Federal Reserve Bank of New York

May 2013
Volume 19 Number 1

Economic Policy Review

1 TBA Trading and Liquidity in the Agency MBS Market
James Vickery and Joshua Wright

19 Unintended Consequences of School Accountability Policies: Evidence from Florida and Implications for New York
Rajashri Chakrabarti and Noah Schwartz

45 Trading Activity and Price Transparency in the Inflation Swap Market
Michael J. Fleming and John R. Sporn

Economic Policy Review

Editor
Kenneth D. Garbade

Coeditors
Meta Brown
Richard Crump
Marco Del Negro
Jan Groen
Andrew Haughwout
Stavros Peristiani

Editorial Staff
Valerie LaPorte
Michelle Bailer
Karen Carter
Mike De Mott
Anna Snider

Production Staff
Jessica Iannuzzi
David Rosenberg
Jane Urry

The *Economic Policy Review* is published by the Research and Statistics Group of the Federal Reserve Bank of New York. Articles undergo a comprehensive refereeing process prior to their acceptance in the *Review*. The views expressed are those of the individual authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

www.newyorkfed.org/research

Follow us on Twitter: *@NYFedResearch*

Federal Reserve Bank of New York
Economic Policy Review

May 2013
Volume 19 Number 1

Contents

Articles:

1 TBA Trading and Liquidity in the Agency MBS Market
James Vickery and Joshua Wright

Mortgage-backed securities in the United States are generally traded on a "to-be-announced," or TBA, basis. The key feature of a TBA trade is that the identity of the securities to be delivered to the buyer is not specified exactly at the time of the trade, facilitating a liquid forward market. This article describes the main features of the TBA market. It also presents evidence on the liquidity of this market during the financial crisis period. Using variation in TBA eligibility rules, the authors' estimates suggest that the liquidity benefits associated with the TBA market are of the order of 10 to 25 basis points during 2009 and 2010, and magnified during periods of market stress. The estimates further suggest that the presence of a government credit guarantee alone does not appear to be sufficient explanation for the liquidity of agency MBS.

19 Unintended Consequences of School Accountability Policies: Evidence from Florida and Implications for New York
Rajashri Chakrabarti and Noah Schwartz

Over the past two decades, state and federal education policies have tried to hold schools more accountable for educating students by tying rewards and sanctions to test scores and other measurable outcomes. A common criticism of these policies is that they may induce schools to

"game the system" along with—or instead of—making genuine educational improvements. One such strategic response may be to classify low-performing students into categories that are excluded from grade computation in an effort to artificially inflate scores. This article analyzes school responses to an influential accountability-tied voucher program in Florida. The authors find evidence of increased classification into "excluded" categories in failing schools following the program's inception. Their findings have important implications for New York City's Progress Reports program and New York's implementation of the federal No Child Left Behind Act. While these policies were modeled after the Florida program, they contain important design differences that are likely to discourage this type of gaming, although they may encourage other strategic classifications.

45 Trading Activity and Price Transparency in the Inflation Swap Market

Michael J. Fleming and John R. Sporn

The issues of liquidity and price transparency in derivatives markets have taken on greater import given regulatory efforts under way to improve their transparency. To date, the lack of transaction data has impeded the understanding of how the inflation swap and other derivatives markets operate. This article broadens that understanding by using a novel transaction data set to examine trading activity and price transparency in the quickly growing U.S. inflation swap market. The authors find that the market appears reasonably liquid and transparent, despite its over-the-counter nature and modest level of trading activity. Specifically, they find that transaction prices are typically quite close to widely available end-of-day quoted prices and that realized bid-ask spreads are modest, even though the reasonably comprehensive data set from 2010 contains just over two trades per day on average. The authors also identify concentrations of activity in certain tenors (ten years) and trade sizes ($25 million) and among certain market participants, as well as various attributes that help explain trade sizes and price deviations. Their study can serve as a resource for policymakers considering public reporting and other regulatory initiatives and for market participants and observers more generally interested in the workings of the inflation swap market.

James Vickery and Joshua Wright

TBA Trading and Liquidity in the Agency MBS Market

- While mortgage securitization by private financial institutions has declined to low levels since 2007, issuance of *agency* mortgage-backed-securities (MBS) has remained robust.
- A key feature of agency MBS is that each bond carries a credit guarantee by Fannie Mae, Freddie Mac, or Ginnie Mae.
- More than 90 percent of agency MBS trading occurs in the to-be-announced (TBA) forward market. In a TBA trade, the exact securities to be delivered to the buyer are chosen just before delivery, rather than at the time of the original trade.
- This study describes the key institutional features of the TBA market, highlighting recent trends and changes in market structure.
- It presents suggestive evidence that the liquidity associated with TBA eligibility increases MBS prices and lowers mortgage interest rates.

1. Introduction

The U.S. residential mortgage market has experienced significant turmoil in recent years, leading to important shifts in the way mortgages are funded. Mortgage securitization by private financial institutions declined to negligible levels during the financial crisis that began in August 2007, and remains low today. In contrast, throughout the crisis there continued to be significant ongoing securitization in the *agency* mortgage-backed-securities (MBS) market, consisting of MBS with a credit guarantee by Fannie Mae, Freddie Mac, or Ginnie Mae.[1] Agency MBS in the amount of $2.89 trillion were issued in 2008 and 2009, but no non-agency securitizations of new loans occurred during this period. The outstanding stock of agency MBS also increased significantly during the crisis period, from $3.99 trillion at June 2007 to $5.27 trillion by December 2009.[2]

[1] Fannie Mae and Freddie Mac are the common names for the Federal National Mortgage Association and Federal Home Loan Mortgage Corporation, respectively, the government-sponsored enterprises (GSEs) that securitize and guarantee certain types of residential mortgages. Ginnie Mae, shorthand for the Government National Mortgage Association, is a wholly-owned government corporation within the Department of Housing and Urban Development. See Section 2 for more details.

[2] Data on MBS issuance are from the Securities Industry and Financial Markets Association (SIFMA) and the Inside Mortgage Finance *Mortgage Market Statistical Annual*. Data on agency MBS outstanding are from the Federal Reserve Statistical Release Z.1, "Flow of Funds Accounts of the United States," Table L.125. Throughout this article, unless otherwise noted, we use the term MBS to refer to residential MBS, not to securities backed by commercial mortgages.

James Vickery is a senior economist at the Federal Reserve Bank of New York; Joshua Wright is a policy and markets analyst on the open market trading desk at the Federal Reserve Bank of New York.

james.vickery@ny.frb.org
joshua.wright@ny.frb.org

The authors thank Kenneth Garbade, two anonymous referees, Marco Cipriani, David Finkelstein, Michael Fleming, Ed Hohmann, Dwight Jaffee, Patricia Mosser, and market participants for their insights and help with institutional details, and Diego Aragon and Steven Burnett for outstanding research assistance. The views expressed 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.

A key distinguishing feature of agency MBS is that each bond either carries an explicit government credit guarantee or is perceived to carry an implicit one, protecting investors from credit losses in case of defaults on the underlying mortgages.[3] This government backing has been the subject of a long-running academic and political debate. A second, less widely recognized feature is the existence of a liquid forward market for trading agency MBS, out to a horizon of several months.[4] The liquidity of this market improves market functioning and helps mortgage lenders manage risk, since it allows them to "lock in" sale prices for new loans as, or even before, those mortgages are originated.

The liquidity of [the TBA] market improves market functioning and helps mortgage lenders manage risk, since it allows them to "lock in" sale prices for new loans as, or even before, those mortgages are originated.

More than 90 percent of agency MBS trading volume occurs in this forward market, which is known as the TBA (to-be-announced) market. In a TBA trade, the seller of MBS agrees to a sale price, but does not specify which particular securities will be delivered to the buyer on settlement day. Instead, only a few basic characteristics of the securities are agreed upon, such as the coupon rate, the issuer, and the approximate face value of the bonds to be delivered. While the agency MBS market consists of thousands of heterogeneous MBS pools backed by millions of individual mortgages, the TBA trading convention allows trading to be concentrated in only a small number of liquid forward contracts. TBA prices, which are observable to market participants, also serve as the basis for pricing and hedging a variety of other MBS, which themselves would not be delivered into a TBA trade and may not even be eligible for TBA delivery.

The main goal of this article is to describe the basic features and mechanics of the TBA market, and to review recent legislative changes that have affected the types of mortgages eligible for TBA trading. The article also presents some preliminary evidence suggesting that the liquidity benefits associated with TBA eligibility increase MBS prices and reduce mortgage interest rates. Our analysis exploits changes in legislation to help disentangle the effects of TBA eligibility from other characteristics of agency MBS. In particular, we study pricing for "super-conforming" mortgages that became eligible for agency MBS securitization through legislation in 2008, but that were ruled ineligible to be delivered to settle TBA trades. We show that MBS backed by super-conforming mortgages trade at a persistent price discount in the secondary market, and also that interest rates on such loans are correspondingly higher in the primary mortgage market. Preliminary evidence suggests that these stylized facts are not fully explained by differences in prepayment risk. We interpret our estimates to suggest that the liquidity benefits of TBA eligibility may be of the order of 10 to 25 basis points on average in 2009 and 2010, and are larger during periods of greater market stress.

Our institutional discussion and empirical results support the view that the TBA market serves a valuable role in the mortgage finance system. This finding suggests that evaluations of proposed reforms to U.S. housing finance should take into account potential effects of those reforms on the operation of the TBA market and its liquidity.

2. Background

Most residential mortgages in the United States are *securitized*, rather than held as whole loans by the original lender.[5] Securitized loans are pooled in a separate legal trust, which then issues the MBS and passes on mortgage payments to the MBS investors after deducting mortgage servicing fees and other expenses. These MBS are actively traded and held by a wide range of fixed-income investors.

Even in the wake of the subprime mortgage crisis, securitization remains central to the U.S. mortgage finance system because of continuing large issuance volumes of *agency* MBS. In the agency market, each MBS carries a credit

[3] MBS guaranteed by Ginnie Mae carry an explicit federal government guarantee of the timely payment of mortgage principal and interest. Securities issued by Fannie Mae and Freddie Mac carry a credit guarantee from the issuer; although this guarantee is not explicitly backed by the federal government, it is very widely believed that the government would not allow Fannie Mae and Freddie Mac to default on their guarantee obligations. Consistent with this view, the U.S. Treasury has committed to support Fannie Mae and Freddie Mac since September 2008, when they were placed in conservatorship by their primary regulator, the Federal Housing Financing Agency (FHFA). (See Section 2 for a further discussion of this conservatorship.)

[4] In a forward contract, the security and cash payment for that security are not exchanged until after the date on which the terms of the trade are contractually agreed upon. The date the trade is agreed upon is called the "trade date." The date the cash and securities change hands is called the "settlement date."

[5] As of December 2011, 67 percent of home mortgage debt was either securitized through agency or non-agency MBS or held on the balance sheets of Fannie Mae and Freddie Mac (source: Federal Reserve Statistical Release Z.1, "Flow of Funds Accounts of the United States," Table L.218).

guarantee from either Fannie Mae or Freddie Mac, two housing GSEs currently under public conservatorship,[6] or from Ginnie Mae. (Hereafter, we sometimes refer to these three institutions as "the agencies.") In return for monthly guarantee fees, the guarantor promises to forward payments of mortgage principal and interest to MBS investors, even if there are prolonged delinquencies among the underlying mortgages.[7] In other words, mortgage *credit risk* is borne by the guarantor, not by investors. However, investors are still subject to uncertainty about when the underlying borrowers will prepay their mortgages. This *prepayment risk* is the primary source of differences in fundamental value among agency MBS.

Only mortgages that meet certain size and credit quality criteria are eligible for inclusion in mortgage pools guaranteed by Fannie Mae, Freddie Mac, or Ginnie Mae. The charters of Fannie Mae and Freddie Mac restrict the types of loans that may be securitized; these limits include a set of loan size restrictions known as "conforming loan limits."[8] Mortgages exceeding these size limits are referred to as "jumbo" loans; such mortgages can be securitized only by private financial institutions and do not receive an explicit or implicit government credit guarantee. Ginnie Mae MBS include only loans that are explicitly federally insured or guaranteed, mainly loans insured by the Federal Housing Administration (FHA) or guaranteed by the Department of Veterans Affairs.

[6] This period of conservatorship began on September 7, 2008. As of this date, the FHFA—Fannie Mae and Freddie Mac's primary regulator—assumed control of major operating decisions made by these two firms. This was accompanied by an injection of preferred stock by the U.S. Treasury and the establishment of a secured lending credit facility with the Treasury. These steps were made necessary by Fannie Mae's and Freddie Mac's deteriorating financial condition, attributable to mortgage-related credit losses. (For more details, see www.ustreas.gov/press/releases/hp1129.htm.)

[7] The timing of these payments can differ depending on the class of security. For example, Freddie Mac's Gold PCs (participation certificates), which have a forty-five-day payment delay, promise timely payment of both principal and interest, but Freddie Mac's adjustable-rate-mortgage PCs, which have a seventy-five-day payment delay, promise timely payment of interest and *ultimate* payment of principal. For both agencies, a loan that is seriously delinquent is eventually removed from the MBS pool, in exchange for a payment of the remaining principal at par. Thus, a mortgage default is effectively equivalent to a prepayment from the MBS investor's point of view, since the investor receives an early return of principal, but does not suffer any credit losses.

[8] Until 2008, the one-family conforming loan limit for loans securitized through Fannie Mae and Freddie Mac was $417,000, with higher limits applying to two-to-four-family mortgages and loans from Alaska, Hawaii, Guam, and the U.S. Virgin Islands. Lower size limits applied to loans guaranteed by Ginnie Mae. These conforming size limits were raised significantly in high-cost housing areas in 2008, as discussed in Section 4.1.

Table 1

Daily Average Trading Volumes in Major U.S. Bond Markets

Billions of dollars

Year	Municipal Bonds	Treasury Securities	Agency Mortgage-Backed Securities	Corporate Bonds
2005	16.9	554.5	251.8	16.7
2006	22.5	524.7	254.6	16.9
2007	25.2	570.2	320.2	16.4
2008	19.4	553.1	344.9	11.8
2009	12.5	407.9	299.9	16.8
2010	13.3	523.2	320.6	16.3

Source: Federal Reserve Bank of New York.

Notes: Figures are based on purchases and sales of securities reported by primary dealers (see www.newyorkfed.org/markets/gsds/search.cfm). Figures for corporate bonds refer only to securities with a maturity greater than one year.

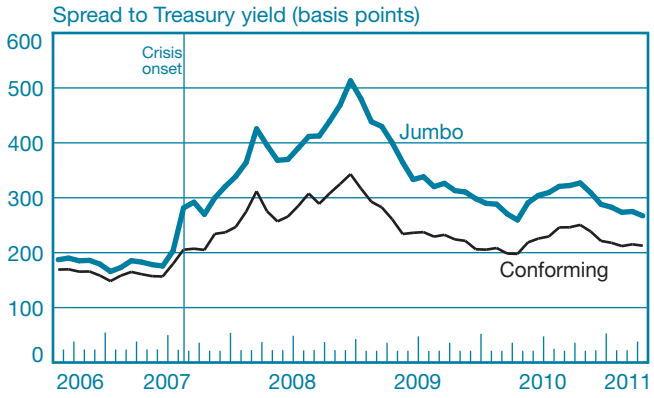
2.1 Mortgage Interest Rates during the Financial Crisis

Only a small number of non-agency residential MBS have been issued since mid-2007 and, during this period, secondary markets for trading non-agency MBS have been extremely illiquid. In contrast, issuance and trading volumes in the agency MBS market remained relatively robust throughout the crisis period. Providing evidence of this market liquidity, Table 1 presents data on daily average trading volumes for different types of U.S. bonds. Agency MBS daily trading volumes have averaged around $300 billion from 2005 to 2010, a level that did not decline significantly during the financial crisis of 2007-09. While in each year MBS trading volumes were lower than Treasury volumes, they were of a larger order of magnitude than corporate bonds or municipal bonds.

The effects of this divergence between the agency and non-agency MBS markets on primary mortgage rates can be seen in Chart 1, which shows the evolution of interest rates on jumbo and conforming mortgages between 2007 and mid-2011. Rates on both loan types are expressed as a spread to Treasury yields. Both spreads increased during the financial crisis, but the increase was much more pronounced for jumbo loans. Before the crisis, interest rates on jumbo loans were only around 25 basis points higher than rates on conforming mortgages; this "jumbo-conforming spread" increased to 150 basis points or more during the crisis. While the jumbo-conforming spread has narrowed more recently, it still significantly exceeds pre-crisis levels, as of mid-2011.

Chart 1

Thirty-Year Fixed-Rate Jumbo and Conforming Mortgage Rates



Source: HSH Associates.

Notes: Mortgage rates are expressed as a spread to the average of the five- and ten-year Treasury yield. Crisis onset is marked at August 2007, the month that BNP Paribas suspended convertibility for two hedge funds, reflecting problems in the subprime MBS markets.

Why were interest rates on conforming mortgages eligible for agency MBS securitization relatively more stable during the financial crisis? Several factors were likely at play: 1) From an investor's perspective, MBS backed by jumbo loans have much greater *credit risk* because, unlike agency MBS, they do not carry a credit guarantee. The price impact of this difference in risk was heightened during the crisis because of high mortgage default rates and an amplification of credit risk premia. 2) Jumbo loans have greater *prepayment risk*, because refinancing by jumbo borrowers is more responsive to the availability of profitable refinancing opportunities.[9] 3) The difference in *liquidity* between conforming and jumbo mortgages became significantly larger and more valuable to investors as the crisis deepened due to the collapse of the non-agency MBS market. 4) From late 2008 to March 2010, the Federal Reserve bought large quantities of agency debt and agency MBS under its large-scale asset purchase (LSAP) programs, helping to lower conforming mortgage rates.[10]

[9] For example, Green and LaCour-Little (1999) find that among mortgages originated in the late 1980s, smaller-balance non-jumbo loans are generally less likely to be prepaid during a later period of sharply falling interest rates, when refinancing was almost certainly optimal from a borrower's perspective. There are several explanations why jumbo borrowers exercise their prepayment options more profitably; for example, they are likely to be more educated, and high-principal mortgages involve smaller per-dollar fixed transaction costs and search costs. See also Schwartz (2006) for evidence that wealthy and educated households display more "rational" and profitable prepayment behavior.

[10] The Federal Reserve purchased $1.25 trillion in agency MBS and nearly $175 billion in agency debt between late 2008 and first-quarter 2010. For an analysis of the purchases' effects, see Gagnon et al. (2010).

(This is unlikely to be the dominant explanation, however, since the jumbo-conforming spread was extremely elevated even before the announcement of the LSAP programs on November 25, 2008.)

2.2 Liquidity Premia and the Jumbo-Conforming Spread

Consistent with the view that liquidity effects were important during this period, the timing of the increase in the jumbo-conforming spread corresponds closely to the collapse in non-agency MBS liquidity and mortgage securitization during the second half of 2007. Furthermore, this spread remains elevated even today, despite normalization of many measures of credit risk premia.

There is a scholarly literature on the size and source of the jumbo-conforming spread during the pre-crisis period; however, that literature focuses on the debate over the value of the GSEs' implicit public subsidy and the extent to which this subsidy has been passed on to consumers in the form of lower interest rates.[11] In most cases, these studies do not attempt to decompose the credit risk and liquidity risk components of this spread. Nonetheless, Passmore, Sherlund, and Burgess (2005) do find that the size of the jumbo-conforming spread moves inversely with jumbo MBS liquidity and with factors affecting MBS demand and supply, consistent with the view that liquidity differences are an important determinant of the spread between jumbo and conforming loans.

This still leaves open the question of *why* the agency MBS market is so liquid, given that it consists of literally tens of thousands of unique securities. One hypothesis is that the implied government credit guarantee for agency MBS alone is sufficient to ensure market liquidity. However, earlier academic literature shows significant differences in liquidity and pricing even among different government-guaranteed instruments of the same maturity. For example, on-the-run U.S. Treasury securities trade at a significant premium to off-the-run Treasuries (Krishnamurthy 2002).[12] Treasury securities also trade at a premium to government-guaranteed corporate debt, such as debt issued under the Federal Deposit Insurance Corporation's Temporary Liquidity Guarantee Program in 2008-09 or by the Resolution Funding Corporation in 1989-91 (Longstaff 2004; Schwarz 2009). Another example is the attempts by Fannie Mae and Freddie Mac to issue quarter-

[11] Examples include Passmore, Sherlund, and Burgess (2005); Ambrose, LaCour-Little, and Sanders (2004); and Torregrosa (2001). See McKenzie (2002) for a literature review.

[12] See also Amihud and Mendelson (1991) and Fleming (2002). Also related, Nothaft, Pearce, and Stevanovic (2002) find that yield spreads between GSEs and other corporations are associated with issuance volumes, a proxy for liquidity.

coupon MBS and participation certificates. Even for the same guarantor and loan term, these quarter-coupon securities have traded at wider bid-ask spreads and have had higher average yields than neighboring whole- and half-coupon securities, which are more liquid. These examples, as well as the literature on the jumbo-conforming spread, are relatively consistent in suggesting that a pure liquidity premium for the most liquid government or government-like securities may be in the range of 10 to 30 basis points under "normal" financial market conditions and significantly larger during periods of market disruption, such as those experienced during the financial crisis.[13]

Thus, the presence of a government credit guarantee alone does not appear to be sufficient explanation for the liquidity of agency MBS and the wedge between jumbo and conforming mortgage rates. The sheer aggregate size of the agency MBS market no doubt contributes to its liquidity, but this does not account for why agency MBS are more liquid than corporate bonds, whose market is similar in total size. The agency MBS market is substantially more homogenous than the corporate bond market, however, and TBA trading helps homogenize the market further, at least for trading purposes. The TBA market has received relatively little attention in the academic literature, and the mechanics of this market are not well understood by many non-specialist observers.[14] To help fill this gap, we now turn to a detailed description of the TBA market.

3. The TBA Market

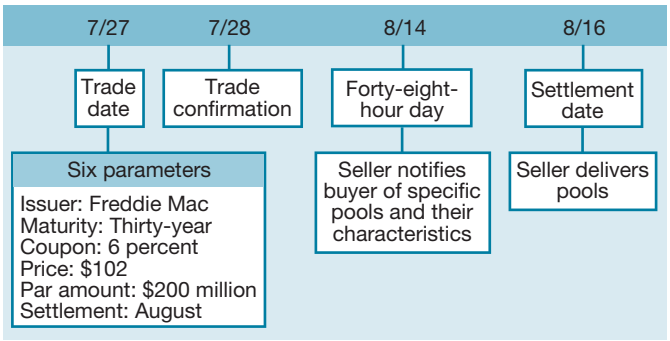
In a TBA trade, similar to other forward contracts, the two parties agree upon a price for delivering a given volume of agency MBS at a specified future date.[15] The characteristic feature of a TBA trade is that the actual identity (that is, the particular CUSIPs) of the securities to be delivered at settlement is not specified on the trade date. Instead, participants agree upon only six general parameters of the securities to be delivered: issuer, maturity, coupon, price, par amount, and settlement date. Coupon rates vary in 50-basis-point increments, in keeping with the underlying MBS.

[13] See Beber, Brandt, and Kavajecz (2009) for a discussion of liquidity premia amid flight-to-quality flows.

[14] Many GSE reform commentaries have similarly made little mention of the TBA market. Exceptions include SIFMA and the Mortgage Bankers Association.

[15] Note that all TBA-eligible securities involve a so-called "pass-through" structure, whereby the underlying mortgage principal and interest payments are forwarded to securityholders on a pro rata basis, with no tranching or structuring of cash flows. Collateralized mortgage obligations (CMOs) are not TBA-eligible.

Timeline for a TBA Trade



Source: Salomon Smith Barney.

A smaller but still significant portion of agency MBS trading volume occurs outside of the TBA market. This is known as "specified pool" trading, because the identity of the securities to be delivered is specified at the time of the trade, much like in other securities markets. Some of these pools are ineligible for TBA trading because the underlying loans have nonstandard features. Others, however, trade outside the TBA market by choice, because they are backed by loans with more favorable prepayment characteristics from an investor's point of view, allowing them to achieve a higher price, as described below. Similarly, some TBA trades will involve additional stipulations, or "stips," beyond the six characteristics listed above, such as restrictions on the seasoning, number of pools, or geographic composition of the pools to be delivered.

3.1 Mechanics of a TBA Trade

A timeline for a typical TBA trade, including three key dates, is shown in the exhibit. The detailed conventions that have developed around TBA trading are encoded in the "good delivery guidelines" determined by the Securities Industry and Financial Markets Association, an industry trade group whose members include broker-dealers and asset managers, as part of its *Uniform Practices for the Clearance and Settlement of Mortgage-Backed Securities and Other Related Securities*. These conventions were developed as the MBS market emerged in the 1970s and became more detailed and formalized in the ensuing decades.

Trade day. The buyer and seller establish the six trade parameters listed above. In the example shown in the exhibit, a TBA contract agreed upon in July will be settled in August, for a security issued by Freddie Mac with a thirty-year maturity, a 6 percent annual coupon, and a par amount of $200 million at a price of $102 per $100 of par amount, for a total price of $204 million. TBA trades generally settle within three months, with volumes and liquidity concentrated in the two nearest months. To facilitate the logistics of selecting and delivering securities from the sellers' inventory, SIFMA sets a single settlement date each month for each of several types of agency MBS.[16] Thus, depending on when it falls in the monthly cycle of settlements, the trade date will usually precede settlement by between two and sixty days.

