, etc.).

More general information on the credit report includes the following:

- indicators for whether the individual has a foreclosure or bankruptcy, both within twenty-four months and ever, on the report,
- the number of collection accounts and the amount of collection, and
- Equifax's credit score, analogous to the well-known FICO score.

In the present study, we use the primary sample members and associated household members to establish a representative sample of all U.S. households in which at least one adult has a credit record. Owing to computational demands, the findings reported in this article are based on a random subsample of CCP households: we retain a randomly determined 10 percent of CCP households.[13] Thus, for example, the estimation sample for 2007 contains 1,090,880 households.

All figures reported below from the two data sources are denominated in 2010 U.S. dollars.

3.3 Comparability

An immediate difficulty arises from the fact that, while the (weighted) SCF is representative of all U.S. households, the CCP is a representative sample of U.S. households in which at least one adult has a credit record. According to Jacob and Schneider (2006), 10 percent of U.S. adults had no credit record in 2005.

We observe that 75 percent of SCF households claim debts that would generally appear on a credit report, and 84 percent of CCP households' collective reports include positive debt levels. Begin by assuming that these two groups represent the same population, namely U.S. households with any conventional debts. Further note that the CCP data represent two populations, those with conventional debts and credit reports (84 percent) and those without conventional debts but with credit reports (16 percent). The SCF represents the former population through the 75 percent of SCF households with conventional debts and credit reports. Define x as the percentage of SCF households without conventional debts but with credit reports. The SCF also represents those with neither conventional debts nor credit reports, who constitute $25 - x$ percent of the sample.

From the assumption that both the CCP and the SCF contain representative shares of both households with conventional debts and credit reports and households without conventional debts but with credit reports, we infer that the ratio of the sizes of the conventional debt and credit report and the no conventional debt and credit report populations must be the same in the two samples. This inference allows us to solve the relationship $\frac{16}{84} = \frac{x}{75}$ for x, which is the share of SCF households with credit reports but no standard debts. If 84 percent of CCP households have reports and debt and 16 percent have reports and no debt, and 75 percent of SCF households have reports and debt, then it must be the case that 14.3 percent of SCF households have reports and no debt. The residual, 10.6 percent of SCF households, must then have no credit reports.[14] Note that this figure is near the rate calculated by Jacob and Schneider.[15]

One difficulty remains: Whether SCF respondents report all of their debt, and hence all of their credit-report-generating debt, is precisely the question at hand. To establish methods based on an inference that assumes SCF reporting to be accurate threatens the credibility of our findings. Let us consider the consequences of assuming reporting accuracy in the above calculations in the event that SCF households in fact underreport their debt. Assuming some SCF households

[12] This field is known as the "high credit" amount in the credit report data. It refers to either the credit limit (for credit cards, HELOCs, and other revolving debt) or the highest balance (for mortgages, vehicle loans, and other installment debt). Credit limits on some revolving accounts are unreported, in which case the high credit variable reflects the historical high credit level for the account. Avery et al. (2003) and Hunt (2002) point out that the reporting of credit limits in credit reports has improved considerably in recent years.

[13] Though sampling is done at the individual level, which would generate overrepresentation of larger households, we reweight the sample based on probability of inclusion so as to be representative at the household level.

[14] Figures are rounded for ease of discussion, and hence contain some rounding error.

[15] Assuming households do not sort perfectly on the presence or absence of credit reports, we would expect the household-level rate of missing credit reports to be smaller. For the 2007 waves of the two data sets, which are considerably closer to Jacob and Schneider's period of observation, we find a missing report rate of 8.33 percent, a figure in line with our expectations under imperfect sorting.

that have credit-report-generating debt report having none, 75 percent is an underestimate of the proportion of the sample with credit-report-generating debt. Suppose that the rate of underreporting in percentage terms is $r > 0$. Then $75 + r$ percent actually have credit-report-generating debt. We seek the percentage of SCF households with no credit-report-generating debt but with credit reports, x, that solves the expression $\frac{16}{84} = \frac{x}{75 + r}$. At r = 0, x = 14.3. From there, x increases with r. Hence, the share of SCF households with no conventional debt but with credit reports increases from 14.3 percent when SCF respondents underreport debt, and the residual share with no conventional debt and no credit reports has an upper bound of 10.6 percent.

Alternatively, one could attempt to infer the proportion of SCF households with no debt and no credit reports based on available SCF measures. For example, if we assume that only the 2010 wave SCF households that have no conventional debts, do not include property owners, and have no household member who reports holding a credit card, including store cards, have no credit reports, then we arrive at a no-credit-report rate below 10.6 percent. Since the validity criteria for this type of approach are unclear, we again focus on the 10.6 percent figure as an upper bound.

In the analysis that follows, we estimate aggregate debt levels, as well as debt holding rates and conditional median and mean balances, for total debt and various debt categories using the SCF and CCP data.[16] The distinction between SCF non-debtors with and without credit reports is clearly irrelevant to our comparison of aggregate debt levels and of conditional mean and median debt levels; each category of non-debtors contributes zero to the aggregate and is omitted from the conditional calculations.

However, the proportion of SCF non-debtor households not represented in the CCP is crucial in the comparison of the rates at which households hold various types of debt. In what follows, we compare SCF and CCP debt rates in two ways: with no adjustment for households without credit reports, and after removing 10.6 percentage points' worth of non-debtor households from the SCF calculations. Should underreporting of debt render the 10.6 percent figure an overestimate of the true rate at which SCF households have no credit reports, this method would cause the rate at which SCF households hold debt to be inflated relative to the rate at which CCP households hold debt.

In the interest of establishing comparable dates of observation, we select CCP data for the third quarter of 2001, 2004, 2007, and 2010. The fielding dates of the SCF are roughly May to December of each survey year; our CCP data are drawn at the midpoint of this range of months, which we hope maximizes comparability. An alternative approach would be to average CCP figures for the final three quarters of each relevant year. The drawback to this method is that it would require constructing a short panel on each household, and the composition of those households could change from quarter to quarter. To avoid this issue, we have adopted a single-quarter approach, though we believe that each method has appealing features.

An additional comparability issue is who, exactly, constitutes the household. While the CCP includes all adults with credit reports living at the primary sample member's address (up to an apartment number), most SCF debt questions concern the debt holdings of the "primary economic unit" (PEU) of the household. A PEU consists of the primary earner, the earner's partner, and any agents dependent on this unit. Children or elderly parents dependent on a primary earning couple, for example, would be PEU members. However, households also at times contain non-PEU members, such as roommates and boarders.

The debts of these non-PEU members would appear in the CCP but not the SCF. We have limited opportunity to infer non-PEU members' debts by category and add them into the household debt calculations, given the data collected on non-PEU members. However, it is possible to determine the overall level of debt held by non-PEU members, and hence to infer the likelihood that such debt changes could influence our conclusions.[17] We return to this issue later in the article.

Other comparability issues related to specific debt categories and associated survey questions or credit reporting are addressed as they arise in the course of the analysis below. In general, we endeavor to make all appropriate adjustments where possible to ensure that the household debts in question are comparable across the two data sources. Where this is not possible, we attempt to understand the likely direction of the resulting bias in our comparison, and its likely effect on our conclusions.

[16] Note that by "conditional mean," we mean the average debt balance among those who hold positive balances in the debt category under consideration. "Conditional median," similarly, is the median debt balance among those who hold positive balances in the debt category under consideration.

[17] Note that other observable characteristics of non-PEU members tend to be associated with low debt levels.

Table 1

A Comparison of SCF and CCP Aggregate Balances by Debt Category

Billions of 2010 U.S. Dollars

		Aggregate Balance	
	Year	SCF	CCP
Total debt	2004	10,192	10,158
	2007	11,800	12,740
	2010	11,512	11,844
Home-secured debt	2004	8,522	7,631
	2007	10,012	10,034
	2010	9,648	9,282
Auto debt	2004	747	864
	2007	785	859
	2010	596	710
Student debt	2004	291	380
	2007	397	555
	2010	578	778
Credit card debt	2004	424	812
	2007	519	858
	2010	440	731
Other debt	2004	448	472
	2007	360	434
	2010	449	343

Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

4. Findings

4.1 The Match between SCF- and CCP-Derived Estimates of Aggregate Debt and Household-Level Debt Distributions Is Close

Though the data collection methods and respondent incentives in the SCF and CCP differ greatly, the primary insight that arises from their comparison is that the two sources generate strikingly similar debt patterns.

Aggregate Debt Estimates

The overall debt figure for 2010 is quite similar in the two sources (Table 1), at $11.51 trillion in the SCF and $11.84 trillion in the CCP. Home-secured debt estimates are nearly as close, at $9.65 trillion for the SCF and $9.28 trillion for the CCP, indicating that the accuracy in mortgage reporting demonstrated by Bucks and Pence (2008) continues to hold in 2010, and holds for comparisons using multiple lender sources.[18]

Vehicle installment loan estimates are $596 billion for the SCF and $710 billion for the CCP. The CCP, as with credit reports in general, includes leased vehicles in its vehicle loan figures, while SCF respondents are likely to report leases separately from vehicle loan debt. According to Experian, 12.1 percent of vehicles that were financed in the first quarter of 2008 were leased. We attempt to remedy this discrepancy by adding SCF vehicle lease balances to the SCF vehicle debt calculation. Though the SCF does not supply public data on the make, model, and year of leased vehicles, its public data do include the value of the leased vehicle based on an industry guidebook estimate. We take this value as an approximation of the remaining balance of the (implicit) loan that would be reflected in the lessee's credit report. To the extent that the industry guidebook value is an overestimate of the principal remaining after the conclusion of the lease payments, this approach will exaggerate the vehicle loan balance we infer from the SCF. We find that, even with a generous allowance for lease balances, the aggregate vehicle debt implied by the SCF is approximately 16 percent lower than that in the CCP. Hence, the vehicle debt balances implied by the borrower- and lender-sourced data are fairly similar, but not perfectly matched.

Credit card balances are estimated at $440 billion in the SCF and $731 billion in the CCP. We analyze what proportion of the gap may be attributable to simple measurement and reporting differences, and what proportion appears to be the result of true underreporting, in Section 4.2.

Household-level student debt balances that rely on current measurement practices are unavailable in the CCP for periods preceding the third quarter of 2011. However, we do have individual-level student debt measurements based on current practices for the third quarters of 2010, 2007, and 2004. We compare the aggregate student debt implied by household-level SCF data to the aggregate student debt implied by individual-level CCP data in Table 1. Assuming representativeness in each case, these measures should be comparable. We find that the debt balances reported by SCF households imply an aggregate student loan balance in 2010 of $578 billion. Individual credit reports in the CCP, however, imply an aggregate student debt balance of $778 billion. Once again, we infer a higher aggregate balance

[18] The estimates of aggregate home-secured debt for 2007 are nearly identical, at $10.0 trillion in each source.

using lender-side data than we do using borrower-side data; in this case, borrowers appear to report 25.7 percent less debt than lenders do.

Other available measures of aggregate student debt for 2010 are limited but tend to be similar to the CCP figure. One estimate published in a *Wall Street Journal* economics blog in the summer of 2010 put aggregate student debt at roughly $830 billion.[19] The Consumer Financial Protection Bureau estimates that aggregate student debt crossed the trillion-dollar threshold in late 2011 (Chopra 2012). The portfolio overseen by the Office of Federal Student Aid (FSA) at the start of 2011 was $722 billion (U.S. Department of Education 2011).

Household Debt Distributions by Debt Category

Table 2 demonstrates the correspondence between SCF and CCP debt distributions across households, both overall and for the five major debt categories. Panels A and B of Table 2 are identical, with the exception that the debt frequencies in panel A are raw frequencies that use the full sample and standard weights in each case, while those in panel B are adjusted to remove SCF households with no credit reports, in the interest of comparability. The adjustment removes the 10.6 percent of SCF households we approximate to be non-debtors without credit reports. We begin by summarizing the similarities and differences we observe between household-level lender- and borrower-reported debt. Next, we consider some potential sources of the differences between lender and borrower reports that we observe for credit card and student debt.

Overall, the figures in Table 2 reflect similar rates of debt holding and similar mean debt levels among households with positive debt, both in total and across debt categories. Adjusted HELOC debt rates are 8.1 and 9.2 percent in the SCF and CCP, respectively. Adjusted vehicle installment loan rates are 36.6 and 38.3 percent, respectively. The overall conditional mean household debt level is $130,700 in the SCF and $114,900 in the CCP. The conditional median and mean HELOC level comparisons are $26,400 in the SCF versus $34,700 in the CCP, and $54,500 in the SCF versus $62,700 in the CCP. For vehicle installment loans, conditional median and mean balance comparisons are $11,000 for the SCF versus $12,400 for the CCP and $15,500 for the SCF versus $16,200 for the CCP. Mortgage and home equity installment loan balances have a conditional median of $110,000 in the SCF and $130,100 in the CCP. The difference in the means, however, is more substantial and presumably reflects a difference in the reporting of vacation and investment property between the two data sets.[20]

Some modest differences are worth noting. The prevalence of home-secured debt is 52.6 percent in the SCF and 42.6 percent in the CCP after adjustment for SCF households without credit reports (the raw comparison is 47.0 versus 42.6 percent).[21] It is not clear why we would observe a somewhat higher rate of home-secured debt in the SCF than in the CCP. The conditional median total debt level in the SCF is $71,900, while the conditional median for the CCP is $42,500. The means are closer together. It is also not clear why, relative to the CCP, the SCF would show a higher incidence at moderate debt levels and lower incidence at high debt levels.

Turning to credit card debt, we note that such debt is generally observed at the end of the billing cycle for each report-holder in the CCP. Hence, our CCP measure contains both carried balances and some share of new charges that will be repaid during the billing cycle, before any interest accrues. We refer to the latter as the convenience uses of credit cards.

The SCF asks respondents for two separate credit card debt amounts. First, regarding standard credit card accounts, respondents are asked, "On your last bill, how much were the new charges made to (this account/these accounts)?" If all new charges are repaid during the billing cycle, then this amount represents the convenience use of the card. If some are carried into future billing cycles, however, this figure represents a combination of carried and convenience balances. Next, respondents are asked, "After the last payment(s) (was/were) made, what was the total balance still owed on (this account/all these accounts)?" We expect this measure to reflect the borrower's recollection of the carried balance on each card. The interviewer advises the respondent to exclude any business use of personal credit cards.

We generate an upper bound measure of the amount of credit card debt observed in the SCF by adding together the convenience use and carried balance figures, as measured by the above two questions, so that our measure of SCF credit card debt consists of all carried balances currently held by

[19] Mark Kantrowitz, cited in "Student-Loan Debt Surpasses Credit Cards," Wall Street Journal's *Real Time Economics* blog, August 9, 2010.

[20] While credit reports cannot typically distinguish between primary residence and other types of properties, and hence the CCP must pool all residential mortgages, the SCF asks separate questions about loans collateralized by the primary residence and by other residential real estate. The SCF questions on loans collateralized by other residential real estate do not allow us to distinguish among mortgages, home equity loans, and HELOCs. As a result, our SCF estimates for the residential real estate debt subcategories do not contain vacation and investment property debt. Of 2010 SCF households, 5.3 percent report some residential debt not secured by the primary residence.

[21] By the "prevalence" of a type of debt, we mean the share of the population holding positive balances in that type of debt.

Table 2

SCF and CCP Household Debt by Account Type, 2010

	Percent of Households		Median (U.S. Dollars)		Mean (U.S. Dollars)	
	SCF	CCP	SCF	CCP	SCF	CCP
Panel A: Raw Frequencies						
Overall debt	75.1	84.0	71,900	42,500	130,700	114,900
Overall home-secured debt	47.0	42.6	109,600	123,400	154,300	181,400
Mortgages or home equity loans	45.2	40.3	110,000	130,100	151,800	186,700
Home equity lines of credit	7.2	9.2	26,400	34,700	54,500	62,700
Vehicle installment loans	32.7	38.3	11,000	12,400	15,500	16,200
Education installment loans	19.2	—	13,000	—	5,500	7,500
Credit card balances	66.2	73.6	2,000	3,500	5,700	9,600
Panel B: Corrected Prevalence						
Overall debt	84.0	84.0	71,900	42,500	130,700	114,900
Overall home-secured debt	52.6	42.6	109,600	123,400	154,300	181,400
Mortgages or home equity loans	50.6	40.3	110,000	130,100	151,800	186,700
Home equity lines of credit	8.1	9.2	26,400	34,700	54,500	62,700
Vehicle installment loans	36.6	38.3	11,000	12,400	15,500	16,200
Education installment loans	21.5	—	13,000	—	5,500	7,500
Credit card balances	74.0	73.6	2,000	3,500	5,700	9,600

Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

Note: The per capita student loan balance for the CCP is calculated by dividing the aggregate student balance measured for the third quarter of 2010 by the number of households represented by the CCP in that quarter. It is an unconditional figure, and hence is compared with the unconditional per household student debt in the SCF.

borrowers plus all new charges from the last completed billing cycle on each card held by the borrower. The total card debt inferred in this manner may, therefore, contain some double-counting.[22,23] However, to the extent that respondents interpret the phrase "new charges" to indicate spending on the card but not finance charges, the measure obtained may understate the total balances one would expect lenders to report. The above approach is used in generating the Table 1 aggregate balances and the Table 2 distributional characteristics.

[22] The authors thank Joanne Hsu and Kevin Moore for suggesting this approach.

[23] We infer that this approach is generous from other SCF data. The 2007 SCF asks respondents with credit cards whether they "always or almost always," "sometimes," or "hardly ever" pay off the full billing cycle balance on their credit cards. Among households with credit cards, the answers were 68 percent, 15 percent, and 17 percent, respectively. These rates are at odds with the 46.1 percent of SCF households that report positive credit card balances following their most recent payments.

The credit card debt rates, conditional median, and conditional mean comparisons suggest greater agreement between the borrower- and lender-side measures than one might infer from the aggregates. Table 2, panel B, indicates that, once we correct for SCF households without credit reports, 74.0 percent of SCF households and 73.6 percent of CCP households hold some credit card debt. The conditional medians and means reflect some difference in balances, however, with \$2,000 (SCF) versus \$3,500 (CCP) in credit card debt at the median, and \$5,700 versus \$9,600 at the mean. So it appears that less credit card debt is reported in the SCF than in the CCP, and that the major source of the difference in reporting (and presumably the difference in the aggregates evident in Table 1) is the low balances reported by SCF credit card users (or high balances reported by CCP lenders), as opposed to a failure among SCF credit card users to report any credit card use at all.

Turning to student debt, we note that our ability to compare student debt distributions in the two sources suffers from the above-mentioned restrictions in the availability of household-level student debt measures in the CCP. We have chosen to generate the aggregate U.S. student debt balances implied by the SCF household-level and CCP individual-level observations for 2010. We infer from these measures, and the numbers of households represented by the two data sets, the household-level mean student loan balance (Table 2). Unlike other figures in Table 2, these reported means are not conditioned on holding positive debt in the category, because we are unable to determine the proportion of CCP households with positive student debt balances in the third quarter of 2010.

The SCF respondents are asked, "Do you (and your family living here) owe any money or have any loans for educational expenses?" The interviewer instructs the respondents to exclude any credit card or other loans previously recorded in the survey.[24] The respondents are then asked to supply estimates of the balance on each of their first six reported education loans, and the total balance on the seventh and any additional loans: "How much (in total) is still owed on (this loan/all the remaining loans)?" The sum of these values constitutes our student debt measure for the household.

The SCF question format may generate smaller debt responses than those we observe from lenders because while it asks about the debt of all primary economic unit members, the question is put *only* to the respondent. We explore the possibility of underreporting in large families later in the article.

As suggested by the 25.7-percentage-point gap in aggregate student debt between the SCF and CCP, the mean household-level student debt we infer for the SCF in 2010 is markedly lower than the debt we infer for the CCP in 2010. When households without credit reports are removed from the calculation, SCF households claim $5,500 in student debt balances on average, while CCP households show an average balance of $7,500.

Though the discussion in this section emphasizes prevalence, medians, and means, other moments of the SCF and CCP debt distributions may be of interest. Appendix Chart A1 depicts the mortgage and credit card balance densities in the SCF and CCP, after adjusting the SCF data for the 10.6 percent of households whose members have no credit reports. The results are fairly similar in the two data sources, except for a higher reported mortgage prevalence in the SCF than in the CCP, and a lower credit card debt in the SCF than in the CCP, as described above.

