

Federal Reserve Bank of New York
Staff Reports

Assessing the Quality of “Furfine-based” Algorithms

Olivier Armantier
Adam Copeland

Staff Report No. 575
October 2012



This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

Assessing the Quality of “Furfine-based” Algorithms

Olivier Armantier and Adam Copeland

Federal Reserve Bank of New York Staff Reports, no. 575

October 2012

JEL classification: G10, C81

Abstract

To conduct academic research on the federal funds (fed funds) market, historically one of the most important financial markets in the U.S., some empirical economists have used market level measures published by the Markets Group at the Federal Reserve Bank of New York (FRBNY). To obtain more disaggregate data, some researchers have relied on a separate source of information: individual transactions inferred indirectly from an algorithm based on the work of Furfine (1999). To date, however, the accuracy of this algorithm has not been formally established. In this paper, we conduct a test aimed at assessing the ability of the algorithm to identify correctly individual overnight fed funds transactions conducted by two banks, which are among the most active in the fed funds market. The results are discouraging. We estimate the average type I and type II errors from 2007 to 2011 to be 81% and 23%, respectively. Furthermore, we argue that these errors i) apply to almost half of the algorithm’s output, ii) introduce systematic biases, and iii) may not subside when the algorithm’s output is aggregated. Our results therefore raise serious concerns about the appropriateness of using the algorithm’s output to study the fed funds market. Because the FRBNY Markets Group relies on a different source of data than the algorithm output, our results have no bearing on their understanding of the fed funds market and their calculation of market level measures, including the effective fed funds rate.

Key words: federal funds market, data quality

Armantier, Copeland: Federal Reserve Bank of New York (e-mail: olivier.armantier@ny.frb.org, adam.copeland@ny.frb.org). The authors thank Gara Afonso, Todd Keister, Anna Kovner, Antoine Martin, Jamie McAndrews, David Skeie, and James Vickery for their help on this project, as well as Sha Lu and Michael Walker for valuable research assistance. The authors also thank seminar participants at the Federal Reserve Bank of New York. The views expressed in this paper 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.

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 the Federal Reserve.¹ 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 fund market has been a key financial market with major macroeconomic and monetary policy implications. In particular, the average rate on the fed funds market, 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 set by the Federal Open Market Committee (FOMC).

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 Markets Group at 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 statistics, in particular the effective fed funds rate, are used widely by policy makers, financial market participants, and researchers in academia.

An alternative source of data, used exclusively by academic researchers, is inferred from an algorithm based on the work of Furfine (1999). Although there are now different versions of this algorithm, they all seem to rely on the same principles. As further explained below, we are only concerned with assessing the accuracy of these Furfine-based algorithms. The results presented in this paper do not extend to the traditional source of data collected by the FRBNY Markets Group. In particular, the results in this paper have no bearing on the ability of the Markets Group to understand the fed funds market and to calculate accurately market level measures, including the effective fed funds rate.

The Money and Payments Studies function (MaPS) in the Research group at the FRBNY uses a version of this algorithm to conduct some of its academic research. This MaPS algorithm exploits the fact that privately traded fed funds transactions are often settled over the “Fedwire funds payment system” (Fedwire), the large value real time settlement system operated by the Federal Reserve.² As further explained in Section 2, the MaPS algorithm searches over all

¹ Although other forms of short term interbank lending may be informally referred to as “fed funds,” we are solely concerned in this paper with loans of reserves between eligible counterparties as officially defined as fed funds 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, Preslowsky, and Whitesell (2006), Afonso, Kovner, and Schoar (2011), and Afonso and Lagos (2012a, 2012b).

² Fed funds transactions can be settled over Fedwire, possibly over CHIPS (another high value payments settlement system), or conducted on a bank’s books. Based on 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. To the best of our knowledge, however, the exact extent to which Fed funds are primarily settled over Fedwire has not been established formally.

payments sent over Fedwire to identify the pairs of payments which 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 an algorithm such as the one used by MaPS correctly identifies fed funds transactions, then its output would be extremely useful to academic economists to study what has been one of the most important financial markets in the U.S. Indeed, it would provide data at the lowest level of aggregation (i.e. individual transactions between specific pairs of banks) that could help understand the underpinnings of the U.S. fed funds market. The algorithm’s output is especially attractive when trying to explain whether and why the fed funds market failed during the 2008-2009 financial crisis, as well as the specific role played by individual banks. In fact, a number of academic papers have been recently written using the algorithm’s output, some of which have been published in top-ranked journals.³

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

As further explained in Section 3, the basic methodology underlying the test may be explained 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 banks, among the more active 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 MaPS 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.

The outcome of the test is discouraging. We estimate that in the first quarter of 2007, the type I and type II errors produced by the algorithm are 64% and 24%, respectively. In other words,

³ The following papers use a version of the MaPS 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 (2009), Acharya and Skeie (2011), Ashcraft, McAndrews, Skeie (2011), Bech, Bergstrom, Garratt, and Rosvall (2011) Afonso, Kovner, Schoar (2011, 2012) Afonso and Lagos (2012a, 2012b), Armantier, Ghysels, Sarkar, and Shrader (2012). 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.

64% of all pairs of payments identified by the algorithm are not fed funds transactions conducted by the two banks. This negative result seems to be robust to the time period considered. Going forward to the first quarter of 2011, the type I error is estimated to be 93%, while the type II error is estimated to be 17%. Although our results may not extend to every bank, we argue 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 outputted by the algorithm over 2007 to 2011. Second, if we assume that our 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 outputted by the algorithm. Consequently, our results cast substantial doubt about the ability of the algorithm to produce transaction level measures that characterize accurately and comprehensively the fed funds market.

There is an additional, perhaps insurmountable, problem with the MaPS algorithm: Even if it could correctly find every fed funds transaction, there is no guarantee that the algorithm 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 clients 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.