Two days before settlement. No later than 3 p.m. two business days prior to settlement ("forty-eight-hour day"), the seller provides the buyer with the identity of the pools it intends to deliver on settlement day. If two counterparties have offsetting trades for the same TBA contract, these trades will be netted out.

Settlement day. The seller delivers the securities specified two days prior and receives the cash specified on the trade date. Amid the trading, lending, analysis, selection, and settling of thousands of individual securities each month, operational or accounting problems can arise—the resolution of which relies on a detailed set of conventions developed by SIFMA.

3.2 "Cheapest-to-Deliver" Pricing

Similar to Treasury futures, TBAs trade on a "cheapest-to-deliver" basis. On a forty-eight-hour day, the seller selects which MBS in its inventory will be delivered to the buyer at settlement. The seller has a clear incentive to deliver the lowest-value securities that satisfy the terms of the trade (recall that differences in value across securities are driven by pool characteristics affecting prepayment risk, such as past prepayment rates, or the geographic composition of the pool). This incentive is well understood by the TBA buyer, who expects to receive a security of lower value than average and accordingly adjusts downward the price it is willing to pay in the TBA market at the time of the trade. This is an example of a market phenomenon known to economists as "adverse selection."[17] Compounding this cheapest-to-deliver effect, the fact that the TBA seller effectively receives a valuable option well before settlement date to *choose* at settlement which bonds will be delivered, after additional information about the value of each security has been realized, further reduces the equilibrium price of the TBA contract relative to the value of an average MBS deliverable into that contract.

3.3 Temporary Fungibility and TBA Liquidity

TBA trading effectively applies a common cheapest-to-deliver price level to an intrinsically diverse set of securities and underlying mortgages. While the practice also occurs in the Treasury futures market, this homogenization seems more striking in the context of agency MBS because of the greater heterogeneity of the underlying assets. For trading purposes, groups of MBS that share the six general characteristics listed above may be treated as fungible, in the sense that any could be delivered into a given TBA trade. This fungibility is only temporary, however, because after physical settlement the buyer observes additional characteristics of each pool that it has received (one or more of hundreds deliverable into the relevant TBA contract), which provide information about prepayment behavior and hence value.

TBA trading effectively applies a common cheapest-to-deliver price level to an intrinsically diverse set of securities and underlying mortgages.

Thus, while the agency MBS market consists of tens of thousands of pools, backed by millions of individual mortgages, trading is concentrated in only a few dozen TBA contracts spread across three maturity points (thirty-year, twenty-year, and fifteen-year mortgages). For each maturity point, there are usually only three or four coupons in active production at any time. TBA trading may occur across a larger number of coupons, reflecting the broader range of coupons in the outstanding stock of agency MBS, which itself reflects the previous path of interest rates. We computed some simple trading summary statistics for calendar years 2010 and 2011 using data from TradeWeb, an agency MBS trading platform (discussed in more detail in Section 3.6) for outright Fannie Mae thirty-year TBAs. For this product, there is positive trading volume for an average of 6.6 different coupons on any given trading day. The most active coupon on each day contributes 49 percent of total trading volume.

Due in part to the concentration of trading in a small number of contracts, market participants are able to place TBA trades in amounts of as much as $100 million to $200 million or

[16] A full calendar of future settlement dates can be found at www.sifma.org.

[17] For evidence of how adverse selection affects the types of securities resecuritized into multiclass MBS, see Downing, Jaffee, and Wallace (2009).

more (for securities backed by individual loans of several hundred thousand dollars each) with a high degree of liquidity. This is reflected in the high trading volumes in the agency MBS market (several hundred billion dollars per day, as reported in Table 1), as well as relatively narrow bid-ask spreads.[18] See Section 3.6 for further discussion of trading volume data in the agency MBS market.

Similar to the MBS pooling process itself, TBA trading simplifies the analytical and risk management challenges for participants in agency MBS markets. Rather than attempting to value each individual security, participants need only to analyze the more tractable set of risks associated with the parameters of each TBA contract. This helps encourage market participation from a broader group of investors, notably foreign central banks and a variety of mutual funds and hedge funds, translating into a greater supply of capital for financing mortgages and presumably lower rates for homeowners.

The treatment of TBA pools as fungible is sustainable in part because a significant degree of actual homogeneity is present among the securities deliverable into any particular TBA contract.

The treatment of TBA pools as fungible is sustainable in part because a significant degree of *actual* homogeneity is present among the securities deliverable into any particular TBA contract. Most notably, each TBA-eligible security carries the same high-quality, GSE-backed credit guarantee on the underlying mortgage cash flows, which essentially eliminates credit risk. However, standardization of underwriting and securitization practices in the agency MBS market contributes meaningfully to homogeneity as well. At the loan level, the standardization of lending criteria for loans eligible for agency MBS constrains the variation among the borrowers and properties underlying the MBS. At the security level, homogenizing factors include the geographic diversification incorporated into the pooling process, the limited number of issuers, the simple structure of "pass-through" security features, and the restriction of the range of interest rates on loans deliverable into a single security. The GSEs' pooling criteria also help assure that pools are relatively homogenous. These criteria include mortgage contract rate ranges (limits on mortgage contract rates, defined relative to the MBS coupon rate) and limits on the distribution of loan age.

3.4 Adverse Selection without Market Failure

Because of the incentives associated with cheapest-to-deliver pricing, not all eligible MBS pools actually trade on a TBA basis. Higher-value pools (those with the most advantageous prepayment characteristics from an investor's point of view) can command a higher price in the less liquid specified pool market.[19] Specified pool trading, as well as the use of "stips," is generally more common for seasoned pools than for newly issued pools, reflecting their lower prepayment risk and therefore higher value. However, specified pools are much less liquid, largely because of the much greater fragmentation of the market.[20]

According to conversations with market participants, a significant volume of physical delivery of securities occurs through the TBA market because, for many securities, the liquidity value of TBA trading generally exceeds any adverse-selection discount implied by cheapest-to-deliver pricing. In part, this is because the significant level of homogeneity in the underwriting and pooling process constrains the variation in value among securities deliverable into a given TBA contract. Paradoxically, the limits on information disclosure inherent in the TBA market seem to actually *increase* the market's liquidity by creating fungibility across securities and reducing information acquisition costs for buyers of MBS. A similar argument explains why DeBeers diamond auctions involve selling pools of diamonds in unmarked bags that cannot be inspected by potential buyers. More generally, the idea that limited information can reduce adverse selection and increase trade is known to economists as the "Hirshleifer paradox" (Hirshleifer 1971).[21]

[18] Emphasizing the lower trading costs and greater liquidity of the TBA market, recent research by Friewald, Jankowitsch, and Subrahmanyam (2012) finds that the average transaction cost of a round-trip trade in the TBA market is only 5 basis points, compared with 48 basis points for MBS specified pool trades.

[19] Note that the term "specified pool" can also apply to an agency MBS that is not deliverable into a TBA contract because it does not meet the good delivery guidelines set by SIFMA. These include pools backed by high-balance mortgages, forty-year mortgages, and interest-only mortgages. These ineligible pools may trade at lower values than do TBAs.

[20] In calendar year 2011, TBA trades (including dollar rolls) made up 94 percent of agency MBS trading, based on TRACE data, including a large volume of both customer trades and dealer-to-dealer trades. This figure includes trading across a wide range of coupons on any given day. See Section 3.6 for more information on TRACE.

[21] See French and McCormick (1984) for a discussion of the DeBeers example. Glaeser and Kallal (1997) present a formal model demonstrating how restricting the set of information provided to MBS investors may enhance liquidity by decreasing information asymmetries and hence opportunities for adverse selection. Dang, Gorton, and Holmström (2009) present a related model in which shocks to fundamentals can generate adverse selection and market freezes.

3.5 Hedging and Financing Mortgages through TBAs

TBAs also facilitate hedging and funding by allowing lenders to prearrange prices for mortgages that they are still in the process of originating, thereby hedging their exposure to interest rate risk. In the United States, lenders frequently give successful mortgage applicants the option to lock in a mortgage rate for a period of thirty to ninety days. Lenders are exposed to the risk that the market price will fluctuate in the period from the time the rate lock is set to when the loan is eventually sold in the secondary market. The ability to sell mortgages forward through the TBA market hedges originators against this risk. It is important for originators to offer applicants fixed-rate loan terms before a mortgage actually closes, which greatly facilitates the final negotiations of house purchases and the overall viability of the thirty-year fixed-rate mortgage as a business line.

TBA trading has . . . led to the development of a funding and hedging mechanism unique to agency MBS: the dollar roll.

Although this price risk could also be hedged with other instruments, TBAs provide superior hedging benefits because of their lower basis risk. Confirming this view, Atanasov and Merrick (2012) find that prices of specified pool trades for TBA-deliverable securities co-move almost perfectly with prices for the corresponding TBA contract, except for trades small (below $25,000) in size. Price movements of Treasury futures, in contrast, can diverge significantly from those of MBS because of movements in prepayment risk premia or changes in relative supply. Mortgage option contracts are more expensive than TBAs, less liquid, and only available for short time horizons (these options are instead used to hedge against variation in the *fraction* of rate locks subsequently utilized by borrowers). While a mortgage futures contract might provide some of the benefits of TBAs, historical attempts to establish a mortgage futures contract in the United States have been unsuccessful (see Nothaft, Lekkas, and Wang [1995] and Johnston and McConnell [1989]). The hedging benefits provided by TBAs will likely be passed on to mortgage borrowers in the form of lower interest rates because of competition among lenders.

TBA trading has also led to the development of a funding and hedging mechanism unique to agency MBS: the dollar roll. A dollar roll is simply the combination of one TBA trade with a simultaneous and offsetting TBA trade settling on a different date. This mechanism allows investors and market makers great flexibility in adjusting their positions for either economic or operational reasons. For example, an investor who has purchased a TBA, but faces operational concerns about receiving delivery as scheduled, could sell an offsetting TBA on that date and simultaneously buy another TBA due one month later, effectively avoiding the operational issue but retaining his economic exposure. An investor could also obtain what amounts to a short-term loan at a favorable rate by selling a TBA for one date and buying another TBA for a later one. For market makers on the other side of such trades, dollar rolls provide an efficient means for maintaining a neutral position while providing liquidity.

A dollar roll transaction is similar to a repurchase agreement (repo), in which two parties simultaneously agree to exchange a security for cash in the near term and to reverse the exchange at a later date.[22] Dollar rolls facilitate financing by providing an alternative and cheaper financing vehicle to the MBS repo, drawing in market participants whose preferences are better suited to the idiosyncrasies of the dollar roll.[23] Note that dollar rolls also simplify the adjustment of originators', servicers', and other market participants' TBA commitments and hedges by reducing not only the total cost but also the cash outlay associated with hedging, because the cost of a dollar roll is only the difference between the prices of two different TBAs.

The ability to lock in TBA forward prices may be particularly useful for smaller originators, who have less access to complex risk management tools that would otherwise be needed to hedge price risk. Some smaller banks already can and do engage in "correspondent" relationships, whereby they sell some or all of their whole loans to larger banks, which then arrange securitization and may be able to negotiate more attractive prices from GSEs. In the absence of a TBA market, this practice might become more widespread. A further consequence could be an increase in the overall share of mortgages originated by the largest commercial banks.[24]

[22] As with a repurchase agreement, these are two separate purchase/sale transactions, but the economic effect is equivalent to secured borrowing/lending. Since the initial exchange of cash and security is reversed, the economic impact is measured by the difference in the prices of the two transactions and the allocation of principal and interest payments over the term of the dollar roll. One fundamental difference is that while the second leg in a repo (reversing the original exchange) requires the return of the original security, in a dollar roll only a "substantially similar" security needs to be delivered, consistent with a definition of substantially similar directly tied to SIFMA's "good delivery" guidelines for TBAs.

[23] For instance, dollar rolls can be used to transfer prepayment risk, since, unlike MBS repos, dollar rolls transfer rights to principal and interest payments over the term of the transaction.

[24] Currently, the four largest commercial banks originate more than half of all mortgages (source: www.mortgagestats.com), a sharp increase compared with the banks' market share before the onset of the financial crisis.

3.6 Price Discovery and Transparency

TBA trading occurs electronically on an over-the-counter basis, primarily through two platforms, DealerWeb (for interdealer trades) and TradeWeb (for customer trades).[25] Quotes on DealerWeb are "live," in the sense that dealers must trade at their posted prices if a counterparty wishes to do so. The TradeWeb platform continuously provides indicative bids and offers (known as Composite Market indicators) for each agency MBS coupon, offering investors "real-time" estimates of the prices at which trades can be executed. While these quotes are indicative, internal Federal Reserve analysis shows that the quotes generally track prices of completed transactions closely.

Since May 2011, market participants that are members of the securities self-regulatory body FINRA (the Financial Industry Regulatory Authority) have been required to report agency MBS trades to the FINRA TRACE (Trade Reporting and Compliance Engine) system. After each trading day, FINRA publicly reports summary statistics of trading volumes and prices for trades completed during the day, such as the weighted average transaction price for different coupons, issuers, and settlement months, and the number and volume of trades.[26] Current coupon MBS prices and spreads between yields on MBS and other assets are also available on Bloomberg and Reuters. These different data sources allow market participants to obtain timely estimates of current market prices for TBA contracts.

TBA trading occurs electronically on an over-the-counter basis.

The TRACE data also illustrate the concentration of MBS trading activity in the TBA segment. According to these data, for calendar year 2011 TBA trading volume (including stips and dollar rolls) is sixteen times larger than trading in specified pools (including pass-through and collateralized mortgage obligation pools), and 187 times larger than trading in non-agency MBS. It is also common to observe trades in the TRACE data exceeding $100 million.[27]

Within the TBA segment, FINRA's TRACE summary reports do not break down the relative volumes of dollar rolls and outright trades; however, as a guide, TradeWeb data analyzed by the Federal Reserve Bank of New York suggest that, on average over 2010 and 2011, 58 percent of thirty-year Fannie Mae trading volume was part of a dollar roll or swap transaction.

3.7 Settlement Volumes

In practice, most TBA trades do not ultimately lead to a transfer of physical MBS. In many cases, the seller will either unwind or "roll" an outstanding trade before maturity, rather than physically settle it. Furthermore, as part of the settlement process, a centralized counterparty operated by the Depository Trust and Clearing Corporation nets all offsetting trades that have been registered with it, greatly reducing the value of securities and cash that must change hands between TBA counterparties.

Even so, TBA trading still generates a large volume of physical MBS settlement. We have conducted some preliminary analysis using Fedwire Securities Service data for the first calendar quarter of 2012 to try to quantify these volumes. During this period, average daily agency MBS settlement volume was $94 billion, representing a mix of TBA, dollar roll, stip, and specified pool transactions. Notably, the three dates with the highest settlement volume corresponded exactly to the three Class A TBA settlement dates in the three-month period. Settlement volume on these dates averaged $418 billion, more than four times the overall daily market average. This evidence suggests that, even though the TBA market is by its nature subject to adverse selection, it is still used as a vehicle for transacting large volumes of physical agency MBS—most likely because of its liquidity.

3.8 Legal Basis of TBA Trading

From a legal perspective, the TBA market, as it currently operates, is made possible by Fannie Mae's and Freddie Mac's exemption from the registration requirements of the Securities Act of 1933 with respect to sales of their MBS. This exemption allows newly issued agency MBS to be offered and sold (including in TBAs) without registration statements filed with the Securities and Exchange Commission (SEC). In contrast, public offers and sales of newly issued private-label MBS are subject to the registration requirements of the Securities Act.

Sales of newly issued agency MBS by way of TBA trading would not be possible without such an exemption because, at the time of a TBA trade, the securities that will eventually be delivered often do not exist. Even if they do exist, the buyer is not told the identity of the specific securities that will be delivered until two days before settlement, which is usually significantly after the trade date itself. Indeed, for many MBS delivered to fulfill TBA contracts, the underlying mortgages

[25] Although agency MBS are not exchange-traded, TBA trades are subject to centralized clearing through a centralized counterparty operated by the Depository Trust and Clearing Corporation.

[26] For the most recent daily report, visit www.finra.org/Industry/Compliance/MarketTransparency/TRACE/StructuredProduct/. Historical summary statistics (by trading day) based on these data are reported by SIFMA at www.sifma.org/research/statistics.aspx.

[27] See Atanasov and Merrick (2012) and Friewald et al. (2012) for a more detailed analysis of agency MBS TRACE data.

Table 2
Loan Limit Timeline

Date	Event
February 13, 2008	Economic Stimulus Act (ESA) temporarily expands conforming loan limit for mortgages originated from July 1, 2007, to December 31, 2008, to the higher of 125 percent of the area median house price or the national baseline level of $417,000, but not to exceed $729,750.
February 15, 2008	Securities Industry and Financial Markets Association (SIFMA) announces high-balance loans will not be eligible for TBA (to-be-announced) trading.
May 6, 2008	Fannie Mae announces "TBA flat" pricing for pools of super-conforming loans.
July 14, 2008	Housing and Economic Recovery Act (HERA) permanently increases loan limits in any area for which 115 percent of the area median house price exceeds the national baseline level of $417,000 to the lesser of 115 percent of the area median or $625,500.
August 14, 2008	SIFMA announces that super-conforming loans up to HERA's permanent limit will be eligible to comprise up to 10 percent of a TBA pool, for super-conforming loans originated on or after October 1, 2008, but only for TBAs settling from January 1, 2009, onward.
Fall 2008	Fannie Mae announces TBA flat pricing will expire on December 31, 2008.
November 2008	Federal Housing Financing Agency publishes list of high-cost areas eligible for permanent HERA-based super-conforming loan limits.
January 1, 2009	ESA's temporary higher limit ($729,750) expires; HERA's permanent limit ($625,500) becomes binding.
February 17, 2009	American Recovery and Reinvestment Act re-establishes the temporary $729,750 maximum limit for all super-conforming loans originated during calendar year 2009.
October 30, 2009	H.R. 2996, Pub. L. No. 111-88, extends the temporary $729,750 maximum limit for mortgages originated through the end of calendar year 2010.
September 30, 2010	H.R. 3081, Pub. L. No. 111-242, extends the temporary $729,750 maximum limit for mortgages originated through the end of fiscal year 2011, or September 30, 2011.
September 30, 2011	Temporary limits expire; permanent limits become binding.

have not even been originated as of the trade date (enabling the hedging described in the previous section).

In practice, while offers and sales of GSE MBS are exempt from SEC registration requirements, the agencies do publicly disclose summary information about the composition of each pool. This information includes the average loan-to-valuation ratio, debt-to-income ratio, borrower credit score, the number and value of mortgages from each U.S. state, weighted average mortgage coupon rates and maturities, and broker versus non-broker origination channels. Nevertheless, at the time of trade, the TBA buyer lacks access to this information simply because it does not know which securities it will receive.

4. TBA Eligibility of Super-Conforming Loans

4.1 Increases in Conforming Loan Limits in 2008

Recent changes in the conforming loan limits provide a useful natural experiment to study the price impact of TBA eligibility, even for agency MBS pools that already enjoy a credit guarantee. As discussed, Fannie Mae and Freddie Mac are prohibited from purchasing mortgages larger than a set of conforming loan limits set by the Federal Home Financing Agency. The FHFA adjusts these limits annually in line with the general level of home prices.[28] As the U.S. housing market deteriorated in 2007 and mortgage market stresses increased, market participants and policymakers looked to the GSEs to support the housing sector in a variety of ways, for example, by expanding their retained portfolios, and raised the conforming loan limit to allow the GSEs to support a broader range of residential mortgages, particularly the prime jumbo market.

The ensuing debate culminated in the Economic Stimulus Act of 2008 (ESA), passed on February 13, which temporarily raised the conforming loan limit in designated "high-cost" areas through December 31, 2008, to as much as $729,750 from a previous national level of $417,000.[29] Maximum FHA limits in high-cost markets were also temporarily increased to the same levels as those applying to Fannie Mae and Freddie Mac. Further permanent changes to conforming loan limits were announced later in 2008, as described below and presented in Table 2.

[28] Under the 2008 Housing and Economic Recovery Act (HERA), the national conforming loan limit is set according to changes in average home prices over the previous year, but it cannot decline from year to year.

[29] "High-cost areas" are designated by the FHFA based on median home values in a given county as estimated by the FHA. The figures given above are for single-family homes; higher limits apply to multifamily dwellings.

4.2 TBA Deliverability of Super-Conforming Mortgages

While the GSEs' purchases are authorized by Congress and regulated by the FHFA, TBA trading conventions are set by SIFMA. Two days after enactment of the ESA, SIFMA announced that high-balance loans ("super-conforming loans") between $417,000 and the new, higher conforming loan limits would not be eligible for inclusion in TBA-eligible pools.[30] Instead, these pools could only be securitized as a new category of specialty products and traded as specified pools. In testimony to Congress in May 2008, SIFMA explained its opposition to allowing the new super-conforming loans to be included in TBA-eligible pools.[31] Two main concerns were cited: First, the initial increases in conforming loan limits were temporary, expiring at the end of 2008. SIFMA judged that the addition and subtraction of super-conforming loans from TBA pools over such a short horizon could cause significant market disruption. Second, including super-conforming loans would undermine the homogeneity underpinning the TBA market. SIFMA noted that mortgages with high principal balances tend to be prepaid more efficiently, reflecting the greater sophistication of the underlying borrowers and the larger dollar amount of incentive for optimal exercise of the prepayment option (given the larger loan balance). This could therefore establish a new and lower cheapest-to-deliver price for TBAs, making it less attractive to deliver standard conforming pools into TBA trades, thereby reducing the liquidity of these standard pools. The inclusion of super-conforming pools could also make TBAs a less effective tool for hedging price risk for other MBS pools.

To support the super-conforming market in the face of this lack of TBA eligibility, Fannie Mae announced on May 6, 2008, that it would purchase pools of super-conforming loans at a price on par with TBA-eligible pools throughout the remainder of 2008.[32] Supported by this announcement, the issuance of super-conforming specified pools increased over the summer of 2008, and the underlying loans were originated at primary mortgage rates close to those for standard conforming loans.

[30] To our knowledge, there is no generally accepted term to describe loans between the national conforming loan limit and high-cost housing area limits. Other terms sometimes used to describe these mortgages are "high-balance conforming" loans and "jumbo-conforming" loans. Both these names are potentially confusing: the first because loans near to but below the national limit are also sometimes called high-balance conforming loans; the second because the term jumbo-conforming could also be interpreted to mean prime jumbo loans. For this reason, we use the term "super-conforming" to refer to these mortgages.

[31] Written testimony by SIFMA Vice Chairman Thomas Hamilton to the House Committee on Financial Services, May 22, 2008 (www.sifma.org/legislative/testimony/pdf/Hamilton-052208.pdf).

[32] This commitment expired in December 2008. Yet in October 2008, Fannie Mae, in an effort to ease the transition to market-based pricing, promised to continue in 2009 to purchase super-conforming mortgages originated in 2008, but with a 175-basis-point fee added to the TBA mortgage rates.

Nevertheless, the U.S. housing market continued to deteriorate, and on July 14, 2008, Congress passed the Housing and Economic Recovery Act (HERA), permanently increasing to $625,500 the conforming loan limit in high-cost areas.[33]

Noting the permanent nature of this change, SIFMA announced a month later that super-conforming loans up to the HERA limit would be TBA eligible. However, it imposed a *de minimis* limit—that is, super-conforming loans could represent at most 10 percent of a TBA pool. The announcement had little immediate market impact because of Fannie Mae's previous commitment to purchase super-conforming loans at "TBA flat" pricing. However, it proved critical in 2009 as Fannie Mae's price support expired.

4.3 Further Adjustments to Conforming Loan Limits

The temporary conforming loan limits (up to $729,750) established under the ESA expired at the end of 2008. However, in February 2009 these temporary limits were reestablished, and in November 2009 they were extended until the end of 2010. On September 30, 2010, the temporary limits were extended for another year. They finally expired on September 30, 2011.

5. Effects on the MBS Market

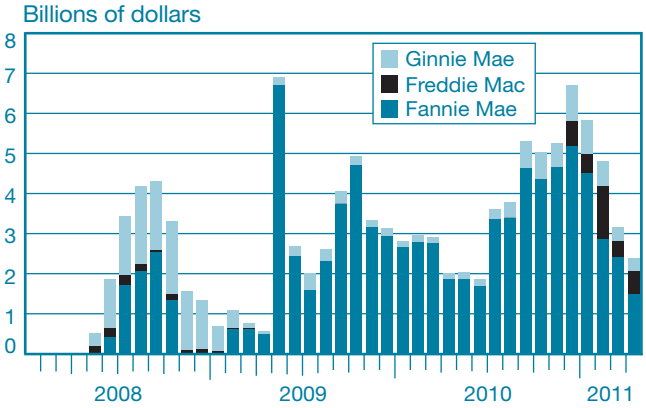
5.1 Issuance of Super-Conforming MBS Pools

Chart 2 presents data on the issuance of super-conforming MBS since the ESA raised the agency loan limits in February 2008. As the chart shows, issuance of super-conforming pools has been volatile. There was little issuance of MBS backed by super-conforming mortgages in the months immediately following passage of the ESA, reflecting the TBA ineligibility of these loans and the time needed by issuers to set up their super-conforming securitization program. Spurred by Fannie Mae's announcement that it would purchase pools backed by super-conforming loans at par to TBA pricing, issuance of super-conforming specified pools grew during summer 2008, concentrated in Fannie Mae and Ginnie Mae pools.