We conclude that the prevalence of consumer use of each major debt category is similar in the two sources. The pattern of conditional median and mean balances is also similar. However, reported household balances tend to be lower in the borrower-sourced data than in the lender-sourced data. The two categories in which we observe substantial mean balance gaps between the SCF and the CCP are credit card and student loan debt. Even under our most inclusive assumptions regarding SCF debt levels, unconditional mean credit card balances are 40 percent lower in the SCF than in the CCP, and unconditional mean student loan balances are 27 percent lower in the SCF than in the CCP.

Patterns by Age, Region, and Year

Credit reports contain limited demographic information, and hence we are unable to use a more detailed household-level matching estimator to examine the difference between debt reported in the SCF and in the CCP. But the reports do contain the date of observation, the borrower's location, and in many instances, the borrower's age, and we exploit these data to produce a more granular comparison of the debt distributions in the two samples.

First, we consider age. In the SCF, we are able to identify a household head (defined to be the single adult in PEUs with one adult, the male partner in male-female couple PEUs, and the older member of the pair in same-sex PEUs). The SCF data contain ages of household members, and so we have a self-reported age of the household head available. In the CCP, as in credit reports, we cannot identify a household head. But we do have ages of household members. In response, we experiment with a variety of rules for predicting household head and evaluate their effectiveness in the SCF data. The most effective simple rule we developed was to assign the household head age as the median age among adult household members (implying the age of the one adult household member in single-headed households, the average of the two ages in two-adult households, the middle of three ages in three-adult households, and so on). This approach generates the age of household head distribution reported for the third-quarter 2007 CCP in Table 3.[25] Table 3 then compares this household head age distribution with the actual age of household head distributions in both the

[24] The debt questions preceding this one cover financial institutions, credit cards, the principal residence, other housing lines of credit, investment and vacation properties, businesses, and vehicles; hence, the exclusion of previously reported loans is quite comprehensive. Of course, these other categories typically do not include explicit school loans.

[25] As elsewhere in this article, we use household weights in the comparison of CCP household head ages to those in the SCF and Census.

TABLE 3

Breakdown by Age of Household Head in the SCF, CCP, and Census

Age Group	SCF (Percent)	CCP[a] (Percent)	Census (Percent)
< 35	21.7	20.64	20.70
35-44	19.6	24.21	20.27
45-54	20.8	21.84	21.69
55-64	16.8	15.34	16.84
65-74	10.5	8.89	20.50[b]
75+	10.6	7.56	

Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

[a] Age of household head is inferred from the median age household member.

[b] The census age category is 65+.

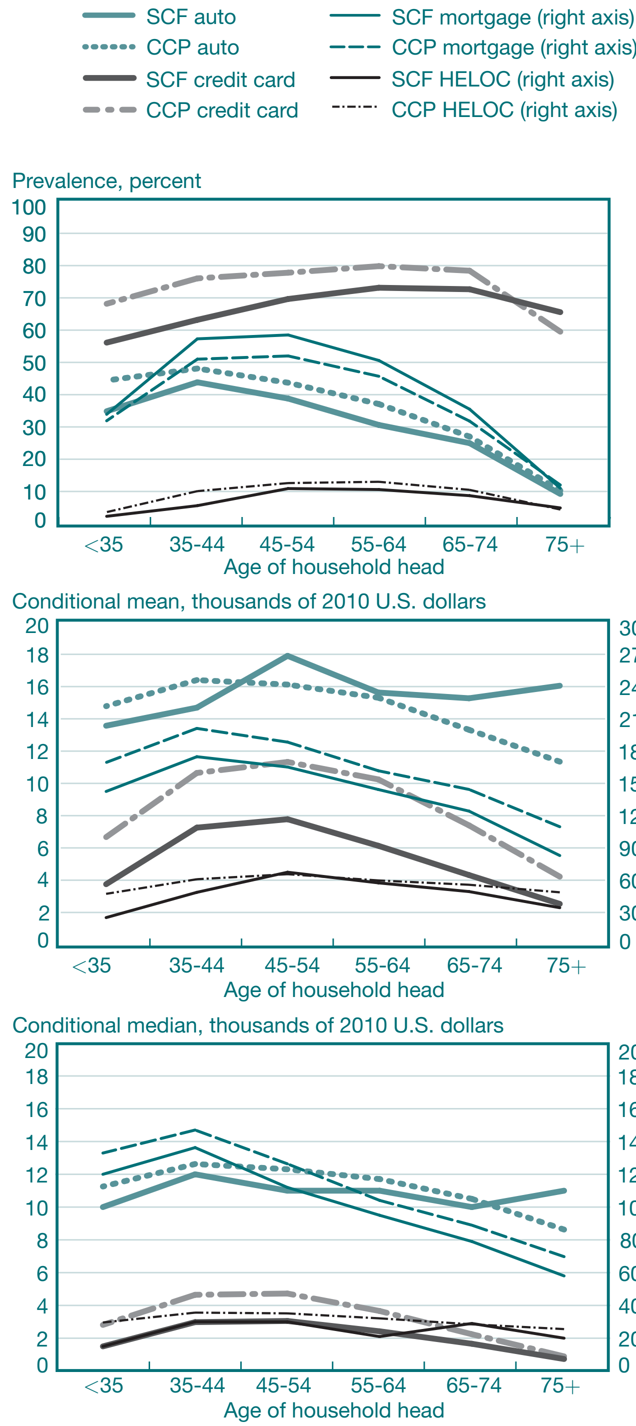
weighted 2007 SCF and the U.S. Census projections for 2007. The distributions are quite similar, with perhaps a slight underrepresentation of older households and a slight overrepresentation of middle-aged households in the CCP. We use our household head prediction method to predict household head ages in both the CCP and the SCF, and we compare features of the distribution of household debt across six age categories (Chart 1).

Chart 1 depicts debt prevalence, conditional mean, and conditional median by debt type and age, comparing estimates from the SCF and CCP. Households are grouped by age of head into six bins (under 35, 35-44, 45-54, 55-64, 65-74, and 75 and over), shown along the horizontal axis. The vertical axis of the first panel of the chart represents the percentage of the sample with any debt in a given category. We examine four debt categories in this and the following charts: mortgage, HELOC, vehicle loan, and credit card debt. Debt categories are distinguished by the color and style of the lines. The age trajectories for each debt category are traced by a solid line representing SCF estimates and a dashed line representing CCP estimates. A perfect match between the SCF and CCP across all age groups for a given debt category would be represented by coincident solid and dashed curves of the same color.

In Chart 1, panel A, we see that the mortgage, HELOC, vehicle loan, and credit card debt prevalences follow similar age patterns in the two data sets. Younger households appear to report slightly lower rates of credit card debt and vehicle loans in the SCF than in the CCP, but, overall, each pair of lines remains quite close over the full age distribution. The

CHART 1

SCF and CCP Consumer Debt by Age, 2010



Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

Note: HELOC is home equity line of credit.

Table 4
A Comparison of Average Student Debt Balances in the SCF and CCP

	Unconditional Mean Balance per Household (2010 U.S. Dollars)		
Year	SCF	CCP	SCF as Percentage of CCP
2004	2,592	3,419	0.76
2007	3,420	4,850	0.71
2010	4,915	7,496	0.66

Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

prevalence for credit card debt shows the widest discrepancy. The differences in reported credit card debt rates range from -6 to 13 percentage points for the various age groups, and conventional tests of means reject, with high degrees of confidence, the null hypothesis that credit card debt prevalence is the same in the two sources for most age groups. However, the economic significance of the largest observed differences in debt rates is comparatively modest, and the similarity in the levels and shapes of each pair of age profiles is striking.

For the conditional mean and median debt levels in the two samples, several of the line pairs are nearly coincident (Chart 1, panels B and C). The SCF mortgage and HELOC amounts lie below the CCP amounts for most age groups, but these differences are of a magnitude that may largely be explained by the exclusion of vacation and investment properties from the SCF measures.[26] The age patterns of conditional debt balances are remarkably similar in the two data sets. The single exception to this pattern is credit card debt, whose levels again differ meaningfully in the two sources.[27]

When comparing data by year (Chart 2), we find that the levels and time trends in the prevalence and sizes of the various debt categories match well in the two data sets. Some minor variations in mortgage and HELOC patterns arise from the data sets' differing treatment of vacation and investment property: mortgage prevalence is a bit higher in the CCP, and recent increases in the dollar amounts of mortgages and HELOCs in the CCP are muted in the SCF. However, we find that the majority of the difference in each of these cases does not appear in the case of total home-secured debt, where we are able to account for vacation and investment properties more comparably.[28] Vehicle debt was significantly more prevalent in the SCF in 2001, and then significantly more prevalent in the CCP in 2010. Credit card amounts in the SCF remain well below those in the CCP. By and large, however, the time trends in the two data sets are quite similar.

We can infer mean household student debt from CCP aggregates and the number of households represented by the CCP in each of the years 2004, 2007, and 2010, and therefore we are able to compare the time paths of unconditional mean student debt in the CCP and SCF. Since the patterns in the unconditional means would be obscured by the scale of Chart 2, panels A-C, we present the unconditional student debt means on their own in panel D. While the proportional gap between SCF and CCP aggregate student debt estimates in Table 1 is reasonably stable over time, the unconditional mean student debt we estimate at the household level in the two data sets diverges over this period. In 2004, the SCF student debt mean estimate is 76 percent of the CCP value. In 2007, it is 71 percent, and by 2010, the SCF estimate is only 66 percent of the CCP estimate (Table 4).

The widening difference in student debt estimates has various potential explanations. The difference in the populations represented by the two sources as a result of the presence or absence of credit reports should play little role, because most student debts generate reports. There is the possibility that not all student loan servicers report all student debts to Equifax, but this should reduce the CCP means and hence the measured gap with the SCF. The omission of institutional populations from the SCF sample may lead to the omission of debt held by students living away from home. The SCF's use of household-level financial reporting by a single respondent may lead to undercounting of student debts held by grown children or other household members that are not fully known to the respondent. And, of course, respondents may not be fully aware of their own current debt balances. A combination of the latter three factors could produce the type of balance gaps we observe in Chart 2, panel D.

The SCF patterns by region (Chart 3) are derived from Bricker et al. (2012), since Census region is not available in the public data set. As a result, we are unable to adjust

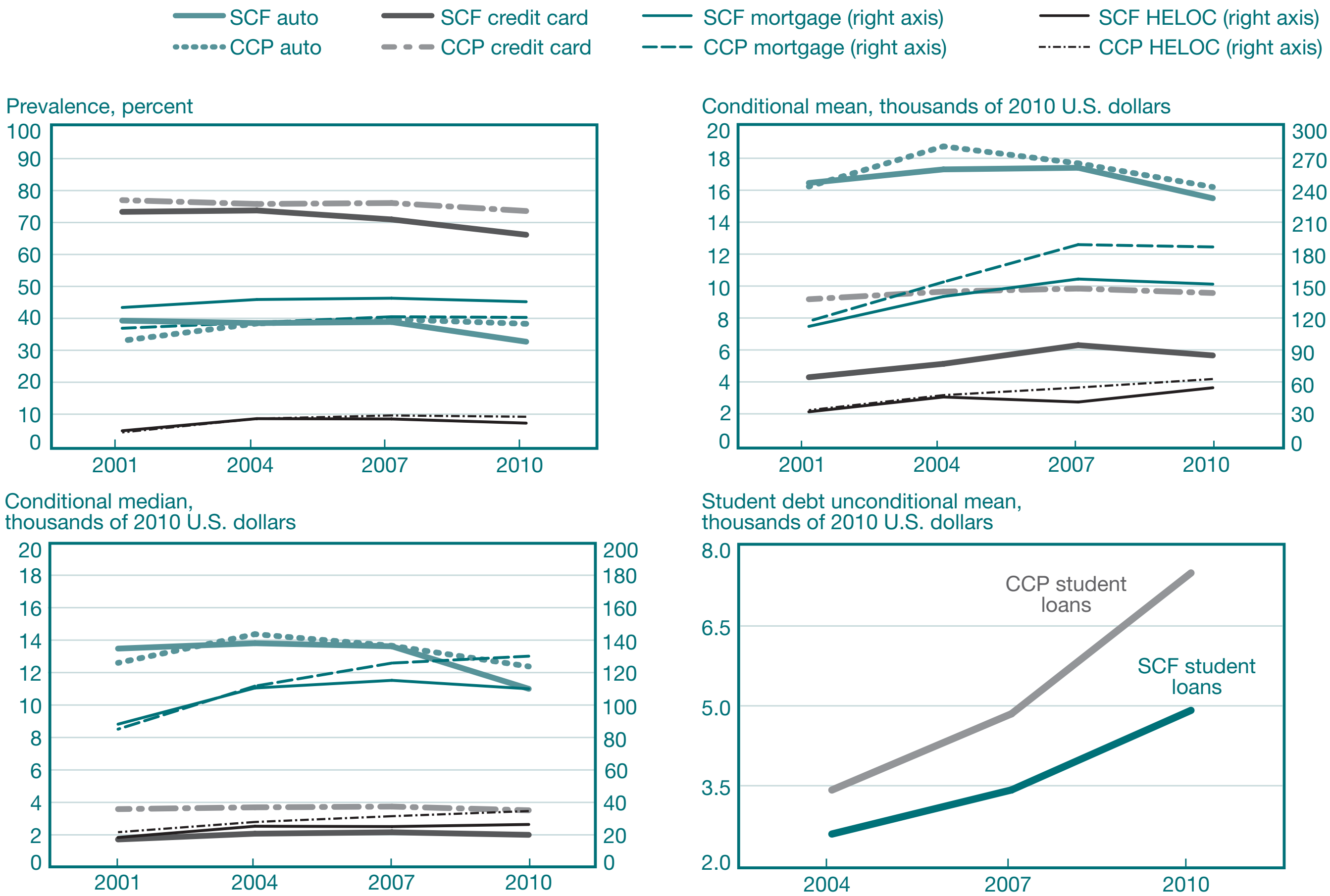
[26] However, the mortgage differences are approximately constant across the age groups, a profile somewhat at odds with what we expect for vacation and investment properties.

[27] Appendix Chart A2 demonstrates very similar age profiles of debt for 2007, indicating a high degree of stability of the age dependence of debt, and of the SCF-CCP similarity in these patterns, over the three years.

[28] This pattern is somewhat similar to the one evident in Table 2, in which the difference between overall home-secured debt balances is smaller than the difference between mortgage balances in the SCF and CCP, owing to the similar treatment of vacation and investment properties.

Chart 2

SCF and CCP Consumer Debt by Year



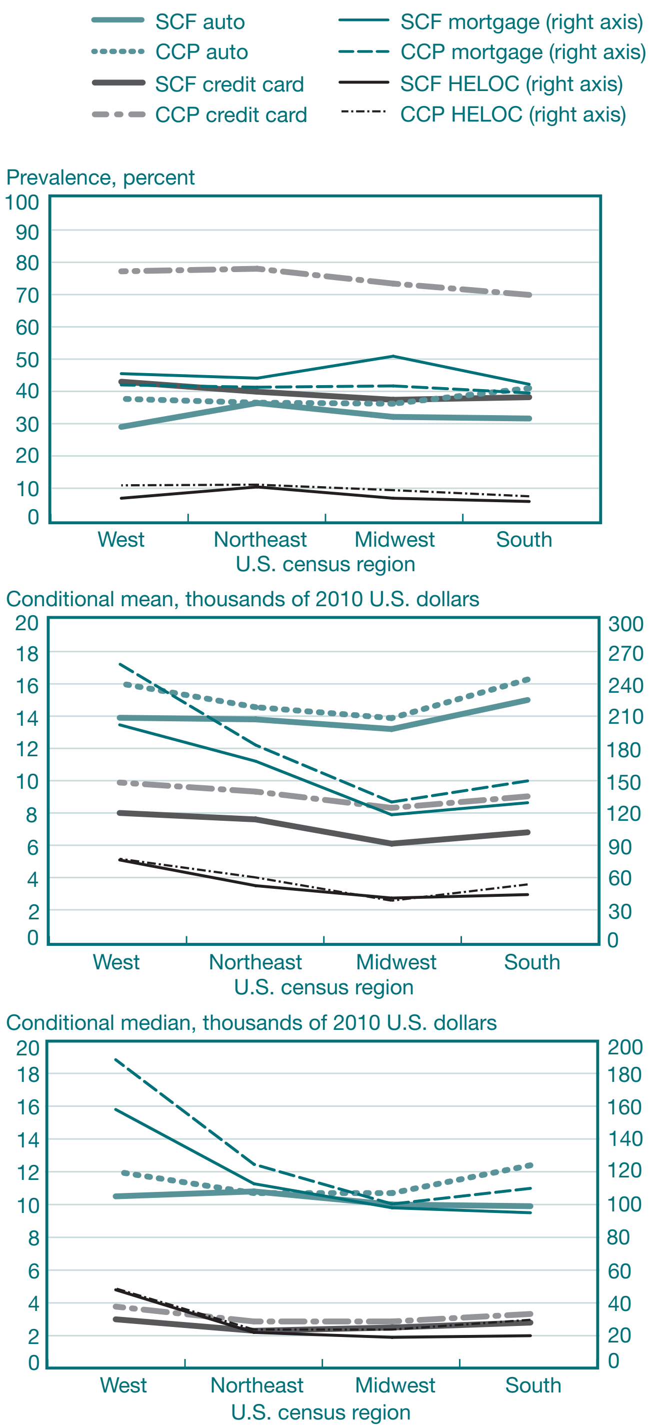
Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

Note: HELOC is home equity line of credit.

Bricker et al.'s SCF credit card debt use and balances to add new charges on the last bill to the balance after the last card payment—a constraint that reduces the credit card debt prevalence and balances substantially relative to Charts 1 and 2. Further, we are unable to add lease balances to the vehicle debt measures in Bricker et al., leading to slightly lower vehicle debt prevalence and balances. Nonetheless, the regional variation in the two samples is comparable for most debt categories. Again, exceptions in home-secured debt categories arise from, and are largely reconciled by, the treatment of vacation and investment property, and, as always, credit card debt is greater in the CCP.

The removal of new credit card charges required by the limited availability of the SCF regional data allows us to demonstrate the effect of new charges on our credit card debt comparisons. Without new charges, the credit card debt prevalence shown in Chart 3 is much lower for the SCF than for the CCP. Differences by region vary from 32 to 38 percentage points. However, balances conditional on positive debt are now approximately coincident for the CCP and SCF. Hence, the inferred source of the measured gap in credit card debt between the borrower- and lender-side data depends heavily on one's treatment of new charges. If one includes all SCF new charges in credit card debt, the difference is attributed almost entirely to reported balances. However, if

Chart 3
SCF and CCP Consumer Debt by Region, 2010



Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

Note: HELOC is home equity line of credit.

one omits new charges from SCF balances, then the difference is attributed almost entirely to the rate at which borrowers report any credit card use.

Our tests of the pairwise difference in means (Table 2, panel B, and Charts 1-3) generally reinforce the observations above.[29] Differences in the mean balances for credit card and student loan debt are large and statistically significant. Given sample sizes, most other prevalence and mean comparisons in Table 2, panel B, and Charts 1-3 meet standard significance criteria. In other words, credit card and student debt balances aside, the differences are both small (as the point estimates indicate) and precisely measured. Examples of the rare cases in which the difference in means is insignificant include the prevalence of credit card debt and vehicle loans (Table 2, panel B) and the prevalence of HELOC debt in 2001 and 2004 (Chart 2, panel A).