While our work focuses on the MaPS algorithm, slightly different versions of this algorithm are used by researchers outside the FRBNY. Economists at the Board of Governors of the Federal Reserve (FRB) use the same proprietary Fedwire data and a similar algorithm to create measures of overnight fed funds activity. We therefore expect the FRB algorithm's output to suffer from the same large type I and type II errors.⁵ Beyond fed funds, researchers have used algorithms based on Furfine (1999) to construct estimates of unsecured interbank lending. For instance, a similar algorithm was applied to Canadian and U.K. data to identify overnight loans.⁶ Further, Kuo, Skeie, and Vickery (2012) have expanded the MaPS algorithm to identify loans with maturities longer than overnight. To the best of our knowledge, however, no publically available document exists in which the ability of these algorithms to correctly identify interbank loans is assessed formally.

⁴ E.g., foreign banks often have nostro accounts at domestic banks.

⁵ A paper that uses the FRB algorithm's output is Demiralp, Preslopsky, and Whitesell (2006).

⁶ Hendry and Kamhi (2009), Allen, Chapman, Echenique, and Shum (2012), and Allen, Hortacsu, and Kastl (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 UK payments data.

The test conducted in this paper only demonstrates the general inability of the MaPS 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. We do not formally test this hypothesis, but in the conclusion we provide separate preliminary evidence that the algorithm may characterize accurately the bank-level fed funds activity of some Government Sponsored Entities. It is also possible that when the output of the MaPS algorithm is aggregated to the bank-to-bank, to the 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 the bank levels, and we provide 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 it may be too difficult for the MaPS algorithm to systematically recognize that a given pair of payments corresponds to an overnight interbank loan between two specific banks.

The remainder of the paper is structured as follows. In section 2, we describe the MaPS algorithm and we discuss the potential problems it may encounter when trying to identify individual fed funds transactions. In section 3, we present the methodology underlying our test and report the outcome of this test. We conclude in section 4 with a discussion of our results' implications.

2 The MaPS Algorithm

2.1 Background

The Fedwire Funds Service (Fedwire) is a real-time gross settlement system operated by the Federal Reserve. It enables depository institutions and certain financial institutions to make large value payments that are immediate and final.⁷ To initiate a transfer through Fedwire, a participant must fill electronically a number of fields specifying in particular the identity of the sending and receiving parties, and the amount sent.

While data from Fedwire are not publically 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 in any given day. However, we are only allowed to observe a subset of the message fields corresponding to the payment. Specifically, we observe the American Banking 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

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

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 which has been slightly adapted over the years by researchers at MaPS and FRB. The current algorithm used by the MaPS to produce some of its reports follows these general steps:

1. Transfers from or to a settlement institution (i.e. CHIPS, CLS, or DTC) are dropped because loans to or from these institutions are not considered fed funds loans as defined by Regulation D.
2. On a given 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 over \$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 day. 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 to the range $[\underline{l}, \bar{l}]$, where \underline{l} (respectively \bar{l}) is the minimum (respectively maximum) fed funds rate reported by the FRBNY Markets Group at date t minus (respectively plus) 50 basis points.^{9,10} If the implied interest rate is within the range $[\underline{l}, \bar{l}]$, 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 (i.e. $Y(x_{ijt}) = \emptyset$), then x_{ijt} is not considered part of an overnight loan. Second, if there is a unique matching return payment (i.e. $Y(x_{ijt})$ 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 candidates return payments (i.e. $\dim[Y(x_{ijt})] > 1$), then the algorithm first computes the median interest rate implied by all the candidate payments in $Y(x_{ijt})$. The algorithm then chooses the

⁸ This interest rate is equal to $((y_{jit+1} - x_{ijt}) / x_{ijt}) * (360)$, where 360 is used to annualize an overnight loan, per convention in the fed funds market.

⁹ Every day, the FRBNY Markets group conducts a survey of the four largest fed funds brokers. As mentioned in the introduction, using this source of data the FRBNY publishes the mean, standard deviation, minimum, and maximum interest rates of brokered fed funds transactions for the prior day.

¹⁰ Currently, the minimum bound on an interest rate is the maximum of 0.9 basis points and the minimum fed funds rate reported by the Markets group 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 Markets group minus 50 basis points. The absolute lower bound was pushed down from 1/32 to 0.009% because the extremely low nominal rates in recent times made interest rates below 1/32 plausible.

return leg of the overnight loan with an implied interest rate that is closest to the median rate from above.¹¹ 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} .¹²

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 classified as an overnight Eurodollars loan.¹³ Otherwise, the pair of transactions is classified as an overnight fed funds loan.¹⁴

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% 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 MaPS 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 it has a variety of uses, researchers in MaPS often use the algorithm’s output to calculate summary statistics which describe features of the fed funds markets (e.g. average rates, volumes) at the bank-to-bank, bank, and market levels.

2.2 Potential Problems

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

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 think of 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 discussed in the introduction, the algorithm cannot distinguish

¹¹ 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.

¹² 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 ordered of payments effects the algorithm’s output.

¹³ “CTR” stands for customer transfer, and is meant to designate that the beneficiary of the payment is not a bank.

¹⁴ The motivation for using the CTR business code to differentiate fed funds loans from Eurodollars loans is based on internal work by McAndrews (2009).

between a bank sending or receiving a payment on its own behalf versus doing so on behalf of a client. 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 the share of the total fed funds activity for which it accounts.

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 capture exclusively 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 fed funds. This payment activity could be picked up in the algorithm and incorrectly labeled as a fed funds transaction.

Fourth, the MaPS algorithm could identify a legitimate fed funds loan, but incorrectly link one of the two payments related to that transaction. Such an instance 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 (i.e. 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 which are not identified by the algorithm. There are at least two ways the constraints embedded in the algorithm could produce such errors. First, the algorithm imposes that the principal amount of fed funds loans must be greater than \$1 million and in an increment of \$100,000. Actual fed funds activity where 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 across banks, the plus/minus 50 basis point range around the minimum and maximum fed funds rate reported by the FRBNY Markets Group might rule out actual fed funds activity.