[33] HERA uses a slightly different calculation methodology from ESA for identifying high-housing-cost areas, complicating comparison of the two sets of high-balance loan limits.

Chart 2

Issuance of Non-TBA-Eligible Super-Conforming Mortgage-Backed Securities (MBS)



Source: Federal Reserve Bank of New York.

Notes: The chart shows the monthly issuance of non-TBA-eligible MBS backed by loans above the national conforming loan limit of $417,000 sponsored by Fannie Mae, Freddie Mac, and Ginnie Mae. Note that after August 2008, the Securities Industry and Financial Markets Association allowed super-conforming loans to be securitized in TBA-eligible pools in *de minimis* amounts (up to 10 percent of the pool balance). The chart does not include super-conforming securitizations through these TBA-eligible pools.

Issuance of super-conforming pools dropped sharply in fall 2008. This decrease likely reflected both the overall turmoil in financial markets during this period as well as uncertainties specific to super-conforming loans that may have discouraged originators from extending such loans. First, lenders faced significant regulatory risk because the FHFA did not publish until November 2008 its list of "high-cost" census tracts eligible for the permanent higher-loan limits. In addition, market participants were uncertain how prices for super-conforming loans would respond to the expiration of Fannie Mae's commitment to TBA-equivalent pricing, and some originators may have simply waited to deliver their super-conforming loans into TBA pools starting in January 2009.

Issuance of super-conforming pools remained low between January and April 2009, reflecting the withdrawal of Fannie Mae's price support for this market.[34] Super-conforming issuance rose more steeply in summer 2009, likely for two reasons: the sharp rise in mortgage rates during this period led many borrowers to close on pending mortgages out of fear that rates would rise further, and bank-driven demand for short-duration CMO tranches rose during that period, increasing demand for faster-prepaying agency MBS.[35]

[34] Although overall issuance was low during this period, super-conforming issuance rose modestly in February 2009, as pressures on financial institutions' balance sheets began to subside.

[35] A CMO is a structured MBS that distributes payments and prepayments of mortgage principal across a number of different tranches in order of seniority. Banks tend to demand more short-duration CMO tranches in steep yield curve environments to avoid an asset-liability mismatch when rates rise.

5.2 Secondary-Market Pricing of Super-Conforming MBS Pools

Table 3 presents data on the price premium (or discount) for Fannie Mae super-conforming pools relative to standard TBA-eligible pools between first-quarter 2009 and first-quarter 2010.[36] During this span, corresponding to the period after Fannie Mae's price support of super-conforming loans expired in December 2008, super-conforming pools consistently traded at a significant discount to the corresponding TBA contract. The average discount is 1.1 percent of MBS par value, averaging through time and across securities with different coupons (the "coupon stack"). Applying a simple "rule-of-thumb" that MBS have an approximate duration of four years, we see that this figure corresponds to an average difference in *yield* of 27.5 basis points.[37]

Three possible explanations for this price discount are: 1) the price differential reflects an illiquidity discount for super-conforming pools, since these pools trade on a specified pool basis, rather than in the TBA market; 2) the price discount reflects greater prepayment risk for super-conforming pools; 3) the higher price for TBA pools reflects the effects of the Federal Reserve MBS purchase program, which purchased only TBA-eligible MBS.

It is difficult, and beyond the scope of this article, to fully disentangle the prepayment and liquidity risk explanations. However, we note that during this period the super-conforming price discount was persistent and relatively homogenous across the coupon stack. This is notable, because differences in prepayment risk would be expected to have a larger price impact on securities trading further from par (that is, when the coupon rate is significantly different from the market yield). We view the relative consistency of the discount across the coupon stack as evidence suggesting that illiquidity, and not just differences in prepayment characteristics, is likely to be an important explanation for the spreads observed in Table 3. Furthermore, these super-conforming pools were sought after as collateral for the growing CMO market precisely because of their higher prepayment rates, suggesting that the price discount reflected in Table 3 may be lower than would otherwise have been the case.

It is easier to rule out the effects of the Federal Reserve's purchases of TBAs as an explanation for the discounts in Table 3, since the Fed's purchases were completed in March 2010. The fact that we observe little change in the price discount around

[36] While the table focuses on Fannie Mae pools, data for Ginnie Mae super-conforming pools indicate a similar price discount. The magnitude of the price discount is less uniform than it is for Fannie Mae pools, however, likely reflecting the lower issuance volumes and consequent lack of liquidity.

[37] Duration is a measure of the maturity of a fixed-rate security or, equivalently, its sensitivity to movements in interest rates. A duration of four years implies that a 1 percent change in yields is associated with a 4 percent change in price. Note that this market rule-of-thumb estimate of MBS duration is approximate—because future prepayment rates are unknown, the expected duration of an MBS will fluctuate over time because of variation in market conditions and the term structure of interest rates.

Table 3
Price Discounts on Fannie Mae Super-Conforming Pools
Percent

	2010				2009				Average
Coupon	Q4	Q3	Q2	Q1	Q4	Q3	Q2	Q1	
3.5	-1.1	-0.8							-0.9
4.0	-0.8	-1.1	-1.3	-1.2	-1.1	-1.6	-1.0		-1.1
4.5	-1.4	-0.8	-0.5	-0.5	-0.6	-1.2	-0.9	-1.6	-0.9
5.0	-1.7	-1.2	-0.6	-0.8	-0.9	-1.0	-0.9	-1.4	-1.1
5.5		-1.3	-1.3	-1.3	-1.1	-1.2	-1.0	-1.3	-1.2
6.0					-1.3	-1.2	-1.1	-1.1	-1.2
6.5					-1.3	-1.3	-1.3	-1.3	-1.3
Average	-1.3	-1.0	-0.9	-0.9	-1.1	-1.2	-1.0	-1.3	-1.1

Sources: Federal Reserve Bank of New York; Fannie Mae; authors' calculations.

Notes: Pools of Fannie Mae super-conforming loans are marketed with a "CK" CUSIP prefix, in contrast to the "CL" prefix used to reference a benchmark Fannie Mae fixed-rate thirty-year TBA-eligible pool. Data show the indicative price difference between CK and CL pools, obtained from the trading desks of two significant market participants, measured as a percentage of MBS par value. A negative value indicates a price discount for CK pools for the coupon and quarter indicated.

this time suggests that Fed MBS purchases are not likely to be an important source of the price differential between TBA-eligible and TBA-ineligible securities.[38]

6. Effects on Primary Market Mortgage Supply

6.1 Effects on Mortgage Pricing

The secondary-market price discount for super-conforming pools, shown in Table 3, also translated into higher interest rates for mortgage borrowers. Chart 3 shows how mortgage rates on super-conforming mortgages compare with jumbo and standard conforming rates during the crisis period. Overall, rates for super-conforming loans were quite close to conforming rates over the period when the government-sponsored enterprises were permitted to securitize such loans (suggesting that the credit guarantee provided by the GSEs is the primary driver of the difference in mortgage rates between the jumbo and conforming markets). However, the rates did not fully converge: Super-conforming rates remained above those for standard conforming loans over this entire period, consistent with the secondary-market price discounts shown in Table 3.

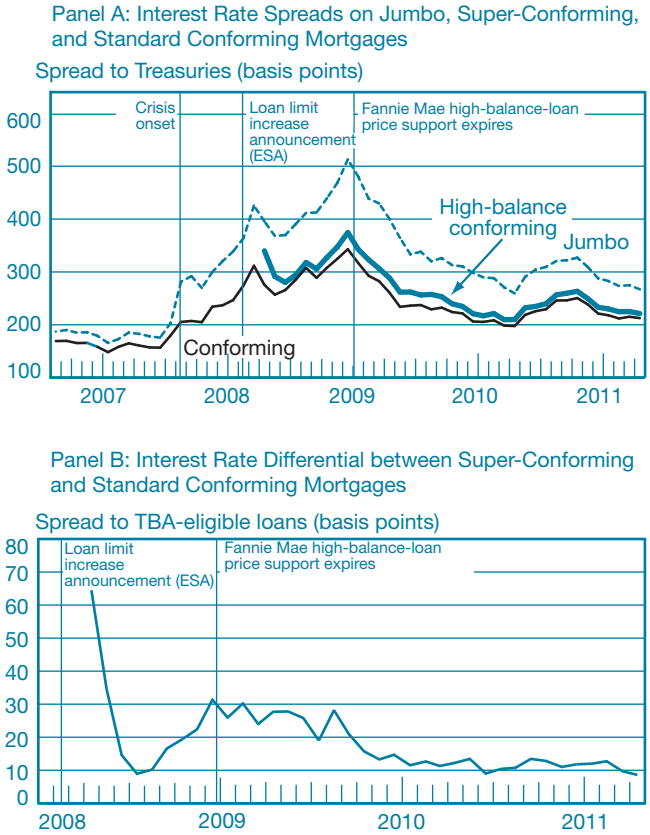
Panel B of Chart 3 focuses on trends in the interest rate spread between super-conforming mortgages and standard-conforming mortgages. The spread declined sharply after Fannie Mae announced that it would begin purchasing super-conforming mortgages at par to TBA prices. It rose to around 30 basis points toward the end of 2008 and early 2009, reflecting the rise in liquidity premia during the financial crisis, as well as the expiration of Fannie Mae's price support for the super-conforming market. The interest rate premium on super-conforming loans then declined over 2009 and 2010, as market conditions normalized to around 12 basis points by mid-2010.

One limitation of our results is that the primary-market interest rate spread is a useful but imperfect measure of the liquidity premium associated with TBA eligibility. First, mortgages above the conforming loan limit are still partially TBA eligible, since they can be included in *de minimis* amounts (up to 10 percent of the total pool size) in TBA pools. This would lead the spread in Chart 3, panel B, to underestimate the benefits of TBA eligibility (since we are not comparing TBA-eligible and TBA-ineligible loans, but instead eligible and partially eligible loans). Second, however, loans in super-conforming pools may have different prepayment characteristics, or have different transaction costs because of their larger size, driving part of the difference in primary market yields. While the uniformity of the secondary-market price discount across the coupon stack suggests that this prepayment risk explanation is not dominant, it is difficult to state definitively how large a role it plays.

[38] One explanation why the Federal Reserve's LSAP programs may not lead to a price differential between TBA-eligible and TBA-ineligible securities is the presence of "portfolio balance" effects, namely, that the programs affect prices for securities purchased and securities that are close substitutes. Gagnon et al. (2010) present evidence consistent with portfolio balance effects for the LSAP programs.

Chart 3

Mortgage Spreads on Jumbo, Super-Conforming, and Standard Conforming Loans

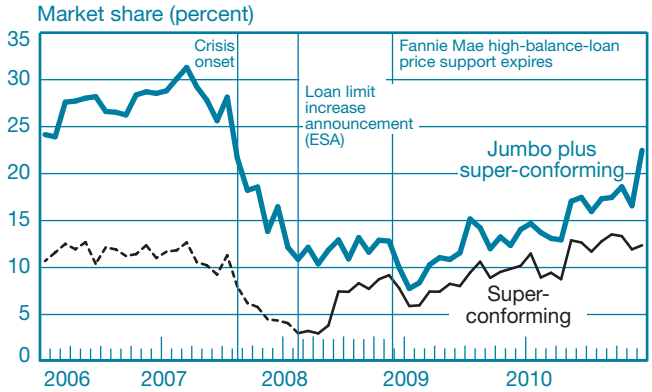


Source: HSH Associates.

Notes: Mortgage rates are expressed as a spread to the average of the five- and ten-year Treasury yield. ESA is the Economic Stimulus Act.

Chart 4

Share of Mortgage Originations in the Super-Conforming Segment



Source: Lender Processing Services.

Note: The chart plots the total value-weighted fraction of mortgage originations above $417,000 (blue line) and the fraction of "super-conforming" mortgage originations between $417,000 and the temporary loan limits established under the Economic Stimulus Act (black line). Recall that under the Act, conforming loan limits in high-housing-cost areas were increased to as much as $729,750. The dashed segment of the black line represents the fraction of loans that fell between $417,000 and the high-balance limits in the period before passage of the Act.

6.2 Effects on the Quantity of Credit

Chart 4 shows the fraction of the dollar volume of new mortgages whose size exceeds the national single-family conforming loan limit of $417,000 as well as the fraction of mortgages between the national conforming limit and the higher super-conforming loan limits introduced under the ESA.[39]

Also shown in Chart 4, the origination share of super-conforming mortgages (those with principal amounts above $417,000) decreased sharply in the second half of 2007, as both non-agency MBS markets and bank balance sheets came under extreme stress and house prices declined. Strikingly, however, after the conforming loan limits were raised in February 2008, the share of loans between $417,000 and these super-conforming limits began to rise significantly, from less than 5 percent in early 2008 to nearly 15 percent by the end of 2010. In contrast, the market share of jumbo mortgages above the super-conforming limits (measured as the difference between the two lines plotted in Chart 4) remains far below pre-crisis levels, even through late 2010.

Together, Charts 3 and 4 suggest that the decision to make super-conforming loans eligible for agency securitization significantly increased secondary-market demand for this class of mortgages; this correspondingly increased the supply of mortgage credit for the super-conforming market segment, increasing the quantity of loans that eligible homeowners could obtain and reducing mortgage interest rates. The majority of this increase in mortgage supply reflects the direct effect of the government guarantee. But the fact that super-conforming rates did not fully converge to standard conforming rates, as well as the evidence presented in Section 5, suggests that secondary-market MBS liquidity also influences the availability and affordability of mortgage credit.[40]

[39] These shares are calculated using loan-level data from Lender Processing Services (LPS). To calculate the share of loans between $417,000 and the new super-conforming loan limits, we geographically match each mortgage in the LPS data to the conforming loan limits applicable in that county at the time the mortgage was originated.

[40] See also Fuster and Vickery (2013) for detailed evidence of how access to securitization affected mortgage supply for different types of loans during this episode, based on loan-level data and difference-in-differences methods.

6.3 Interpretation

The evidence presented above suggests that the liquidity associated with TBA eligibility increases MBS prices and lowers mortgage interest rates, consistent with evidence in other fixed-income markets, such as the "old bond" illiquidity discount in the Treasury market documented by Krishnamurthy (2002) and others. We strive to be somewhat cautious in our interpretation,

The evidence . . . suggests that the liquidity associated with TBA eligibility increases MBS prices and lowers mortgage interest rates, consistent with evidence in other fixed-income markets.

however, because pricing differences between conforming and super-conforming loans may also reflect differences in prepayment risk, at least in part. While we present some evidence on this point, our analysis does not allow us to fully quantify the relative importance of prepayment risk. Conducting a more detailed statistical analysis—for example, using loan-level data to exploit variation in loan size around the TBA-eligibility limits—would be an interesting topic for future research.

With this caveat in mind, our preliminary assessment of this evidence is that: the premium associated with TBA eligibility is likely about 10 to 25 basis points on average over 2009 and 2010, and this premium is magnified during periods of market stress or disruption, consistent with evidence from other fixed-income markets (recall Section 2.2). For example, the primary-market spread between TBA-eligible and TBA-ineligible mortgages was as large as 65 basis points (when the conforming loan limit was first raised in March 2008 and there was no secondary market at all for super-conforming mortgages) and 25 to 30 basis points at the start of 2009 (when Fannie Mae's price support for super-conforming loans first expired and the financial crisis was still near its peak). The spread then declined steadily over 2009 and 2010, to around 9 to 12 basis points, as financial market conditions gradually normalized.

7. Prospects for the TBA Market amid Housing Finance Reform

Congress and the U.S. Treasury Department continue to consider different options for reshaping the housing GSEs Fannie Mae and Freddie Mac. As part of this process, the Treasury has published a paper discussing a number of prominent policy options (Department of the Treasury and Department of Housing and Urban Development 2011). Market observers, as well as Federal Reserve Chairman Ben Bernanke and former Treasury Secretary Henry Paulson, have considered a spectrum of GSE reform options ranging from full privatization to full nationalization. Intermediate options between these extremes include an industry-owned mortgage cooperative,[41] the introduction of a public tail-risk insurer, covered bonds, and the conversion of Fannie Mae and Freddie Mac into "public utilities."

Perhaps surprisingly, many discussions of mortgage finance reform make little mention of the TBA market or secondary-market trading more generally. Preservation of a liquid TBA market in something akin to its present form is likely compatible with a number of different market structures and should not be viewed as a reason to avoid reform *per se*. However, given the central role currently played by the TBA market, it is important to consider how different reform options could affect the operation and liquidity of this market. There is little consensus on exactly how much actual homogeneity in the underlying mortgages and securities is necessary to support the fungibility and liquidity created by the TBA market, as demonstrated by SIFMA's concerns regarding super-conforming loan eligibility and other revisions to TBA delivery guidelines. However, beyond some unknown point, fragmentation of the MBS market through greater diversity of loan and MBS features would likely reduce liquidity. In contrast, standardization of documentation, structuring, and mortgage underwriting criteria within the TBA-eligible universe is likely important to help maintain fungibility across securities, and thus promote market liquidity.

As a matter of law, a fully private TBA market might be possible with sufficient amendments to current securities law. The key would be to provide exceptions to the Securities Act of 1933 for private mortgage securities, such that commitments to purchase mortgage pools could become binding before the receipt of the pool's prospectus. However, such changes could be challenging given the current trend in securities law toward greater disclosure. In addition, it is unclear whether greater disclosure could itself impair the operation of the TBA market, by increasing sellers' ability to discriminate value among MBS pools and leading to greater adverse selection, siphoning off the most valuable securities into the specified pool market.

The history of the TBA market illustrates that the consequences of changes to market structure are unpredictable and sometimes negative. One example is the failure of mortgage

[41] See Dechario et al. (2010) for one proposed design of a cooperative model.

futures contracts that have been launched several times over recent decades (Johnston and McConnell 1989). In another example, Freddie Mac's decision to alter the timing of payments to MBS holders was poorly received by market participants, contributing to a negative spread between Freddie Mac and Fannie Mae MBS that persists more than twenty years later.

In conclusion, this article has described the mechanics of the TBA market and presented summary statistics documenting its substantial size and liquidity. We have also provided preliminary evidence suggesting that its liquidity raises market prices and lowers mortgage interest rates for TBA-eligible loans. Our interpretation of the existing evidence is that these liquidity effects are of the order of 10 to 25 basis points on average during 2009 and 2010, and are magnified during periods of greater market stress. These estimates are consistent with statistical estimates in the academic literature for liquidity premia on other government-guaranteed bonds. Our discussion and preliminary evidence therefore suggest that agency MBS liquidity is not solely attributable to implicit government guarantees, and that the structure of secondary markets can significantly affect MBS liquidity and thereby influence borrowing rates paid by households. This in turn suggests that evaluations of proposed reforms to the U.S. housing finance system should take into account the potential effects of those reforms on the operation of the TBA market and its liquidity.

References

Ambrose, B. W., M. LaCour-Little, and A. B. Sanders. 2004. "The Effect of Conforming Loan Status on Mortgage Yield Spreads: A Loan-Level Analysis." Real Estate Economics 32, no. 4 (December): 541-69.

Amihud, Y., and H. Mendelson. 1991. "Liquidity, Maturity, and the Yields on U.S. Treasury Securities." Journal of Finance 46, no. 4 (September): 1411-25.

Atanasov, V., and J. Merrick Jr. 2012. "Liquidity and Value in the Deep vs. Shallow Ends of Mortgage-Backed Securities Pools." Unpublished paper. Available at papers.ssrn.com/sol3/papers.cfm?abstract_id=2023779.

Beber, A., M. W. Brandt, and K. A. Kavajecz. 2009. "Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market." Review of Financial Studies 22, no. 3 (March): 925-57.

Dang, T. V., G. Gorton, and B. Holmström. 2009. "Opacity and the Optimality of Debt for Liquidity Provision." Unpublished paper, Yale University.

Dechario, T., P. Mosser, J. Tracy, J. Vickery, and J. Wright. 2010. "A Private Lender Cooperative Model for Residential Mortgage Finance." Federal Reserve Bank of New York Staff Reports, no. 466, August.

Department of the Treasury and Department of Housing and Urban Development. 2011. Reforming America's Housing Finance Market: A Report to Congress. Available at: www.treasury.gov/initiatives/documents/reforming%20america's%20housing%20finance%20market.pdf.

Downing, C., D. Jaffee, and N. Wallace. 2009. "Is the Market for Mortgage-Backed Securities a Market for Lemons?" Review of Financial Studies 22, no. 7 (July): 2457-94.

Fleming, M. 2002. "Are Larger Treasury Issues More Liquid? Evidence from Bill Reopenings." Journal of Money, Credit, and Banking 34, no. 3 (August): 707-35.

French, K. R., and R. E. McCormick. 1984. "Sealed Bids, Sunk Costs, and the Process of Competition." Journal of Business 57, no. 4 (October): 417-41.

Friewald, N., R. Jankowitsch, and M. G. Subrahmanyam. 2012. "Liquidity, Transparency, and Disclosure in the Securitized Product Market." Unpublished paper, New York University Stern School of Business.

Fuster, A., and J. Vickery. 2013. "Securitization and the Fixed-Rate Mortgage." Federal Reserve Bank of New York Staff Reports, no. 594, January.

Gagnon, J., M. Raskin, J. Remache, and B. Sack. 2010. "Large-Scale Asset Purchases by the Federal Reserve: Did They Work?" Federal Reserve Bank of New York Staff Reports, no. 441, March.

Glaeser, E. L., and H. D. Kallal. 1997. "Thin Markets, Asymmetric Information, and Mortgage-Backed Securities." Journal of Financial Intermediation 6, no. 1 (January): 64-86.

Green, R. K., and M. LaCour-Little. 1999. "Some Truths about Ostriches: Who Doesn't Prepay Their Mortgages and Why They Don't." Journal of Housing Economics 8, no. 3 (September): 233-48.

Hirshleifer, J. 1971. "The Private and Social Value of Information and the Reward to Inventive Activity." American Economic Review 61, no. 4 (September): 561-74.

Johnston, E. T., and J. J. McConnell. 1989. "Requiem for a Market: An Analysis of the Rise and Fall of a Financial Futures Contract." Review of Financial Studies 2, no. 1 (January): 1-23.

Krishnamurthy, A. 2002. "The Bond/Old-Bond Spread." Journal of Financial Economics 66, nos. 2-3 (November-December): 463-506.

Longstaff, F. A. 2004. "The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices." Journal of Business 77, no. 3 (July): 511-26.

McKenzie, J. A. 2002. "A Reconsideration of the Jumbo/Non-Jumbo Mortgage Rate Differential." Journal of Real Estate Finance and Economics 25, nos. 2-3 (September-December): 197-213.

Nothaft, F. E., V. Lekkas, and G. H. K. Wang. 1995. "The Failure of the Mortgage-Backed Futures Contract." Journal of Futures Markets 15, no. 5 (August): 585-603.

Nothaft, F. E., J. E. Pearce, and S. Stevanovic. 2002. "Debt Spreads between GSEs and Other Corporations." Journal of Real Estate Finance and Economics 25, nos. 2-3 (September-December): 151-72.

Passmore, W., S. M. Sherlund, and G. Burgess. 2005. "The Effect of Housing Government-Sponsored Enterprises on Mortgage Rates." Real Estate Economics 33, no. 3 (September): 427-63.

Schwartz, A. 2006. "Household Refinancing Behavior in Fixed-Rate Mortgages." Unpublished paper, Harvard University.

Schwarz, K. 2009. "Mind the Gap: Disentangling Credit and Liquidity in Risk Spreads." Unpublished paper, University of Pennsylvania. Available at ssrn.com/abstract=1486240.

Torregrosa, D. 2001. "Interest Rate Differentials between Jumbo and Conforming Mortgages, 1995-2000." Congressional Budget Office *CBO Paper*, May.

Rajashri Chakrabarti and Noah Schwartz

Unintended Consequences of School Accountability Policies: Evidence from Florida and Implications for New York

- A key question for educators is whether accountability policies linked to measurable performance outcomes induce schools to "game the system," rather than make genuine improvements.
- This study of an influential Florida program allowing students from failing schools to transfer to better ones suggests that the failing schools engaged in differential classifications of students into exempt categories to artificially boost accountability.
- The finding that schools resort to strategic classifications offers lessons for the design of accountability programs elsewhere, including New York City's Progress Reports program and New York's implementation of the federal No Child Left Behind Act.

1. Introduction

Over the past two decades, state and federal education policies have increasingly emphasized school accountability. This approach focuses on the assignment of rewards and sanctions for schools based on measurable outcomes, usually student performance on standardized tests. A common criticism of accountability policies is that they may induce schools to "game the system" along with—or instead of—making genuine educational improvements. This article investigates whether schools resorted to such strategic behavior in response to the Florida Opportunity Scholarship Program (FOSP), an influential accountability policy that made students from low-performing schools eligible for vouchers to transfer to better ones. Our findings have important implications for New York City's Progress Reports program and New York's implementation of the federal No Child Left Behind (NCLB) Act, which were modeled on the Florida program but contain crucial design changes.