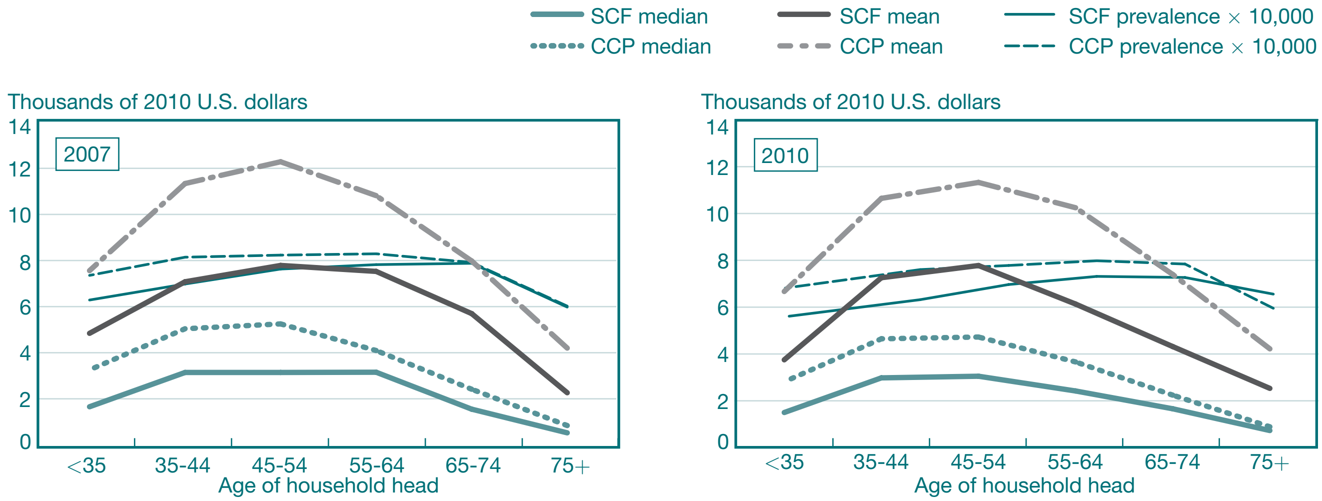
4.2 Borrower-Reported Credit Card Debt in the SCF Is Substantially Lower Than Lender-Reported Credit Card Debt in the CCP

As in Zinman (2009), our empirical findings indicate a large difference between credit card debt as reported in the SCF and credit card debt as reported in lender-derived administrative data. The raw CCP-SCF difference in aggregate credit card debt is roughly 40 percent of the CCP estimate (Table 1).[30] We see that the major reporting discrepancy is in balances, with SCF households reporting only 40 percent of the balances that appear on CCP households' credit reports. As noted earlier, the prevalence of credit card use inferred from each source is quite similar. Although the underreporting of credit card debt balances is apparently universal, it is greatest among prime-age households. Borrowers under 35 and over 75 show the closest match (Chart 4). This pattern appears to be stable over time, but we observe a substantial improvement in the SCF-CCP match for borrowers aged 45-54 from 2007 to 2010, and a somewhat weakening match for borrowers nearing retirement.

[29] Since Census region is not publicly available in the SCF, SCF sample sizes for the difference in means tests of comparisons in Chart 3 have been inferred from population densities in the regions and SCF national sample sizes.

[30] This gap is already smaller than the gap discussed in Zinman, which was more than 50 percent. In the following subsection, we discuss the time trend in this gap since Zinman's study.

Chart 4
SCF and CCP Credit Card Debt by Age, 2007 and 2010



Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

A factor that we have not yet taken into account is that some part of the household credit card debt evident in the CCP is generated by small business use of personal credit cards. Such use may or may not be reported by SCF respondents in response to the following questions: "Do you or anyone in your family living here have any credit cards or charge cards?" "After the last payment was made, roughly what was the balance still owed on this account?" And "On your last bill(s), how much were the new charges made to (this account/these accounts)?"[31]

However, as described above, the interviewer is instructed to tell respondents not to report any cards used entirely for business.

Data from the Survey of Small Business Finances (SSBF) shed light on the prevalence and amount of borrowing for business purposes on personal credit cards. In the most recent wave of the survey, fielded in 2003, 46.5 percent of businesses with fifty or fewer employees used personal credit cards for business transactions (Board of Governors of the Federal Reserve System 2010, Table 1). The SSBF sample represents, among others, a population of 9,493,732 businesses with fifty or fewer employees. Assuming that each of these firms borrows on the personal credit cards of only one household, that none of this business borrowing on personal cards was reported in the SCF, and that personal credit card borrowing was identical in 2003 and 2010, this generates an estimate of the prevalence of unreported business borrowing on personal cards in the 2010 SCF of 3.81 percent.[32]

Regarding balances, the SSBF shows that among the 46.5 percent of small businesses using personal cards, the average monthly transaction total on personal cards is $2,161. Further, 13.3 percent of small businesses carry balances on personal cards for business purposes, and these balances average $9,353.[33] Assuming that balance carriers are among the 46.5 percent with any transactions, and that their average carried balance excludes transaction uses, we infer that the sum of average transactions plus debt balance on small business personal cards was $2,249. Distributing this amount of business borrowing among the full population represented by the 2010 SCF, and inflating to 2010 dollars, we calculate a contribution to average SCF credit card debt of $218.

[31] We thank Neil Bhutta for data on the magnitude of business use of personal credit cards.

[32] Some of these 3.81 percent of households with small business credit card debts would also hold personal credit card debts, and the change in the prevalence of credit card borrowing that we measure in the SCF would be less than 3.81 percent.

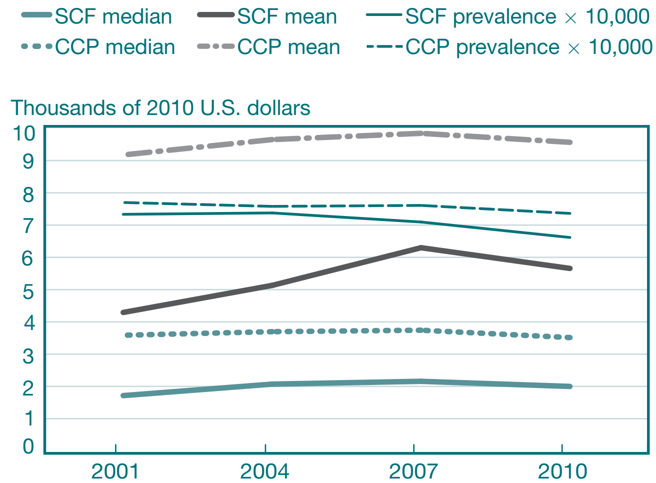
[33] Small businesses here are again defined as those with fifty or fewer employees (Board of Governors of the Federal Reserve System 2010).

Adding this generous estimate of small business usage, and removing the inferred portion of SCF households without credit reports, results in a 2010 SCF unconditional mean credit card balance of $4,437, which may be compared with the CCP unconditional mean of $7,066. This calculation leaves a gap of 37 percent between the SCF and CCP mean household balances.

One final possibility worth mentioning, noted by a lead SCF investigator, is that SCF respondents do not report debt in long-dormant accounts, which they may regard as no longer relevant or may have forgotten. This is not a measurement explanation, but rather an aspect of underreporting. The CCP data include information on accounts that have been updated by the creditor within three months of the date on which the quarter's data were collected. This standard may result in the inclusion of some dormant account balances that lenders continue to report, and the exclusion of other dormant account balances lenders no longer report.[34] This inconsistency may explain some of the difference in aggregate balances. It does not address the question of what consumer behaviors generate dormant, forgotten accounts.

Chart 5

SCF and CCP Credit Card Debt by Year



Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

Note: HELOC is home equity line of credit.

4.3 The Gap between SCF and CCP Credit Card Debt Narrowed from 2001 to 2007

Zinman (2009) demonstrates a widening gap between aggregate credit card debt estimates from the SCF and the G.19 consumer credit data over the 1989-2004 period. We are able to revisit the question for 2001-10 in terms both of household-level debt distribution characteristics and of aggregates. While the SCF-CCP matches between credit card prevalence and conditional median balance are quite stable over time, the difference in conditional mean balances narrowed from 53 to 36 percent of the CCP value between 2001 and 2007 (Chart 5). By 2010, however, the gap had risen to 41 percent. The overall trend in the similarity of lender- and borrower-reported credit card balances is encouraging.

[34] See Lee and van der Klaauw (2010) for further detail on inactive accounts.

4.4 Evidence of Reporting Heterogeneity in 2010 Data Is Limited

One method of correcting for the apparently low level of credit card debt measured by the SCF in research on net worth and consumer balance sheets has been to multiply observed credit card debt by a common factor for each SCF household.[35] This is an appropriate correction if the underreporting of credit card debt is relatively homogenous within the sample. However, based on his finding that SCF-G.19 credit card debt discrepancies grew over time from 1989 to 2004, Zinman (2009) raised the concern that marginal entrants to the credit card market, who likely differed in important ways from previous credit card users, were reporting credit card debt less effectively. This would suggest the presence of meaningful heterogeneity in the quality of credit card debt reporting, which in turn suggests that homogenous corrections for underreported credit card debt are inappropriate.

Our results show relatively homogenous underreporting of unconditional credit card balances by region and age, with the exception of retirees, who under all measures maintain

[35] Examples include Bertaut, Haliassos, and Reiter (2009), Gross and Souleles (2002a), Telyukova (2008), Telyukova and Wright (2008), and Zinman (2007).

low credit card balances. Further, we find these patterns to be very stable over time. Though these findings fall far short of being sufficient to rule out all (observable and unobservable) types of reporting heterogeneity, we fail to find evidence that making a common adjustment for SCF credit card debt underreporting is inappropriate.

4.5 Bankruptcy

The two prominent potential explanations for the remaining gap between SCF and CCP credit card debt levels are the possibility of social stigma attaching to the use of uncollateralized debt, and the possibility that borrowers are not well informed about their credit card debt levels. In 2007, 64 percent of SCF interviews were conducted in person and the remainder over the phone.[36] In both types of interview, the respondent interacts over a long period of time with an interviewer, who grows increasingly familiar with the respondent's personal and financial circumstances. If the respondent suspects that credit card debt, or other consumer attributes, might be looked upon unfavorably by the interviewer, then the respondent may have reason to answer questions regarding such attributes inaccurately. As in most surveys, respondents in the SCF incur no material cost for responding inaccurately. These factors together could lead to inaccurately low reports of credit card debt.

Being uninformed could result from several factors, including willful ignorance, given that large credit card balances are not welcome information; difficulty understanding the growth of credit card balances, as described in Lusardi and Tufano (2009); limited information on other household members' debts; or other cognition and information costs. While stigma issues in reporting are primarily a data quality concern, being uninformed regarding one's debt position may have consequences both for data quality and for the effectiveness of consumers' decision-making. Therefore, it would be valuable to find a way to distinguish between responding to a stigma and being uninformed.

Bankruptcy is a consumer behavior that is both memorable and relatively likely to be stigmatized. Hence, we may be able to learn something about the importance of stigma in debt reporting in the SCF by assessing the accuracy of the survey's bankruptcy figures.

A new literature has emerged on consumers' post-bankruptcy experiences, an increasingly important issue as rates of consumer bankruptcy by 2010 approached levels observed prior to bankruptcy reform.[37] Han and Li (2011) look at post-bankruptcy access to credit using the SCF. Cohen-Cole, Duygan-Bump, and Montoriol-Garriga (2009) examine post-bankruptcy experiences using credit bureau data. We believe that information on the relative quality of bankruptcy measures in the two data sources would be of value to this discussion.

Past default is possibly the most relevant consumer behavior for potential lenders, and hence the accurate reporting of bankruptcy is a leading concern of credit reporting agencies. Given the care taken in recording and reporting bankruptcies, we believe the bankruptcy data in the CCP are fairly accurate. In this section, we examine the similarity between self-reported bankruptcy in the SCF and credit-bureau-reported bankruptcy in the CCP.

One difficulty we face in comparing bankruptcy rates in the two surveys is a difference in the terms of measurement. The SCF asks whether the respondent or the respondent's spouse/partner has filed for bankruptcy, and if so how long ago. The publicly available SCF data report time frames of less than one year as -1, and then round all durations since bankruptcy to the nearest odd integer. Hence, we can identify the proportion of responding individuals or couples who have declared bankruptcy less than two years ago, less than four years ago, and so on. If respondents answer in years, then this allows us to identify the proportion who have declared bankruptcy in the past year, past three years, and so on. The CCP, on the other hand, reports whether an individual has filed for bankruptcy within the past twenty-four months. We can aggregate these CCP individuals into households but, as noted above, we cannot identify the relationships among the household members. Therefore, we are unable to restrict household-level bankruptcies to those of a single household head or married/partnered couple.

We find that the SCF three-year bankruptcy rates—2.90, 2.91, 2.25, and 2.70 in 2001, 2004, 2007, and 2010, respectively—are very similar to the twenty-four-month household bankruptcy rates in the CCP of 2.70, 2.98, 1.97, and 2.65 (Table 5). This appears to indicate that bankruptcy is underreported in the SCF. Significantly, however, this comparison does not account for the difference in the members of the household whose bankruptcy experiences are being reported. When we restrict each sample to households with either one or two adult members, we find little change in the SCF three-year bankruptcy rates. Presumably this is because the SCF asks only about bankruptcies experienced

[36] The unweighted figure is 55 percent in person.

[37] See Federal Reserve Bank of New York, "Quarterly Report on Household Debt and Credit" (2011).

Table 5
Bankruptcy Filing Rates for Consumers or Households in the SCF and CCP

	All Household Sizes		
Year	SCF One-Year Rate (Percent)	SCF Three-Year Rate (Percent)	CCP Two-Year Rate (Percent)
2001	1.18	2.90	2.70
2004	1.20	2.91	2.98
2007	0.93	2.25	1.97
2010	1.45	2.70	2.65

	One or Two Adults in Household			Individual
	SCF One-Year Rate (Percent)	SCF Three-Year Rate (Percent)	CCP Two-Year Rate (Percent)	CCP Two-Year Rate (Percent)
2001	1.21	2.97	2.06	1.74
2004	1.17	2.87	2.34	1.88
2007	0.96	2.34	1.61	1.20
2010	1.27	2.47	2.17	1.59

Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

by the respondent and the respondent's spouse/partner. The CCP's twenty-four-month household bankruptcy rates, however, fall to 2.06, 2.34, 1.61, and 2.17, respectively. Further, the respective CCP individual twenty-four-month bankruptcy rates are 1.74, 1.88, 1.20, and 1.59. These results suggest both that members of large households have relatively high collective bankruptcy rates, and that households with only one or two adult members are a selected group with particularly low bankruptcy rates.

Taken together, the bankruptcy rate estimates in Table 5 suggest little if any underreporting of bankruptcy in the SCF. CCP two-year rates fall squarely between the SCF one- and three-year rates for one-to-two-adult households. The evidence we are able to assemble on bankruptcy reporting in the two sources does not indicate that stigma plays an important role in the collection of survey data on bankruptcy.[38]

[38] Kennickell, in private discussion, notes that bankruptcy questions are fielded late in the SCF survey. At this point, the interviewer and respondent may have built a level of familiarity, and the interviewer has a great deal of information about the respondent's personal and financial position. These factors, he hypothesizes, may contribute to the accuracy of bankruptcy reporting.

Given that bankruptcy is arguably a more stigmatized consumer behavior than credit card borrowing, the lack of evidence of stigma in bankruptcy reporting might suggest that being uninformed, rather than stigma, drives the remaining borrower-lender credit card debt reporting gap.[39] One caveat, however, comes from the marketing literature on conditions under which subjects are likely to lie. Evidence there indicates that subjects tolerate committing dishonesty of limited magnitude without altering their self-concept, but more serious dishonesty may not be tolerable to them (Mazar, Amir, and Ariely 2008). If an inaccurate report of a low credit card balance or the omission of a small credit card balance is perceived as a more tolerable lie than omitting a bankruptcy, then evidence that SCF respondents avoid big lies about bankruptcy, despite stigma, may not be decisive regarding the importance of stigma in the reporting of credit card usage.[40]

4.6 Singles versus Couples

The data also allow us to compare SCF and CCP debt patterns by household size, and this comparison proves informative regarding the ability of a lone respondent to report the debt reliance of all household members. We determine household size by the number of adults, as children are not present in CCP data. Roughly 10 percent of U.S. adults are without credit reports, and thus not included in the CCP. Therefore, some CCP households that truly contain two adults will be miscategorized as single households, some with three adults will be miscategorized as having two, and so on. One might expect this process to inflate CCP debt estimates for a given household size relative to SCF estimates, if slightly.[41]

We do see evidence of slightly more prevalent and higher debt in the CCP estimates than in the SCF estimates by household size as measured by the number of adults (Chart 6). However, as average debt levels are higher in the CCP overall,

[39] Given the evidence that credit card debt reporting has improved over the past decade, one might also seek evidence on trends in knowledge of debt and the stigmatization of uncollateralized borrowing in order to distinguish between the two explanations.

[40] We thank Dean Karlan for this observation.

[41] The logic behind this expectation is as follows: the overall consumption of two-member households tends to be greater than the overall consumption of one-member households. Though the household members represented in the SCF but not in the CCP are missing from the CCP precisely because of their lack of standard consumer debts, the household member who does hold consumer debt may have used it to fund the greater consumption of a larger household. Hence, household size miscategorization in the CCP may lead to inflated average household debts in the CCP relative to the SCF at any given household size.

it is not entirely surprising to see this to be true for any given household size. The main insight from these measures of debt patterns by household size and debt type, however, is that the CCP and the SCF show a similar relationship between debt balances and household size.

Finally, we see some evidence that the match between debt estimates is closer for single households than for larger households. This might be expected, given the standard survey practice of collecting information on household debts from a single respondent: respondents may be better informed about their own debts than those of other household members. This effect appears to be stronger for vehicle and credit card debt.[42]

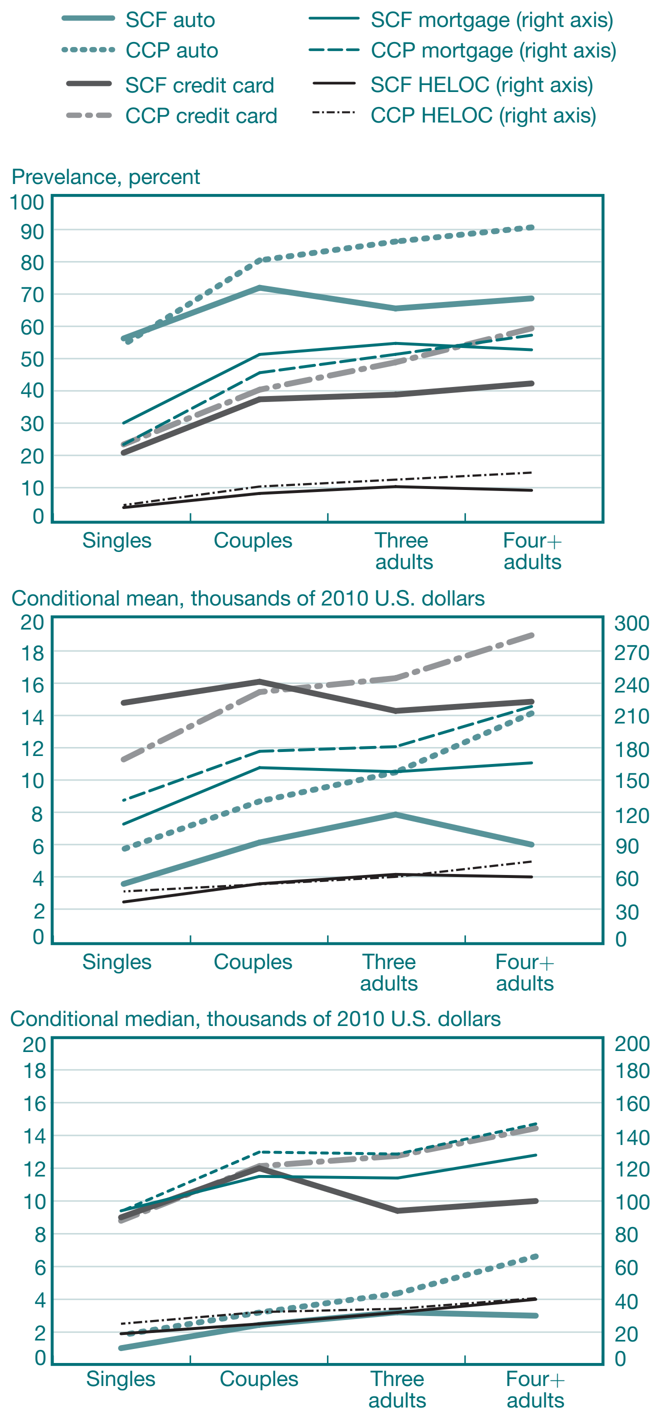
Given this growth in the discrepancy between borrower- and lender-reported debt with household size, we make a back-of-the-envelope calculation of the proportion of the gaps in aggregate debt inferred from the 2010 CCP and SCF that can be explained by reporting challenges in larger households. For both credit card and auto debt, begin with the ratio of the unconditional mean debt of the two-member households in the SCF to that of the two-member households in the CCP. Next, suppose that households with three or more members in the SCF report unconditional mean debts that amount to the same share of the CCP unconditional mean debts for households with three or more members that we observe for two-member households. In other words, suppose that larger SCF households have the same reporting accuracy as two-member households (and that the CCP debt balances reflect the true debt). Finally, sum these inflated SCF three-plus-member households' debts with the observed one- and two-member debts, weighted for the sample shares of each household size. The unconditional mean SCF vehicle debt derived in this manner is 8 percent greater than the observed SCF vehicle debt, and the derived credit card debt is 5 percent greater than the observed SCF credit card debt. This adjustment for reporting quality by family size accounts for 42 percent of the 2010 aggregate vehicle debt gap between the SCF and CCP, but only 8 percent of the (comparatively large) credit card gap between the SCF and the CCP.[43]

[42] We thank Robert Pollak and participants at the Midwest Economic Association session for suggesting a household size comparison. Sierminska, Michaud, and Rohwedder (2008) discuss family size and wealth reporting accuracy. Johnson and Li (2009) find differences between the SCF and Consumer Expenditure Survey housing debt measures that differ more for married than for single households. An additional possible source of difference between single and larger households is that, while relationship types are not an issue in single households, the CCP cannot distinguish among relationship types in larger households. This may lead to categorization of some non-PEU household members as, effectively, PEU members, to borrow SCF terms, and may lead the debt of two-or-more-person CCP households to deviate more from the debt of two-or-more-person SCF households.