In addition to these systematic, potential problems with the MaPS algorithm, there are also idiosyncratic difficulties. 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

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

due to 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 MaPS 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 MaPS Algorithm.

3.1 The Test's Methodology

From the perspective of a given bank, each of its fed funds transaction 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, while the receive leg precedes the send leg when the bank purchases fed funds from a counterparty.

Every day, banks may send and receive a large number of Fedwire payments (more than 150,000 in some cases). To recognize automatically the receive leg of a fed funds transaction within the flow of Fedwire payments, some banks require their counterparties to incorporate a unique identifier into the message portion of the Fedwire payment. Two of these banks voluntarily gave us access to their unique identifiers. As a result, we can identify unambiguously the receive leg of every fed funds transaction the two banks have conducted by searching for the unique identifier within the message fields of every Fedwire payments they receive. 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 fund 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 MaPS algorithm identifies overnight fed funds transactions conducted by the two banks over Fedwire. To do so we consider every possible pairs of payments $\{x_{ijt}, y_{jit+1}\}$ on consecutive 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 fund loan. Because the presence of the unique identifier flags unambiguously which receive leg is and is 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 used to construct these estimates consists of three steps (see figure 1).

First, on a given day, we run the MaPS algorithm for the two banks. This gives us a list of paired payments, each consisting of two legs, a send 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 that day. This reference list therefore consists of receive legs identifying every fed funds payments the banks received that day. 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 figure 1, 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 of every pair in the algorithm list without a match in the reference list. Finally, the “false negative group” consists of every pair 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.¹⁶

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 for 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 they may seriously affect some characteristics of the fed funds transactions (e.g. the exact duration of the loan), in general it should not affect substantially the more important characteristics (i.e. 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 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

¹⁶ 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 loans conditional on the receiving leg being part of a fed funds loan.

payments corresponding to any fed funds transactions which includes 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 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. Although we cannot prove this hypothesis, we were able to find independent evidence to support it: a third bank confirmed that it includes the unique identifiers when sending a fed funds payment to the two banks on which we base our test.

3.2 Type I and Type II Errors

The results reported in table I are discouraging. In the first quarter of 2007, the type I error produced by the algorithm is estimated to be 64% (18,633 / 29,077). While much lower, the estimated type II error, at 24% (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 II).¹⁷ The type I error is estimated to be higher as we go forward in time, reaching 93% 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% in the first quarter of 2011.¹⁸ On average, the type I error is estimated to be 81.4% from 2007 to 2011, and the average type II error is estimated to be 23.0%.

[Insert table I and II here]

As discussed earlier, Type I errors may be the result of several factors (e.g. the algorithm matches two completely unrelated payments or identifies some form of a loan different from 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

¹⁷ Because of technical limitations, the furthest back we can go to test the algorithm is 2007.

¹⁸ We do not know why there are opposing trends in our estimates of the type I and type II error. 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 over \$1 million sent and received by these two banks over Fedwire is also roughly flat, except for a decline of 20% from the first quarter of 2008 to the first quarter of 2009.

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%, had been discarded by the algorithm as being Eurodollars.¹⁹ 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% of the 3,211 false negatives. These small overlooked fed funds transactions, however, account for only 0.07% 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 for 128 of the 3,211 (or 4%) false negatives. Fourth and finally, while 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. In the first quarter of 2007, 1,458, or 45%, of the transactions fall into this category.

3.3 Is the Output of the MaPS Algorithm Biased?

Given the high rates of type I and type II errors, it would appear that the algorithm's transaction level output is unfit 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 found the algorithm to produce similar biases in the first quarter of 2011.

We first examine the set of counterparties for both fed funds sold and purchased by the two banks in the first quarter of 2007.²⁰ 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.²¹ For both banks, only 3 of the top 10 counterparties in the algorithm list also appear in the top 10 counterparties in the reference list. When ranking counterparties by the total value of their transactions, for both banks we find that 5 of the top 10 counterparties in the algorithm list also appear in the equivalent top 10 counterparties in the reference list. This comparison illustrates the poor performance by the algorithm in correctly identifying the two banks most important counterparties.

¹⁹ 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% were in fact fed funds transactions. Consistent with McAndrews (2009), our results therefore support the hypothesis that the "CTR" business code is an effective way to distinguish Eurodollar from fed funds loans.

²⁰ Recall that neither the MaPS 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 be actually comparing the correspondent bank of the true counterparty to the fed funds loan.

²¹ 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 they conduct and one-half of their total value of fed funds activity.

We now turn to quantities. In the reference list, we observe the amount of the receive legs of the federal 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 payments linked by the algorithm. As illustrated in figure 2, 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 and \$72.5 million, versus \$50 and \$143.8 million in the algorithm list. Using the Mann-Whitney U test we can reject at the 1% 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. As illustrated in Appendix A2, similar biases are identified when we consider separately the amount of fed funds purchased and the amount of fed funds sold by the two banks.

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 reference list which can be matched to a receive leg in the algorithm list. We can then compare the inferred interest rates from this set of transactions to the inferred interest rates in the algorithm list. In figures 3 and 4, we plot the interest rate distributions for the fed funds sold and purchased by the two banks. Similar to the result of our analysis of amounts, we find the distributions of rates produced by the algorithm to be different from the distributions of rates of the true positives. In particular, the median rate of fed fund sold and purchased are respectively 537 and 519 basis points for the true positives, while the median rate of fed fund sold and purchased are respectively 525 and 523 basis points for the algorithm list.²² Using a Man Mann-Whitney U test we can reject at the 1% significance level the null hypothesis that these distributions of rates are equal (the Z-score is -26.3 for fed funds sold, and is -33.3 for fed funds purchased). Observe that the algorithm is biased upward for fed funds sold and downward for fed funds purchased. As a result, these biases partially offset each other when the interest rates of fed funds sold and purchased by the two banks are combined. Nevertheless, as illustrated in Appendix A2, there remain significant differences between the distributions of interest rates inferred from the algorithm and from true positives even when fed funds sold and purchased are combined.