Rajashri Chakrabarti is an economist at the Federal Reserve Bank of New York; Noah Schwartz is a former assistant economist at the Bank.
Correspondence: rajashri.chakrabarti@ny.frb.org

The authors thank David Figlio, Sarah Turner, and participants at Duke University, Harvard University, the Massachusetts Institute of Technology, the American Economic Association, the American Education Finance Association, the Econometric Society, and the Society of Labor Economists conferences for helpful discussions, the Florida Department of Education for data, and Brandi Coates for excellent research assistance. The views expressed 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.

Starting in the 1998-99 school year,[1] Florida began assigning letter grades to schools on a scale of A to F based on student performance on statewide standardized tests.[2] The Florida Opportunity Scholarship Program, introduced in June 1999,

Did the exemptions for certain LEP [limited-English-proficient] and ESE [exceptional student education] students induce schools to classify some weaker students into these excluded categories to remove them from school-grade calculations and artificially boost scores?

embedded a voucher program within this accountability system. It made students from low-performing schools eligible for vouchers to transfer to private schools and higher-performing public schools. Specifically, students from any school receiving two F grades in four years were made eligible for vouchers. These vouchers were funded by public school revenue, with funds following students to their new schools. Thus, FOSP can be viewed as a "threat of vouchers" program—schools receiving an F grade for the first time were at risk of being subjected to vouchers, but vouchers were actually issued only if the school received another F grade in the next three years.

Consider the incentives faced by a school threatened by vouchers after receiving its first F grade. As the lowest grade, that mark was associated with stigma, especially because of the publicity and visibility these grades drew. In addition, vouchers were associated with a loss of revenue and shame. As a result, threatened schools had strong incentives to avoid receiving another F grade. This article studies how schools may have responded to this risk, given the features of the program.

Under Florida rules, the test scores of certain high-needs students were excluded from the calculation of school grades, presumably to avoid penalizing schools with large numbers of such students. One exempted category was limited-English-proficient (LEP) students who were in an English-for-speakers-of-other-languages (ESOL) program for less than two years. Several types of special-education (exceptional student education, or ESE) students were also exempted, as we discuss.

[1] Going forward, we refer to school years by the calendar year of the spring semester.

[2] Florida had a different accountability system in place before 1999. This system assigned numeric grades of I-IV (I-lowest, IV-highest) to schools based on test scores.

The features of this program motivate an important question: Did the exemptions for certain LEP and ESE students induce schools to classify some weaker students into these excluded categories to remove them from school-grade calculations and artificially boost scores?

Using data from the Florida Department of Education and a regression-discontinuity estimation strategy, we look for any evidence of increased classifications of students into these excluded categories after the introduction of the program. The regression-discontinuity approach essentially entails comparing schools that just barely avoided an F with ones that just barely received an F. Arguably, these two groups are very similar, and only differ in that the first was not threatened by the program while the second was. So, a comparison is expected to yield a causal estimate of the effect of FOSP. Employing this technique, we find that the program led to increased classification of students into the excluded LEP category in the high-stakes grade 4 and in grade 3, the entry grade for that high-stakes year, following the program's inception. Specifically, schools threatened by the program elected to classify as excluded LEP an additional 0.31 percent of students in grade 4 and an additional 0.36 percent of students in grade 3 in the year after the program was implemented. In contrast, we find no evidence that the threatened schools resorted to increased classification into excluded ESE categories in

[Our] findings suggest the use of strategic classifications into excluded categories by the failing schools after the inception of the [Florida Opportunity Scholarship Program (FOSP)].

that school year. As we discuss, ESE classification was associated with substantial costs during this period,[3] which might have discouraged this form of classification. These findings suggest the use of strategic classifications into excluded categories by the failing schools after the inception of the program.[4]

This article is related to two strands of literature. The first studies the effect on public school performance of voucher programs, "threat of voucher" programs, and programs that incorporate threat of vouchers and stigma. This literature generally finds positive effects of school accountability

[3] We argue that Florida's McKay Scholarship program for students with disabilities acted as a major disincentive to such classification. Since it made every student with a disability in Florida public schools eligible for vouchers, schools that classified students into ESE categories risked losing these students and the corresponding per-pupil funding.

programs on public school performance in the United States.[5] The second strand investigates whether schools facing accountability systems respond by gaming the system. Researchers have presented evidence of various types of strategic behaviors: reclassification of weaker students into exempted disability categories, suspensions of weaker students during the testing period, teacher cheating, increased focus on high-stakes marginal students, and even strategic boosting of the caloric content of school lunches on testing days.[6]

[Our findings] from Florida have important implications for other programs, including the major school accountability policies in the New York region.

Despite the wealth of literature on gaming behaviors of public schools facing accountability systems, it is not immediately obvious that schools facing accountability-tied sanctions will behave in a similar way. Understanding the incentives and behaviors of public schools in such systems is becoming more relevant in today's world due to the shift toward education policies incorporating sanctions as their centerpiece. This article diverges from and advances this literature by analyzing whether accountability-tied sanctions (specifically vouchers) induce schools to behave in similar strategic ways.[7]

Our findings from Florida have important implications for other programs, including the major school accountability policies in the New York region. New York City's Progress Reports program and New York's implementation of the federal No Child Left Behind Act were both modeled in part on the Florida program, tying sanctions (including school choice) and rewards to student test scores and other measurable outcomes. Importantly, though, both policies contain design differences that should discourage the type of gaming that might have occurred in Florida. These programs incorporate into accountability measures the performance of all students, including limited-English-proficient, special education, and other subgroups. In fact, New York City even gives "extra credit" to schools for achieving progress with English-language learners, special education students, and other high-needs groups. Therefore, schools have no adverse incentives to resort to strategic reclassification of low-performing students into special education and limited-English-proficient categories. We do note, though, that these rules can cause their own type of gaming, perhaps inducing schools to classify their higher-performing students into these groups in an effort to artificially boost their scores and grades.

2. Program Details

The Florida Opportunity Scholarship Program, introduced in June 1999, made students from the worst-performing public schools eligible for vouchers ("opportunity scholarships") to attend private schools and higher-performing public schools. Under the program, all students of a public school became eligible for vouchers if the school received two F grades in a period of four years. A school receiving an F grade for the first time was exposed to the threat of vouchers, but vouchers were not implemented unless and until it received a second F within the next three years. Vouchers resulted in loss of revenue and negative publicity. Moreover, the F grade, being the lowest-performing grade, was associated with stigma and shame.

The Florida Opportunity Scholarship Program . . . made students from the worst-performing public schools eligible for vouchers . . . to attend private schools and higher-performing public schools.

School grades were based on student performance on the Florida Comprehensive Assessment Test (FCAT). The FCAT writing test was first administered in 1993. Following a field test in 1997, the FCAT reading and math tests were first administered in 1998. The reading and writing tests were given in grades 4, 8, and 10, and the math tests in grades 5, 8, and 10.

[4] It is worth considering how such classification might affect the students involved. One the one hand, strategic placements into LEP categories can potentially have a demoralizing effect on students and might expose them to weaker student groups. On the other hand, such placements might expose them to more resources with a positive effect on learning. Hanushek, Kain, and Rivkin (2002) study the effect of placement of students with disabilities into special education programs. They find that the programs led to significant gains in math achievement, especially for learning-disabled and emotionally handicapped students. But they do not look at the effect of placement into LEP categories, nor the impact of strategic placement into these categories. Unfortunately, there is virtually no literature on the impact of such strategic placement into exempt categories, making this question an avenue for important future research.

[5] See Greene (2001), Hoxby (2003a, 2003b), Greene and Winters (2003), Figlio and Rouse (2006), West and Peterson (2006), Rouse et al. (2007), Chakrabarti (2008a, 2008b), Chiang (2009), and Figlio and Hart (2010).

[6] See Jacob and Levitt (2003), Jacob (2005), Figlio and Winicki (2005), Cullen and Reback (2006), Figlio and Getzler (2006), Figlio (2006), Reback (2008), Neal and Schanzenbach (2010), and Chakrabarti (2013).

[7] The only exception is Chakrabarti (2013), who studies the behavior of public schools facing accountability-tied vouchers on other types of strategic behaviors, such as whether threatened schools focus more on high-stakes marginal students and subject areas.

The system of assigning letter grades to schools on a scale of A through F started in 1999. The state assigned a school an F grade if it failed to achieve the minimum criteria in all three FCAT subjects (reading, math, and writing), a D grade if it failed the minimum criteria in only one or two subject areas, and a C grade if it passed the minimum criteria in all three. To pass the minimum criteria in reading and math, a school needed to have at least 60 percent of its students score at level 2 or above in the respective subject; to pass the minimum criteria in writing, at least 50 percent had to score at level 3 or above.[8]

While the test scores of all regular students were included in the calculation of school grades, the scores of students in some limited-English-proficient and exceptional student education categories were excluded. Specifically, scores of LEP students who were in an ESOL program for less than two years were not included in the computation of grades, nor were scores of ESE students in eighteen ESE categories. Only LEP students with two or more years in an ESOL program and ESE students in speech-impaired, gifted, and hospital/homebound categories were included in school grade computations.[9]

Henceforth, we refer to the less than two years in an ESOL program category as the "excluded" LEP category and the two years or more in an ESOL program category as the "included" LEP category. Similarly, we refer to the speech-impaired, gifted, and hospital/homebound categories as "included" ESE categories and to the other ESE categories as "excluded" ESE categories.

3. Data

We obtained all data for this study from the Florida Department of Education. The information includes grade-level data on LEP enrollment in grades 2, 3, 4, and 5 for 1999 and 2000 as of February in each year (just before the tests were administered). We also know the number of students in an ESOL program for less than two years and the number of students in an ESOL program for two years or more in each of these grades in each year under consideration.

School-level data on enrollment in the various ESE categories were also obtained. In addition to total ESE enrollment, these data report enrollment in each of the ESE categories in each Florida school for 1999 and 2000.

The third type of data we retrieved was the distribution of students across grades K-12 in each Florida school in 1999 and 2000. We also had access to data on various socioeconomic characteristics of schools, including gender composition, racial composition, and the percentage of students eligible for free or reduced-price lunch. Finally, we obtained several measures of school-level and district-level per-pupil expenditures for both years under consideration.

4. Empirical Strategy

Under the Florida Opportunity Scholarship Program, schools that received an F grade in 1999 were directly threatened with stigma and vouchers since all of their students would be eligible for vouchers if the school received another F grade in the next three years. We refer to these schools as "F" schools. The schools that received a D grade in 1999 were closest to the "F" schools in terms of grade, but were not directly threatened by the program. We refer to them as "D" schools. Our empirical strategy essentially compares schools that barely received an F to those that barely received a D, as we explain below.

Because grades were not randomly assigned to schools, the schools that received an F grade in 1999 were likely to be quite different from those that did not, both in terms of observable and unobservable characteristics. These differences may themselves affect the outcome of interest—whether schools engage in strategic ESE or LEP classification. Thus, simply comparing the outcomes of "F" schools to those of "D" schools will not yield a causal estimate of the effect of the program; there are many confounding variables besides the program that could explain any differences we observe.

By comparing the schools that fell just below the cutoff ("F" schools) with those just above ("D" schools), we get an estimate of the effect of the [FOSP].

[8] We mainly focus on the responses of the schools that just received an F versus those that just received a D in 1999. In Section 6.4, we study the response of the "D" schools relative to the "C" schools as well. While the "D" schools did not face any direct threat of vouchers, they may have faced an indirect threat as they were close to an F grade and might have also faced stigma by being one of the lowest-performing groups. Correspondingly, we focus on the criteria for F, D, and C grades. Detailed descriptions of the criteria for the other grades are available at schoolgrades.fldoe.org.

[9] Florida classified ESE students into twenty-one ESE categories in total: educable mentally handicapped, trainable mentally handicapped, orthopedically handicapped, occupational therapy, physical therapy, speech-impaired, language-impaired, deaf or hard of hearing, visually impaired, emotionally handicapped, specific learning disabled, gifted, hospital/homebound, profoundly mentally handicapped, dual-sensory-impaired, autistic, severely emotionally disturbed, traumatic brain injured, developmentally delayed, established conditions, and other health-impaired.

To minimize the influence of confounding variables, we use a regression-discontinuity strategy (Hahn, Todd, and van der Klaauw 2001; van der Klaauw 2002; Imbens and Lemieux 2008) to analyze the effect of the program. The analysis essentially entails comparing the response of schools that barely failed to that of schools that barely passed. The institutional structure of the Florida program allows us to follow this strategy. We exploit the fact that there was a sharp discontinuity in how the F grade was assigned. Schools that scored below a fixed cutoff received an F, and thus the threat, while schools that scored above the cutoff did not. By comparing the schools that fell just below the cutoff ("F" schools) with those just above ("D" schools), we get an estimate of the effect of the program. Presumably, these two groups of schools were nearly identical in terms of socioeconomic and demographic characteristics (a testable assumption that we examine later), and the only difference between them was that one group was subjected to stigma and the threat of vouchers while the other was not.

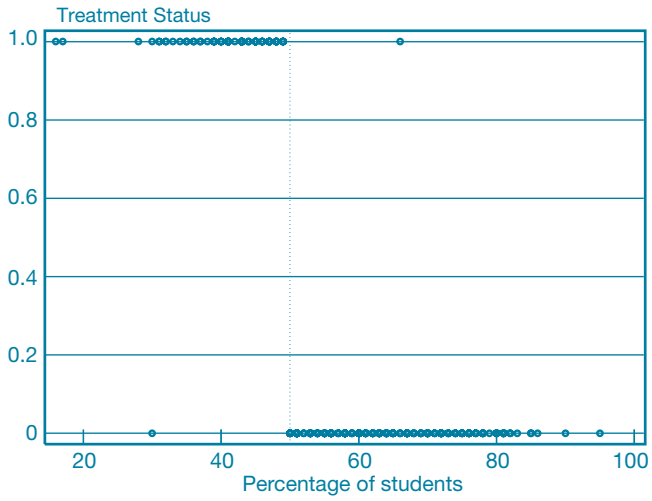
We focus on the sample of "F" and "D" schools that failed both reading and math in 1999. In this sample, according to the Florida grading rules, only the schools that also failed writing would receive an F, while the schools that passed writing would receive a D. Therefore, in this sample, schools that had less than 50 percent of their students pass the 1999 writing FCAT would receive an F and face a direct threat, while schools at or above 50 percent on the writing portion would not.

In the rest of this article, we refer to schools receiving an F grade in 1999 as being in the "treatment" group. Treated schools were exposed to the threat of vouchers and sanctions. Using the sample of "F" and "D" schools that failed both reading and math in 1999, we illustrate in Chart 1 the relationship between treatment status (those receiving an F in 1999) and the schools' percentages of students scoring at or above level 3 in FCAT writing, or the "running variable" (r_i) in the regression-discontinuity literature. There are 269 schools in this sample, with 65 falling below the cutoff of 50 percent on the writing portion and 204 schools at or above the cutoff. The chart shows that all but one of the schools in this sample that had less than 50 percent of their students scoring at or above level 3 actually received an F grade. Similarly, all but one that had 50 percent or more of their students scoring at or above level 3 were assigned a D grade. The result demonstrates that, in this sample, the percentage of students scoring at or above level 3 in writing indeed uniquely predicts assignment to treatment for all but two schools, and there is a sharp increase in the probability of treatment at the 50 percent mark. In fact, the estimated discontinuity is 1 and highly significant; there was a perfect correlation between falling below 50 percent and receiving an F. Using this sample ("F" and "D" schools that failed in reading and math in 1999), we rank schools in terms of the percentage of students scoring at or above level 3 in FCAT writing and then pick schools that are close to the cutoff. Our analysis uses this set of schools.

The percentage of students scoring at or above level 3 in writing indeed uniquely predicts assignment to treatment for all but two schools, and there is a sharp increase in the probability of treatment at the 50 percent mark.

Chart 1

Relationship between Treatment Status and Percentage of Students Scoring at or above Level 3 in 1999 FCAT Writing



Source: Authors' calculations.

Notes: Treatment status is 1 if a school received a grade of "F" and 0 if it received a grade of "D." FCAT is the Florida Comprehensive Assessment Test.

We also consider two alternate samples in which both "F" and "D" schools fail reading and writing or math and writing. (According to the Florida rules, "F" schools would also fail math [reading], unlike "D" schools.) We find that indeed in these samples, the probability of treatment increases sharply when less than 60 percent of a school's students scored at or above level 2 in math (reading). The sizes of these samples, however, are considerably smaller than those of the first sample we described, and these samples are considerably less dense in the vicinity of the cutoff. So, we focus on the first sample above, in which the "D" schools passed the writing cutoff and the "F"

schools missed it, and both groups of schools missed the cutoffs in the other two subject areas. Note, though, that the results from the alternate samples are qualitatively similar. Also, as a robustness check, we present in section 6.2 estimates from a combined sample in which we pool the three samples.

Consider the following model, where Y_i is school i's outcome,[10] T_i equals 1 if school i received an F grade in 1999 and $f(r_i)$ is a function of the running variable r_i. Recall that the running variable here is the percentage of students scoring at or above level 3 in FCAT writing:

$$(1) \qquad Y_i = \gamma_0 + \gamma_1 T_i + f(r_i) + \varepsilon_i .$$

Hahn, Todd, and van der Klaauw (2001) show that γ_1 is identified by the difference in average outcomes of schools that just missed the cutoff and those that just made it, provided that the conditional expectations of the other determinants of Y are smooth through the cutoff. Here, γ_1 identifies the local average treatment effect (LATE) or the effect of getting an F at the cutoff.

The estimation can be done in many ways. We use local linear regressions with a triangular kernel and a rule-of-thumb bandwidth, as suggested by Silverman (1986). We also allow for flexibility on both sides of the cutoff by using a linear spline functional form that enables us to include an interaction term between the running variable and a dummy indicating whether or not the school falls below the cutoff (see equation 2 below). We estimate alternate specifications that do not include controls as well as those that use them.[11] Assuming the covariates are balanced on both sides of the cutoff (we formally test this assumption below), the purpose of including covariates is variance reduction. They are not required for the consistency of γ_1. Thus, our preferred specification is:

$$(2) \qquad Y_i = \alpha_0 + \alpha_1 T_i + \alpha_2 r_i + \alpha_3 (T_i \times r_i) + (\Sigma_k \alpha_{4k} X_{ik}) + \varepsilon_i ,$$

where $f(r_t) = r_i + (T_i \times r_i)$ denotes the linear spline functional form; $\Sigma_k X_{ik}$ denotes the set of covariates (or controls) and includes racial composition (percentage black, Hispanic, Asian, American Indian, multiracial; percentage white serves as the excluded category), gender composition (percentage male), percentage of students eligible for free or reduced-price lunch, and real per-pupil expenditures.

To test the robustness of our results, we also experiment with alternative bandwidths. The results remain qualitatively similar, and are available on request. We also conduct a parametric estimation in which we include a third-order polynomial in the percentage of students scoring at or above level 3 in writing and interactions of the polynomial with a dummy indicating whether or not the school falls below the cutoff. We also estimate alternative functional forms that include a fifth-order polynomial instead of a third-order polynomial and the corresponding interactions.[12] The results are very similar in each case, and are available on request.

An advantage of a regression-discontinuity analysis is that identification relies on a discontinuous jump in the probability of treatment at the cutoff. Consequently, mean reversion—a potentially confounding factor in other settings—is not apt to be important here, as it likely varies continuously with the running variable (r_i) at the cutoff. Also, regression-discontinuity analysis essentially entails comparison of schools that are very similar, even virtually identical, except that the schools to the left of the cutoff faced a discrete increase in the probability of treatment. As a result, another potentially confounding factor—existence of differential preprogram trends—is not likely to be important here.

4.1 Testing the Validity of the Regression-Discontinuity Analysis

We now investigate whether the underlying assumptions governing the validity of the regression-discontinuity design are satisfied in this context. First, we check whether schools just below the cutoff differed from those just above it in terms of preprogram characteristics. Recall that any such differences would confound our attempt to attribute a difference in outcomes to the program. There is not much reason to expect any differences between these groups. For such differences to arise, certain types of schools would need to strategically manipulate their test scores in an effort to fall on one side of the cutoff. However, the program was announced in June 1999, while the tests were given a few months before (in January and February), making it unlikely that Florida's schools had the necessary information and time to resort to such manipulation.

Nevertheless, we check for discontinuities in predetermined characteristics of schools at the cutoff. For the regression-discontinuity strategy to be valid, preexisting characteristics should vary continuously through the cutoff. The only factor that should vary discontinuously is the probability of treatment. In such a case, any discontinuity in student

[10] In most of this article, Y_i refers to schools' percentages of students in various ESE and LEP categories. Exceptions are in sections 4.1 and 6.1, where Y_i also refers to various demographic and socioeconomic characteristics of the schools. See those sections for more details.

[11] Covariates used as controls include racial composition of schools, gender composition of schools, percentage of students eligible for free or reduced-price lunches, and real per-pupil expenditures.

[12] We use odd-order polynomials because they are more efficient (Fan and Gijbels 1996) and are not subject to boundary bias problems, as even-order polynomials are.

Table 1

Testing Validity of Regression-Discontinuity Analysis: Looking for Discontinuities in Preprogram Characteristics at Cutoff

	Percentage				
Panel A	(1) White	(2) Black	(3) Hispanic	(4) Asian	(5) American Indian
	2.92	-5.06	2.43	0.09	-0.16
	(7.24)	(11.39)	(6.73)	(0.28)	(0.06)
Panel B	Percentage Multiracial	Percentage Male	Percentage Free/ Reduced-Price Lunch	Enrollment	Real Per-Pupil Expenditure
	-0.23	-1.21	-5.97	-14.45	-1.97
	(0.26)	(1.44)	(5.36)	(60.32)	(2.29)
	Percentage				
Panel C	Exceptional Student Education (ESE)	Excluded ESE	Included ESE	Learning-Disabled	Emotionally Handicapped
	-2.92	-2.89	-0.03	0.05	-0.63
	(1.87)	(1.83)	(0.78)	(0.79)	(0.56)
	Percentage Excluded Limited-English-Proficient (LEP)				
Panel D	Grade 2	Grade 3	Grade 4	Grade 5	
	0.03	0.30	0.24	0.30	
	(0.18)	(0.20)	(0.22)	(0.18)	
	Percentage Included LEP				
Panel E	Grade 2	Grade 3	Grade 4	Grade 5	
	-0.54	0.06	-0.09	0.26	
	(0.51)	(0.56)	(0.28)	(0.41)	

Source: Authors' calculations.

Note: Robust standard errors adjusted for clustering using the running variable are in parentheses.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

classification (into excluded or included ESE and LEP categories) at the cutoff can be attributed to the discontinuity in the probability of treatment, or, in other words, to the program. The discontinuity estimates for preprogram characteristics (using the regression-discontinuity strategy described above) are presented in Table 1. As expected, they are small and never statistically distinguishable from zero.

Following McCrary (2008), we also use a density test to investigate whether there is selection at the cutoff. The idea is that if schools strategically placed themselves on one side of the cutoff, we would expect to see a clustering close to it, and consequently an unusual spike in the density of the running variable (the percentage of students at or above level 3 in writing). However, as Table 2 shows, we find no evidence of discontinuity in the density of the running variable at the cutoff.

5. Results

Having established that a regression-discontinuity approach in this setting is valid, we now look at the program's behavioral

Table 2

Testing Validity of Regression-Discontinuity Analysis: Looking for Discontinuities in Density of Running Variable

	1999
Difference	-0.01
	(0.01)

Source: Authors' calculations.

Notes: The table shows the percentage of students at or above FCAT level 3 in writing. Standard error is in parentheses and is clustered using the running variable (percentage of students at or above writing cutoff). FCAT is the Florida Comprehensive Assessment Test.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

effect on threatened schools. We focus on the elementary grades; grades 4 and 5 were the tested grades during this period in Florida.

For reference, we first look at the behavior of the schools in our sample in the preprogram period. Table 1 (panels C-E) shows classification into excluded and included LEP and ESE categories in 1998-99, the school year just before the program started. Each entry shows the average difference between the soon-to-be-threatened and the nonthreatened schools. There is no evidence that the schools that would be threatened the next year behaved any differently than the nonthreatened schools in terms of excluded or included LEP classification in any of the high- or low-stakes grades. We also see no evidence of differential classification into excluded or included ESE categories in 1999. The picture in the post-program period, however, is very different.

Table 3 examines the effect of the FOSP on the percentage of students classified into the excluded and included LEP categories in grades 2-5 in 1999-2000, the first school year after the program went into effect.[13] Again, each entry in the table shows the difference between the LEP percentages of threatened versus nonthreatened schools.

Consider the excluded LEP category in the top panel. In the year after the program's inception, there was a positive and statistically significant difference between threatened and nonthreatened schools in terms of the percentage of students classified as excluded LEP in the high-stakes grade 4 and the entry grade 3. In contrast, there is no evidence of a statistically significant difference in the low-stakes grade 2 or the high-stakes grade 5. Of note, though, is that while the grade 2 and

[13] These variables are defined as enrollment in excluded and included LEP categories in each grade as a percentage of total school enrollment.