[43] We thank a referee for suggesting this calculation.

Chart 6

SCF and CCP Consumer Debt by Household Size, 2010



Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

Note: HELOC is home equity line of credit.

Finally, note that it is not obvious that the CCP reflects the truth, and the SCF a less accurate self-report, when it comes to the debt of larger households. While the sense that a single household respondent answering on behalf of several consumers might overlook some obligations is intuitive, here, as elsewhere in this article, we should keep in mind the potential limitations of the CCP. Chief among these for the case of larger households is the possibility that addresses may be updated imperfectly by lenders, while SCF survey respondents presumably have an accurate picture of the current members of a household. Grown children who leave their parents' homes, for example, may take time to update their addresses with their lenders, and lenders may take time to report the changes to the credit bureaus. This would shift some two-person households into the three-or-more-person category in the CCP and lead to inaccuracy in the measurement of debt by household size, though the direction of the bias in measurement for each group is unclear.[44]

4.7 Primary Economic Unit (PEU) Members

One remaining comparability issue is that, while the CCP data contain debt information for all adults with credit reports residing at a given address, the SCF data typically exclude the debt of non-Primary Economic Unit (PEU) members, where PEU members are as described in Section 3 on data and comparability. The SCF does ask about the presence and amount of any debt held by non-PEU members, and whether the respondent included any of this debt in his or her previous debt responses. The answer to the latter question is not included in the public access SCF data, and hence we are not able to correct even total debt figures for the subset of non-PEU debts that were previously unreported. However, we can use the reported prevalence and amounts of non-PEU members' debt to infer the effect of omitting it on our central conclusions.

We find that 4.4 percent of 2007 SCF households contain a non-PEU member with positive debt. The unconditional mean of non-PEU member debt among our SCF households is $619. Hence, non-PEU member debt is a concern where our conclusions regarding debt comparisons might be swayed by the addition of $619 to the SCF debt level in question or 4.4 percentage points to the relevant debt prevalence. We find that such instances are rare.[45]

5. Implications of Reporting Accuracy for Debt Repayment

As explained above, the match between borrower and lender credit card and student loan debt reports is shown to be weak relative to other debt categories in our SCF-CCP comparison, and elsewhere. Credit card and student debt are generally recognized to be of relatively low repayment quality.[46] Mortgages, HELOCs, and vehicle loans carry substantially lower delinquency rates. Given that reporting quality for credit card and student debt appears to be substantially worse than reporting quality for mortgages, HELOCs, and vehicle loans, the relationship we observe between reporting quality and repayment quality by debt type is consistent with a claim that inaccurate debt reporting is associated with poor repayment outcomes.

One might also consider reporting and delinquency by borrower characteristics. In Chart 1, panels A-C, we observe debt reporting matches that, in many cases, strengthen slightly with age. In the CCP, as well as other sources, we see that delinquency declines almost monotonically with the age of the household head, or the age of the borrower. These observations may suggest a modest positive association between debt reporting accuracy and repayment, when comparisons are made across consumer age groups. But the association is modest indeed. On net, there appears to be some evidence of a positive association between debt reporting quality and repayment. This may be unsurprising, given that one expects borrowers with limited knowledge of their debts to have more difficulties with financial decision making.

[44] Further, while it is clear that this measurement concern regarding the CCP complicates the analysis of debt by household size, we do not believe that it should have a substantial influence on the comparison of aggregate debt. The young adults changing residence, and their debt, should be caught by both the CCP and the SCF sampling scheme in one household or another.

[45] We focus on 2007 in determining the possible magnitude of non-PEU members' debt because it is near the peak of consumer debt for the 2001-10 period. However, the 2010 figures are similar to those for 2007.

[46] Gross and Souleles (2002b), for example, report an 8.2 percent three-cycle delinquency rate among a large, representative pool of 1995 U.S. credit card accounts. Further evidence is available in Federal Reserve Bank of New York, "Quarterly Report on Household Debt and Credit" (2011).

6. Conclusion

This article reports the results of the most complete vetting of SCF debt information to date, to our knowledge. Our central finding is the surprising similarity in the patterns of debt-holding evident in the borrower-reported SCF and lender-reported CCP, both in the aggregate and by debt category, year, region, age, and household structure.

Nevertheless, we also find a substantial gap in credit card debt reporting between the SCF and the CCP, with the raw gap equal to roughly 40 percent of the lender-reported debt level. Generous accounting for differences in the two data sources' sampling design and for small business uses of credit cards narrows the difference in unconditional average household credit card debt to 37 percent of the lender-reported debt level. However, more realistic assumptions would presumably leave a somewhat larger difference, and these adjustments stop far short of reconciling the two measures.

We also find a noteworthy gap in the lender- and borrower-reported levels of the other major uncollateralized debt category, student loans. Aggregate student loans inferred from the SCF are 25.7 percent lower than those inferred from the CCP. This gap may be explained by various measurement differences that would lead debts evident in the CCP not to appear in the SCF.[47] Outside measures of aggregate student debt, though limited, tend to be similar to, or greater than, the CCP figure, and hence far larger than the SCF figure.

[47] However, any limitation in servicer reporting could result in the omission from the CCP of some debts that appear in the SCF.

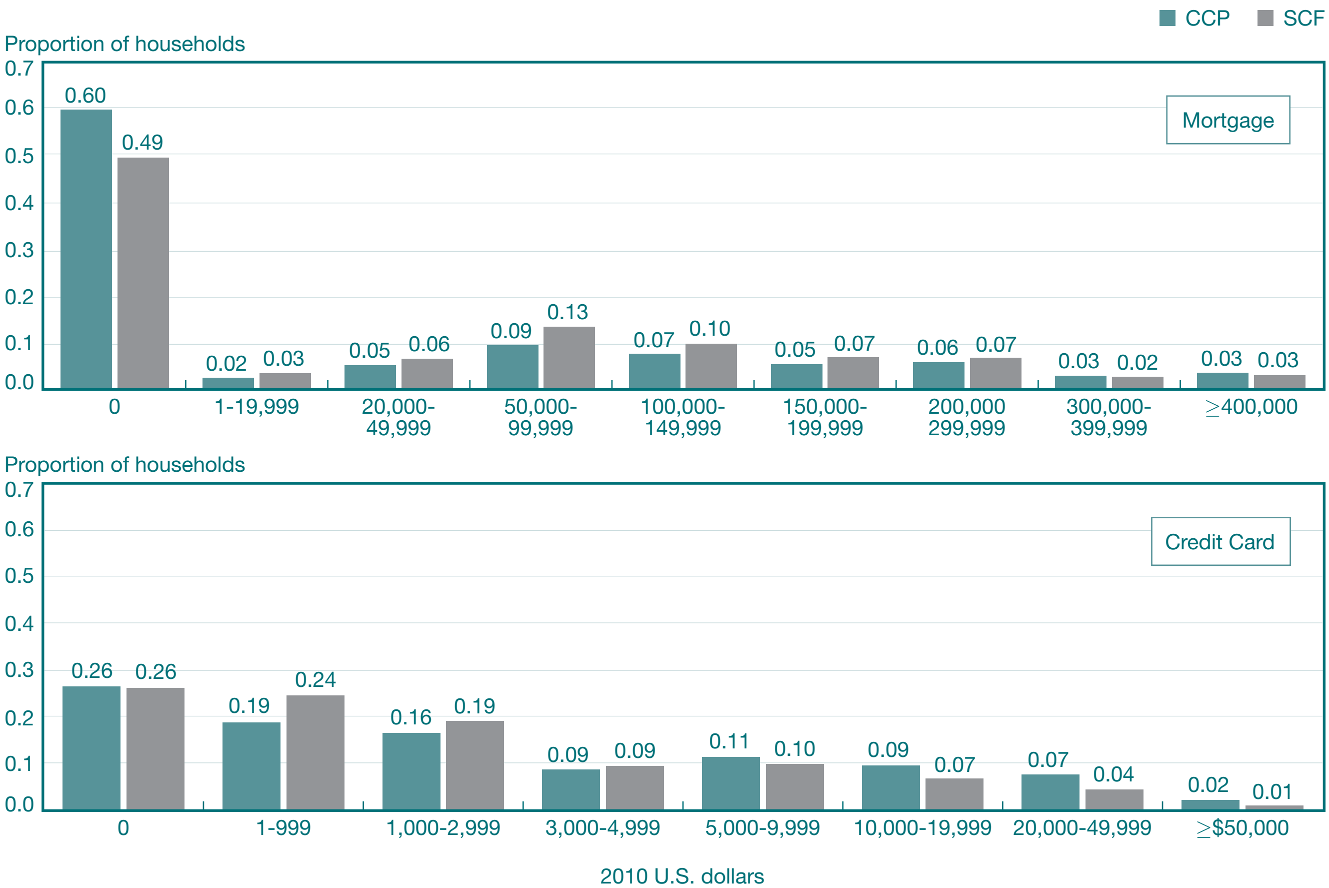
Overall, we observe a pattern of (evident) underreporting of uncollateralized debts, along with comparatively reliable reporting of collateralized debts. The poorer repayment rates we observe for uncollateralized debts may suggest an association between debt awareness and debt repayment quality.

Bankruptcy, like heavy reliance on uncollateralized debt, is arguably a stigmatized consumer behavior. Despite the mismatch in credit card debt reporting, SCF borrowers and CCP lenders report recent personal bankruptcy filings at similar rates (though differences in available measures of bankruptcy in the two data sets impose some qualifications on this claim). We infer from this finding that not all stigmatized consumer behaviors are similarly underreported. Whether this indicates that something other than stigma, such as ignorance of debt positions, underlies the credit card debt discrepancy, or that consumers feel differently about reporting major life events, such as bankruptcy, in contrast to more marginal financial position changes, remains an open question.

Clearly all of this analysis relies on the validity of comparisons at the distributional level. It would be preferable to make the lender-borrower debt report comparison at the level of the household or individual. Therefore, we continue to seek opportunities to observe linked consumer self-reports and lender-reported data.[48] Until such data are available, however, the detailed comparisons permitted by the rich SCF and CCP data provide our most complete picture of the reliability of debt reporting. Finally, while existing survey data provide limited opportunity to separate unwillingness to report financial information from lack of knowledge of financial information, experimental data might permit a distinction between knowledge of debt and willingness to report debt.

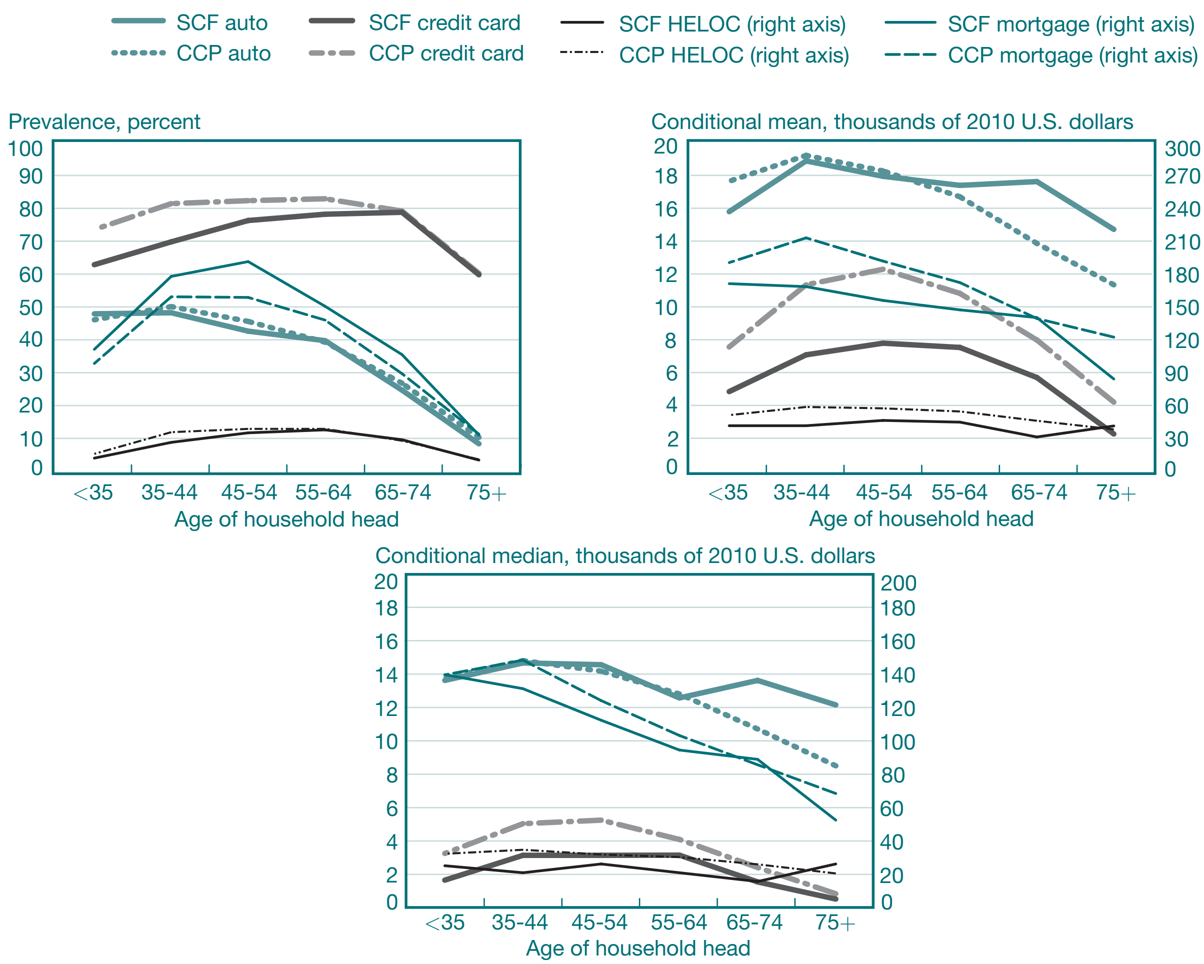
[48] Unfortunately, even a direct match of CCP to SCF households would be of limited value, because coverage by the CCP of the 4,422-6,492 SCF households in each wave would be restricted to a small sample representing somewhat more than a 5 percent match rate.

Chart A1
Mortgage and Credit Card Densities in the SCF and CCP, 2010



Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

Chart A2
SCF and CCP Consumer Debt by Age, 2007



Sources: Board of Governors of the Federal Reserve System, Survey of Consumer Finances; Federal Reserve Bank of New York Consumer Credit Panel/Equifax.

Note: HELOC is home equity line of credit.

References

Antoniewicz, R. L. 2000. "A Comparison of the Household Sector from the Flow of Funds Accounts and the Survey of Consumer Finances." Board of Governors of the Federal Reserve System Occasional Staff Studies working paper.

Avery, R. B., G. E. Elliehausen, and A. B. Kennickell. 1988. "Measuring Wealth with Survey Data: An Evaluation of the 1983 Survey of Consumer Finances." Review of Income and Wealth 34, no. 4: 339–69.

Avery, R. B., P. S. Calem, G. B. Canner, and R.W. Bostic. 2003. "An Overview of Consumer Data and Credit Reporting." Federal Reserve Bulletin 89, no. 2 (February): 47-73.

Bertaut, C., M. Haliassos, and M. Reiter. 2009. "Credit Card Debt Puzzles and Debt Revolvers for Self Control." Review of Finance 13, no. 4: 657-92.

Board of Governors of the Federal Reserve System. 2007. Survey of Consumer Finances Chartbook.

Board of Governors of the Federal Reserve System. 2010. "Report to the Congress on the Use of Credit Cards by Small Businesses and the Credit Card Market for Small Businesses." May. Available at http://www.federalreserve.gov/BoardDocs/RptCongress/smallbusinesscredit/smallbusinesscredit.pdf.

Bound, J. 2001. "Measurement Error in Survey Data." In J. Heckman and E. Leamer, eds., Handbook of Econometrics. Vol. 5. Amsterdam: Elsevier.

Bricker, J., A. B. Kennickell, K. B. Moore, and J. Sabelhaus. 2012. "Changes in U.S. Family Finances from 2007 to 2010: Evidence from the Survey of Consumer Finances." Federal Reserve Bulletin 98, no. 2 (February): 1-80.

Bruine de Bruin, W., W. van der Klaauw, J. S. Downs, B. Fischhoff, G. Topa, and O. Armantier. 2010. "Expectations of Inflation: The Role of Financial Literacy and Demographic Variables." Journal of Consumer Affairs 44, no. 2 (Summer): 381-402.

Bucks, B. K., A. B. Kennickell, T. L. Mach, and K. B. Moore. 2009. "Changes in U.S. Family Finances from 2004 to 2007: Evidence from the Survey of Consumer Finances." Federal Reserve Bulletin 95 (February): A1-A55.

Bucks, B. K., and K. M. Pence. 2008. "Do Borrowers Know Their Mortgage Terms?" Journal of Urban Economics 64, no. 2: 218-33.

Cagetti, M., and M. DeNardi. 2006. "Entrepreneurship, Frictions, and Wealth." Journal of Political Economy 114, no. 5 (October): 835-70.

——. 2008. "Wealth Inequality: Data and Models." Macroeconomic Dynamics 12, no. S2: 285-313.

Castañeda, A., J. Díaz-Giménez, J. V. Ríos-Rull. 2003. "Accounting for Wealth and Income Inequality." Journal of Political Economy 111, no. 4 (August): 818-57.

Chopra, R. 2012. "Too Big to Fail: Student Debt Hits a Trillion." Consumer Financial Protection Bureau blog. Available at http://www.consumerfinance.gov/blog/too-big-to-fail-student-debt-hits-a-trillion (last visited August 29, 2013).

Cohen-Cole, E., B. Duygan-Bump, and J. Montoriol-Garriga. 2009. "Forgive and Forget: Who Gets Credit after Bankruptcy and Why." Federal Reserve Bank of Boston Working Paper no. QAU09-2, July.

Cox, D., and T. Jappelli. 1993. "The Effect of Borrowing Constraints on Consumer Liabilities." Journal of Money, Credit, and Banking 25, no. 2 (May): 197-213.

Davies, J. B., and A. F. Shorrocks. 1999. "The Distribution of Wealth." In A. B. Atkinson and F. Bourguignon, eds., Handbook of Income Distribution. Vol. 1. Amsterdam: Elsevier.

De Nardi, M. 2004. "Wealth Inequality and Intergenerational Links." Review of Economic Studies 71, no. 3 (July): 743-68.

Federal Reserve Bank of New York. 2011. *Quarterly Report on Household Debt and Credit.* August.