These three comparisons provide statistical evidence that the set of counterparties as well as the distributions of amounts and interest rates inferred from the algorithm's output for our two banks are significantly biased. 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.

²² In the first quarter of 2007, the target fed funds rate was 525 basis points.

4 Discussion

Because the fed funds market has been one of the key financial markets in the U.S., it has attracted attention from researchers, especially after the 2008-2009 financial crisis. Empirical analyses of this market have typically relied on transactions inferred by an algorithm comparable to the one used by MaPS. There is no guarantee, however, that this algorithm correctly identifies individual fed funds transactions. In this paper, we reported on a test aimed at assessing the transaction level quality of the MaPS algorithm. For two large banks, among the more active on the fed funds market, we found the type I and type II errors to be large, averaging 81% and 23%, 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 rate and volume of fed funds activity, as well as in the set of counterparties. Before we discuss a few points raised by our analysis we want to acknowledge that our study may have possible limitations. In particular, our test only applies to fed funds as defined under Regulation D and it is based only on two banks. Despite these possible limitations, we argue below that our results have important implications.

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 on 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 our results are in fact quite general. First, the two banks that provided their unique identifier are either senders or receivers for a sizable share all pairs of transactions outputted by the algorithm. Over the 2007 to 2011 period, the two banks were involved, on average, with 29.4% of the algorithm's output. Our results, then, directly concern a large fraction of the algorithm's output. Second, it is reasonable to assume that our results are applicable to other large banks with similar Fedwire activity. We define large banks as those which receive or send over 800,000 payments a quarter (in a typical quarter, only 9 or 10 banks met this criterion). Assuming our type I and type II errors generalize to these banks implies that, on average, 44.3% of the algorithm's output is impacted (see table III).

The MaPS algorithm, however, may perform better for smaller banks. Indeed, these banks send fewer payments over Fedwire, and these payments may correspond to 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 this hypothesis formally at this point, it finds some support in the fact that there is separate evidence that the algorithm may perform well for some Government Sponsored Entities (GSEs). A preliminary comparison suggests that the algorithm could accurately measure the total amount of fed funds sold by each Federal Home Loan Bank (FHLB) over a quarter (see table AI in the appendix). More work should be done, however, to see how well the algorithm's output matches the FHLB data over a longer time series, and more generally, to test if the algorithm performs better for small banks.

If it is established that the algorithm is only inaccurate for a few large banks, then a possible remedy could be to exclude these banks from any empirical analysis. We see two problems with this approach. First, ignoring at least a third of all transactions outputted by the algorithm would prevent any comprehensive analysis of the fed funds market. Second, 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.

Does aggregating the algorithm's output make it more precise?

Our test suggests that the algorithm is unlikely to identify correctly individual fed funds transactions. 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 the algorithm can provide, in general, meaningful measures at the bank-to-bank or the bank level. In particular, the algorithm will attribute i) more transactions to large banks which serve as intermediaries, and ii) fewer transactions to small banks using correspondent banking. Because the fed funds rate extended to small and large banks may be different, the average rate identified by the algorithm for those banks may be biased.

Our analysis provides little evidence that the MaPS 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 Markets Group, because it would cover fed funds transactions arranged both through brokers and privately between banks. To confirm this hypothesis, however, further work needs to be conducted to test whether the algorithm's type I error are almost exclusively produced by correspondent banking.

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, it may capture 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 (e.g., loans to financial institutions other than banks). This hypothesis finds support in the fact that 89% of the transactions paired by the algorithm in first quarter of 2007 are found to have inferred interest rates which, once rounded, can be considered

to be in whole basis point or 32nds of an interest rate.²³ Discussions with market participants suggest that overnight unsecured loans are typically traded in these discrete amounts, suggesting the pairing of transactions by the algorithm is not random.²⁴

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 the banks' clients which are outside the banking 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 for correspondent banking. More generally, the hypothesis that the MaPS algorithm's output captures overnight interbank loans would need to be formally tested in order to be validated. Until then, researchers and policy makers should be reluctant in using the algorithm's output as a proxy for interbank lending.

In conclusion, our results raise serious concerns about the appropriateness of using the MaPS 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 could serve as a cautionary tale for applied economists. Indeed, it emphasizes the prudence with which researchers should treat data that cannot be considered direct observations, and it underscores the need for due diligence to establish, prior to any analysis, that indirect inferences produced by an algorithm truly correspond to what the researchers intend.

²³ The dollar amount a bank can send to another bank over Fedwire Funds is constrained to be rounded to the nearest cent. Due to 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 principle and interest payment amount is increased by 1 cent and then when the amount is decreased by 1 cent. If these two inferred interest rates straddle an interest rate in whole basis points or 32nds of an interest rate, then we say 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.

²⁴ 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.