Table 3

Effect of Program on Classification in Excluded and Included Limited-English-Proficient Categories

	(1) Grade 2	(2) Grade 3	(3) Grade 4	(4) Grade 5
Percentage excluded	0.29	0.36**	0.31**	0.27
	(0.23)	(0.18)	(0.12)	(0.25)
Observations	123	121	119	116
R^2	0.53	0.54	0.40	0.43
Percentage included	0.11	-0.42	0.04	0.01
	(0.30)	(0.48)	(0.31)	(0.39)
Observations	123	121	119	116
R^2	0.66	0.57	0.53	0.33

Source: Authors' calculations.

Notes: Robust standard errors adjusted for clustering using the running variable are in parentheses. All regressions control for racial composition, gender composition, percentage of students eligible for free/reduced-price lunch, and real per-pupil expenditure.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

grade 5 effects are not statistically significant, they are positive and not statistically different from the grade 3 or grade 4 effects.

The estimates suggest that in the first year of the program, schools facing stigma and the threat of vouchers classified an additional 0.31 percent of students into the excluded LEP category in grade 4 and an additional 0.36 percent in grade 3.

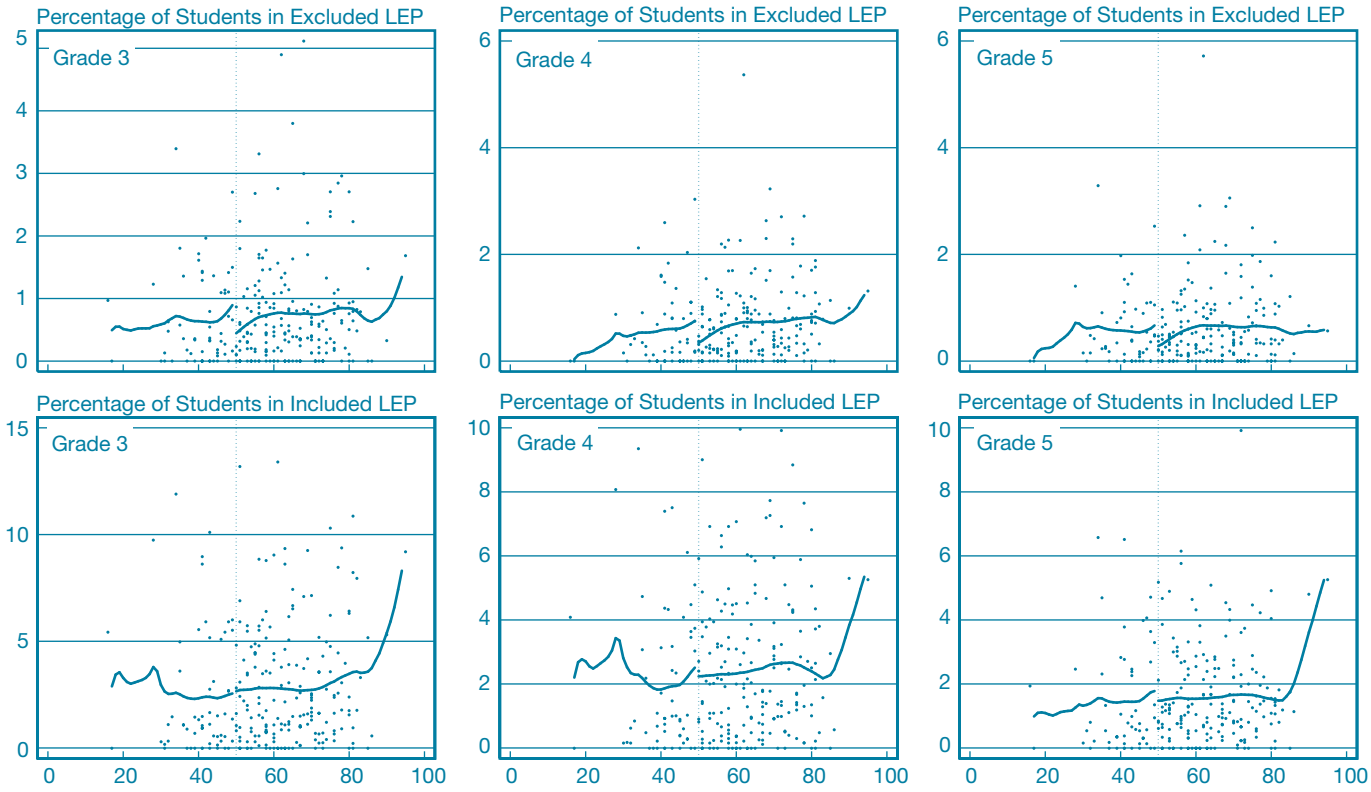
In the year after the [FOSP's] inception, there was a positive and statistically significant difference between threatened and nonthreatened schools in terms of the percentage of students classified as excluded LEP in the high-stakes grade 4 and the entry grade 3.

To put these numbers in perspective, we note that the average enrollment of these schools in the immediate preprogram period was approximately 713 students. Thus, the threatened schools classified an additional 53 percent of their excluded LEP students in grade 4 and an additional 55 percent of their excluded LEP students in grade 3. The results are, in turn,

Chart 2

Effect of Program on Classification in Excluded and Included Limited-English-Proficient (LEP) Categories

Regression-Discontinuity Estimates; February 2000 Survey



Source: Authors' calculations.

Notes: The x-axis in each panel depicts the percentage of students at or above level 3 in FCAT (Florida Comprehensive Assessment Test) writing.

equivalent to classification of an additional 2.6 students in grade 4 and 2.3 students in grade 3 into the excluded LEP category.

The lower half of Table 3 presents the program's effects on the percentage of students in the included LEP category. There is no evidence that the program led to differential classification into included LEP in any grade in the first year after the program; the discontinuities are small and statistically insignificant.[14] Chart 2 illustrates the impact on classifications into excluded and included categories.[15] Consistent with the above findings, the chart provides evidence in favor of increased classifications into excluded LEP categories in grades 3 and 4 (and these discontinuities are statistically significant). There is evidence of a smaller (statistically insignificant) discontinuity in grade 5, but none in favor of any differential classification into included LEP categories.

Tables 4 and 5 examine the effect of the program on ESE classification. Table 4, column 1, shows the effect on total ESE classification. The dependent variable for this analysis is percentage ESE enrollment (total ESE enrollment as a share of total enrollment). The estimates show no evidence of any differential classification in the threatened schools at the cutoff.

While trends in total ESE classification provide a summary picture, they are unlikely to provide a conclusive look at whether the "F" schools resorted to such classification. Yet in our view, the absence of shifts in total ESE classification does not rule out the possibility of shifts in certain ESE categories.

[14] Of note here is that neither the excluded LEP effects nor the included LEP effects are statistically different across grades.

[15] While the regression-discontinuity estimates in the tables were obtained from specifications that included all covariates mentioned above, the estimates in the charts were obtained from specifications that did not include any covariate. The similarity of the two sets of estimates attests to the robustness of the estimates.

Table 4

Effect of Program on Classification in Exceptional Student Education (ESE) Categories

	Percentage		
	(1) Students in ESE	(2) Students in Excluded ESE	(3) Students in Included ESE
	0.44	0.70	-0.24
	(0.40)	(0.56)	(0.29)
Observations	130	130	130
R^2	0.92	0.92	0.84

Source: Authors' calculations.

Notes: Robust standard errors adjusted for clustering using the running variable are in parentheses. All regressions control for racial composition, gender composition, percentage of students eligible for free/reduced-price lunch, real per-pupil expenditure, and preprogram (1999) percentage of students in All (Column 1), Excluded (Column 2), or Included (Column 3) ESE categories.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

Table 5

Effect of Program on Classification in Learning-Disabled and Emotionally Handicapped Categories

	Percentage	
	(1) Students in Learning-Disabled	(2) Students in Emotionally Handicapped
	-0.18	0.08
	(0.26)	(0.16)
Observations	130	130
R^2	0.80	0.93

Source: Authors' calculations.

Notes: Robust standard errors adjusted for clustering using the running variable are in parentheses. All regressions control for racial composition, gender composition, percentage of students eligible for free/reduced-price lunch, real per-pupil expenditure, and preprogram (1999) percentage of students in All (Column 1), Excluded (Column 2), or Included (Column 3) ESE categories.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

To offer a closer look, Table 4 also displays the effect of the program on classification into excluded and included ESE categories. The dependent variables here are the percentages of total enrollment classified into excluded (column 2) and included (column 3) categories. The estimates show no evidence that the threatened schools resorted to greater classification into excluded ESE categories in the first year of the program. The effects are not at all statistically significant, nor are they economically significant. There is also no statistically or economically significant evidence of differential classification out of (or into) the included categories.[16] Consistent with this evidence, Chart 3 offers no evidence of (statistically significant) differential classification into excluded or included ESE categories.

The various ESE categorizations differ in the extent of their severity, and consequently it may be easier to reclassify students into some categories than others. While some categories such as those involving observable or severe disabilities or physical handicaps are comparatively nonmutable, others such as learning disabled and emotionally handicapped are often mild and comparatively mutable. Classification into these latter categories often has a large subjective element and, as such, could be prone to manipulation. While the above analysis does not find evidence of differential classification into excluded categories as a whole, it does not rule out the possibility of increased classification into certain categories that are more easily manipulated on the spectrum of special needs.

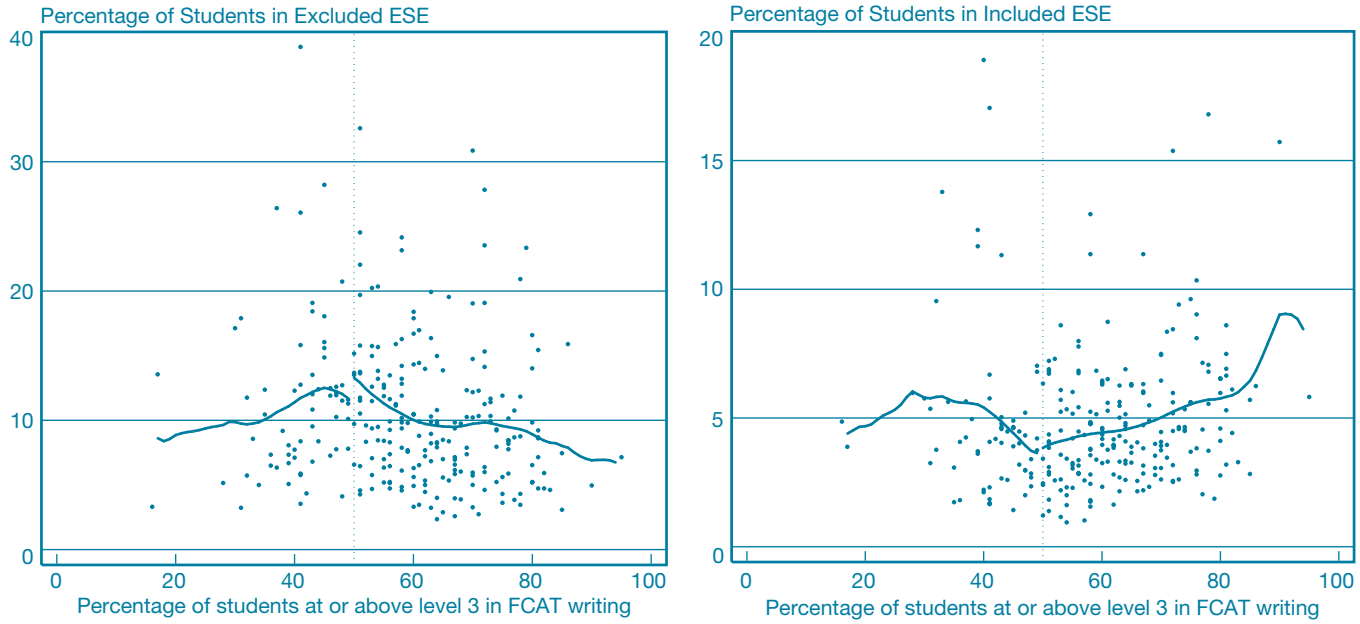
To investigate this possibility, we examine the effect of the program on classification into two mutable excluded categories: learning disabled (column 1) and emotionally handicapped (column 2). We find no evidence that the threatened schools tended to differentially classify students into either of these categories; the discontinuities are small and not statistically significant.

Our next step is to ask what might be driving these classification patterns that we do see. It is worth considering two explanations: 1) the "wake-up-call" hypothesis and 2) the "strategic-classifications" hypothesis.

[16] Recall that these are school-level effects, unlike grade-level effects for LEP. Also of note here is that the excluded LEP effect (computed from data aggregated over the available grades to generate a school-level measure for easier comparison) is both economically and statistically different from the excluded ESE effect. However, the included LEP effect is not statistically different from the included ESE effect.

Chart 3

Effect of Program on Classification in Excluded and Included Exceptional Student Education (ESE) Categories
Regression-Discontinuity Estimates, 2000



Source: Authors' calculations.

Note: FCAT is the Florida Comprehensive Assessment Test.

To summarize, we observe that the program led to statistically significant increased classifications into excluded LEP categories in high-stakes grade 4 and entry grade 3 in the threatened schools. Yet we find no evidence of any difference in classifications into included LEP categories. Neither do we find evidence of any difference in classification into ESE categories (excluded or included) in the threatened schools. Our next step is to ask what might be driving these classification patterns that we do see. It is worth considering two potential explanations: 1) the "wake-up-call" hypothesis and 2) the "strategic-classifications" hypothesis.

Under a wake-up-call hypothesis, one might argue that the F grade served as a wake-up call for these schools and led them to proactively classify their low-performing students into LEP or ESE groups to ensure greater and more specialized support for these students. Under a strategic-classifications hypothesis, an opposing argument can be made that the threatened schools tended to classify their low-performing students into excluded categories in a strategic effort to boost their scores and grades.

While the data do not allow us to pinpoint the exact cause of such classifications, there seems to be somewhat more evidence that strategic classifications are the more likely driver of the results. One would expect the wake-up call to manifest itself in increased classifications in all grades symmetrically, with a school acting on a genuine desire to help weaker students in each grade. It is not clear why such classification into an LEP track would be more prominent in high-stakes grade 4, and the entry to that high-stakes year, grade 3. Also the wake-up-call hypothesis would predict classifications into both ESE and LEP categories, perhaps more into ESE, as ESE categories provide more resources as well as more specialized help to students.

While the data do not allow us to pinpoint the exact cause of such classifications, there seems to be somewhat more evidence that strategic classifications are the more likely driver of the results.

In contrast, a strategic-classifications hypothesis would point to schools classifying students into excluded LEP in the high-stakes grades or entry grades. Specifically, students

classified into the excluded LEP category in grade 4 would not count toward school grades either in the current year or in the following year, when these students would advance to grade 5, another high-stakes grade. Note, though, that doing the additional classification all at once may have been difficult, which is why the administrators may have chosen to spread out the process to the entry grade 3.

Strategic classifications would also tend to result in classifications only into excluded LEP, but not excluded ESE categories, since there were considerable costs associated with reclassification into ESE categories. ESE designations had to be approved by the parents and a group of experts (such as physicians and psychologists). But the steepest cost to ESE classification was posed by the McKay Scholarship program. Created in 1999 and fully implemented statewide in the 2000-01 school year,[17] this program made every student with disabilities in Florida public schools eligible for vouchers to move to a private school or to another public school. Thus, reclassification of students into special education categories was associated with a risk of losing the students and their corresponding per-pupil funding. Moreover, because special education students were more expensive to educate than regular students, McKay vouchers cost more than Opportunity Scholarships—approximately $7,000 versus $3,500 per student on average. This fact meant that schools were likely to lose more funding with the departure of an ESE student under the McKay program than with the loss of a regular student under the FOSP. Consequently, the McKay Scholarship program acted as a strong disincentive to this sort of reclassification. The strategic-classifications view, therefore, seems to be more compelling in this scenario, as it matches better the patterns observed in the data.[18] However, the implication that strategic classifications play a role should only be taken as suggestive, and not conclusive. A further caveat is worth mentioning here. As with any regression-discontinuity analysis, the estimates obtained above are all local average treatment effects, meaning that the effects obtained are local to the cutoff only. These results should not be generalized to the whole sample.

The strategic-classifications view . . . seems to be more compelling in this scenario, as it matches better the patterns observed in the data.

[17] The McKay program was run as a small pilot in the 1999-2000 school year with only one school and two students participating in the program.

6. Sensitivity Checks

6.1 Compositional Changes of Schools and Sorting

If differential student sorting or compositional changes occurred in the treated schools, then the above effects may in part be driven by those changes.[19] To investigate this issue, we examine whether the FOSP led to a differential change in the demographic composition in the treated schools. We use the same regression-discontinuity strategy outlined above, but the dependent variables are now demographic (the percentages of students that are white, black, Hispanic, Asian, American Indian, multiracial, male, eligible for free or reduced-price lunch, as well as enrollment). We find no evidence of differential shifts in the treated schools in these characteristics after the introduction of the program. (These results are not reported here, but are available on request.) Thus, it seems safe to conclude that the results described above are not being driven by differential changes in the composition of schools or student sorting.

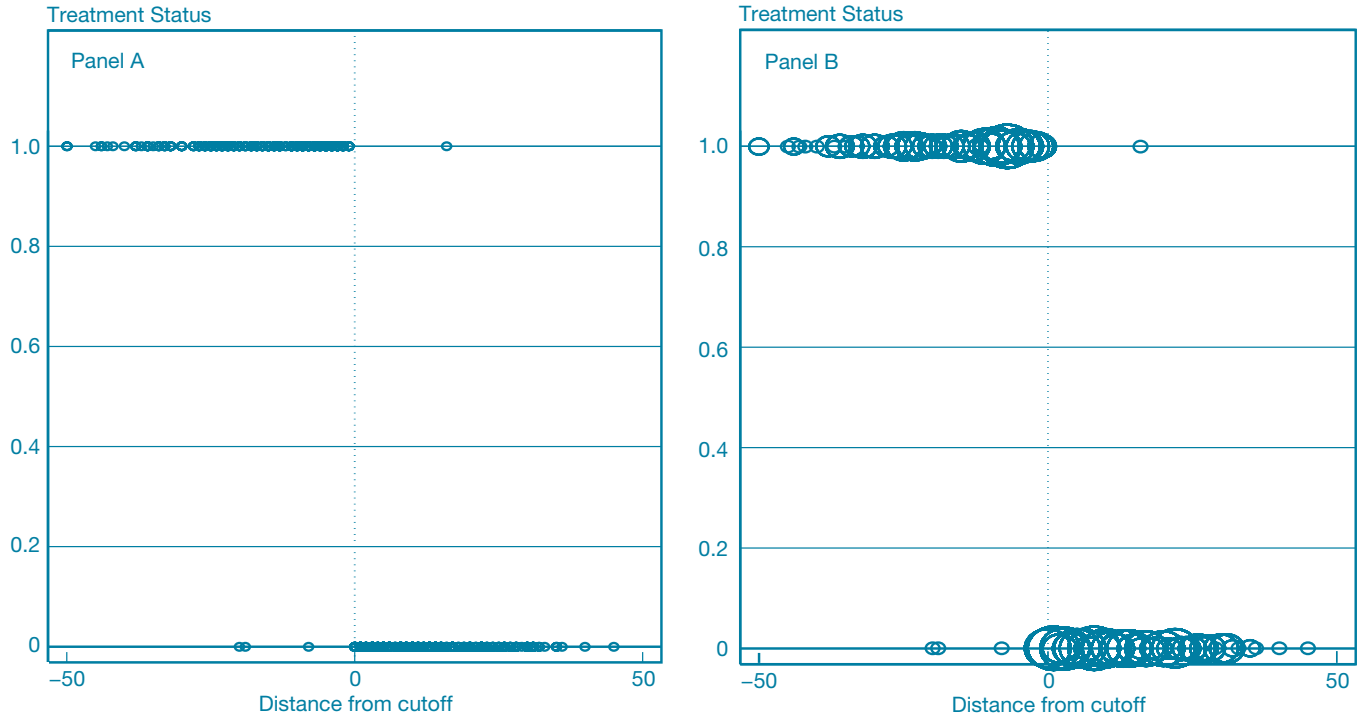
6.2 Does Combining the Three Discontinuity Samples Affect Results?

To broaden our analysis, we also apply an alternative regression-discontinuity strategy in which we combine the three samples described in section 4: the sample that failed in reading and math, but just passed or failed in writing (F/D writing sample); the sample that failed in reading and writing, but just passed or failed in math (F/D math sample); and the sample of schools that failed in math and writing, but just passed or failed in reading (F/D reading sample). In the F/D reading (math) sample, according to Florida rules, schools with

[18] A question worth considering here is whether such classification was enough for an "F" school to escape an F grade in the near future. Note that the percentages of students classified into the excluded LEP category were not small (53 percent and 55 percent). The additional classification in terms of numbers of students of between two and three in grade 3 and grade 4 does not appear to be big. However, these were marginal schools located close to the cutoff that only barely failed to make the cutoff. So, for such schools, even such a small classification could potentially make a difference. Also, consider that schools may not respond in only one margin. Such classifications along with responses along other margins could together make a difference in terms of grade.

[19] None of the threatened schools was subjected to vouchers in the 1999-2000 school year, so the concern about vouchers leading to sorting is not applicable here. However, the F and D grades alone (exposing schools to the threat of vouchers) could lead to differential sorting of students in these two types of schools. Figlio and Lucas (2004) find that following the first assignment of school grades in Florida, the better students differentially entered schools receiving A grades, though this differential sorting tapered off over time.

Chart 4
Relationship between Treatment Status and Distance from Cutoff (Combining the Three Discontinuity Samples)



Source: Authors' calculations.

just under 60 percent of their students scoring at or above level 2 in FCAT reading (math) should receive an F, while schools with just above (or exactly) 60 percent should receive a D. In the F/D writing sample, schools with just under 50 percent of their students scoring at or above level 3 in FCAT writing should receive an F, while schools with just above (or exactly) 50 percent of their students scoring at or above level 3 should receive a D. Centering these running variables at their respective cutoffs (60 percent or 50 percent), we pool the three samples to improve efficiency. We first examine the relationship between treatment status and the running variable in each of these samples as well as in the pooled sample. Chart 4 illustrates this relationship for the pooled sample—specifically, between probability of treatment and the respective running variable centered at the cutoff (marking essentially the distance from the relevant cutoff). In Chart 4, panel B is the same as panel A, except that the sizes of the bubbles are proportional to the number of schools at that point. In each of the individual samples (Chart 1 for the writing sample; others available on request) as well as in the pooled sample (Chart 4), there is a sharp discontinuity at the cutoff, with an estimated discontinuity size of 1. The underlying validity assumptions (continuity of preexisting observables and continuity of density) are also satisfied for each of the individual samples and the pooled sample (estimates available on request).

There is no evidence of any increased classification into either the total ESE or excluded/included ESE categories, nor is there evidence of any change in classification into learning-disabled or emotionally handicapped categories.

The results for the LEP categories using the combined sample are reported in Table 6. The picture depicted in the table is very similar to that obtained above, both quantitatively

Table 6

Effect of Program on Classification in Excluded and Included Limited-English-Proficient Categories: A Regression-Discontinuity Analysis Combining the Three Discontinuity Samples

	(1) Grade 2	(2) Grade 3	(3) Grade 4	(4) Grade 5
Percentage excluded	0.19	0.34**	0.30**	0.26
	(0.26)	(0.16)	(0.12)	(0.23)
Observations	215	216	213	205
R^2	0.03	0.05	0.03	0.04
Percentage included	0.12	-0.04	0.18	0.08
	(0.95)	(0.66)	(0.57)	(0.52)
Observations	215	216	213	205
R^2	0.02	0.05	0.02	0.02

Source: Authors' calculations.

Notes: Robust standard errors adjusted for clustering using the running variable are in parentheses. All regressions control for racial composition, gender composition, percentage of students eligible for free/reduced-price lunch, and real per-pupil expenditure, and include sample dummies to control for the respective sample from which the observation is obtained.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

and qualitatively. The estimates suggest that the "F" schools tended to classify an additional 0.34 percent of their total students into the excluded LEP category in the entry grade 3 and an additional 0.30 percent of their total students into the excluded LEP category in the high-stakes grade 4. These effects are statistically significant and equivalent to classifying as LEP an additional 2.37 students in grade 3 and an additional 2.1 students in grade 4. There is no statistically significant evidence of any change in classification in either the low-stakes grade 2 or high-stakes grade 5.

The results for ESE using the combined sample are reported in Tables 7 and 8. Like before, there is no evidence of any increased classification into either the total ESE or excluded/included ESE categories, nor is there evidence of any change in classification into learning-disabled or emotionally handicapped categories.

Table 7

Effect of Program on Classification in Exceptional Student Education (ESE) Categories: A Regression-Discontinuity Analysis Combining the Three Discontinuity Samples

	Percentage		
	(1) Students in ESE	(2) Students in Excluded ESE	(3) Students in Included ESE
	-0.94	-1.01	0.34
	(1.40)	(1.61)	(0.77)
Observations	241	241	241
R^2	0.04	0.02	0.06

Source: Authors' calculations.