Fissel, G. S., and T. Jappelli. 1990. "Do Liquidity Constraints Vary over Time? Evidence from Survey and Panel Data: Note." Journal of Money, Credit, and Banking 22, no. 2 (May): 253-62.

Gokhale, J., L. J. Kotlikoff, J. Sefton, and M. Wheale. 2001. "Simulating the Transmission of Inequality across Generations." Journal of Public Economics 79, no. 1 (January): 93-128.

Gross, D., and N. Souleles. 2002a. "Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data." Quarterly Journal of Economics 117, no. 1 (February): 149–85.

——. 2002b. "An Empirical Analysis of Personal Bankruptcy and Delinquency." Review of Financial Studies 15, no. 1 (Spring): 319-347.

Han, S., and G. Li. 2011. "Household Borrowing after Personal Bankruptcy." Journal of Money, Credit, and Banking 43, no. 2-3 (March-April): 491-517.

Hunt, R. M. 2002. "What's in the File? The Economics and Law of Consumer Credit Bureaus." Federal Reserve Bank of Philadelphia Business Review. Second Quarter: 17–24.

Iacoviello, M. 2008. "Household Debt and Income Inequality, 1963–2003." Journal of Money, Credit, and Banking 40, no. 5 (August): 929–65.

Jacob, K., and R. Schneider. 2006. Market Interest in Alternative Data Sources and Credit Scoring. Center for Financial Services Innovation. December.

Jappelli, T. 1990. "Who Is Credit Constrained in the U. S. Economy?" Quarterly Journal of Economics 105, no. 1: 219-34.

Jappelli, T., J. S. Pischke, and N. S. Souleles. 1998. "Testing for Liquidity Constraints in Euler Equations with Complementary Data Sources." Review of Economics and Statistics 80, no. 2 (February): 251-62.

Johnson, B., and K. Moore. 2005. "Consider the Source: Differences in Estimates of Income and Wealth from Survey and Tax Data." Board of Governors of the Federal Reserve System Occasional Staff Studies working paper.

Johnson, K. W. 2007. "The Transactions Demand for Credit Cards." *B.E. Journal of Economic Analysis and Policy* 7, no. 1 (March): Article 16.

Johnson, K. W., and G. Li. 2009. "Household Liability Data in the Consumer Expenditure Survey." Monthly Labor Review 132, no. 12 (December): 18-27.

——. 2010. "The Debt-Payment-to-Income Ratio as an Indicator of Borrowing Constraints: Evidence from Two Household Surveys." Journal of Money, Credit, and Banking 42, no. 7 (October): 1373-90.

Karlan, D., and J. Zinman. 2008. "Lying About Borrowing." Journal of the European Economic Association 6, no. 2-3 (April-May): 510–21.

Keister, L. A. 2000. Wealth in America: Trends in Wealth Inequality. Cambridge: Cambridge University Press.

Kennickell, A. B. 1991. "Imputation of the 1989 Survey of Consumer Finances," Proceedings of the Section on Survey Research Methods. 1990 Joint Statistical Meetings, Atlanta, Georgia.

——. 1998. "Multiple Imputation in the Survey of Consumer Finances." Board of Governors of the Federal Reserve System Occasional Staff Studies working paper. Available at http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Keys, B. J. 2010. "The Credit Market Consequences of Job Displacement." Board of Governors of the Federal Reserve System Finance and Economics Discussion Series, no. 24, March.

Kiyotaki, N., A. Michaelides, and K. Nokolov. 2011. "Winners and Losers in Housing Markets." Journal of Money, Credit, and Banking 43, no. 2-3 (March-April): 255–96.

Kreider, R. M., and D. B. Elliott. 2009. "America's Families and Living Arrangements: 2007." Current Population Reports. U.S. Bureau of the Census, September.

Lee, D., and W. van der Klaauw. 2010. "An Introduction to the FRBNY Consumer Credit Panel." Federal Reserve Bank of New York Staff Reports, no. 479, November.

Lusardi, A., and P. Tufano. 2009. "Debt Literacy, Financial Experiences, and Overindebtedness." NBER Working Paper no. 14808, March.

References (Continued)

Mazar, N., O. Amir, and D. Ariely. 2008. "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance." Journal of Marketing Research 45, no. 6 (December): 633-44.

Meier, S., and C. Sprenger. 2010. "Present-Biased Preferences and Credit Card Borrowing." American Economic Journal: Applied Economics 2, no. 1 (January): 193–210.

Scholz, J. K., and A. Seshadri. 2009. "The Assets and Liabilities Held By Low-Income Families." In R. Blank and M. Barr, eds., Insufficient Funds. New York: Russell Sage Foundation.

Sierminska, E., P. C. Michaud, and S. Rohwedder. 2008. "Measuring Wealth Holdings of Older Households in the U.S.: A Comparison using the HRS, PSID, and SCF." Paper presented at the Pensions, Private Accounts, and Retirement Savings over the Life Course workshop, Ann Arbor, Michigan. Available at http://psidonline.isr.umich.edu/Publications/Workshops/2008/LC/MRS_WealthCompsv10.pdf.

Sullivan, J. X. 2008. "Borrowing during Unemployment: Unsecured Debt as a Safety Net." Journal of Human Resources 43, no. 2 (Spring): 383-412.

Telyukova, I. 2008. "Household Need for Liquidity and the Credit Card Debt Puzzle." University of California at San Diego Economics Working Paper Series, December.

Telyukova, I., and R. Wright. 2008. "A Model of Money and Credit, with Application to the Credit Card Debt Puzzle." Review of Economic Studies 75, no. 2 (April): 629–47.

U.S. Bureau of the Census. 2010. "Projections of the Number of Households and Families in the United States: 1995 to 2010." Current Population Reports, no. P25-1129.

U.S. Department of Education. 2011. FY 2011 Federal Student Aid Annual Report. November.

Wolff, E. N. 1992. "Changing Inequality of Wealth." American Economic Review 82, no. 2 (May): 552-8.

Zinman, J. 2007. "Household Borrowing High and Lending Low under No-Arbitrage." Dartmouth College working paper.

——. 2009. "Where is the Missing Credit Card Debt?" Review of Income and Wealth 55, no. 2 (June): 249-65.

Rajashri Chakrabarti, Max Livingston, and Elizabeth Setren

The Great Recession's Impact on School District Finances in New York State

- Researchers have explored the effects of the Great Recession on different parts of the economy, but little research exists on the impact of the Great Recession on schools.
- Property, income, and sales tax revenue were all hurt by the financial crisis and recession, and these declines limited the ability of state and local governments to fund school districts.
- An analysis of school financing in New York State from 2004 to 2010 finds that total funding and expenditures were maintained in line with pre-recession trends, but that the composition of each changed in significant ways.
- On the funding side, the federal stimulus offset cuts in local and, especially, state financing. On the expenditure side, instructional spending was maintained on trend while noninstructional spending—transportation, activities, utilities—suffered. Affluent districts saw larger drops than poorer districts, while the New York City metro area was hit harder than other areas of the state.

1. Introduction

The financial crisis and the Great Recession that followed led to declining tax revenues, which, in turn, strained state and local government finances. Property, income, and sales tax revenue were all hurt by the bursting of the housing bubble and a weakened labor market, and these decreases in revenue limited state and local governments' ability to fund school districts. Starting in the fall of 2009, the federal government, through the American Recovery and Reinvestment Act (ARRA), allocated $100 billion to states for education in an effort to lessen the impact of decreased state and local funding and stave off serious budget cuts. New York State received $5.6 billion of the ARRA stimulus funding and an additional $700 million from the Race to the Top Competition.[1]

Because schools are an indispensable part of our economy and society and have an undisputed role in human capital formation and the shaping of the nation's future, it is

[1] Race to the Top is a competitive grant program created by the U.S. Department of Education that rewards states on the basis of reforms and innovation in K-12 education.

Rajashri Chakrabarti is a senior economist at the Federal Reserve Bank of New York. Max Livingston is a former research analyst and Elizabeth Setren a former assistant economist at the Bank.

Correspondence: rajashri.chakrabarti@ny.frb.org

The authors would like to thank Jason Bram, Erica Groshen, Andrew Haughwout, James Orr, Joydeep Roy, Amy Ellen Schwartz, Giorgio Topa, and seminar participants at the Federal Reserve Bank of New York and the NYU-FRBNY Fiscal Breakfast for their valuable insight and feedback. We are grateful to Jennifer Bell-Ellwanger and Katie Brohawn of the New York City Department of Education, Deborah Cunningham, Darlene Tegza, and Margaret Zollo of the New York State Department of Education, and Theresa Hunt of the New York State Comptroller's Office for their generous help with the data and for patiently answering numerous questions. We would also like to thank Sarah Sutherland for her assistance. The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

essential to understand how the Great Recession affected schools and what, if any, repercussions the recession might have on school funding and spending and hence the delivery of educational services and student learning. While a slowly emerging literature seeks to understand how the Great Recession affected other parts of the economy, there is surprisingly little literature on how it affected schools (Chakrabarti and Sutherland 2013). This article starts to fill the gap. Here, we study the ways in which New York State's school funding and expenditures, as well as the composition of each, were affected by the recession and the federal stimulus. In addition to investigating aggregate trends, we analyze whether there were variations in these patterns across metropolitan areas, poverty levels, district sizes, and urban status (urban, suburban, or rural). New York is of interest primarily because it includes New York City, the country's largest school district. In addition, New York's is the third-largest state school system, serving 5.6 percent of the nation's students.[2] Also notable is the state's diversity: it contains a range of urban, suburban, and rural districts, with a wide distribution of income levels.

Some interesting findings emerge. There is no evidence of any statistically significant shift—relative to trend—in either total funding per pupil or total expenditure per pupil after the recession.[3] But while we find no evidence of overall shifts, there is robust evidence of compositional shifts within both funding and expenditures. With the infusion of federal stimulus funds, state aid shifted downward (relative to trend), and so did local funding. Meaningful shifts are also observed in the composition of expenditures. Instructional expenditures, the key category that most directly affects student learning, remained on trend. In contrast, noninstructional categories such as student activities, student services, transportation, and utilities and maintenance ("utilities") experienced cutbacks (relative to trend), although the effects were not always statistically significant. See Table 1 for descriptions of the various expenditure categories.

In addition to these overall patterns, we find considerable variations within the state. Affluent districts were the worst hit in terms of both funding and expenditure (relative to trend). Noninstructional expenditures fell the most in these districts, and unlike high- and medium-poverty districts, affluent districts exhibited a fall in instructional expenditures as well. Analysis by metro area reveals that Nassau County experienced sizable downward shifts both in total expenditure and in its various components. New York City also experienced some declines, though they were considerably smaller economically than those in Nassau County. There were heterogeneities by urban status as well. Urban districts exhibited the largest declines in both instructional and noninstructional expenditures, although these declines were not always statistically significant. (Note that all these changes are relative to trend of the corresponding variable.)

The patterns suggest that, in the face of budget cuts, school districts focused on maintaining instructional expenditures on trend. Across the board, noninstructional categories were affected much more adversely than instructional expenditures, while in most cases, instructional expenditures were maintained on trend. In the small number of cases where there were declines, they were economically and statistically small.

A caveat relating to our analysis is worth noting here. We use a trend shift analysis: we look for a shift in various school finance indicators from their

Table 1
Definitions of Expenditure Components

Instruction	
Instructional expenditures	All expenditures associated with direct classroom instruction, including teacher salaries and benefits, classroom supplies, and instructional training
Noninstruction	
Instructional support	All support service expenditures designed to assess and improve students' well-being, including food services, educational television, library, and computer costs
Student services	Psychological, social work, guidance, and health services
Utilities and maintenance	Heating, lighting, water, and sewage; operation and maintenance
Transportation	Total expenditures on student transportation services
Student activities	Extracurricular activities, including physical education, publications, clubs, and band

[2] This statistic is based on authors' calculations using the Common Core of Data of the National Center for Education Statistics for the 2008-09 school year.

[3] While there is evidence of small declines in total funding per pupil (especially in the 2009-10 school year), these effects are never statistically different from zero.

pre-existing trends to two subsequent time frames: the first school year after the start of the recession (2008-09) and the school year during which school districts received the infusion of federal stimulus funds (2009-10). We attribute any such shifts in the school year just after recession to the recession and any shift in the following year to a combination of recession and federal stimulus. Note, though, that if there were shocks during these two years that affected our school finance indicators independently of the recession, our estimates would be biased. So we look upon our estimates as strongly suggestive but not necessarily causal. Although this caveat should be kept in mind while interpreting the results of this article, we did an extensive search for such potentially confounding shocks and found none. Moreover, the Great Recession was not a marginal shock at all, but rather a highly discontinuous one. So even if there were small shocks during these two years, they would, by far, be overpowered by the enormous shock of the Great Recession.

2. Overview of the Literature

This article is related to the literature that studies school district funding. Stiefel and Schwartz (2011), analyzing school finance patterns in New York City from 2002 to 2008, find evidence of large increases in per pupil funding during this period. Rubenstein et al. (2007), studying schools in New York City, Cleveland, and Columbus, Ohio, find that schools with higher poverty levels receive more funding per student. Baker (2009), studying schools in Texas and Ohio, finds that resources vary according to student needs within districts. But this article is most closely related to the literature that studies the impact of recessions on schools. Studying the 2001 recession and regressing the percentage change in property taxes per capita on the change in state aid per capita as a percentage of property taxes per capita, Dye and Reschovsky (2008) find that state funding cuts were partially offset by increased property tax funding. Studying funding and expenditure patterns for New Jersey following the Great Recession, Chakrabarti and Sutherland (2013) find that New Jersey districts faced declines in state funding (relative to trend). Interestingly, this decline prompted compositional shifts in expenditures in favor of categories linked most closely to instruction, while expenditures in several noninstructional categories, including transportation and utilities, declined.

It follows from the above discussion that while there is research on school funding and resource allocation within and across districts, the literature on the impact of recessions, especially the Great Recession, on schools is woefully sparse. This article takes a step toward filling that gap by studying the impact on school finances in New York State. Understanding how school districts fared during the Great Recession promises to improve current understanding of schools' financial situations and response to financial stress, and will aid future policy decisions.

3. Background

3.1 Financial Crisis and Federal Stimulus Funding

The burst of the housing bubble and the onset of the recession in 2007 strained the finances of state and local governments as their funding slowed. The housing market began cooling in 2005 and 2006 as foreclosures increased. In 2007, as subprime lenders declared bankruptcy and credit for home equity loans dried up, the housing market crashed. According to the CoreLogic Home Price Index, the United States as a whole saw a 29.4 percent drop in housing values from October 2006 to February 2009. The decline in New York State, at 13.5 percent, was less drastic. Local governments nationwide, which typically derive a large percentage of their total revenue from property taxes, faced falling revenues as a result of declines in the housing market.

State governments also saw a decline in funds, owing both to reduced income tax revenues from increased unemployment and reduced sales tax revenues from lower consumption. New York's unemployment rate increased from 4.6 percent in 2006 to 8.5 percent in 2010, though the state fared better than the nation, which had the same unemployment rate in 2006 and 9.6 percent unemployment in 2010.[4] State tax revenue fell 8 percent in New York from 2007 to 2009, similar to the national state average, which declined 9 percent.

[4] Authors' calculations based on the Current Population Survey and Local Area Unemployment Statistics, U.S. Bureau of Labor Statistics. Accessed via Haver Analytics.

The financial downturn limited state and local governments' ability to fund school districts and resulted in difficult budget decisions. According to the Center on Budget and Policy Priorities, at least forty-six states and the District of Columbia worked to close budget shortfalls entering the 2011 fiscal year. K-12 education derives more than half of its funding from state revenue, so these budget gaps had significant implications for education financing. To stave off serious budget cuts, the federal government allocated $100 billion to states for education through the American Recovery and Reinvestment Act (ARRA). The funds were available starting in the 2009-10 school year and running through the fall of 2011.

The ARRA money lessened the impact of decreased state and local funding on school budgets. Approximately $5.6 billion of the ARRA funds went to New York schools.[5] Nationwide, districts were directed to use the ARRA funds to save and create jobs, to boost student achievement and bridge student achievement gaps, and to improve accountability and performance reporting. The funds were distributed using the states' formulas for distributing education aid. New York won an additional $700 million from the Race to the Top competition.[6]

3.2 Budget Cuts

When faced with tight budgets, school districts tend to trim spending that does not affect core subjects (Cavanagh 2011). Common cuts include extracurricular activities, art and music programs, maintenance, purchases, transportation, and equipment upgrades. After these initial cuts, more severe options are considered, such as increased class size, decreased staff, and reductions in instruction hours, benefits, professional development, and bonuses.

[5] These estimates include State Fiscal Stabilization Funds; Title I Part A–Supporting Low-Income Schools; IDEA Grants, Parts B & C–Improving Special Education Programs; and Education Technology Grants. This number does not include competitive grants such as Race to the Top. Source: http://www2.ed.gov/policy/gen/leg/recovery/state-fact-sheets/index.html

[6] Race to the Top (RTT) awards were announced in April 2010 and distributed starting in the 2010-11 school year, running to the fall of 2014, so these RTT funds were not available during the school years discussed in this article.

3.3 New York State School Funding Overview

Funding for public schools in the United States comes from three main sources: the federal government, the state government, and local funding. The last item, local funding, reflects locally raised revenue within a school district, mostly from property taxes. In the 2007-08 school year—which we take as the immediate pre-recession year because budgets were set in spring 2007, before the recession began—New York State districts received approximately 3 percent of their funding from federal aid, 40 percent from the state, and 57 percent from local funding. By 2009-10, reliance on federal aid increased to approximately 7 percent, and the share of funding from state and local sources fell to 38 percent and 55 percent, respectively. The bulk of federal school aid goes to Title I funding to support low-income students and students with disabilities.

State aid for education primarily comes from the State General Fund, which is financed by state income and sales taxes. Some additional funding comes from the Special Funding account supported by lottery receipts (State Department of Education 2009). State aid to school districts is based on a variety of characteristics of the school districts, including enrollment, regional labor market costs, the percentage of low-income students, and the percentage of students with limited proficiency in English.

In New York State, 90 percent of local funding comes from residential and commercial property tax receipts. The largest school districts—Buffalo, New York City, Rochester, Syracuse, and Yonkers—fund their schools from city budgets instead of linking funding directly to property tax revenue. New York City, which accounts for about half of the New York State student population, has undergone important finance policy changes in recent years. The Children First initiative, which started in 2003, increased teachers' salaries and boosted financial incentives to work in high-need schools and subject areas with teacher shortages (Goertz, Loeb, and Wyckoff 2011). In 2008, the Fair Student Funding program aimed to improve the distribution of resources by allocating school funds based on the number of low-income, special education, and low-achieving students, as well as the number of English language learners. According to some, but not all, measures, this policy resulted in increased spending on students with greater needs (Stiefel and Schwartz 2011).

4. Data

We use school district financial report data from the New York Office of the State Comptroller. The data cover the 2004-05 to 2009-10 school years and the 714 school districts in New York State. Student demographic data and the percentage of students eligible for free or reduced-price lunches from 2004-05 to 2009-10 are available from the New York State Department of Education.

The school finance data set includes funding, expenditure, and enrollment information, as well as individual components of funding and expenditure. Funding information includes data on total funding, on the amount of aid received from federal and state sources, and on local funding, including property tax funding. Expenditure information includes total expenditures, as well as detailed data on instructional expenditures, instructional support expenditures, student services, transportation, and utilities. The definition of each of these variables is provided in Table 1. The data set includes total fall student enrollment figures for each school year in the covered period.

We categorize districts as high-poverty, medium-poverty, or low-poverty, based on the percentage of students who received free or reduced-price lunch in the 2007-08 school year. Districts that fall within the top 75th percentile (that is, those in which 42 percent or more of students were receiving free or reduced-price lunch) are categorized as high-poverty districts. We categorize the bottom 25th percentile, or those districts with 13 percent or less of students in the lunch program, as low-poverty. The rest of the districts are referred to as medium-poverty.

We use the National Center for Education Statistics (NCES) Common Core of Data (CCD) designations of urban status in 2007-08 to categorize districts as urban, suburban, or rural. Districts inside urbanized areas or inside urban clusters less than thirty-five miles from urbanized areas are categorized as urban. Districts outside principal cities and towns but close to urbanized areas make up the suburban districts. The NCES categorizes areas that have fewer than 2,500 inhabitants and are outside of an urban area as rural.