Bibliography

- Acharya V. and O. Merrouche, 2011, "Precautionary Hoarding of Liquidity and Inter-Bank Markets: Evidence from the Sub-Prime Crisis," *conditionally accepted at the Review of Finance*.
- Acharya V. and D. Skeie, 2011, "A Model of Liquidity Hoarding and Term Premia in Inter-Bank Markets," *Journal of Monetary Economics*, 58(5), 436-447.
- Afonso G., Kovner A. and A. Schoar, 2011, "Stressed, not Frozen: The Federal Funds Market in the Financial Crisis," *Journal of Finance*, 66 (4), 1109-1139.
- Afonso G., Kovner A. and A. Schoar, 2012, "The Importance of Trading Relationships in the Fed Funds Market," working paper.
- Afonso G. and R. Lagos, 2012a, "Trade Dynamics in the Market for Federal funds," *FRBNY Staff Report 549*.
- Afonso G. and R. Lagos, 2012b, "An Empirical Study of Trade Dynamics in the Fed Funds Market," *FRBNY Staff Report 550*.
- Allen J., Chapman J., Echenique F. and M. Shum, 2012, "Efficiency and Bargaining Power in the Interbank Loan Market," *Bank of Canada Working Paper 2012-29*.
- Allen J., Kastl J. and A. Hortacsu, 2011, "Crisis Management: Analyzing Default Risk and Liquidity Demand during Financial Stress," *Bank of Canada working paper 2011-17*.
- Armantier O., Ghysels E., Sarkar A. and J. Shrader, 2011, "Stigma in Financial Markets: Evidence from Liquidity Auctions and Discount Window Borrowing during the Crisis" *FRBNY Staff Reports 483*.
- Ashcraft A. and H. Bleakley, 2006, "On the Market Discipline on Informationally Opaque Firms: Evidence from Bank Borrowers in the Federal funds Market," *FRBNY Staff Report 257*.
- Ashcraft A. and D. Duffie, 2007, "Systemic Illiquidity in the Federal Funds Market," *American Economic Review, Papers and Proceedings*, 97:2, 221-225.
- Ashcraft A., Skeie D. and J. McAndrews, 2011, "Precautionary Reserves and the Interbank Market," *Journal of Money, Credit and Banking*, 43(7), 311-348.
- Atalay E. and M. Bech, 2010, "The Topology of the Federal Funds Market," *Physica A: Statistical Mechanics and its Applications*, Vol. 389(22).
- Bartolini L., Hilton S. and J. McAndrews, 2008, "Settlement Delays in the Money Market," *FRBNY Staff Report 319*.
- Bech M., Bergstrom C., Garratt R. and M. Rosvall, 2012, "Mapping Change in the Federal funds Market," *FRBNY Staff Report 507*.
- Copeland A., Martin A. and M. Walker, 2010, "The Tri-Party Repo Market before the 2010 Reforms", *FRBNY Staff Report 477*.
- Demiralp S., Preslowsky B. and W. Whitesell, 2006, "Overnight Interbank Loan Markets," *Journal of Economics and Business*, Vol. 58, 67-83.
- Furfine C., 1999, "The Microstructure of the Federal Funds Market," *Financial Markets, Institutions, and Instruments*, 8(5), 24-44.

- Furfine C., 2001, "Banks Monitoring Banks: Evidence from the Overnight Federal funds Market," *Journal of Business*, 74(1), 33-58
- Furfine C., 2003, "Interbank Exposures: Quantifying the Risk of Contagion," *Journal of Money, Credit, and Banking*, 35(1), 111-128.
- Hamilton J., 1996, "The Daily Market for Federal Funds," *Journal of Political Economy*, 104(1), 26-56.
- Hamilton J., 1997, "Measuring the Liquidity Effect," *American Economic Review*, 87(1), 80-97.
- Hendry S. and N. Kamhi, 2009, "Uncollateralized Overnight Loans Settled in LVTS", *Applied Finance Economics*, 869-880.
- Keister T., Martin A. and J. McAndrews, 2008, "Divorcing Money from Monetary Policy," *Economic Policy Review*, Vol. 14, No. 2.
- Kuo D., Skeie D. and J. Vickery (2012). "A Comparison of Libor to Other Measures of Bank Borrowing Costs", FRBNY work in progress.
- McAndrews J. (2009). "Segmentation in the U.S. Dollar Money Markets During the Financial Crisis," *FRBNY working paper*.
- 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*, 42-47.

Tables

Table I: Estimates of type I and type II error for the first quarter of 2007:

The Algorithm List		The 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%		Type II error: 17%	

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.

Table II: Estimates of type I and type II errors over time.

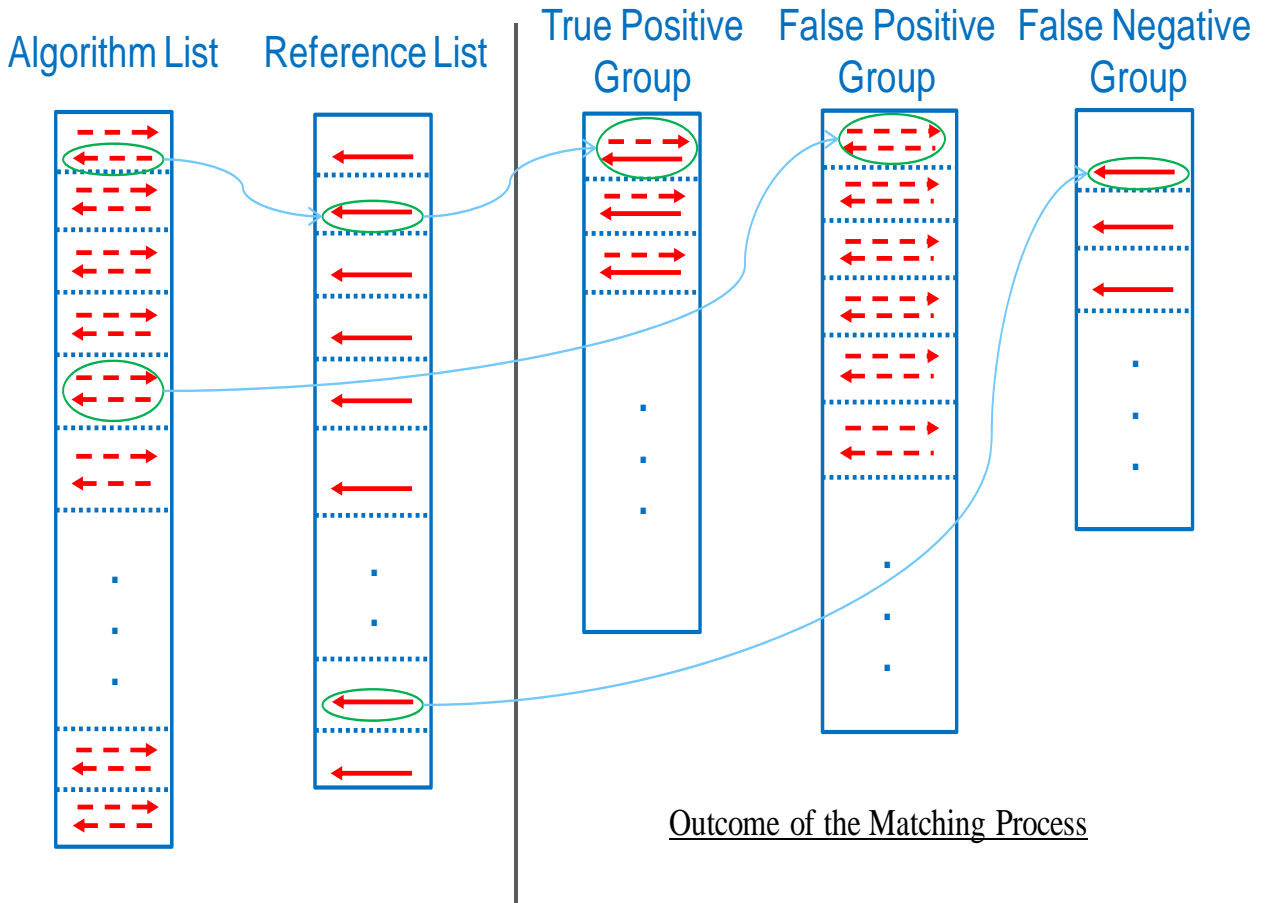
	2007:Q1	2008:Q1	2009:Q1	2010:Q1	2011:Q1	Average
Type I	64%	72%	85%	93%	93%	81.4%
Type II	24%	28%	27%	19%	17%	23.0%