Notes: Robust standard errors adjusted for clustering using the running variable are in parentheses. All regressions control for racial composition, gender composition, percentage of students eligible for free/reduced-price lunch, real per-pupil expenditure, and include sample dummies to control for the respective sample from which the observation is obtained.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

6.3 Are the Results Robust to Expressing the LEP Share as Percentage of Grade Enrollment?

Recall from footnote 13 that the various LEP or ESE shares (or percentages) are computed as percentages of total school enrollment. Since all ESE data are available at the school level, it is natural to divide ESE enrollment by total school enrollment to get the corresponding ESE percentage. However, since LEP data are available at the grade level, there are two alternatives: expressing excluded and included LEP as percentages of grade enrollment or as percentages of total school enrollment. In the above analysis, we take the latter route to be consistent with the definitions of various ESE percentages and to facilitate comparison with the ESE results. One disadvantage of using this definition, though, is that grade-specific LEP shares are also affected by enrollment changes in other grades.[20]

[20] Note, though, that when one divides by grade enrollment, grade-level LEP shares will change if non-LEP enrollment of that grade changes, even though LEP enrollment does not. Such a change will also be reflected in the first definition, in which we divide by total school enrollment, but dividing by total enrollment will dampen the effect of the change of the non-LEP share of the grade. Each measure, therefore, has its advantages and disadvantages.

Table 8

Program Effects on Classification in Learning-Disabled and Emotionally Handicapped Categories: A Regression-Discontinuity Analysis Combining the Three Discontinuity Samples

	Percentage	
	(1) Students in Learning-Disabled	(2) Students in Emotionally Handicapped
	-0.23	-0.38
	(0.60)	(0.46)
Observations	241	241
R^2	0.06	0.03

Source: Authors' calculations.

Notes: Robust standard errors adjusted for clustering using the running variable are in parentheses. All regressions control for racial composition, gender composition, percentage of students eligible for free/reduced-price lunch, real per-pupil expenditure, and include sample dummies to control for the respective sample from which the observation is obtained.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

Table 9

Program Effects on Classification in Excluded and Included Limited-English-Proficient (LEP) Categories: A Regression-Discontinuity Analysis Using Excluded and Included LEP as Percentages of Grade-Level Enrollment

	(1) Grade 2	(2) Grade 3	(3) Grade 4	(4) Grade 5
Percentage excluded	1.91	2.48**	1.62***	1.39
	(1.34)	(1.18)	(0.55)	(1.76)
Observations	123	121	119	116
R^2	0.53	0.51	0.42	0.43
Percentage included	0.28	-3.25	-1.13	-1.98
	(2.18)	(2.80)	(1.60)	(2.71)
Observations	123	121	119	116
R^2	0.66	0.57	0.55	0.35

Source: Authors' calculations.

Notes: Robust standard errors adjusted for clustering using the running variable are in parentheses. All regressions control for racial composition, gender composition, percentage of students eligible for free/reduced-price lunch, and real per-pupil expenditure.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

To ensure that changes in enrollment in other grades are not driving the results above, and that they are robust to the definition of percentage (or share) used, we reestimate the above regression-discontinuity specifications for LEP using the alternative definition. In this section, percentage LEP is defined as LEP enrollment in that grade divided by total enrollment in the same grade.

The results for LEP are presented in Table 9 and are similar to those obtained above. There is evidence of increased classification into excluded LEP in both the entry grade 3 and high-stakes grade 4. To put the effects below in perspective, we note that in the immediate preprogram period (1999), average grade 3 and grade 4 enrollments of the schools under consideration were 125 and 124, respectively. Facing the threat of vouchers and stigma, the "F" schools resorted to an additional classification of 2.48 percent of their grade 3 students into the excluded LEP category in that grade and 1.62 percent of their grade 4 students into the excluded LEP category in grade 4. We observed that the coefficients here are bigger than earlier because of the difference in the definition of LEP share (excluded LEP expressed as a percentage of grade enrollment rather than school enrollment). These figures are equivalent to an increase of 2.87 students in grade 3 and 2.0 students in grade 4 and are similar to those obtained above. Moreover, there is no statistically significant evidence of a change in classification into either excluded categories in low-stakes grade 3 or high-stakes grade 5 nor is there evidence of any change in classification into included categories in any of the grades.

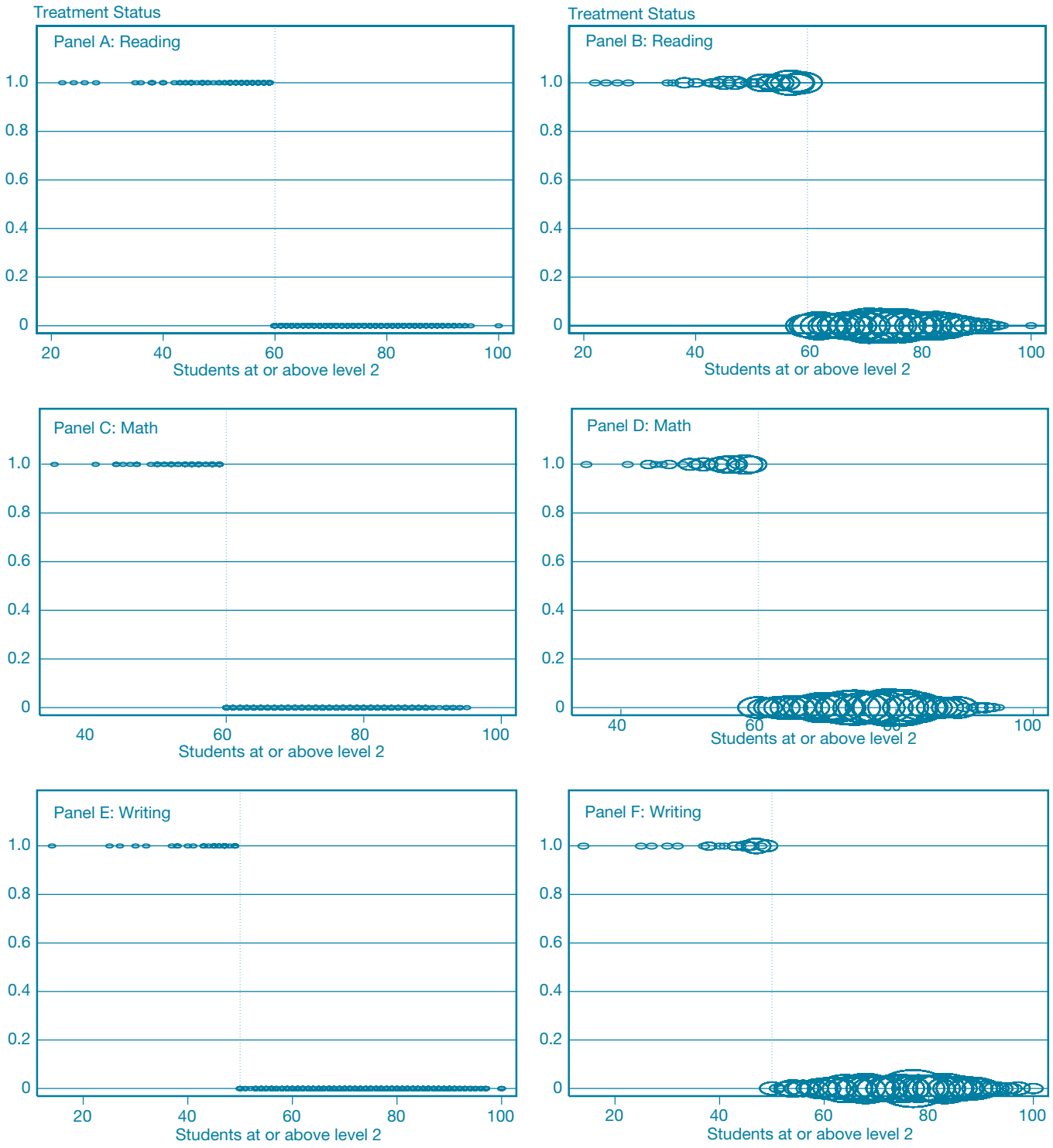
6.4 How "D" Schools Responded Relative to "C" Schools: A Regression-Discontinuity Analysis at the C/D Cutoff

A related question is whether the "D" schools exhibited any strategic behavior in terms of additional classification into excluded LEP and ESE categories. "D" schools did not face any direct threat of vouchers or stigma, but they were close to getting an F. Moreover, while they were not the lowest-performing schools, they were one of the lower-performing groups, and hence might have felt stigma to some extent. In this section, we

Chart 5

Relationship between Treatment Status (D) and Running Variable in Reading, Math, and Writing Samples

Regression-Discontinuity Estimates, 2000



Source: Authors' calculations.

Notes: The x-axis in all panels depicts percentages. In this chart, treatment status is 1 if a school received a grade of "D" and 0 if it received a grade of "C."

investigate whether the "D" schools responded differently than the "C" schools, ranking higher in the grade scale.

Once again, we use a regression-discontinuity strategy to study this response. Recall from section 2 that according to Florida rules, a school was assigned a D grade if it passed the minimum criteria in one or two of the three subject areas, while it got a C grade if it passed the minimum criteria in all three subject areas. Consider the group of schools that passed in two of the three subject areas. In this sample of schools, those that failed the third subject area should have received a D, while those that passed the third subject area should have received a C. There are three such possible samples: schools that passed in math and writing, but just passed or failed in reading (reading sample); schools that failed in reading and writing, but just passed or failed in math (math sample); and schools that passed in reading and math, but just passed or failed in writing (writing sample). According to Florida rules, the minimum criteria of each subject area yielded a sharp cutoff. In each of these samples, schools that were just below the cutoff in the third subject area should have received a D, and schools just above should have gotten a C.

Chart 5 illustrates the relationship between treatment status (for the purposes of this section, receiving a D rather than a C)[21] and the running variable for each of the three samples. Panels A and B show the relationship in the reading sample, where the running variable is the percentage of students at or above level 2; panels C and D illustrate the relationship in the math sample, where the running variable is the percentage of students at or above level 2; panels E and F depict the relationship in the writing sample, where the running variable is the percentage of students at or above level 3. For each sample, the second panel (B, D, and F) is the same as the first one (A, B, and C), except that each dot is weighted by the number of schools at that time. The smallest bubble corresponds to one school, while bigger bubbles correspond to larger numbers of schools. Indeed, we find that in the first two samples (Chart 5, panels A-B and panels C-D, respectively), the probability of treatment (getting a D) increases discontinuously at 60 percent as a function of the percentage of students scoring at or above level 2 in reading (math). In the third sample, the probability of treatment increases discontinuously at 50 percent as a function of the percentage of students scoring at or above level 3 in writing. As can perhaps be anticipated from the panels, each of these samples yields an estimated discontinuity of size 1 at the respective cutoffs.

[W]hile the "D" schools may have faced an indirect threat and some stigma since they were close to F status, those issues were not enough to lead to any strategic classifications into ... excluded categories.

To leverage efficiency gains and to build power, we pool these three samples together, centering the running variables at the respective cutoffs. First, we check whether the standard assumptions that govern the validity of regression-discontinuity techniques are satisfied in this context. Specifically, we find that for each of these samples as well as the combined sample, observable preprogram characteristics are indeed smooth through the cutoff. The preprogram results for the reading sample are presented in the appendix;[22] results for the other samples are not reported for lack of space, but are

Table 10

Effect of Program on Classification in Excluded and Included Limited-English-Proficient Categories: A Regression-Discontinuity Analysis Combining the Three Discontinuity Samples for Schools at the C/D Cutoff

	(1) Grade 2	(2) Grade 3	(3) Grade 4	(4) Grade 5
Percentage excluded	-0.09	-0.09	-0.02	-0.22
	(0.06)	(0.06)	(0.06)	(0.14)
Observations	331	327	333	321
R^2	0.45	0.40	0.57	0.42
Percentage included	0.27	0.30	0.20	-0.07
	(0.17)	(0.24)	(0.12)	(0.13)
Observations	311	311	306	294
R^2	0.92	0.90	0.85	0.76

Source: Authors' calculations.

Notes: Robust standard errors adjusted for clustering using the running variable are in parentheses. All regressions control for racial composition, gender composition, percentage of students eligible for free/reduced-price lunch, and real per-pupil expenditure, and include sample dummies to control for the respective sample from which the observation is obtained; regressions in the last three rows also include the lagged dependent variable as an additional covariate (see footnote 20).

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

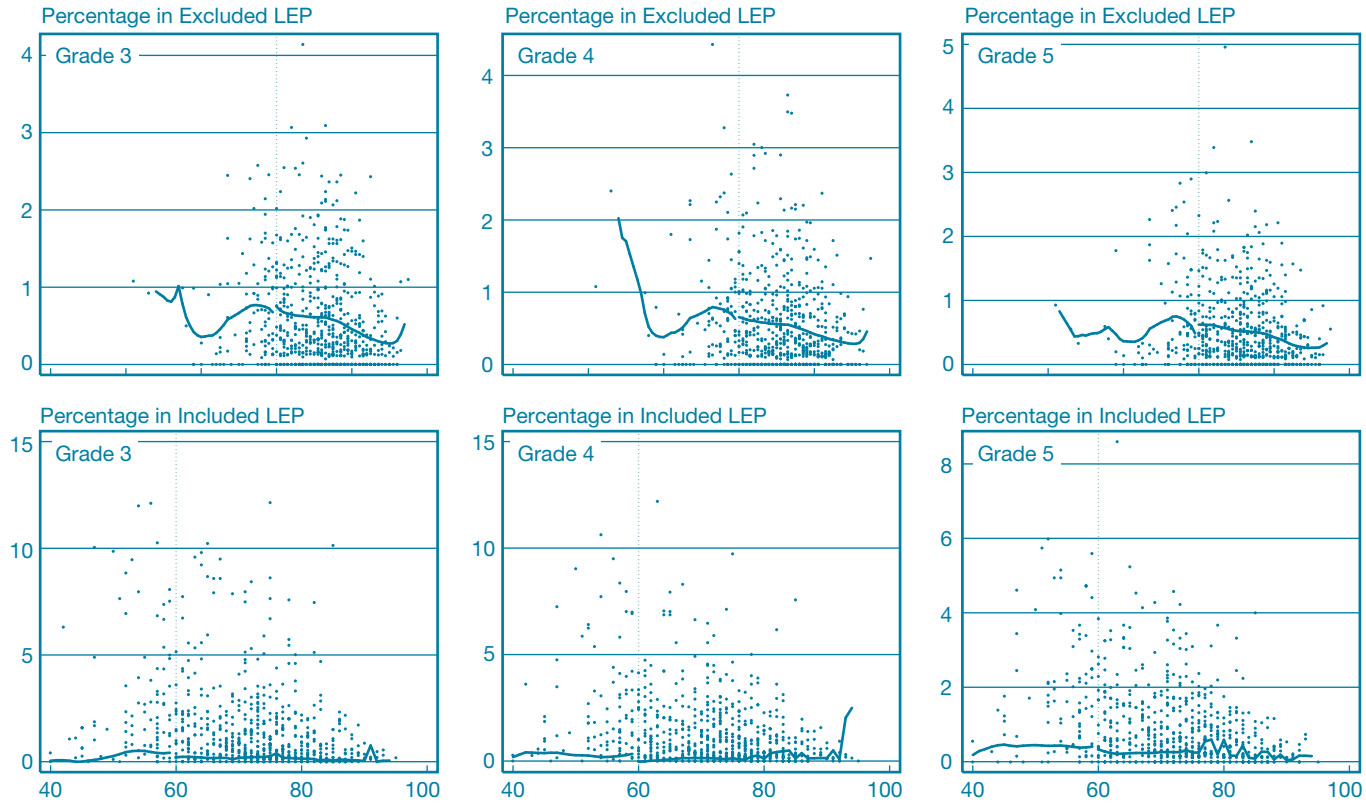
[21] Here, receiving a D in the immediate preprogram year (1999) is considered to be the treatment. In the rest of the article, getting an F in 1999 is the treatment.

[22] One exception is the estimate for included LEP percentage in grade 5, which is statistically different from zero. However, with a large number of differences, it is natural to have a few statistically different from zero, even if by random variation. Still, we observe that even though the coefficients for percentage LEP in the other grades are not statistically different from zero, they are not small. Therefore, in the estimations for included LEP in this subsection, we include the lagged dependent variable as an additional covariate.

Chart 6

Effect of Program on Classification in Excluded and Included Limited-English-Proficient (LEP) Categories on Schools at the C/D Cutoff

Regression-Discontinuity Estimates at C/D Cutoff; February 2000 Survey



Source: Authors' calculations.

Note: The x-axis in each panel depicts the percentage of students at or above level 2 in FCAT (Florida Comprehensive Assessment Test) reading in 1999.

available on request. We also find no evidence of discontinuity in the density of any of the running variables at the cutoff. (These results are also not reported here, but are available on request.)

Having established the validity of regression-discontinuity design in this context, and using the combined sample, we investigate in Table 10 and Chart 6 the effect of the program on classification into excluded and included LEP categories in "D" schools at the cutoff (relative to "C" schools). Interestingly, there is no evidence of any differential classification in the "D" schools at the cutoff into either excluded or included LEP categories in any of the low- or high-stakes grades.

We also look at the effect of getting a D on classification into total ESE, excluded ESE, and included ESE (Table 11 and Chart 7) as well as into more mutable learning-disabled and emotionally handicapped categories (Table 12). Once again, there is no evidence of any differential classification into any of these categories at the cutoff. These results imply that while the "D" schools may have faced an indirect threat and some stigma since they were close to F status, those issues were not enough to lead to any strategic classifications into any of the excluded categories. In contrast, the direct threat of vouchers and the stigma effect associated with the lowest grade led to additional classifications by the "F" schools (at the cutoff) into excluded LEP categories in high-stakes grade 4 and entry grade 3.

Table 11

Effect of Program on Classification in Exceptional Student Education (ESE) Categories: A Regression-Discontinuity Analysis Combining the Three Discontinuity Samples for Schools at the C/D Cutoff

	Percentage		
	(1) Students in ESE	(2) Students in Excluded ESE	(3) Students in Included ESE
	-0.001	-0.001	0.000
	(0.008)	(0.006)	(0.004)
Observations	383	383	383
R^2	0.17	0.20	0.05

Source: Authors' calculations.

Notes: Robust standard errors adjusted for clustering using the running variable are in parentheses. All regressions control for racial composition, gender composition, percentage of students eligible for free/reduced-price lunch, real per-pupil expenditure, and include sample dummies to control for the respective sample from which the observation is obtained.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

Table 12

Effect of Program on Classification in Learning-Disabled and Emotionally Handicapped Categories: A Regression-Discontinuity Analysis Combining the Three Discontinuity Samples of Schools at the C/D Cutoff

	Percentage	
	(1) Students in Learning-Disabled	(2) Students in Emotionally Handicapped
	0.001	-0.001
	(0.003)	(0.003)
Observations	383	383
R^2	0.16	0.07

Source: Authors' calculations.

Notes: Robust standard errors adjusted for clustering using the running variable are in parentheses. All regressions control for racial composition, gender composition, percentage of students eligible for free/reduced-price lunch, real per-pupil expenditure, and include sample dummies to control for the respective sample from which the observation is obtained.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

7. Implications for Education Policies in New York

The Florida experience yields important lessons for school accountability programs elsewhere. These policies include New York City's accountability framework, known as the Progress Reports program, and the federal No Child Left Behind Act as implemented by New York State.

In 2007, the New York City Department of Education introduced a new accountability system centered on annual school progress reports. These publicly available school "report cards" assign each school a letter grade ranging from A to F based on three separate components: school environment, student performance, and student progress (accounting for 15 percent, 30 percent, and 55 percent of the overall score, respectively). The school environment score is based on responses to surveys given to teachers, parents, and students in grade 6 and above. The student-performance and progress scores are based on student performance on statewide standardized math and English language arts tests. The performance score is based on the level of test scores in the current year, while the progress score is based on improvements or declines in test scores compared to previous years.

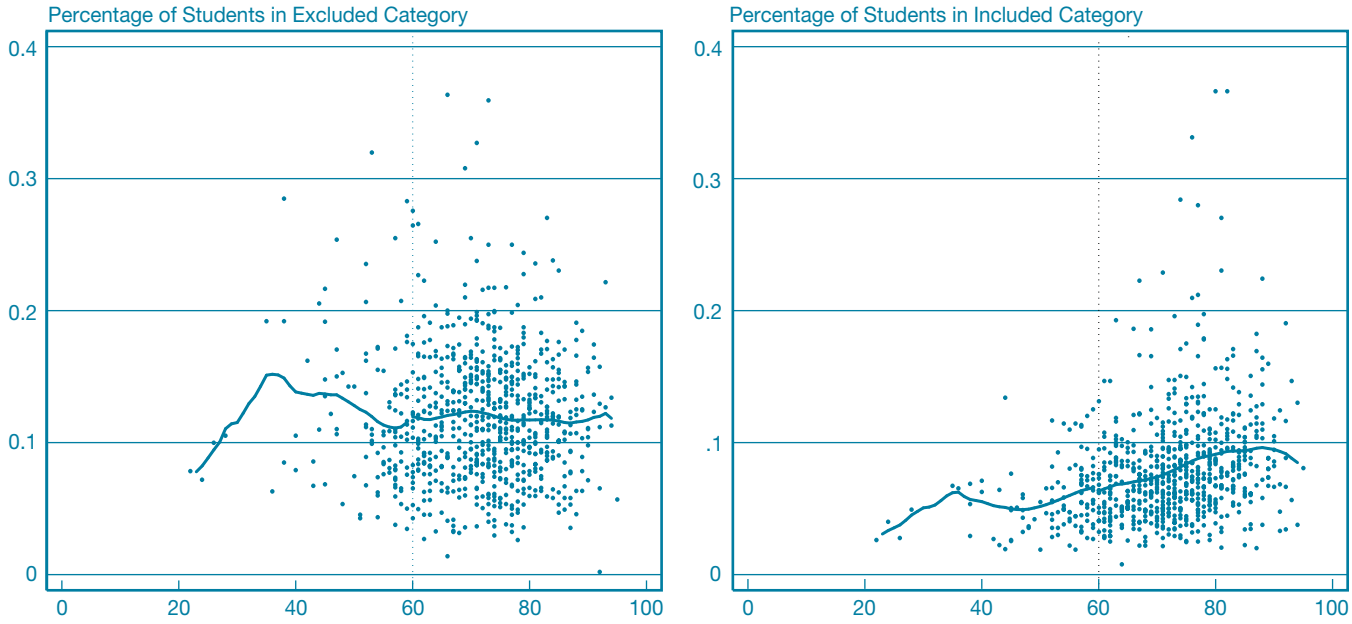
In contrast to the Florida program, New York City's accountability program includes not only high-needs students in grade calculations, but also gives schools additional credit for making achievement gains with particular high-needs groups: English language learners (ELL), special education students, and students performing in the lowest third of all students citywide. Overall scores are calculated as a weighted sum of the scores in each component plus any additional credit received. Letter grades from A to F correspond to specific thresholds on the overall score scale. Thus, additional credit can (and has already often) allowed schools to receive a higher grade.

The Florida experience yields important lessons for school accountability programs elsewhere . . . [including] New York City's accountability framework, known as the Progress Reports program, and the federal No Child Left Behind Act as implemented by New York State.

Chart 7

Effect of Program on Classification in Excluded and Included Exceptional Student Education (ESE) Categories on Schools at the C/D Cutoff

Regression-Discontinuity Estimates, 2000



Source: Authors' calculations.

Note: The x-axis in each panel depicts the percentage of students at or above level 2 in FCAT (Florida Comprehensive Assessment Test) reading in 1999.

This approach attaches clear rewards for high scores and clear sanctions for low scores. Schools receiving high grades are eligible for increases in per-pupil funding, and their principals are eligible for bonuses ranging from $7,000 to $25,000. In contrast, schools receiving low grades (F or D) are threatened with principal dismissal, restructuring, or even closure. This threat is credible and has often been implemented in practice.[23] In addition to the possibility of leadership change or closure, all schools receiving F and D grades (or a C grade three years in a row) are required to implement school improvement measures and target-setting. Finally, students in "F" schools are eligible to transfer to better-performing public schools.

Although the Progress Reports program does not include a voucher element, it is in many ways similar to the Florida voucher program. For example, it assigns schools letter grades based in part on student performance on standardized tests and imposes sanctions on low-performing schools, including allowing students to transfer out of failing schools.[24] But a key difference is that the New York City program includes the test scores of all ELL and special education students in the computation of school grades. In fact, it gives schools extra credit for achieving progress with ELL and special education students as well as other high-needs groups (such as students in the lowest third citywide). This additional credit can be substantial—in 2007, 161 schools received a higher grade due to additional credit (Rockoff and Turner 2010). Consequently, the strategic classification we describe earlier in the Florida context would not be expected to take place in New York City. However, the New York City program rules can generate other adverse incentives for classification. Since the failing schools there can earn additional credit for demonstrating progress of ELL and special education students, they might have an impulse to classify their higher-performing students in these categories in an effort to artificially boost scores.[25] Whether or not this behavior actually happened is a topic of future research.

[23] In December 2007, the New York City Department of Education announced that seven of the forty-two schools receiving F grades and two of the eighty-seven schools receiving D grades would be closed or phased out in the following year (Rockoff and Turner 2010); this sent a clear signal to other low-performing schools that the threat of closure was credible.

[24] Students are eligible to transfer to public schools but do not receive vouchers to transfer to private schools, as they do in Florida.

We now turn to the federal education law—the No Child Left Behind Act—as implemented in New York. Like New York City's Progress Reports program, NCLB establishes an accountability framework modeled on the Florida program, though with important differences.