We perform heterogeneity analysis by metropolitan area. We consider the following metro areas: Albany, Buffalo, Rochester, Syracuse, Ithaca, New York City, and Nassau-Suffolk. The first four are Metropolitan Statistical Areas (MSAs). Since Ithaca's MSA has only a few school districts, we study the Binghamton, Cortland, Elmira, and Ithaca MSAs together and refer to them as the Ithaca Metropolitan Area. While New York City and Nassau-Suffolk constitute one MSA, because of their differences, we study them separately as the New York–White Plains Division and the Nassau County Metropolitan Division ("Nassau").[7] See Exhibit 1 for a map of the areas we examine.

As noted previously, we take the school year 2007-08 to be the immediate pre-recession year. School year budgets are finalized in the preceding spring, meaning that the budget for the 2007-08 school year was set in spring 2007, before the recession hit.

In the rest of the article, we refer to school years by the year of the spring semester.

5. Interpretation of Post-Recession Effects

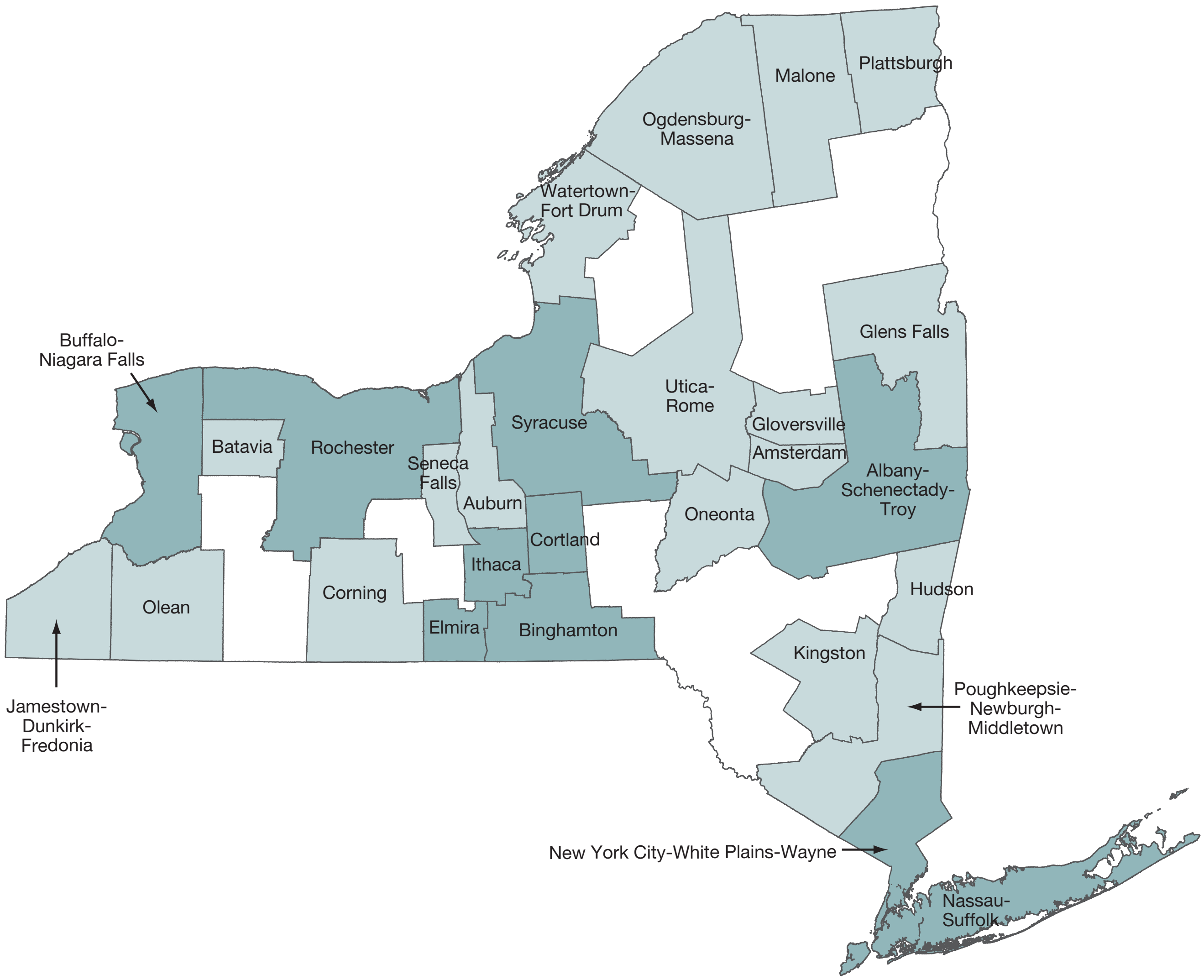
The goal of this article is to investigate whether the Great Recession and the federal stimulus funding period that followed were associated with shifts in education financing in New York State. We conduct a trend shift analysis and use the specification in the Box to analyze these effects. The reasoning behind this methodology is that we expect that school finances would have continued to grow at their pre-recession rate had there been no recession. Thus, post-recession effects ($\propto_2$ and $[\propto_2 + \propto_3]$ in the Box) capture shifts from this trend in the post-recession period in 2009 and 2010, respectively.

To quantify the relative change in each finance variable, we also compute percentage shifts that are obtained by expressing the shifts $\propto_2$ and $\propto_2 + \propto_3$ from the specification as percentages of the pre-recession (2008) base of the corresponding financial variable (Y_{it}). This pre-recession base is simply the average across districts of each variable in the 2008 school year. As noted previously, local, state, and federal governments finalize their budgets in the spring prior to the budgeted year. More specifically, budgets for the 2008 school year were finalized in the spring of 2007, before the recession officially began (December 2007) and before decision makers were aware of the impending

[7] We use ArcGIS mapping technology to represent changes in financial variables spatially. The district and MSA shape files come from the U.S Census Bureau.

Exhibit 1

New York State Metropolitan and Micropolitan Statistical Areas



Notes: The map represents all metropolitan and micropolitan statistical areas in New York State, as defined by the Office of Management and Budget in 2009. A metro area contains a core urban area with a population of 50,000 or more, and a micro area contains an urban core with a population of 10,000 to 50,000. The metro areas that we focus on in our analysis by metro area are Albany, Buffalo, Ithaca, Nassau, New York City, Rochester, and Syracuse. These are shaded dark blue in the map. In the case of Ithaca, we pool four areas (Binghamton, Cortland, Elmira, and Ithaca, all of which are metro areas except Cortland, which is a micro area). In the case of the New York City MSA, we consider its component metropolitan divisions—New York City and Nassau-Suffolk ("Nassau")—as separate metro areas.

Empirical Strategy

We analyze whether the recession and federal stimulus periods were associated with shifts in various school finance indicators from their pre-existing trends. We use the following specification for this purpose:

$$Y_{it} = \alpha_1 t + \alpha_2 v_1 + \alpha_3 v_2 + \alpha_4 X_{it} + f_i + \varepsilon_{it},$$

where Y_{it} is a financial indicator for school district i in year t; t is a time trend variable that equals 0 in the immediate pre-recession year (2008) and increases by 1 for each subsequent year and decreases by 1 for each previous year; v_1 is the recession dummy, $v_1 = 1$ if year $>$ 2008 and 0 otherwise; v_2 is the stimulus dummy, $v_2 = 1$ if year $>$ 2009 and 0 otherwise; X_{it} represents the school district demographic characteristics (racial composition and percentage of students eligible for free or reduced-price lunches); and f_i denotes district fixed effects.

The coefficient on the time trend variable, $\propto_1$, denotes the overall trend in the financial indicator in the pre-recession period. The intercept shift coefficient, $\propto_2$, denotes whether there was an intercept shift (from the pre-recession trend) in the first year after recession, and $\propto_3$ captures any additional shift in 2009-10, the year ARRA was implemented and school districts received an infusion of funds under the federal stimulus. In Tables 2 through 7, we define $\propto_2$ as "recession" and $\propto_3$ as "stimulus." The shifts relative to pre-existing trends in 2009 and 2010 are captured by $\propto_2$ and $(\propto_2 + \propto_3)$, respectively.

All financial variables are inflation-adjusted to 2009 dollars. All regressions reported in the article include district fixed effects. Demographic controls and robust standard errors are used in all regressions. The results are robust, to the inclusion or exclusion of covariates.

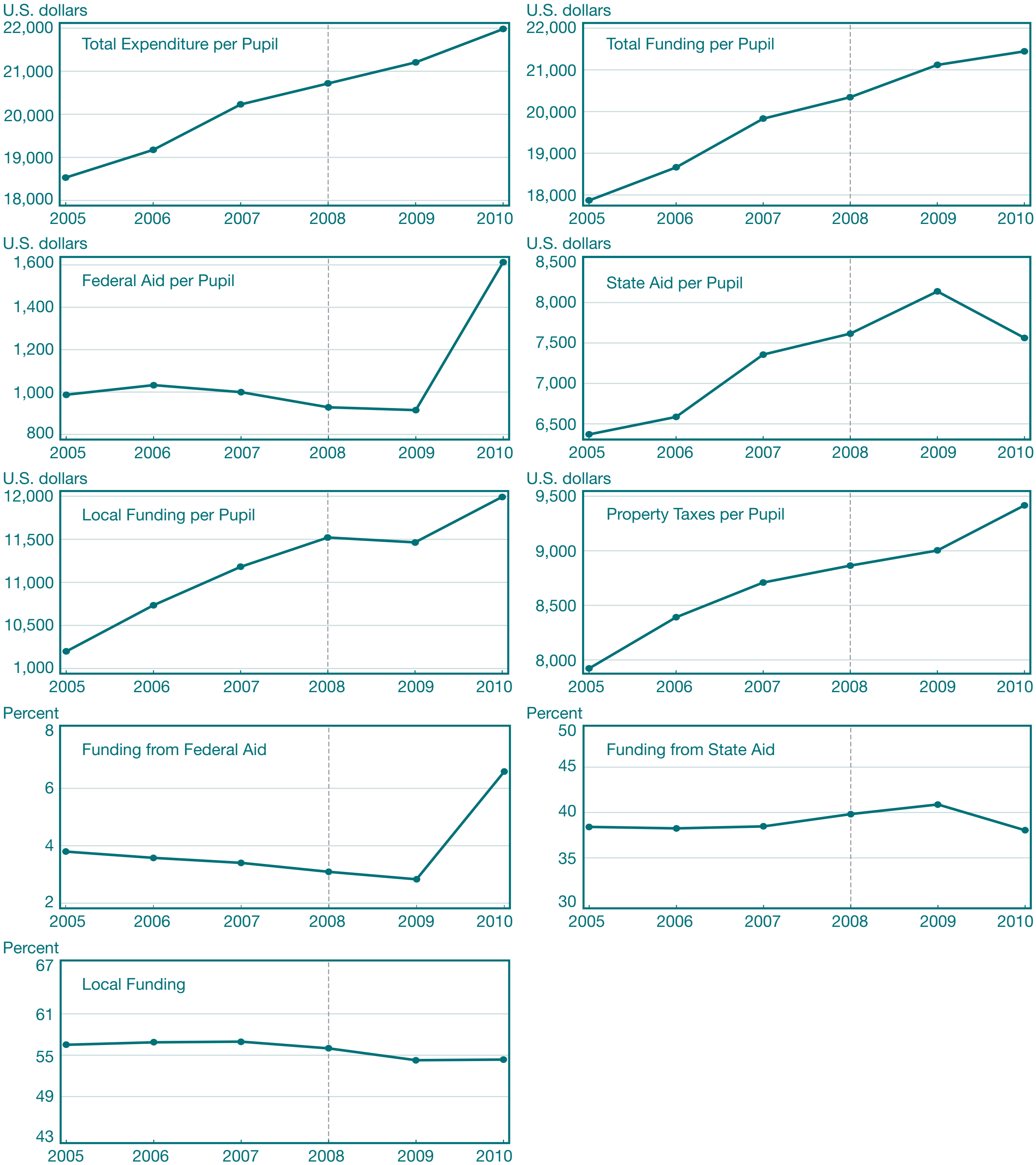
recession. Therefore, 2008 is taken as the last pre-recession year in this article.

These percentage effects allow for a clearer interpretation and are more informative than simply looking at the coefficients ($\propto_2$ and $\propto_3$) because they give an idea about the size of the effects and can be easily compared with one another. In our discussion, we will focus on two percentage shifts: first, the 2009 percentage shift immediately following the recession, calculated as $\frac{\propto_2}{\text{pre-recession base}}$ for each finance variable (Y_{it}); and second, the percentage shift in 2010, calculated as $\frac{\propto_2 + \propto_3}{\text{pre-recession base}}$ for each finance variable (Y_{it}). The first percentage shift captures the effect of the recession in 2009 and the latter captures the combined effect of the recession and the federal stimulus in 2010.

An important caveat relating to the strategy above should be mentioned here. The estimates from the specification capture shifts from the pre-existing trend of the corresponding financial variables. However, these specifications do not control for any other shocks following the recession that might also have affected these financial variables. To the extent that there were such shocks, our estimates would be biased. As a result, we would not like to portray these estimates as causal effects, but as effects that are strongly suggestive of the effects of recession and stimulus on various school finance variables. However, we conducted some research to assess the presence of shocks (for example, policy changes) that might affect our outcome variables of interest independently of the recession and stimulus. We found no evidence of such shocks during this period.

Chart 1

Trends in School Revenues and Expenditures in New York State during the Great Recession



Notes: School years are expressed as the year corresponding to the spring semester. Dotted lines mark the immediate pre-recession (2007-08) school year.

Chart 2

Trends in Composition of Expenditures in New York State during the Great Recession

Notes: School years are expressed as the year corresponding to the spring semester. Dotted lines mark the immediate pre-recession (2007-08) school year.

6. Results

6.1 Overall Patterns

Chart 1 shows trends in various aggregate school finance variables. The dotted vertical line marks the immediate pre-recession (2007-08) school year. There is not much evidence of shifts in expenditure or revenue per pupil; both remained on trend. As expected, federal aid per pupil and the federal share in total funding show a steep increase in 2009-10, the year of the federal stimulus funding. State aid per pupil, as well as the share of state aid, exhibits a decline in 2009-10 as the federal stimulus came in. Local funding per pupil, as well as its share, declined after the recession (relative to trend).

Chart 2 focuses on the various components of expenditure. There is no evidence of effects on instructional expenditure; however, several noninstructional categories (transportation, student services per pupil, and student activities per pupil) show some flattening after the recession. Next, we investigate whether these patterns hold up in a more formal trend shift analysis.

Table 2 presents results from estimation of the specification. The setup of the table establishes the pattern for the five tables that follow. The top part of each panel presents the percentage shifts, while the lower part presents the regression estimations from which the percentage shifts were derived. Our discussion of results will focus on these percentage shifts. The first row presents the percentage shift in

TABLE 2

Funding and Expenditures per Pupil during the Financial Crisis and the Federal Stimulus Period

Panel A	Total Expenditure per Pupil	Total Funding per Pupil	Federal Aid per Pupil	State Aid per Pupil	Local Funding per Pupil	Property Taxes per Pupil
Percentage shift in 2008-09	-0.410	-0.743	5.674	3.377***	-5.060***	-2.517 *
Percentage shift in 2009-10	1.985	-2.348	126.844***	-6.285***	-6.673***	-2.467
Pre-recession base	23,580.53	22,724.17	705.01	7,883.87	13,914.50	10,172.06
Trend	940.3***	1035.8***	-4.3	412.5***	629.632***	420.400***
	(125.4)	(120.0)	(12.8)	(14.0)	(112.738)	(92.307)
Recession	-96.7	-168.9	40.0	266.3***	-704.125***	-256.041*
	(310.8)	(275.3)	(42.9)	(50.2)	(222.578)	(138.652)
Stimulus	564.7	-364.7	854.3***	-761.7***	-224.341	5.074
	(369.4)	(317.7)	(63.6)	(59.3)	(245.825)	(185.666)
Observations	4,146	4,146	4,146	4,146	4,146	4,146
R^2	0.88	0.91	0.85	0.96	0.94	0.96

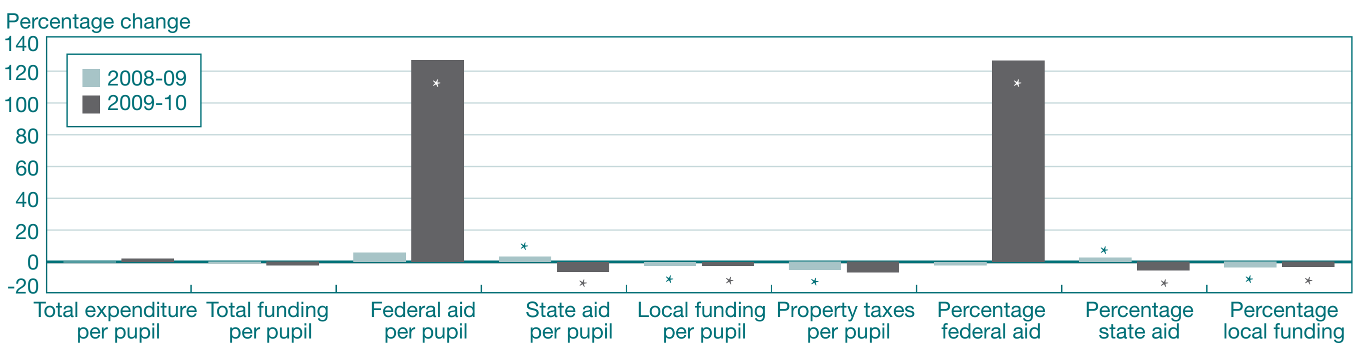
Panel B	Percentage Federal Aid	Percentage State Aid	Percentage Local Funding	Total Number of Students
Percentage shift in 2008-09	-2.134	2.664***	-3.512***	-0.163
Percentage shift in 2009-10	126.798***	-5.509***	-3.154***	1.151
Pre-recession base	3.09	39.83	56.00	3889.72
Trend	-0.229***	0.402***	-0.100***	-37.653***
	(0.018)	(0.037)	(0.038)	(9.337)
Recession	-0.066	1.061***	-1.967***	-6.342
	(0.053)	(0.112)	(0.116)	(30.835)
Stimulus	3.987***	-3.255***	0.201 *	51.128
	(0.070)	(0.116)	(0.109)	(38.394)
Observations	4,146	4,146	4,146	4,146
R^2	0.90	0.99	0.99	1.00

Notes: Robust standard errors are in parentheses. All regressions include school district fixed effects and control for racial composition and percentage of students eligible for free or reduced-price lunch. The pre-recession base is expressed in 2009 constant dollars.

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

CHART 3

Changes in Revenue during the Financial Crisis and the Federal Stimulus Period



Note: Asterisk (*) denotes significance at the 10, 5, or 1 percent level.

2009 $\left(\frac{\propto_2}{\text{pre-recession base}}\right)$ and captures the effect of the recession. The second row gives the percentage shift in 2010 $\left(\frac{\propto_2 + \propto_3}{\text{pre-recession base}}\right)$ and captures the combined effect of federal stimulus funding and the recession. The third row shows the district average pre-recession base of the relevant dependent variable. The bottom section of each panel shows the regression estimation results that are used to calculate the percentage shifts. "Trend" corresponds to $\propto_1$, "recession" to $\propto_2$, and "stimulus" to $\propto_3$. For ease of comparison, these percentage shifts are also presented in bar charts.

Table 2 and Chart 3 show that, overall, New York State school districts maintained the trend of total funding and total expenditure per pupil during the recession. The composition of funding changed following the recession. In 2008-09, local funding shifted downward and state aid filled in the gap by shifting upward. Federal aid per pupil more than doubled in the 2009-10 school year relative to the pre-recession trend. This coincided with downward shifts in state and local funding per pupil (relative to the pre-recession trend). Thus, there seems to have been a substitution of funds away from state and local funds and toward federal funds. The increased reliance on federal aid is also evidenced by the maps in Exhibit 2. On average, New York districts received 3 percent of their funding from federal sources in 2007-08. However, they received more than 7 percent of their funding from federal sources after the start of the ARRA money in 2009-10. This uptick in and increased reliance on federal aid stem from the fiscal stimulus, which sought to prevent serious budget cuts given declining state and local funding.