Table III: Percent of the 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%

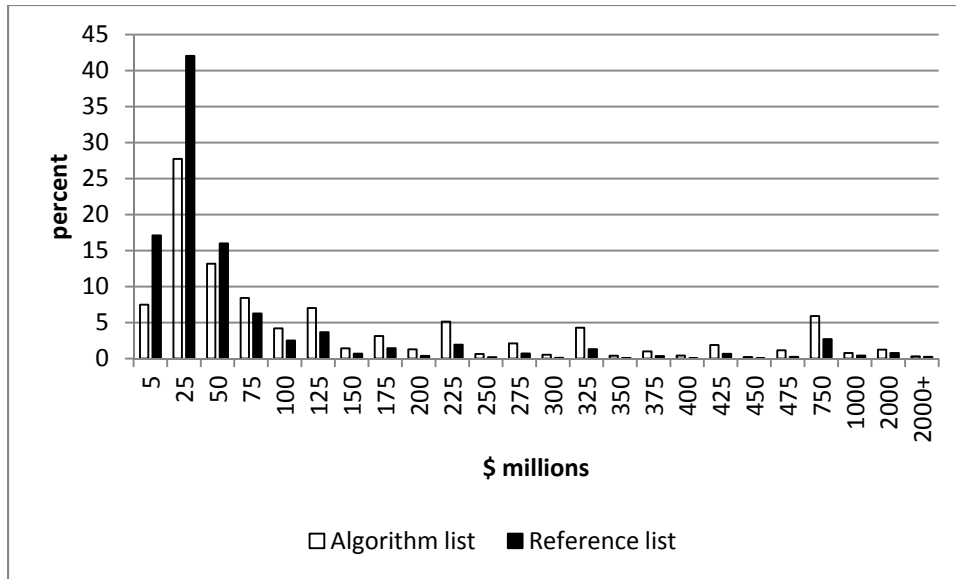
Note: "Two banks" are the two institutions on which our tests are based. "Large banks" are those which sent and received more than 800,000 payments in the relevant quarter. The same nine banks met this criterion for every quarter in the table, including the two banks at the center of our analysis. A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

Figure 1: The Test's Methodology



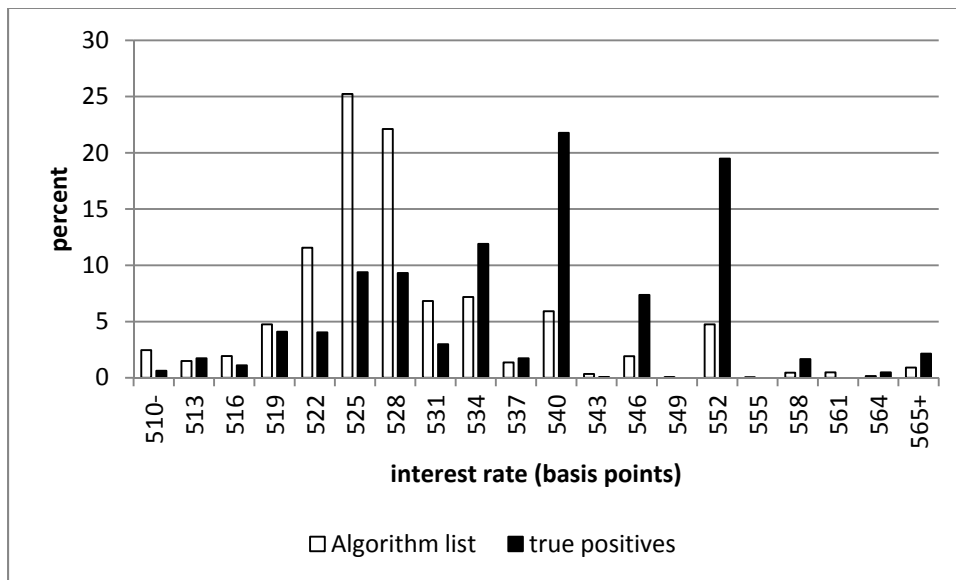
The "Algorithm List" consists of all pairs of payments identified by the algorithm as fed funds loans. The "Reference List" consists of all Fedwire payments with the unique identifier. The "True Positive Group" consists of every pair of payments in the Algorithm List with a match in the Reference List. 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 dash line indicates a send or a receive leg of a fed funds transaction identified by the algorithm. A solid line indicates a receive leg of a fed funds transaction with the unique identifier.