NCLB, a major reform of the Elementary and Secondary Education Act, was signed into law on January 8, 2002. The states, including New York, implemented it soon thereafter. In compliance with the law, New York established targets for adequate yearly progress (AYP). AYP is determined based on each school's progress toward meeting target proficiency levels for all students in English-language arts, mathematics, and science. Schools must achieve these proficiency targets for the student body as a whole, and also for particular subgroups of students. Schools must also have an average of 95 percent of students participating in state tests over two years. Finally, schools must meet a target for attendance rate or, in the case of high schools, for graduation rate. If a school does not meet requirements in any one of these categories, it is said to miss AYP.

In all, the features of both New York City's Progress Reports program and the federal No Child Left Behind Act (as implemented in New York) represent important steps forward in eliminating adverse incentives for the type of strategic reclassification that appears to have taken place in Florida.

Schools that receive Title I federal funds are subject to NCLB sanctions if they miss AYP for two consecutive years. A Title I school missing AYP for two consecutive years is required to provide public school choice to its students. That rule permits students to transfer to better-performing public schools, with per-pupil funding following the students to their new schools. If a school misses AYP for three consecutive years, it is required to provide (and finance) supplemental educational services (such as tutoring) in addition to public school choice. Missing AYP for four consecutive years leads to corrective action in addition to the above sanctions; missing it for five consecutive years leads to restructuring in addition to the sanctions.

Recall that schools must meet AYP not only for the student body as a whole, but for particular subgroups: white, black, Hispanic, Asian, and American Indian students; students with disabilities; students with limited English proficiency; and students from low-income families. If a school fails to meet the target for any subgroup, it is deemed to have missed AYP. Thus, LEP students, students with disabilities, and other subgroups are not only included in the calculation of scores for the "All Students" group, they also separately count toward AYP formation.[26] Therefore, the potential incentives to reclassify weak students into ungraded groups are not present here.

In all, the features of both New York City's Progress Reports program and the federal No Child Left Behind Act (as implemented in New York) represent important steps forward in eliminating adverse incentives for the type of strategic reclassification that appears to have taken place in Florida. These two programs do not permit high-needs students to be excluded from the calculation of school grades.[27] All students count toward grade formation, and, in the case of the New York City program, the weaker categories carry more weight. While this program design can potentially ward off the gaming of the system seen in Florida, it introduces an incentive to move stronger students into high-needs categories as a way to boost scores.

8. Conclusion

This article analyzes the responses of public schools to the Florida Opportunity Scholarship Program, an influential school accountability policy employing vouchers as a sanction for low school achievement. Looking closely at the institutional details of the program, we identify the incentives it establishes and the behavior of public schools responding to it. Under the program, two types of students were excluded from the calculation of school grades: limited-English-proficient students in an ESOL program for less than two years and several categories of special education students. As a result, threatened schools may have had an incentive to reclassify their low-performing students into these exempted categories in order to remove them from school grade calculations and thereby artificially inflate their marks. Did this actually happen in practice?

Using data obtained from the Florida Department of Education and a regression-discontinuity approach, we compare LEP and ESE classification in schools that barely

[25] It is important to note, though, that students have to test into the special education categories. Consequently, it can be relatively difficult to have higher-performing students test into these categories since they are more likely to pass the diagnostic tests.

[26] The only exemption is for any subgroup with less than forty students in a school (less than fifty for the students with disabilities subgroup). Subgroups with small numbers of students are not evaluated separately, but students in these groups are still included in the evaluation of the "All Students" group.

[27] An exception should be noted here: If a school's total enrollment is less than forty, and even a summing of total enrollment over three years does not yield a total of forty, then that school and its students are exempted from AYP determination. But, as might be expected, this is a very rare occurrence.

avoided the threat of vouchers with such classification in schools that barely received the threat. We find robust evidence that the threatened schools classified a greater percentage of their students into the excluded LEP category in high-stakes grade 4 and entry grade 3. We find no evidence of any differential classification into the included LEP category in any of the grades. For reference, there was no evidence of a difference in behavior between threatened versus non-threatened schools before the program. These findings suggest that schools threatened with vouchers and stigma tended to reclassify students into the excluded LEP category in an effort to remove them from the effective test-taking pool in both the current year and the following year.

In contrast, we find no evidence that the program led to greater classification into excluded (or included) ESE categories by the threatened schools. This result is not surprising given the substantial costs associated with ESE classification. The main disincentive to this form of classification was posed by Florida's McKay Scholarship program, which made any student with disabilities in Florida public schools eligible for vouchers to move to a private school or another public school. Under the McKay program, schools that classified students into excluded ESE categories faced losing them and their corresponding per-pupil funding. Since McKay vouchers cost about twice as much on average as FOSP vouchers, schools actually risked losing more funding with a move of an ESE student under the McKay program than with the departure of a regular student under the Florida program. It is likely that threatened schools weighed the costs and benefits of their options and chose to respond in the least costly ways.

These findings have important implications for school accountability policies in the New York region. New York City's Progress Reports program and New York's implementation of the federal No Child Left Behind Act were modeled in part on the Florida program, though both have avoided the types of exemptions that incentivized gaming of the system in Florida. Because the policies hold schools accountable for the performance of all students—including limited-English-proficient and special education students—New York schools do not have adverse incentives to classify weaker students into these categories. Moreover, schools have the motivation to improve the performance of these and other historically low-performing groups since such improvements are tied to better school grades and concomitant rewards. The New York City program rules, however, have the potential to induce schools to classify their high-performing students into these high-needs groups in an effort to earn extra credit and better grades. Whether or not this kind of sorting actually happened is a topic of future research.

The general lesson to take from examining the Florida and New York accountability policies is that policymakers must be careful when designing exemptions, special allowances, or credits for certain groups of students since these accommodations can create adverse incentives and unintended consequences. While accountability policies must acknowledge the challenges schools face in educating students with limited English proficiency, disabilities, and other special needs, excluding them entirely from accountability measures may induce struggling schools to reclassify low-performing students into exempted categories. The danger is that such an approach can lead to strategic sorting rather than genuine improvements to the quality of education for the students whom the programs aimed to help.

Testing Validity of 1999 Regression-Discontinuity Analysis: Looking for Discontinuities in Preprogram Characteristics at the C/D Cutoff (Reading Sample)

	Percentage				
Panel A	(1) White	(2) Black	(3) Hispanic	(4) Asian	(5) American Indian
	5.99	-6.51	3.12	-0.51	-0.18
	(4.074)	(3.959)	(5.560)	(0.310)	(0.126)
Panel B	Percentage Multiracial	Percentage Male	Percentage Free/Reduced-Price Lunch	Enrollment	Real Per-Pupil Expenditure
	0.20	1.67	-1.19	18.66	0.61
	(0.137)	(0.809)	(1.294)	(42.168)	(0.426)
	Percentage				
Panel C	Exceptional Student Education (ESE)	Excluded ESE	Included ESE	Learning-Disabled	Emotionally Handicapped
	-0.002	-0.004	0.002	-0.004	0.001
	(0.008)	(0.008)	(0.005)	(0.004)	(0.004)
	Percentage Excluded Limited-English-Proficient (LEP)				
Panel D	Grade 2	Grade 3	Grade 4	Grade 5	
	0.075	-0.051	-0.197	-0.058	
	(0.084)	(0.094)	(0.115)	(0.196)	
	Percentage Included LEP				
Panel E	Grade 2	Grade 3	Grade 4	Grade 5	
	0.852	0.952	0.442	0.908***	
	(0.531)	(0.608)	(0.456)	(0.289)	

Source: Authors' calculations.

Note: Robust standard errors adjusted for clustering using the running variable are in parentheses.

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

References

Chakrabarti, R. 2008a. "Impact of Voucher Design on Public School Performance: Evidence from Florida and Milwaukee Voucher Programs." Federal Reserve Bank of New York Staff Reports, no. 315, January.

———. 2008b. "Can Increasing Private School Participation and Monetary Loss in a Voucher Program Affect Public School Performance? Evidence from Milwaukee." Journal of Public Economics 92, nos. 5-6 (June): 1371-93.

———. 2013. "Vouchers, Public School Response, and the Role of Incentives: Evidence from Florida." Economic Inquiry 51, no. 1 (January): 500-26.

Chiang, H. 2009. "How Accountability Pressure on Failing Schools Affects Student Achievement." Journal of Public Economics 93, nos. 9-10 (October): 1045-57.

Cullen, J., and R. Reback. 2006. "Tinkering towards Accolades: School Gaming under a Performance Accountability System." In T. J. Gronberg and D. W. Jansen, eds., Improving School Accountability: Check-Ups or Choice. Advances in Applied Microeconomics 14. Amsterdam: Elsevier.

Fan, J., and I. Gijbels. 1996. "Local Polynomial Modeling and Its Applications." Monographs on Statistics and Applied Probability 66. London: Chapman and Hall.

Figlio, D. 2006. "Testing, Crime, and Punishment." Journal of Public Economics 90, nos. 4-5 (May): 837-51.

Figlio, D., and L. Getzler. 2006. "Accountability, Ability, and Disability: Gaming the System?" In T. J. Gronberg and D. W. Jansen, eds., Improving School Accountability: Check-Ups or Choice. Advances in Applied Microeconomics 14. Amsterdam: Elsevier.

Figlio, D., and C. Hart. 2010. "Competitive Effects of Means-Tested Vouchers." NBER Working Paper no. 16056, June.

Figlio, D., and M. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." American Economic Review 94, no. 3 (June): 591-604.

Figlio, D., and C. Rouse. 2006. "Do Accountability and Voucher Threats Improve Low-Performing Schools?" Journal of Public Economics 90, nos. 1-2 (January): 239-55.

Figlio, D., and J. Winicki. 2005. "Food for Thought? The Effects of School Accountability Plans on School Nutrition." Journal of Public Economics 89, nos. 2-3 (February): 381-94.

Greene, J. 2001. "An Evaluation of the Florida A-Plus Accountability and School Choice Program." Manhattan Institute for Policy Research civic report, February.

Greene, J., and M. Winters. 2003. "When Schools Compete: The Effects of Vouchers on Florida Public School Achievement." Manhattan Institute for Policy Research Education Working Paper no. 2, August.

Hahn, J., P. Todd, and W. van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression Discontinuity Design." Econometrica 69, no. 1 (January): 201-9.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin. 2002. "Inferring Program Effects for Special Populations: Does Special Education Raise Achievement for Students with Disabilities?" Review of Economics and Statistics 84, no. 4 (November): 584-99.

Hoxby, C. 2003a. "School Choice and School Productivity: Could School Choice Be a Tide that Lifts All Boats?" In C. Hoxby, ed., The Economics of School Choice. Chicago: University of Chicago Press.

———. 2003b. "School Choice and School Competition: Evidence from the United States." Swedish Economic Policy Review 10: 9-65.

Imbens, G. W., and T. Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." Journal of Econometrics 142, no. 2 (May): 615-35.

Jacob, B. 2005. "Accountability, Incentives, and Behavior: The Impacts of High-Stakes Testing in the Chicago Public Schools." Journal of Public Economics 89, nos. 5-6 (June): 761-96.

Jacob, B., and S. Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." Quarterly Journal of Economics 118, no. 3 (August): 843-77.

McCrary, J. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." Journal of Econometrics 142, no. 2 (February): 698-714.

References (Continued)

Neal, D., and D. W. Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." Review of Economics and Statistics 92, no. 2 (May): 263-83.

Reback, R. 2008. "Teaching to the Rating: School Accountability and Distribution of Student Achievement." Journal of Public Economics 92, nos. 5-6 (June): 1394-415.

Rockoff, J. E., and L. J. Turner. 2010. "Short-Run Impacts of Accountability on School Quality." American Economic Journal: Economic Policy 2, no. 4 (November): 119-47.

Rouse, C. E., J. Hannaway, D. Figlio, and D. Goldhaber. 2007. "Feeling the Florida Heat: How Low-Performing Schools Respond to Voucher and Accountability Pressure." National Center for Analysis of Longitudinal Data in Education Research Working Paper no. 13, November.

Silverman, B. W. 1986. Density Estimation for Statistics and Data Analysis. London: Chapman and Hall.

van der Klaauw, W. 2002. "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach." International Economic Review 43, no. 4 (November): 1249-87.

West, M., and P. Peterson. 2006. "The Efficacy of Choice Threats within School Accountability Systems: Results from Legislatively Induced Experiments." Economic Journal 116, no. 510 (March): 46-62.

Michael J. Fleming and John R. Sporn

Trading Activity and Price Transparency in the Inflation Swap Market

- Liquidity and price transparency in derivatives markets have become increasingly important concerns, yet a lack of transaction data has made it hard to fully understand how the inflation swap and other derivatives markets work.
- This study uses novel transaction data to shed light on trading activity and price transparency in the rapidly growing U.S. inflation swap market.
- It reveals that the market is reasonably liquid and transparent, despite its over-the-counter nature and low level of trading activity. Transaction prices are typically near widely available end-of-day quoted prices and realized bid-ask spreads are modest.
- The authors also identify concentrations of activity in certain tenors and trade sizes and among certain market participants as well as point to various attributes that explain trade sizes and price deviations.

1. Introduction

An inflation swap is a derivative transaction in which one party agrees to swap fixed payments for floating payments tied to the inflation rate, for a given notional amount and period of time. A "buyer" might therefore agree to pay a per annum rate of 2.47 percent on a $25 million notional amount for ten years in order to receive the rate of inflation for that same time period and amount. Inflation swaps are used by market participants to hedge inflation risk and to speculate on the course of inflation and by market observers more broadly to infer inflation expectations.

Several recent studies have compared the inflation swap rate with breakeven inflation as calculated from Treasury inflation-protected securities (TIPS) and nominal Treasury bonds.[1] The two market-based measures of expected inflation should be equal in the absence of market frictions. In practice, inflation swap rates are almost always higher, with the spread exceeding 100 basis points during the recent financial crisis.

Fleckenstein, Longstaff, and Lustig (forthcoming) attribute this differential to the mispricing of TIPS relative to nominal

[1] Other studies have examined how inflation swaps are priced or have utilized the information in swap rates to make inferences about breakeven inflation. Jarrow and Yildirim (2003) propose an approach for valuing inflation derivatives, which is applied to inflation swaps by Mercurio (2005) and Hinnerich (2008). Krishnamurthy and Vissing-Jorgensen (2011) use changes in inflation swap rates as evidence that the Federal Reserve's quantitative easing increased expected inflation. Rodrigues, Steinberg, and Madar (2009) use swaps to examine the effect of news on breakeven inflation.

Michael J. Fleming is a vice president in the Federal Reserve Bank of New York's Research and Statistics Group; John R. Sporn is a senior analyst in the Bank's Markets Group.

Correspondence: michael.fleming@ny.frb.org

The authors thank Laura Braverman, Darrell Duffie, Glenn Haberbush, Ada Li, Wendy Ng, Johanna Schwab, and seminar participants at the Federal Reserve Bank of New York and at the Commodity Futures Trading Commission 2012 Research Conference for helpful comments. The views expressed 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.

Treasury bonds, and not to inflation swaps.[2] In contrast, Christensen and Gillan (2011) argue that the differential comes from a liquidity premium in inflation swaps as well as a liquidity premium in TIPS.[3] While a recent study examines the liquidity of the TIPS market (Fleming and Krishnan 2012), there is virtually no evidence on the liquidity of the inflation swap market.

Aside from past research on inflation swaps, the issues of liquidity and price transparency in derivatives markets more generally have taken on greater import given regulatory efforts under way to improve the transparency of over-the-counter derivatives markets. In particular, the Dodd-Frank Wall Street Reform and Consumer Protection Act calls for the Commodity Futures Trading Commission (CFTC) and Securities and Exchange Commission to promulgate rules that provide for the public availability of over-the-counter derivatives transaction data in real time.[4] To date, the lack of transaction data has impeded the understanding of how the inflation swap and other derivatives markets operate.

In early 2010, the OTC Derivatives Supervisors Group (ODSG), an international body of supervisors with oversight of major over-the-counter derivatives dealers, called for greater post-trade transparency. In response, major derivatives dealers provided the ODSG with access to three months of over-the-counter derivatives transaction data to analyze the implications of enhanced transparency for financial stability. Fleming et al. (2012) examine the data from the interest rate derivatives market, focusing on the four most actively traded products: interest rate swaps, overnight indexed swaps, swaptions, and forward rate agreements.

This article uses the same interest rate derivatives data set to examine trading activity and price transparency in the U.S. inflation swap market. Specifically, we analyze all electronically matched zero-coupon inflation swap trades involving a G14 dealer for a three-month period in 2010.[5] The data source is MarkitSERV, the predominant trade-matching and post-trade processing platform for interest rate derivatives transactions. An analysis of such data can serve as a resource for policymakers considering public reporting and other regulatory initiatives for the derivatives markets and for market participants and observers more generally interested in the workings of the inflation swap market.

We find that relatively few trades occur in the U.S. zero-coupon inflation swap market. Our reasonably comprehensive data set contains only 144 trades (just over two trades per day) over our June 1 to August 31, 2010, sample period. Daily notional trading volume is estimated to average $65 million. In the TIPS market, in comparison, an estimated $5.0 billion per day traded over the same period, on average.[6]

We identify concentrations of activity in certain tenors, with 45 percent of activity at the ten-year tenor, 14 percent at five years, and 1 percent at three years. Trade sizes tend to concentrate as well, with 36 percent of all trades (and 48 percent of "new" trades) having a notional amount of $25 million. Trade sizes are generally larger for new trades and trades that are allocated across subaccounts, and they tend to decrease with tenor. Over half (54 percent) of trades are between G14 dealers, 39 percent are between G14 dealers and other market participants, and 7 percent are between other market participants. The activity in our data set occurs across nine G14 dealers and nine other market participants.

Despite the low level of activity in this over-the-counter market, we find that transaction prices are quite close to widely available end-of-day quoted prices. After we control for tenor and trading day, the standard deviation of rate differences between our transaction rates and the average end-of-day rates quoted by Barclays and Bloomberg is just 3 basis points. The differential tends to decrease with tenor and increase with trade size and for customer trades. Lastly, by comparing trades for which customers pay and receive inflation, we are able to infer a realized bid-ask spread for customers of 3 basis points, which essentially matches the quoted bid-ask spreads reported by dealers.

Our study proceeds as follows. Section 2 describes how inflation swaps work and the market in which they trade. Section 3 discusses the data used in our analysis. Our empirical results are presented in section 4; section 5 concludes.

[2] Haubrich, Pennachi, and Ritchken (2011) similarly conclude that TIPS were underpriced during the financial crisis. Campbell, Shiller, and Viceira (2009) attribute the differential to anomalous liquidity problems in TIPS.

[3] In their argument, the liquidity premium in inflation swaps comes from reduced funding costs for buyers of inflation and hedging costs for sellers of inflation. Lucca and Schaumburg (2011) also note these hedging costs, as well as TIPS liquidity premia, as explanations for the differences in breakeven inflation.

[4] Inflation swaps fall under the jurisdiction of the CFTC, which, as of December 31, 2012, began requiring real-time public reporting of swap transactions.

[5] The G14 dealers are the largest derivatives dealers and, during the period covered by this study, include Bank of America, Barclays, BNP Paribas, Citigroup, Credit Suisse, Deutsche Bank, Goldman Sachs, HSBC, JP Morgan Chase, Morgan Stanley, Royal Bank of Scotland, Société Générale, UBS, and Wells Fargo.

[6] TIPS volume data come from the Federal Reserve's FR 2004 series and cover activity involving the primary government securities dealers (that is, dealers with a trading relationship with the Federal Reserve Bank of New York). Trades between two primary dealers are reported by each dealer and hence are double-counted.

Zero-Coupon Inflation Swap Cash Flows (at Maturity)

$$Notional \times [(1 + swap\ rate)^{tenor} - 1]$$

$$Notional \times \left(\frac{inflation\ index\ at\ maturity}{inflation\ index\ at\ start} - 1 \right)$$

Notes: The exhibit shows the cash flows exchanged at maturity by swap counterparties. No cash flows are exchanged at the initiation of a swap.

2. Inflation Swaps

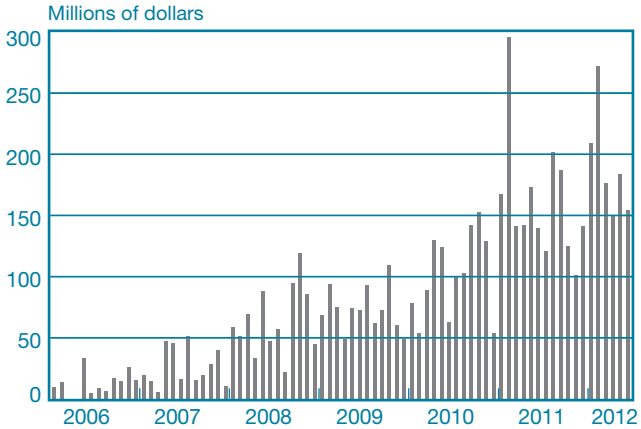
An inflation swap is a bilateral derivatives contract in which one party agrees to swap fixed payments for floating payments tied to the inflation rate, for a given notional amount and period of time. The inflation gauge for U.S. dollar inflation swaps is the nonseasonally adjusted consumer price index for urban consumers, the same gauge used for TIPS. The fixed rate (the swap rate) is negotiated in the market, so that the initial value of a trade is zero. As a result, no cash flows are exchanged at inception of a swap.

The exhibit illustrates the cash flows for a zero-coupon inflation swap—the most common inflation swap in the U.S. market. As the name "zero-coupon" swap implies, cash flows are exchanged at maturity of the contract only. In particular, the inflation payer makes a payment to its counterparty in an amount equal to the contract's notional amount times realized inflation over the term of the contract.[7] The fixed payer, in turn, makes a payment in an amount equal to the contract's notional amount times the annually compounded fixed rate. Technically, cash flows are netted, so that only one party makes a net payment to the other; notional amounts are not exchanged at maturity.

Inflation swaps are used to transfer inflation risk. Entities with obligations exposed to inflation, such as pension funds and insurance companies, can hedge that risk by agreeing to receive inflation. Entities with assets exposed to inflation, such as utility companies, can hedge that risk by agreeing to pay inflation. Other entities may choose to take on inflation risk for speculative or diversification purposes. While inflation risk can also be transferred using securities such as TIPS, inflation swaps can be tailored to more precisely meet investor needs because the swap maturity, notional amount, and other terms are agreed upon at the time of the trade.

[7] To be precise, we note that since changes in the consumer price index are only known with a lag, the floating payment is based on inflation over the period starting three months before the start date and ending three months before the termination date.

Chart 1
Daily Inflation Swap Activity over Time



Source: Authors' calculations, based on data from BGC Partners.

Note: The chart plots average daily brokered inflation swap activity by month.

Inflation swaps trade in a dealer-based over-the-counter market. The predominant market makers are the G14 dealers, which trade with one another and with their customers. In the dealer-customer market, customers can view dealers' indicative two-way prices throughout the day on Bloomberg and receive closing prices from dealers via e-mail. Customers and dealers communicate directly via e-mail and phone and execute trades over the phone.

In the interdealer market, dealers typically trade with one another indirectly via voice brokers. Recently, the brokers have introduced periodic auctions at which dealers can enter their orders to buy or sell contracts of a given tenor at midmarket prices. If a dealer enters an order to buy or sell, other dealers can see that the dealer has expressed interest in trading a particular contract, without knowing if the order is a buy or a sell, and can consider entering their own orders before the auction closes. When the auction closes, contracts for which there is both buying and selling interest are executed at the midpoint between the bid and offer rates in the market.

Evidence suggests that the U.S. inflation swap market has grown quickly in recent years. Data from BGC Partners, a leading broker, indicate that interdealer trading of zero-coupon swaps averaged roughly $100 million per day in 2010, $160 million per day in 2011, and $190 million per day in the first half of 2012 (Chart 1). Data from an informal survey of dealers—accounting for activity with customers as well as activity brokered among dealers—peg the overall market size in April 2012 at roughly $350 million per day.

While the inflation swap market may be modest in size, it is part of a much larger market for transferring inflation risk. This larger market includes other derivatives products as well as more actively traded TIPS and nominal Treasury securities. The broader market provides a vehicle for pricing inflation swaps and for hedging positions taken in the market. As a result, the modest size of the market is not necessarily a good gauge of the market's liquidity or transparency.

3. Data

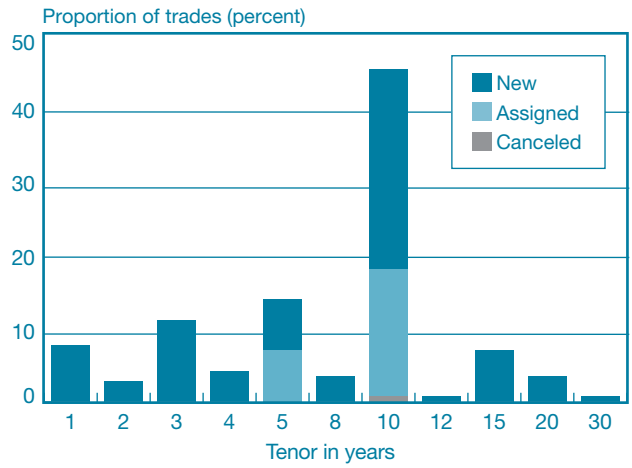
Our primary data set is made up of electronically matched inflation swap transactions between June 1 and August 31, 2010, in which a G14 dealer is on at least one side of the resulting position.[8] The data come from MarkitSERV, the predominant trade-matching and post-trade processing platform for interest rate derivatives. The interest rate derivatives data were provided by the dealers to their primary supervisors so that regulators could assess the derivatives market's conduciveness to trade-level public reporting.