While overall expenditure remained on trend, the composition of expenditure shows interesting changes (Table 3 and Chart 4). Districts maintained instructional and instructional support expenditures on trend.[8] Since classroom expenditures and teachers most directly affect student learning, they are likely to be undesirable targets for budget cuts. Additionally, teachers' salaries make up a large portion of instructional spending, and reducing expenditures in this area is difficult, since it involves contract renegotiations or layoffs.

The noninstructional expenditures per pupil, especially transportation, student activities, and utilities and maintenance, faced cuts in both years after the onset of the recession (relative to the pre-recession trend), especially in 2010. Expenditures for student services also trended downward, but the decline was not statistically significant.[9]

6.2 Examining the Heterogeneity of Effects by Poverty Level

While the above analysis focuses on aggregate patterns, the rest of this article investigates whether there were differences in impact within the state by various characteristics such as poverty level, location, and urban status. To save space, this analysis focuses only on a subset of the finance indicators analyzed above—the various components of expenditure, which are the indicators of greatest interest. This analysis provides valuable insight into how the different types of districts allocated funds and how the students in these districts were affected. Results for the other indicators are available on request.

In this section, we investigate whether there were variations in effects across different poverty levels. As Table 4 and Chart 5 show, instructional expenditure declined (relative to trend) only in the low-poverty districts (and this was statistically significant only in 2009). In contrast, cuts to noninstructional spending were much more widespread. Transportation and utilities suffered significant decreases in both 2009 and 2010 in medium- and low-poverty districts. Student services also decreased in low-poverty districts in both years, but the decline was not statistically significant. Surprisingly, medium-poverty districts experienced statistically significant increases in student services expenditures in both years. None of the three groups of districts experienced a statistically significant shift in instructional support per pupil.

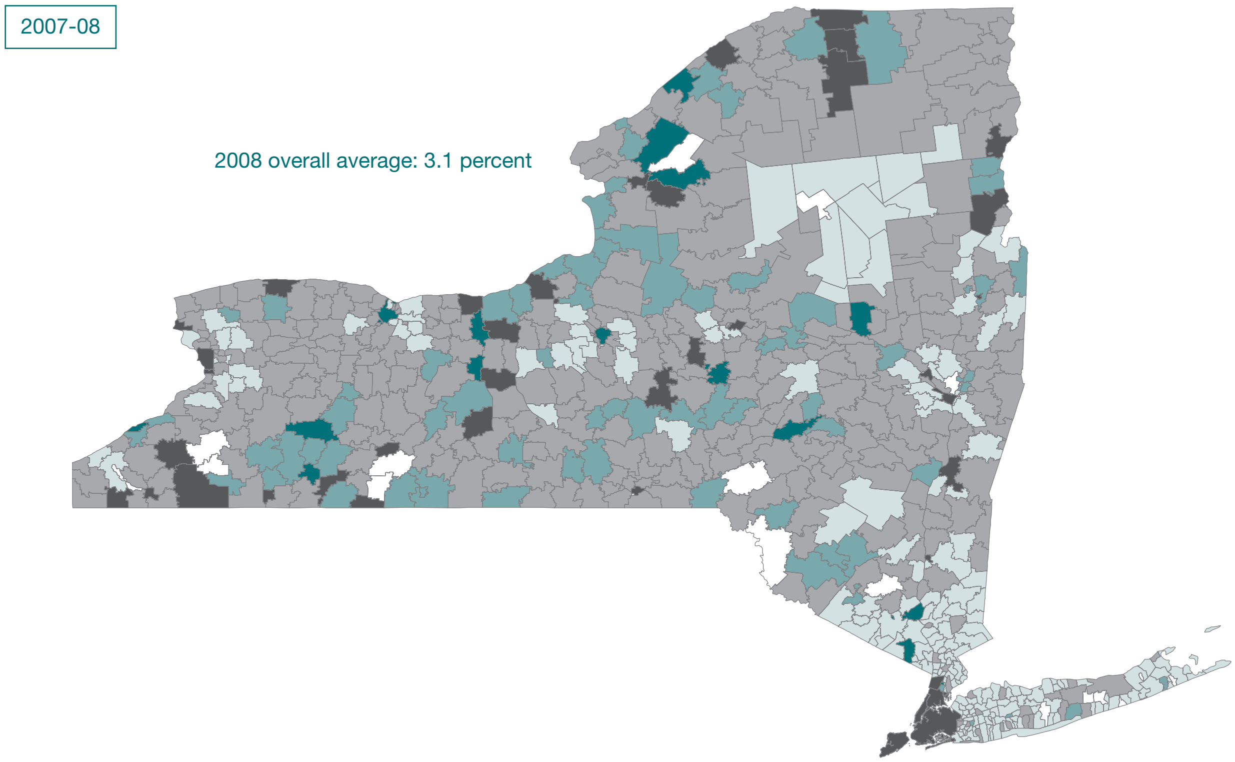
[8] Note that while some of the percentage shifts are negative, they are small and never statistically different from zero.

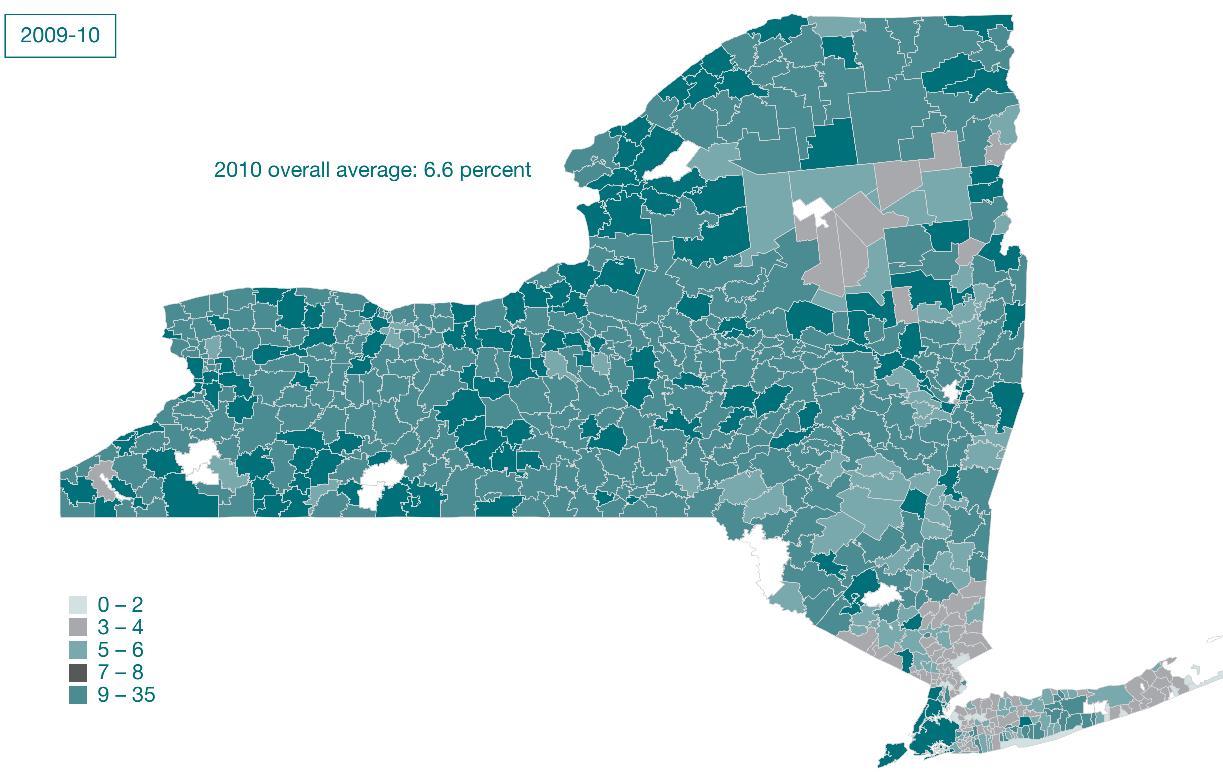
[9] Note that it is not inconsistent that relative to corresponding pre-existing trends, several noninstructional expenditure categories shifted downward but the overall expenditure did not. This is because these shifts are relative to the corresponding variables' pre-existing trends, which, in turn, differed between variables. Additionally, we do see a positive change in instructional expenditure in 2009-10, although it is not statistically significant. Instructional expenditure plays a much larger role in total expenditure than most of the noninstructional components, so when considering the overall effect, we cannot treat the subcomponents equally.

It is worth thinking about why spending in multiple noninstructional categories showed declines (relative to trend), although total expenditure was maintained on trend. This is likely because school districts anticipated future declines in funding and expenditure. Revenues from state and local funding sources declined drastically because of the Great Recession, and the primary reason that school districts' overall funding was maintained on trend was the influx of the federal stimulus aid from ARRA funding. It was widely known that the stimulus funding was temporary and would dry up in a couple of years (which it did). Thus, it is plausible that districts anticipated sharp funding cuts in the near future and responded by cutting spending in nonessential noninstructional categories.

Exhibit 2

Percentage of District Revenue from Federal Sources





Sources: New York Office of the State Comptroller; authors' calculations.

Table 3

Composition of Expenditures during the Financial Crisis and the Federal Stimulus Period

Panel A	Instructional Spending per Pupil	Instructional Support Spending per Pupil	Student Services Spending per Pupil
Percentage shift in 2008-09	-0.245	-0.109	-1.091
Percentage shift in 2009-10	1.131	-0.785	-0.980
Pre-recession base	11,064.65	886.47	652.02
Trend	334.9***	28.8***	17.2***
	(59.6)	(3.2)	(4.7)
Recession	-27.2	-1.0	-7.1
	(123.2)	(8.7)	(12.1)
Stimulus	152.3	-6.0	0.7
	(163.0)	(14.5)	(13.6)
Observations	4,146	4,146	4,146
R^2	0.92	0.88	0.91

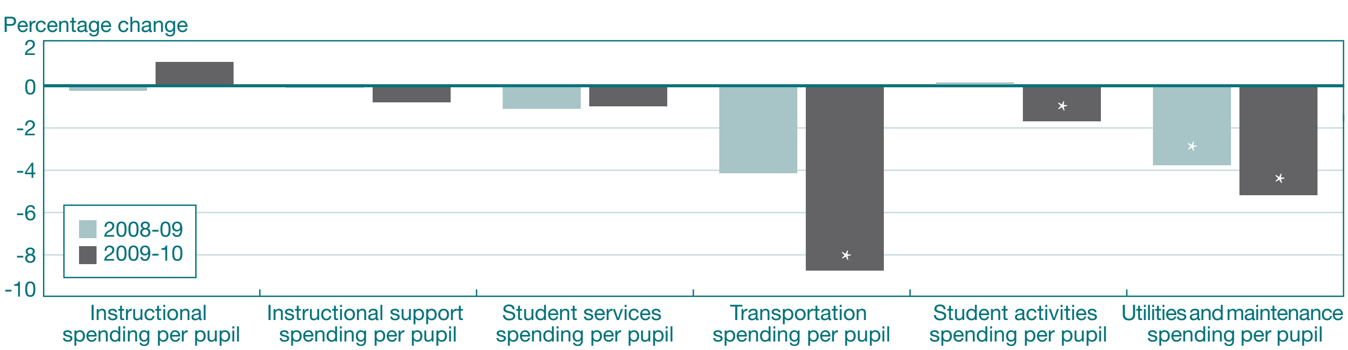
Panel B	Transportation Spending per Pupil	Student Activities Spending per Pupil	Utilities and Maintenance Spending per Pupil
Percentage shift in 2008-09	-4.130	0.151	-3.760**
Percentage shift in 2009-10	-8.753**	-1.676*	-5.188**
Pre-recession base	1,198.24	264.17	5,692.08
Trend	76.9***	9.7***	272.3***
	(20.3)	(0.6)	(63.6)
Recession	-49.5	0.4	-214.0***
	(43.5)	(1.8)	(98.9)
Stimulus	-55.4	-4.8**	-81.2
	(46.8)	(2.0)	(117.9)
Observations	4,146	4,146	4,146
R^2	0.83	0.96	0.95

Notes: Robust standard errors are in parentheses. All regressions include school district fixed effects and control for racial composition and percentage of students eligible for free or reduced-price lunch. The pre-recession base is expressed in 2009 constant dollars.

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

Chart 4

Changes in Expenditures during the Financial Crisis and the Federal Stimulus Period



Note: Asterisk (*) denotes significance at the 10, 5, or 1 percent level.

Table 4

Expenditures by School District Poverty Status

Panel A	Instructional Spending per Pupil			Instructional Support Spending per Pupil			Student Services Spending per Pupil		
	High	Medium	Low	High	Medium	Low	High	Medium	Low
Percentage shift in 2008-09	2.663	0.255	-3.990*	-0.164	0.879	-2.308	0.062	1.554*	-5.020
Percentage shift in 2009-10	5.126	1.940**	-3.954	-1.149	0.918	-4.332	-2.128	2.371*	-3.781
Pre-recession base	11,341.13	9,390.48	13,902.44	924.67	863.43	890.95	719.85	516.88	835.32
Trend	455.4***	214.3***	448.8**	26.2***	29.0***	32.3***	37.1***	10.5***	21.3**
	(106.3)	(17.8)	(188.9)	(5.3)	(2.9)	(8.4)	(13.2)	(1.4)	(10.1)
Recession	302.0	23.9	-554.6*	-1.5	7.6	-20.6	0.4	8.0*	-41.9
	(316.7)	(48.6)	(335.0)	(16.7)	(9.0)	(24.4)	(36.0)	(4.5)	(29.2)
Stimulus	279.3	158.2**	5.0	-9.1	0.3	-18.0	-15.8	4.2	10.3
	(474.8)	(71.0)	(395.0)	(17.5)	(10.8)	(43.7)	(46.3)	(5.5)	(22.7)
Observations	1,059	2,010	1,077	1,059	2,010	1,077	1,059	2,010	1,077
R^2	0.86	0.95	0.94	0.85	0.91	0.88	0.88	0.96	0.95

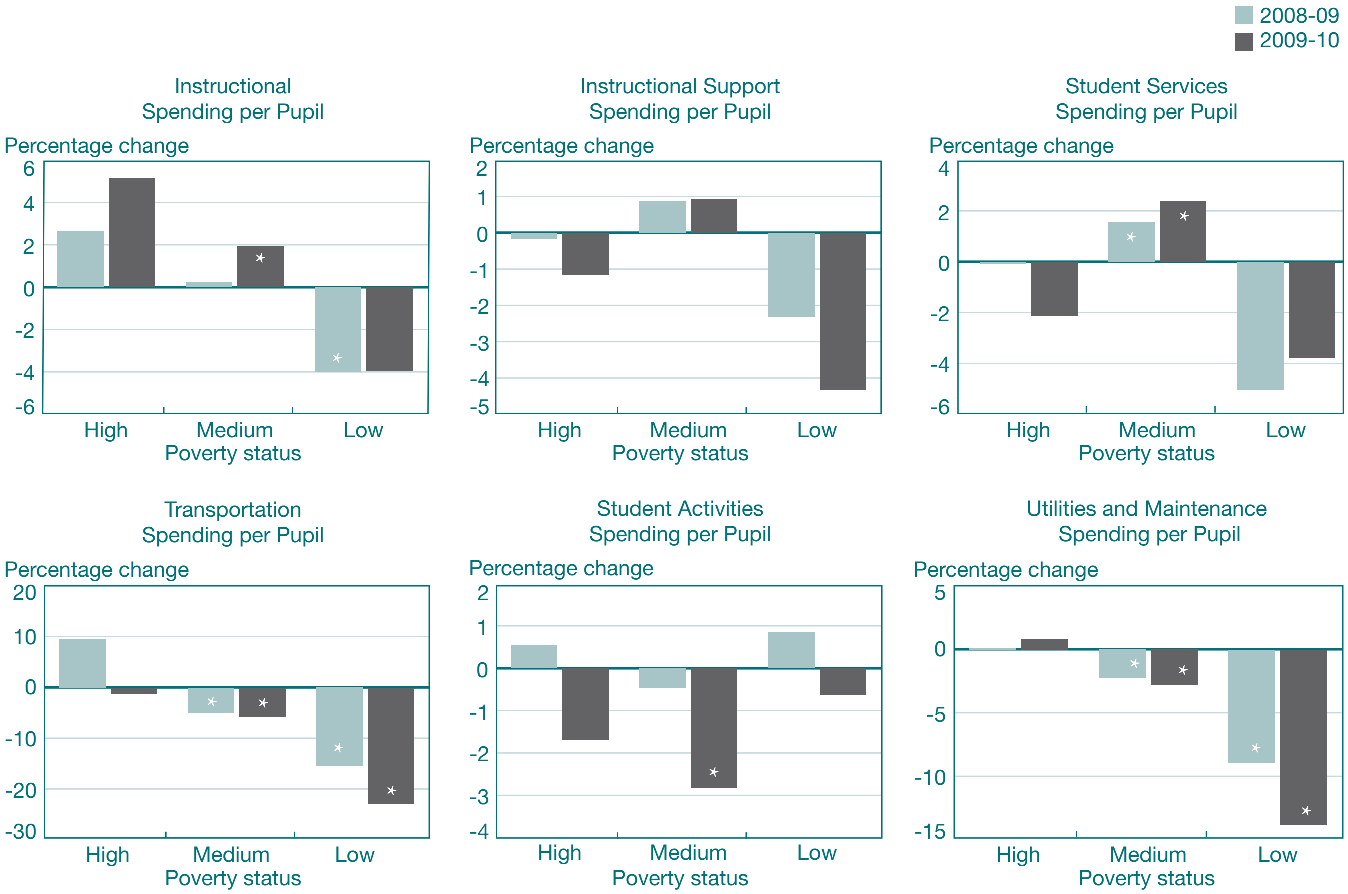
Panel B	Transportation Spending per Pupil			Student Activities Spending per Pupil			Utilities and Maintenance Spending per Pupil		
	High	Medium	Low	High	Medium	Low	High	Medium	Low
Percentage shift in 2008-09	9.609	-4.884***	-15.356***	0.554	-0.468	0.858	0.138	-2.265***	-8.919**
Percentage shift in 2009-10	-1.169	-5.676***	-22.873**	-1.684	-2.820**	-0.628	0.854	-2.756**	-13.804**
Pre-recession base	1,119.58	1,108.20	1,444.92	218.07	262.53	313.60	5703.77	4715.30	7498.22
Trend	62.6**	44.4***	144.6**	7.1***	10.9***	10.5***	247.3***	180.5***	444.2**
	(27.0)	(3.4)	(64.7)	(1.5)	(0.7)	(1.3)	(40.8)	(14.7)	(209.1)
Recession	107.6	-54.1***	-221.9***	1.2	-1.2	2.7	7.9	-106.8***	-668.8**
	(144.7)	(11.1)	(79.7)	(3.5)	(2.4)	(4.2)	(117.0)	(38.7)	(340.2)
Stimulus	-120.7	-8.8	-108.6	-4.9	-6.2**	-4.7	40.9	-23.1	-366.2
	(145.8)	(11.8)	(103.3)	(3.4)	(2.6)	(4.6)	(178.1)	(51.6)	(422.0)
Observations	1,059	2,010	1,077	1,059	2,010	1,077	1,059	2,010	1,077
R^2	0.66	0.93	0.89	0.93	0.95	0.97	0.95	0.96	0.95

Notes: Robust standard errors are in parentheses. All regressions include school district fixed effects and control for racial composition and percentage of students eligible for free or reduced-price lunch. The pre-recession base is expressed in 2009 constant dollars.

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

Chart 5

Changes in Expenditures by School District Poverty Status



Note: Asterisk (*) denotes significance at the 10, 5, or 1 percent level.

To summarize, high-poverty districts were relatively unaffected and did not see a statistically significant change in any expenditure category. Districts in the middle had mixed experiences, boosting instructional expenditure and student services overall while cutting spending for transportation, student activities, and utilities. Low-poverty districts were the most affected, experiencing economically significant declines in all categories, many of which were statistically significant.