Figure 2: Comparison of amounts across the algorithm and reference lists



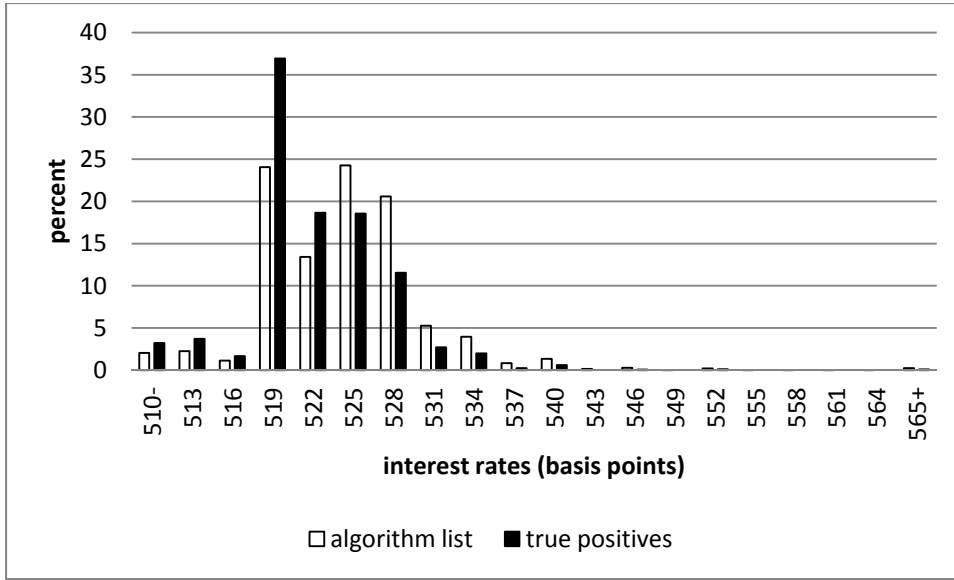
Note: Comparison done for the first quarter of 2007. For the algorithm list, amounts plotted are those in the receive leg of the paired payment transactions. The horizontal axis's label is the amount bin's larger end point, except for "2000+" which denotes the bin with all payments greater than \$2000 million.

Figure 3: Comparison of interest rates for fed funds sold



Note: Comparison done for the first quarter of 2007. As a reference, the fed funds rate targeted by the Federal Open Market Committee in this quarter was 525 basis points. The horizontal axis's 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.

Figure 4: Comparison of interest rates for fed funds purchased



Note: Comparison done for the first quarter of 2007. As a reference, the fed funds rate targeted by the Federal Open Market Committee in this quarter was 525 basis points. The horizontal axis's 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.

Appendix A1:

The federal home loan banks (FHLB), Fannie Mae and Freddie Mac report quarterly fed funds sold in their quarterly filings to the Securities Exchange Commission (SEC). We use these reports to gauge the algorithm's performance. While the reported fed funds sold numbers include term federal funds, and occasionally reverse repos, we believe it provides a relative clean way to test the algorithm's performance at the bank-quarter level (but not at the transaction level as we have done in this paper).

In table AI, we compare the fed funds sold numbers reported in the SEC quarterly reports by the FHLBs, Fannie Mae and Freddie Mac, to the algorithm's output for the first quarter of 2010. Overall, the average interest rate based on the algorithm's output is quite close to those reported by the FHLBs and Freddie Mac, although substantially below the rate reported by Fannie Mae. The amount of federal funds activity based on the algorithm's output is close for most of the FHLBs, with the Chicago, Indianapolis, and San Francisco FHLBs being the notable exceptions. But, the algorithm's output is significantly different from the numbers reported by Fannie Mae and Freddie Mac. This preliminary work suggests that the algorithm's output may characterize accurately the bank-level fed funds activity of most of the Federal Home Loan Banks. Substantial work still needs to be done, however, to accurately determine how well the algorithm performs along this dimension.

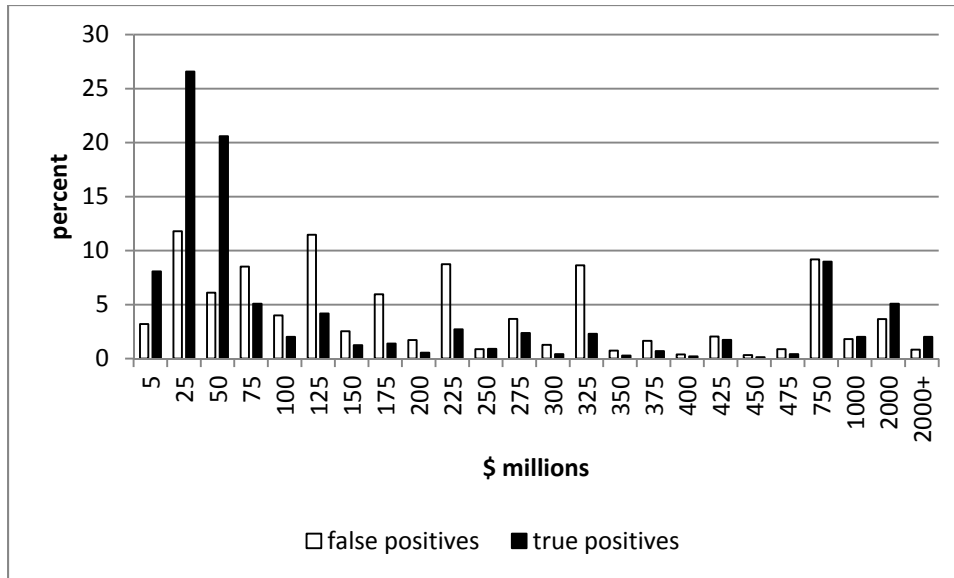
Table AI: GSEs Fed Funds Sold in the First Quarter of 2010