6.3 Did Urban Status Matter?

There were marked differences in how school finances in urban, suburban, and rural districts were affected by the Great Recession. As Table 5 and Chart 6 show, all district types maintained instructional spending: while some of the shifts were negative, they were never statistically different from zero. Additionally, there were no statistically significant shifts in instructional support or student services. Transportation spending fell by a large and statistically significant amount in both urban and rural districts for both years but did not change significantly for suburban districts. Spending on utilities fell in both years in urban and rural districts (although only the 2009 decrease was significant in rural districts).

Table 5

Expenditures by School District Urban Status

Panel A	Instructional Spending per Pupil			Instructional Support Spending per Pupil			Student Services Spending per Pupil		
	Urban	Suburban	Rural	Urban	Suburban	Rural	Urban	Suburban	Rural
Percentage shift in 2008-09	-1.377	0.747	-1.115	0.668	0.426	-1.002	1.501	-2.187	-1.022
Percentage shift in 2009-10	-3.305	3.701	0.855	0.757	-0.871	-1.497	-0.033	-2.619	1.661
Pre-recession base	9,617.69	1,2031.65	10,855.3	7,95.1	808.81	991.42	468.94	826.95	584.16
Trend	189.8*	378.1***	376.9***	22.3***	24.0***	37.1***	10.5***	22.7**	19.7***
	(105.8)	(101.2)	(63.3)	(4.6)	(4.2)	(5.7)	(2.3)	(9.8)	(5.6)
Recession	-132.4	89.9	-121	5.3	3.4	-9.9	7.0	-18.1	-6
	(174.3)	(224.9)	(165.1)	(13.0)	(11.2)	(16.9)	(6.3)	(23.3)	(17.7)
Stimulus	-185.4	355.4	213.8	0.7	-10.5	-4.9	-7.2	-3.6	15.7
	(173.9)	(325.6)	(215.0)	(13.6)	(13.5)	(28.9)	(7.1)	(20.7)	(24.1)
Observations	797	1,511	1,831	797	1,511	1,831	797	1,511	1,831
R^2	0.94	0.91	0.94	0.84	0.86	0.89	0.92	0.95	0.87

Panel B	Transportation Spending per Pupil			Student Activities Spending per Pupil			Utilities and Maintenance Spending per Pupil		
	Urban	Suburban	Rural	Urban	Suburban	Rural	Urban	Suburban	Rural
Percentage shift in 2008-09	-6.541***	6.899	-10.224***	-1.445	0.385	0.26	-4.217**	-2.364	-4.823*
Percentage shift in 2009-10	-10.534***	-3.596	-9.773**	-3.667**	-0.634	-2.21	-7.489**	-1.416	-5.754
Pre-recession base	891.03	1,100.15	1,416.62	231.47	261.93	279.55	4,727.63	5,598.72	6,180.89
Trend	34.6***	71.2**	90.8***	9.3***	6.7***	12.3***	98.4	183.0***	325.0***
	(5.4)	(29.6)	(22.3)	(0.9)	(1.0)	(1.0)	(69.3)	(52.2)	(70.0)
Recession	-58.3***	75.9	-144.8***	-3.3	1.0	0.7	-199.4**	-132.3	-298.1*
	(15.4)	(99.3)	(44.1)	(2.8)	(2.4)	(3.2)	(87.9)	(104.5)	(175.2)
Stimulus	-35.6**	-115.5	6.4	-5.1*	-2.7	-6.9**	-154.7	53	-57.5
	(15.9)	(109.6)	(52.9)	(2.9)	(2.9)	(3.5)	(110.5)	(143.6)	(207.6)
Observations	797	1,511	1,831	797	1,511	1,831	797	1,511	1,831
R^2	0.91	0.63	0.90	0.97	0.97	0.94	0.96	0.97	0.95

Notes: Robust standard errors are in parentheses. All regressions include school district fixed effects and control for racial composition and percentage of students eligible for free or reduced-price lunch. The pre-recession base is expressed in 2009 constant dollars.

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

Chart 6
Changes in Expenditures by School District Urban Status

Note: Asterisk (*) denotes significance at the 10, 5, or 1 percent level.

Urban school districts additionally saw a drop in student activities expenditure in both years (significant only in the latter year). Overall, urban and rural districts experienced stronger declines in noninstructional spending than did suburban districts.

6.4 Examining Spatial Heterogeneities—Were There Variations across Metropolitan Areas?

Next, we investigate whether there were variations in experience across metropolitan areas. The results are presented in Tables 6 and 7 and Charts 7 and 8. All metro areas maintained or increased instructional spending except Nassau, where instructional spending shifted downward. However, while almost all metro areas fared well in terms of instructional expenditures, they saw significant declines in various noninstructional categories. All metro areas experienced economically significant declines in transportation expenditure, and most of these declines were statistically significant. Nassau was particularly hard-hit in noninstructional expenditure as well. It experienced the largest decline in transportation and utilities spending in both years among any of the seven metro areas analyzed. Its expenditures on student activities and student services saw a small and insignificant increase.

After Nassau, New York City was the area that experienced the biggest declines in some noninstructional expenditure categories, particularly student activities and utilities. New York City also experienced a small (statistically insignificant) decline in instructional expenditure in 2009. In 2010, while New York City

Table 6

Expenditures by Metropolitan Area: Albany, Buffalo, New York City, and Syracuse

Panel A	Instructional Spending per Pupil				Instructional Support Spending per Pupil				Student Services Spending per Pupil			
	Albany	Buffalo	NYC	Syracuse	Albany	Buffalo	NYC	Suburban	Albany	Buffalo	NYC	Syracuse
Percentage shift in 2008-09	-0.174	2.050	-0.256	0.506	1.879	-0.472	0.036	-0.292	-0.001	1.420	-0.874	0.252
Percentage shift in 2009-10	3.177	3.861***	2.181	3.595*	-0.528	-1.914	-0.513	-1.014	0.053	3.284	-2.074	0.770
Pre-recession base	9,934.55	8,117.35	12,756.70	8,182.69	728.39	746.29	924.59	833.33	510.45	450.52	850.07	414.48
Trend	131.7	99.8***	190.9***	165.5***	22.2***	21.2***	11.8	17.5***	10.3*	6.0**	36.0***	7.1*
	(110.5)	(25.1)	(44.1)	(39.7)	(7.5)	(4.7)	(7.9)	(5.8)	(5.5)	(2.3)	(6.1)	(3.9)
Recession	-17.3	166.4	-32.7	41.4	13.7	-3.5	0.3	-2.4	-0.0	6.4	-7.4	1.0
	(182.7)	(104.8)	(140.9)	(119.6)	(17.3)	(15.8)	(23.2)	(20.4)	(11.4)	(6.5)	(14.2)	(10.8)
Stimulus	332.9	147.0	310.9	252.7*	-17.5	-10.8	-5.1	-6.0	0.3	8.4	-10.2	2.1
	(257.9)	(111.5)	(228.7)	(138.7)	(20.6)	(16.8)	(33.4)	(20.5)	(11.5)	(7.2)	(14.4)	(10.7)
Observations	372	252	335	257	372	252	335	257	372	252	335	257
R^2	0.97	0.91	0.97	0.89	0.78	0.85	0.88	0.93	0.95	0.92	0.92	0.83

Panel B	Transportation Spending per Pupil				Student Activities Spending per Pupil				Utilities and Maintenance Spending per Pupil			
	Albany	Buffalo	NYC	Syracuse	Albany	Buffalo	NYC	Suburban	Albany	Buffalo	NYC	Syracuse
Percentage shift in 2008-09	-4.382	-4.865**	-1.485	-1.508	1.017	-0.271	-2.110	3.297*	-2.631	-0.702	-5.250*	-1.896
Percentage shift in 2009-10	-7.410**	-7.065***	-4.938**	-4.774	-2.058	-2.281	-6.071***	-1.635	1.999	-0.671	-5.226	-2.361
Pre-recession base	1,117.31	957.95	1,260.53	997.57	189.50	197.11	340.14	255.76	5,552.79	3,937.37	5,976.93	3,859.24
Trend	46.4***	27.0***	29.7***	38.7***	6.7***	7.1***	13.2***	10.3***	87.8	75.8**	147.0***	119.1***
	(9.3)	(8.1)	(7.4)	(8.4)	(0.9)	(1.2)	(2.3)	(1.5)	(126.5)	(33.0)	(50.2)	(24.1)
Recession	-49.0	-46.6**	-18.7	-15.0	1.9	-0.5	-7.2	8.4*	-146.1	-27.6	-313.8*	-73.2
	(32.8)	(20.0)	(18.0)	(27.1)	(3.3)	(3.2)	(5.2)	(5.0)	(186.9)	(65.7)	(176.4)	(75.1)
Stimulus	-33.8	-21.1	-43.5*	-32.6	-5.8	-4.0	-13.5**	-12.6**	257.0	1.2	1.4	-17.9
	(31.0)	(15.2)	(23.9)	(25.6)	(3.8)	(3.4)	(5.5)	(5.8)	(305.2)	(64.6)	(244.1)	(78.6)
Observations	372	252	335	257	372	252	335	257	372	252	335	257
R^2	0.9	0.88	0.98	0.8	0.96	0.96	0.97	0.95	0.98	0.93	0.96	0.95

Notes: Robust standard errors are in parentheses. All regressions include school district fixed effects and control for racial composition and percentage of students eligible for free or reduced-price lunch. The pre-recession base is expressed in 2009 constant dollars.

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

Table 7

Expenditures by Metropolitan Area: Ithaca, Nassau, and Rochester

Panel A	Instructional Spending per Pupil			Instructional Support Spending per Pupil			Student Services Spending per Pupil		
	Ithaca	Nassau	Rochester	Ithaca	Nassau	Rochester	Ithaca	Nassau	Rochester
Percentage shift in 2008-09	1.243	-4.460	1.436	3.430	-3.576	3.808	1.577	0.108	-1.058
Percentage shift in 2009-10	3.109*	-8.322*	3.277**	1.518	-3.388	8.985**	3.666	0.772	-2.771
Pre-recession base	8,395.64	15,971.95	8,438.53	9,15.59	958.42	868.50	417.94	848.25	479.09
Trend	287.2***	672.0***	145.0***	39.2***	46.1***	25.5***	14.2***	26.9**	15.1***
	(35.0)	(259.0)	(26.2)	(7.1)	(13.6)	(8.0)	(2.6)	(10.7)	(3.2)
Recession	104.4	-712.4	121.1	31.4	-34.3	33.1	6.6	0.9	-5.1
	(115.8)	(462.1)	(92.6)	(24.2)	(35.1)	(24.5)	(7.9)	(23.4)	(10.2)
Stimulus	156.7	-616.7	155.4	-17.5	1.8	45.0*	8.7	5.6	-8.2
	(145.8)	(631.9)	(99.4)	(27.2)	(72.0)	(25.9)	(10.7)	(31.4)	(9.9)
Observations	252	703	348	252	703	348	252	703	348
R^2	0.84	0.94	0.87	0.89	0.88	0.81	0.87	0.90	0.88

Panel B	Transportation Spending per Pupil			Student Activities Spending per Pupil			Utilities and Maintenance Spending per Pupil		
	Ithaca	Nassau	Rochester	Ithaca	Nassau	Rochester	Ithaca	Nassau	Rochester
Percentage shift in 2008-09	-5.232**	-18.612***	-3.011	2.726	1.919	-0.200	-1.965	-9.614*	-0.688
Percentage shift in 2009-10	-9.149***	-29.410**	-4.272	0.235	1.232	-0.123	-3.829*	-18.770*	1.224
Pre-recession base	875.37	1,780.46	960.25	255.43	323.51	256.68	4,294.35	8,790.90	4,193.63
Trend	48.9***	215.2**	30.3***	12.0***	11.1***	13.0***	214.0***	605.8**	132.9***
	(7.1)	(89.8)	(6.2)	(1.6)	(1.9)	(1.4)	(23.0)	(294.2)	(16.2)
Recession	-45.8**	-331.4***	-28.9	7.0	6.2	-0.5	-84.4	-845.2*	-28.9
	(22.3)	(125.5)	(23.6)	(5.0)	(6.2)	(3.8)	(79.6)	(504.4)	(50.2)
Stimulus	-34.3	-192.3	-12.1	-6.4	-2.2	0.2	-80.1	-804.8	80.2
	(23.6)	(178.1)	(23.0)	(5.4)	(7.3)	(4.7)	(87.7)	(684.3)	(63.3)
Observations	252	703	348	252	703	348	252	703	348
R^2	0.88	0.89	0.78	0.93	0.96	0.93	0.85	0.94	0.94

Notes: Robust standard errors are in parentheses. All regressions include school district fixed effects and control for racial composition and percentage of students eligible for free or reduced-price lunch. The pre-recession base is expressed in 2009 constant dollars.

* Significant at the 10 percent level.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Chart 7

Changes in Expenditures by Metropolitan Area: Albany, Buffalo, New York City, and Syracuse

Note: Asterisk (*) denotes significance at the 10, 5, or 1 percent level.

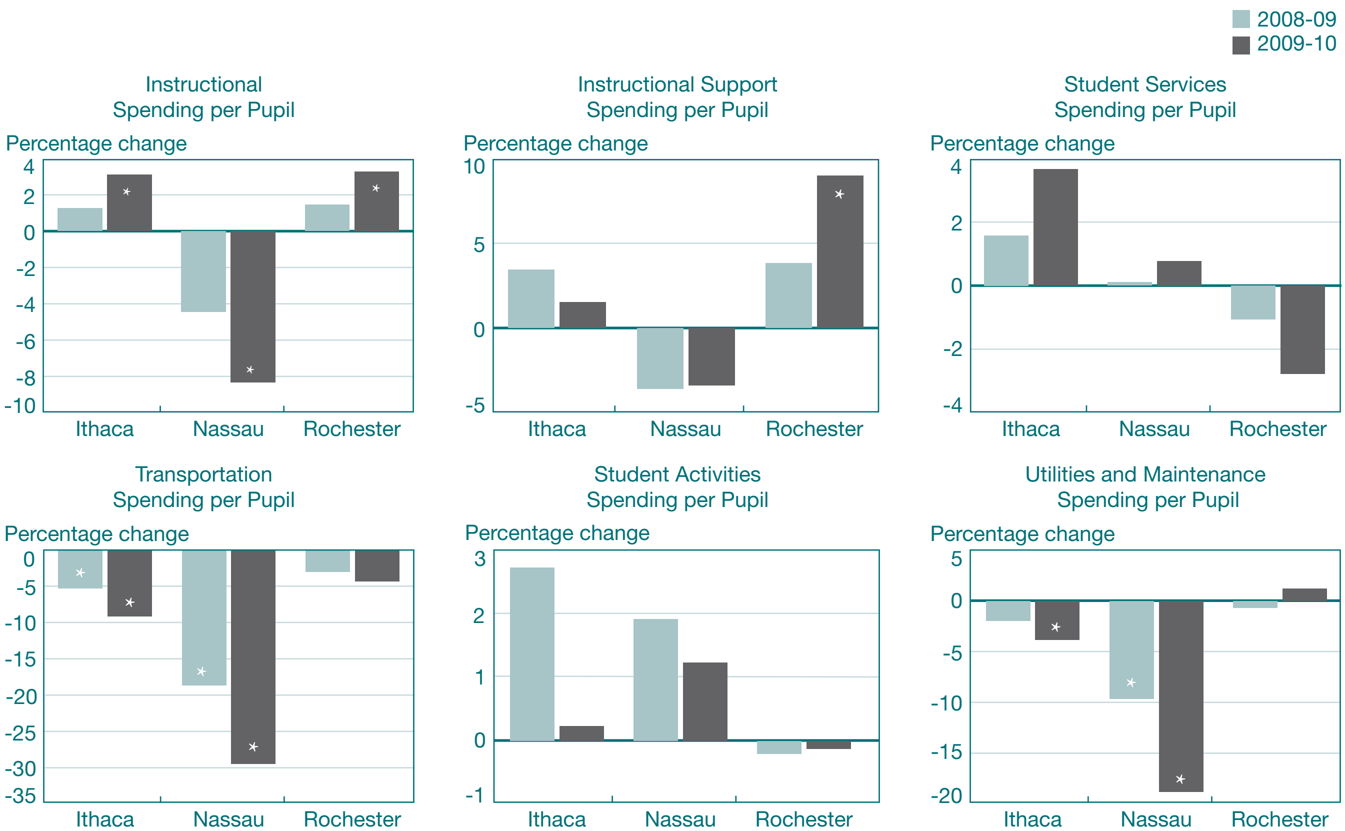
saw an increase in instructional expenditure relative to trend like the other metro areas, this increase was not as large as that experienced by most other metro areas and was not statistically different from zero. Rochester fared relatively well, with a modest (but statistically significant) increase in instructional spending and a significant (both economically and statistically) increase in instructional support in 2010. Rochester did not experience a statistically significant decline in any other noninstructional expenditure category.

7. Conclusion

This article investigates school finance patterns in New York State during the Great Recession and federal stimulus period using a trend shift analysis. We do not find evidence of shifts in total school district funding or expenditure following the Great Recession. However, the composition of funding changed: the share of federal funding increased dramatically, while shares of state and local funding fell when ARRA funding began. The federal stimulus appears to have helped maintain total expenditure and instructional expenditures in the 2009-10 school year. While total expenditure did not show a shift, the composition of total expenditure

Chart 8

Changes in Expenditures by Metropolitan Area: Ithaca, Nassau, and Rochester



Note: Asterisk (*) denotes significance at the 10, 5, or 1 percent level.

changed in interesting ways. Instructional expenditure was maintained on trend, while declines occurred (relative to trend) in noninstructional expenditures, especially in transportation, utilities, and student activities. Thus, districts seem to have protected the expenditures that matter most for student learning, while expenditures in noninstructional categories suffered. In addition to these overall trends, our analysis reveals interesting variations within the state by poverty level, metro area, and urban status. Studying variations by poverty level, we find that low-poverty districts were the most affected in both instructional and noninstructional expenditures. Studying patterns by metro area reveals that New York City, and especially Nassau, were badly hit. Additionally, urban districts suffered the largest declines in funding.

Investing in education is essential to building human capital and improving children's prospects. Recessions can have widespread and long-lasting effects in many aspects of life, far beyond the immediate short-term impact. How, exactly, the recession will affect the economy in the long run remains to be seen, but its impact on human capital development and investment will surely figure importantly in that outcome. The findings of this study should deepen our understanding of how recessions affect schools and the role policy can play in mitigating the consequences.

References

Baker, B. D. 2009. "Within-District Resource Allocation and the Marginal Costs of Providing Equal Educational Opportunity: Evidence from Texas and Ohio." Education Policy Analysis Archives 17, no. 3 (February): 1–31.

Cavanagh, S. 2011. "Educators Regroup in Recession's Aftermath." Education Week, January 13.

Chakrabarti, R., and S. Sutherland. 2013. "Precarious Slopes? The Great Recession, Federal Stimulus, and New Jersey Schools." Federal Reserve Bank of New York Economic Policy Review 19, no. 2: 41-65.

Dye, R., and A. Reschovsky. 2008. "Property Tax Responses to State Aid Cuts in the Recent Fiscal Crisis." Public Budgeting and Finance 28, no. 2 (Summer): 87-111.

Goertz, M., S. Loeb, and J. Wyckoff. 2011. "Recruiting, Evaluating, and Retaining Teachers: The Children First Strategy to Improve New York City's Teachers." In J. O'Day, C. Bitter, and L. Gomez, eds., Education Reform in New York City: Ambitious Change in the Nation's Most Complex School System. Cambridge, Massachusetts: Harvard Education Press.

Klein, J. 2010. "May 6, 2010, Memo [from Joel Klein, New York City Schools Chancellor] to New York City School Principals."

Rubenstein, R., A. E. Schwartz, L. Stiefel, and H. Amor. 2007. "From Districts to Schools: The Distribution of Resources across Schools in Big City School Districts." Economics of Education Review 26, no. 5: 532-45.

State Education Department, Office of State Aid. 2009. "2009-10 State Aid Handbook." Albany, New York.

Stiefel, L., and A. E. Schwartz. 2011. "Financing K-12 Education in the Bloomberg Years, 2002-2008." In J. O'Day, C. Bitter, and L. Gomez, eds., Education Reform in New York City: Ambitious Change in the Nation's Most Complex School System. Cambridge, Massachusetts: Harvard Education Press.

Federal Reserve Bank of New York
Economic Policy Review

Library of Congress Card Number: 77-646559