	SEC Quarterly Reports		Algorithm measure		Level difference		Percent difference	
	Amount	Interest Rate	Amount	Interest Rate	Amount	Interest Rate	Amount	Interest Rate
Atlanta	11.6	16	10.3	16	1.3	0	-11	0
Boston	7.6	14	6.5	14	1.1	0	-14	0
Chicago*	7.5	11	3.0	14	4.5	-3	-60	27
Cincinnati*	7.5	13	7.5	13	0	0	0	0
Dallas	4.3	11	4.3	12	0	-1	0	9
Des Moines	4.9	12	4.5	13	0.5	-1	-8	8
Indianapolis	7.6	14	4.2	14	3.4	0	-45	0
New York	5.4	12	5.1	12	0.3	0	-6	0
Pittsburgh	4.7	11	4.8	13	-0.1	-2	2	18
San Francisco	14.6	14	7.4	12	7.2	2	-49	-14
Seattle	n/a	n/a	6.3	13	n/a	n/a	n/a	n/a
Topeka	2.9	14	2.0	14	0.9	0	-31	0
FHLB System	83.7	14	65.9	13	17.8	1	-21	-7
Freddie Mac*	51.6	12	8.2	13	43.4	-1	-84	8
Fannie Mae*	40.1	21	4.1	15	36	6	-90	-29

Note: SEC Quarterly Reports include overnight and term trades and * indicates that the Quarterly Report includes reverse repurchase agreements. Amounts are in billions of dollars and interest rates are in basis points.

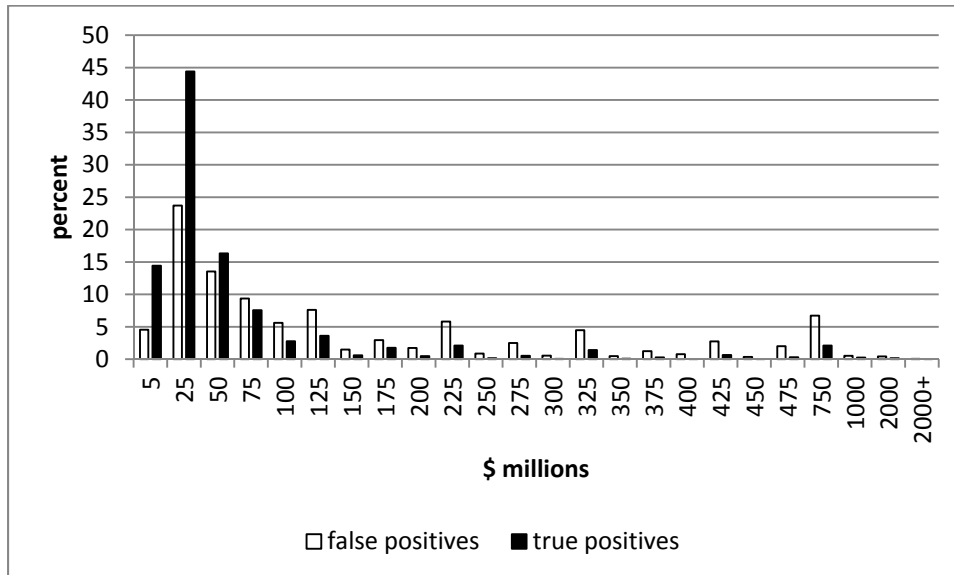
Appendix A2:

Figure A1: Comparison of the amounts of fed funds sold



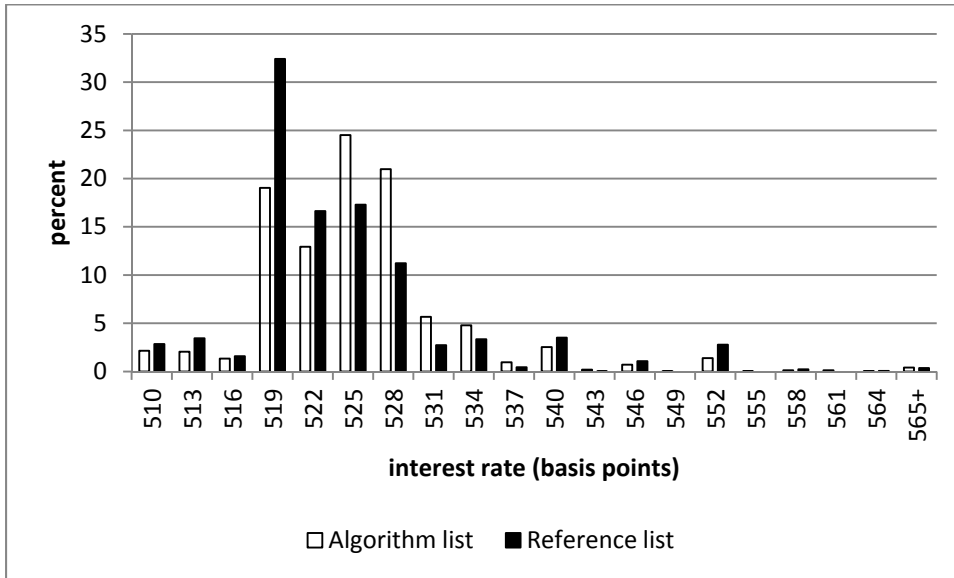
Note: Comparison done for the first quarter of 2007. The principal amount of the fed funds sale is graphed. The horizontal axis's label is the amount bin's larger end point, except for "2000+" which denotes the bin with all payments greater than \$2000 million. A Mann-Whitney U test rejects the null hypothesis that the distribution of amounts across false positive and true positives are equal at the 1% significance level (the Z-scores is -15.0).

Figure A2: Comparison of the amounts of fed funds purchased



Note: Comparison done for the first quarter of 2007. The principal amount of the fed funds purchased is graphed. The horizontal axis's label is the amount bin's larger end point, except for "2000+" which denotes the bin with all payments greater than \$2000 million. A Mann-Whitney U test rejects the null hypothesis that the distribution of amounts across false positive and true positives are equal at the 1% significance level (the Z-scores is -53.0).

Figure A3: Comparison of interest rates across the algorithm and reference lists when fed funds sold and fed funds purchased are combined



Note: Comparison done for the first quarter of 2007. For the reference list, interest rates were inferred for only those transactions in the set of true positives. As a reference, the fed funds rate targeted by the Federal Open Market Committee in this quarter was 525 basis points. The horizontal axis's 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.