The data provided by MarkitSERV are anonymized, with each firm assigned its own code. No information on firm type is provided aside from the code indicating whether a firm is a G14 dealer. Other firms may be customers of G14 dealers, or other dealers not members of the G14. For brevity, we refer to these other firms as "customers."

Our data set is fairly comprehensive, but does not cover every transaction in this over-the-counter market. First, it excludes transactions involving a G14 dealer that are not electronically confirmed, which account for about 22 percent of G14 dealer interest rate derivatives transactions (Fleming et al. 2012). Second, it excludes transactions not involving a G14 dealer, which account for about 11 percent of interest rate derivatives notional activity in MarkitSERV (Fleming et al. 2012). Additional information pertinent to the activity covered by our data set is discussed in the appendix.

Our data set contains 144 U.S. dollar zero-coupon inflation swap transactions, or an average of 2.2 transactions over the 65 trading days in our sample.[9] Daily notional trading volume is estimated to average $65 million. Three-quarters (108/144) of the transactions are new trades, 24 percent (35/144) are assignments of existing transactions (whereby one counterparty to a swap steps out of the deal and assigns its position to a new counterparty), and 1 percent (1/144) are cancelations. One new transaction has a forward start date, for which the accrual period begins two years after the trade date, with the remaining 107 new transactions starting two or three business days after the trade date.

Chart 2
Inflation Swap Trading Frequency by Tenor



Source: Authors' calculations, based on data from MarkitSERV.

We identify concentrations of inflation swap activity in certain tenors (Chart 2). The ten-year tenor alone accounts for 45 percent (65/144) of activity, followed by tenors of five years (14 percent; 20/144), three years (11 percent; 16/144), one year (8 percent; 11/144), and fifteen years (7 percent; 10/144).[10] There are some differences in tenor by transaction type, with every assigned and canceled trade having an original tenor of five or ten years. In every case, the assigned and canceled trades have a start date less than nine months before the transaction date, so the remaining tenors of such contracts are fairly close to their original tenors.

We also identify a concentration of activity among certain market participants. In particular, 54 percent (78/144) of our trades are between G14 dealers, 39 percent (56/144) are between G14 dealers and customers, and 7 percent (10/144) are between customers. Of the new trades between G14 dealers and customers, the G14 dealer receives fixed 63 percent (19/30) of the time and pays fixed 37 percent (11/30) of the time.[11] New trades in which dealers receive fixed are larger, so that dealers receive fixed for 81 percent of new contract volume. That is, dealers are largely paying inflation and receiving fixed in their interactions with customers.

[8] Because the data set is based on a G14 dealer being a counterparty to the resulting position, it includes assignments of existing positions from a non-G14 dealer to a non-G14 dealer in which a G14 dealer is on the other side, but excludes assignments from a G14 dealer to a non-G14 dealer in which a G14 dealer is not on the other side.

[9] MarkitSERV only supports zero-coupon inflation swaps, so all inflation swaps in the data set are of this type.

[10] Note that the original tenor of every trade in our data set is for a round number of years, to the day.

Five of the G14 dealers report no activity over our sample period. The remaining nine dealers transact on both sides of the market. Our data set also shows activity by nine customers, three that trade on both sides of the market, three that only enter transactions to pay fixed, and three that only enter transactions to receive fixed.

Twenty-six (18 percent) of our transactions contain a mutual put break clause. Such clauses provide for set dates at which parties can terminate contracts at current market value, thereby allowing them to mitigate counterparty credit risk associated with mark-to-market balances on long-dated swaps. While 57 percent (82/144) of all trades have a tenor of ten years or more, 85 percent (22/26) of trades with break clauses have such a tenor. G14 dealer trades with customers are more likely to have a break clause (fifteen of fifty-six trades) than are interdealer trades (eleven of seventy-eight).

Seventeen (12 percent) of the trades in our sample period are allocated, whereby a party transacts in a single bulk amount for multiple accounts. All of these allocated trades are new and all involve customers. On average, there are 6.9 allocations related to a primary (or bulk) trade.

Lastly, 55 percent (79/144) of our trades are brokered (accounting for 60 percent of notional volume) and 45 percent (65/144) are executed directly between counterparties. All thirty-six assigned and canceled trades are executed directly, as are twenty-nine of the thirty new customer-dealer trades. All seventy-eight new interdealer trades are brokered, along with one of the thirty new customer-dealer trades.

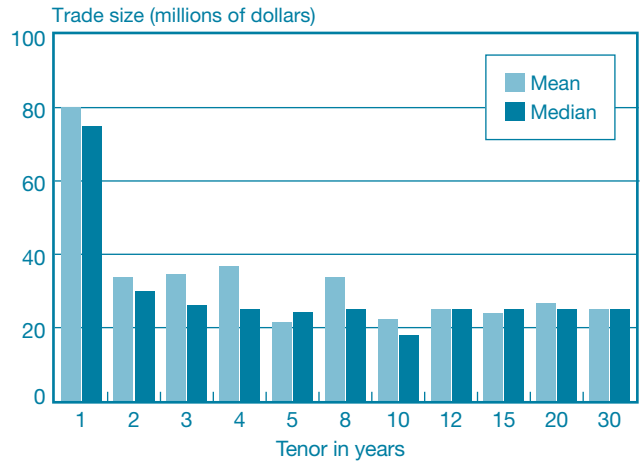
We compare our trading activity figures with figures from BGC Partners as a check on the representativeness of our data set. For our three-month sample period in 2010, BGC reports activity in zero-coupon swaps averaging $89 million per day. Our overall MarkitSERV average is $65 million per day, but the more relevant comparison is brokered activity, which averages $39 million per day. This comparison thereby suggests that our brokered MarkitSERV activity accounts for about 44 percent of all brokered activity (44 percent = $39 million/$89 million).

One other data set we utilize comes from an informal survey of dealers on the liquidity of the zero-coupon inflation swap market. In April 2012, we asked seven primary dealers for information on bid-ask spreads, trade sizes, and trades per day for select tenors and across all tenors in both the customer-dealer and interdealer markets.[12] Our primary interest is in bid-ask spread information, since we lack direct information on bid-ask spreads in our transaction data set, but we are also interested in trade size and trade frequency information as a further check on the representativeness of our MarkitSERV data set.

[11] All thirty-five assignments in our data set involve a customer stepping out of its position. For the twenty-five instances in which the assignment is to a G14 dealer, we are able to infer the dealer's side in fourteen cases. Of those fourteen assignments, the G14 dealer stepped in to receive fixed thirteen times and to pay fixed one time.

[12] All seven primary dealers were members of the G14 during our 2010 sample period.

Chart 3
Inflation Swap Trade Size by Tenor



Source: Authors' calculations, based on data from MarkitSERV.

4. Results

4.1 Trade Size

Inflation swap trade size ranges from $0.2 million to $294 million, with a mean of $29.5 million and a median of $25 million. The most common trade size is $25 million, accounting for 36 percent (52/144) of all trades. An additional 8 percent (12/144) of observations have a trade size of $50 million and 3 percent (4/144) each have trade sizes of $15 million and $100 million. The remaining 50 percent of trades (72/144) occur in fifty-eight different sizes.

One factor explaining trade size is tenor (Chart 3). Trade size tends to decline with tenor, although the largest distinction seems to be between one-year tenors and longer tenors, with only a weak negative relationship past the one-year point. In other securities and interest rate derivatives markets, in contrast, the negative relationship between tenor and trade size appears stronger across the range of tenors and not so

Table 1

Determinants of Inflation Swap Trade Sizes

	Dependent Variable: Inflation Swap Trade Size					
	All Trades					New Trades
Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)
Constant	4.35***	0.61***	3.23***	2.60***	1.26	4.15***
	(0.57)	(0.14)	(0.19)	(0.24)	(1.54)	(0.52)
Tenor	-0.17***				-0.10**	-0.11**
	(0.06)				(0.05)	(0.05)
New trade		3.12***			2.84**	
		(0.38)			(1.23)	
Customer trade			-0.60		0.22	0.21
			(0.62)		(1.24)	(1.24)
Number of allocations				0.43***	0.34***	0.34***
				(0.09)	(0.08)	(0.08)
Adjusted R^2 (percent)	5.0	14.7	0.0	17.6	29.4	17.1
Number of observations	144	144	144	144	144	108

Source: Authors' calculations, based on data from MarkitSERV.

Notes: The table reports results from regressions of inflation swap trade size on tenor, whether a trade is new or not, whether a trade is a customer trade or not, and the number of allocations. Trade size is measured in tens of millions of dollars (notional amount) and tenor is measured in years. Coefficients are reported with heteroskedasticity-consistent (White) standard errors in parentheses.

*Statistically significant at the 10 percent level.
**Statistically significant at the 5 percent level.
***Statistically significant at the 1 percent level.

dependent on a single point (see, for example, Fleming [2003], Fleming and Krishnan [2012], and Fleming et al. [2012]). In general, the negative relationship is likely explained by the higher rate sensitivity of longer-term instruments.

A second factor explaining trade size is trade status. Assigned and canceled trades tend to be smaller and less consistent in size, perhaps because such trades often reduce the amount of—or assign a share of—the original trade. The average trade size for assigned and canceled trades is just $6.1 million, compared with $37.3 million for new trades. The thirty-six assigned and canceled trades occur across thirty different sizes, with none at $25 million or $50 million. In contrast, 48 percent (52/108) of new trades have a size of $25 million and 11 percent (12/108) $50 million. It follows that the relationship between trade size and tenor is more consistently negative if one examines new trades only.

A third factor explaining trade size is whether or not a trade is allocated. Such trades tend to be larger, with an average size of $67.4 million, almost twice as large as the average for new trades overall. Moreover, all three trades in the data set greater than $100 million are allocated as are three of the four trades of exactly $100 million.

We conduct a regression analysis to better understand the relationships between various variables and trade size (Table 1). Our first four regressions are univariate and demonstrate that the relationships between trade size and tenor, trade type, and number of allocations are all statistically significant. On average, an additional year of tenor cuts $1.7 million from trade size, new trades are $31.2 million larger than other trades, and each allocation boosts trade size by $4.3 million. We also test a specification that includes a dummy variable for customer trades, and find such trades to be smaller than interdealer trades (by $6.0 million), but insignificantly so.

We proceed to employ a multiple-regression analysis to show that the previously identified relationships exist independently of one another. That is, the relationships between trade size and tenor, trade type, and number of allocations remain statistically significant, albeit somewhat weaker in magnitude, when we control for the other variables. Results are similar for the subset of transactions that are new. Still further tests suggest that our basic results reasonably characterize the effects of our data set variables on trade size.[13]

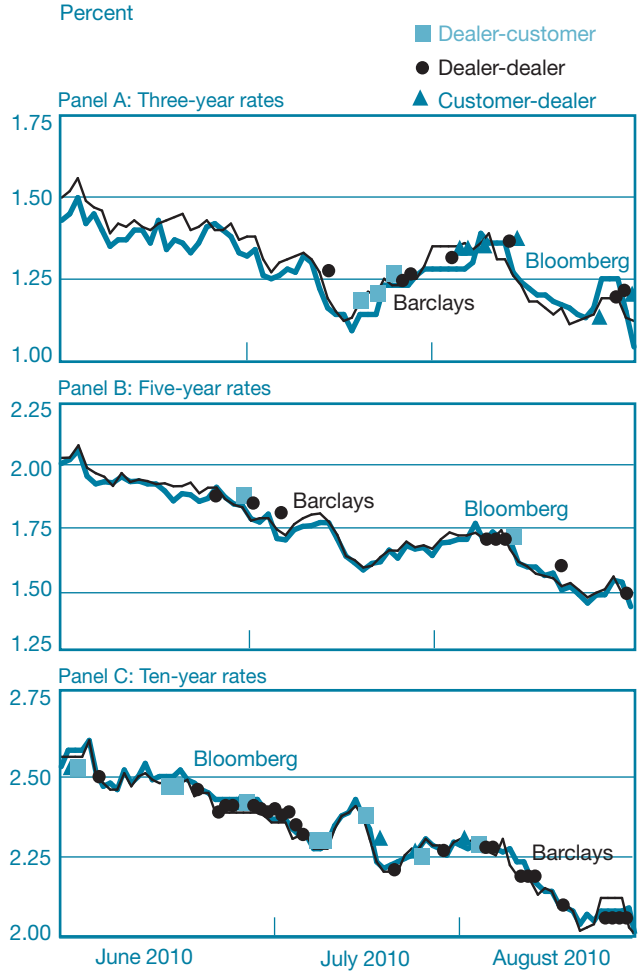
4.2 Price Transparency

Our price transparency analysis examines the relationships among the transaction prices in our data set as well as between the prices in our data set and widely available quoted prices. The purpose of this analysis is three-fold: to understand how close our MarkitSERV transaction prices are to widely available quoted prices, to understand what factors help explain the price differentials, and to provide some insight into the trading costs faced by market participants. We limit this analysis to new trades, which had contract prices negotiated during our sample period, excluding the one new trade with the forward start date.[14]

Visual evidence suggests that the trades in our data set take place at prices close to one another and close to publicly available quoted prices, controlling for tenor and trading day (Chart 4). That is, our MarkitSERV transaction prices look to be within a few basis points of Barclays and Bloomberg quoted prices for a given tenor and trading day. Note that our MarkitSERV prices are from trades throughout the trading day, whereas our Barclays and Bloomberg prices are end-of-day (5 p.m. New York time) midquotes. As a result, one would not expect the MarkitSERV prices to exactly match the other prices even if the inflation swap market were highly transparent and trading costs were negligible.

A more formal look at the data confirms the close relationships among inflation swap prices from the various sources (Table 2). The average differences between MarkitSERV and Barclays, MarkitSERV and Bloomberg, and MarkitSERV and the average of Barclays and Bloomberg are all within 1 basis point after we control for tenor and trading day, with standard deviations ranging from 3 to 5 basis points.[15] The standard deviation is lowest when comparing MarkitSERV with the Barclays/Bloomberg average, suggesting that the average better proxies for transaction prices than either source alone does. Also of note is the fact that the largest differentials among the three sources are observed between Barclays and Bloomberg. The largest differences across sources seem to come from the one-year tenor, with prices much tighter for tenors greater than one year.

Chart 4
Inflation Swap Rates
Percent



Source: Authors' calculations, based on data from Barclays, Bloomberg, and MarkitSERV.

Notes: The chart plots transaction prices from MarkitSERV for select tenors, denoted by whether the trades are between G14 dealers (dealer-dealer); between a G14 dealer and a customer, where the G14 dealer pays fixed (dealer-customer); or between a G14 dealer and a customer, where the customer pays fixed (customer-dealer). End-of-day midquotes from Barclays and Bloomberg are also plotted.

We proceed to assess whether we can explain the deviations that do occur between MarkitSERV transaction prices and other quoted prices. We do this by regressing the absolute difference between the MarkitSERV price and the average of the Barclays and Bloomberg prices (for the same tenor and trading day) on various independent variables. Our independent variables are:

[13] We test a specification with a dummy variable for allocated trades, but the continuous variable better fits the data. We also test specifications including dummy variables for whether there is a break clause and whether a trade is brokered, but neither of these additional variables is significant. Lastly, we test whether the results differ for the subset of transactions with a tenor greater than one year. We find that the coefficient for tenor is cut in half and becomes statistically insignificant in such specifications, the results for new trades are little changed, and the coefficient for number of allocations is little changed (but that the *p*-value for that coefficient increases to about 0.10).

[14] A forward start date could be expected to affect pricing and thus make a contract incomparable to prices for contracts without forward start dates.

[15] The standard deviations are only slightly larger (ranging from 4 to 5.5 basis points) when we compare MarkitSERV transaction prices with Barclays and Bloomberg quoted prices from the preceding trading day.

Table 2

Inflation Swap Rate Differential Statistics

	MarkitSERV-Barclays	MarkitSERV-Bloomberg	MarkitSERV-Barclays/Bloomberg Average	Barclays-Bloomberg
Average deviation	-0.6 [0.6]	0.8 [0.4]	0.2 [0.6]	1.5 [-0.1]
Standard deviation	4.9 [2.8]	3.7 [3.2]	3.0 [2.5]	6.1 [3.3]
Number of observations	106 [95]	107 [96]	106 [95]	106 [95]

Sources: Authors' calculations, based on data from Barclays, Bloomberg, and MarkitSERV.

Notes: The table reports statistics for the difference in inflation swap rates among various sources. The comparisons are made by day and tenor for new transactions, excluding forward transactions. Bracketed figures are based on the subsample of transactions with a tenor greater than one year. Comparisons with Barclays have one fewer observation because we have no Barclays rate for the twelve-year tenor trade in our sample. Differences are in basis points.

- *Tenor:* As noted above, rate dispersion among short-dated tenors seems to be higher, even among widely available data sources.
- *Trade size:* Typical bid-ask spreads are commonly valid only for trades up to a certain size, with larger trades requiring a price concession, so price differences may be positively correlated with trade size.
- *Customer trade:* Customer prices might deviate more from other prices if customers face wider bid-ask spreads than dealers do.
- *Time of trade:* As noted, we have end-of-day quoted prices from Barclays and Bloomberg, but intraday transaction prices from MarkitSERV. Given that prices fluctuate over time, one might expect MarkitSERV prices from trades late in the day to be closer to the end-of-day prices reported by other sources.[16]

Our regression analysis indicates significant univariate relationships between the price deviations and our various variables (Table 3). A one-year increase in tenor is associated with a decrease in the price differential of 0.08 basis point. Each $10 million increase in trade size is associated with an increase in the differential of 0.15 basis point. Customer trades tend to have a differential 0.70 basis point larger than interdealer trades have, and each hour closer to the end of the trading day is associated with a reduction in the differential of 0.09 basis point. A multivariate regression analysis on the full sample of new trades shows that the explanatory variables are independently insignificant (albeit jointly significant) when we control for the other variables.

Given the evidence that price deviations are especially large for contracts with a one-year tenor, we repeat the multivariate analysis on the subsample of trades with a tenor greater than one year. These results show an even weaker effect of tenor, confirming the importance of the one-year trades at explaining the tenor effect. Moreover, trade size and customer trade are significant, and of a similar magnitude as in the univariate regressions, so that larger trades and customer trades tend to occur with larger price differentials for the vast majority of new trades, even after we control for other factors. The time of the trade remains insignificant in the last regression.[17]

4.3 Bid-Ask Spreads

We examine spreads between bid and offer prices in the inflation swap market because they provide a measure of the trading costs faced by market participants. If a customer were to engage in a round-trip trade (that is, enter into a contract to pay fixed as well as a contract to received fixed), for example, it could expect to pay the full bid-ask spread. It follows that a customer engaging in a single buy or sell (that is, entering into a contract to pay fixed or receive fixed, but not both) can expect to pay half of the spread. We assess bid-ask spreads in a couple of different ways.

First, we look at the results of our informal dealer survey. As shown in Table 4, dealers report that bid-ask spreads range from 2 to 3 basis points, depending on tenor. Average trade sizes are estimated to range from $25 million to $50 million in the dealer-customer market and $25 million to $35 million in the interdealer market, consistent with the $29.5 million average we find in our MarkitSERV data. The estimated daily trading frequency of 6 in the customer-dealer market plus 5 in the interdealer market exceeds our overall average of 2.2 by five

[16] Time of trade is measured by the hour of the trading day, based on New York time and a twenty-four-hour clock, so that a trade that occurs at 2:11 p.m. New York time is assigned a value of 14. All but one trade in our data set occurs between 7 a.m. and 5 p.m. New York time, with the exception being 2:14 a.m.

[17] We also test specifications including dummy variables for whether there is a break clause and whether a trade is brokered, but neither of these additional variables is statistically significant.

Table 3

Determinants of Absolute Inflation Swap Rate Differentials

	Dependent Variable: Inflation Swap Rate Differential					
	All New Trades					Greater than One Year
Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)
Constant	2.81***	1.60***	1.90***	3.19***	2.87***	2.36***
	(0.48)	(0.28)	(0.26)	(0.66)	(0.95)	(0.80)
Tenor	-0.08*				-0.05	-0.01
	(0.04)				(0.05)	(0.03)
Trade size		0.15**			0.12	0.13**
		(0.07)			(0.08)	(0.06)
Customer trade			0.70*		0.35	0.96**
			(0.42)		(0.47)	(0.39)
Time of trade				-0.09*	-0.07	-0.09
				(0.05)	(0.07)	(0.06)
Adjusted R^2 (percent)	3.6	5.8	1.3	0.4	6.9	15.8
Number of observations	106	106	106	106	106	95

Source: Authors' calculations, based on data from Barclays, Bloomberg, and MarkitSERV.

Notes: The table reports results from regressions of the absolute inflation swap rate differential on tenor, trade size, whether a trade is a customer trade or not, and the time of the trade. The absolute rate differential is calculated as the absolute value of the difference between the transaction rate from MarkitSERV and the average quoted rate from Barclays and Bloomberg for the same tenor and day. The differential is measured in basis points, tenor is measured in years, trade size is measured in tens of millions of dollars (notional amount), and time of trade is measured in hours. The sample includes new trades only and excludes forward transactions. Coefficients are reported with heteroskedasticity-consistent (White) standard errors in parentheses.

*Statistically significant at the 10 percent level.
**Statistically significant at the 5 percent level.
***Statistically significant at the 1 percent level.

times, likely reflecting growth in the market between 2010 and 2012 and our data set's coverage of less than 100 percent of the market. Overall trading activity per day in April 2012 is estimated to be about $350 million.[18]

A second way in which we look at bid-ask spreads is with the MarkitSERV data. While these data do not contain direct information on bid-ask spreads, such spreads can be inferred from transaction data. In particular, if one knows who initiated a trade, then "realized" bid-ask spreads can be calculated as the difference between the price paid by initiating buyers and initiating sellers. While the MarkitSERV database does not contain information on who initiated a trade, we infer that trades involving customers are initiated by customers (thus, it is dealers making markets for customers and not the reverse).

Suppose, then, that a dealer stands ready to pay 2.00 percent fixed on a ten-year inflation swap and receive 2.03 percent on such a swap. If a customer initiates a transaction with the dealer in which it pays fixed, then it will pay 2.03 percent. If the customer initiates a transaction in which it receives fixed, then it will receive 2.00 percent. The difference in fixed rates between the customer's transactions reflects the dealer's bid-ask spread.

In practice, inflation swap customers rarely buy and sell at the same time. However, by comparing the average rates paid by customers with the average rates received by them, one can obtain a measure of customers' realized bid-ask spreads. Such spreads are often calculated for a particular product and day, because price differences across products and price changes over time add noise to such calculations.

To increase the precision of our estimate, we use the Barclays and Bloomberg prices as reference prices for a given tenor and day. That is, for a given tenor and day, we calculate the difference between the MarkitSERV transaction price and the average of the Barclays and Bloomberg quoted prices. We then generate statistics of these differences for instances in which the customer pays fixed and instances in which the

[18] The $350 million represents the (approximate) median of the market sizes as calculated from each dealer's estimates of trade frequency and trade size for individual tenors.

Table 4

Inflation Swap Dealer Survey Results

Panel A: Customer-Dealer Market

	Three-Year	Five-Year	Ten-Year	All Tenors
Bid-ask spread (basis points)	3	2	2	2.2
Trade size (millions of dollars)	50	50	25	37
Trades per day	1	1	2	6

Panel B: Interdealer Market

	Three-Year	Five-Year	Ten-Year	All Tenors
Bid-ask spread (basis points)	3	2.75	2	2.4
Trade size (millions of dollars)	30	25	25	34
Trades per day	1	2	1	5

Source: Authors' calculations, based on an informal survey of primary dealers.

Notes: The table reports the median responses to an informal survey of seven primary dealers on the liquidity of the zero-coupon inflation swap market in April 2012. For "All Tenors," weighted means are first calculated for each dealer before identifying the median across dealers.

customer receives fixed. As a benchmark, we generate similar statistics for interdealer transactions, for which we have no presumption as to the trade initiator.

As expected, we indeed find that the fixed rate tends to be higher when customers are paying fixed than when they are receiving fixed (Table 5). When a customer pays fixed, the MarkitSERV transaction price is 2.4 basis points higher, on average, than the average of the Barclays and Bloomberg quoted prices. When a customer receives fixed, the MarkitSERV price is 0.4 basis point lower, on average, than the average of the Barclays and Bloomberg prices. The difference—that is, the realized bid-ask spread—is estimated to be 2.8 basis points (2.8 = 2.4 - -0.4) and is statistically different from zero at the 1 percent level.[19] This realized bid-ask spread, calculated for customer-dealer trades, is consistent with the typical bid-ask spreads in the customer-dealer market as reported by dealers.[20]

[19] To assess statistical significance, we regress the price differential on dummy variables for interdealer trades, trades in which the customer pays fixed, and trades in which the customer receives fixed. We then test whether the customer trade coefficients are significantly different from one another, using the heteroskedasticity-consistent (White) covariance matrix. As a robustness test, we repeat this analysis using the previous day's Barclays/Bloomberg average price as the reference, and estimate the realized bid-ask spread to be a slightly larger 3.8 basis points.

Table 5

Inflation Swap Rate Differentials by Trade Type

	Interdealer Trade	Customer Pays Fixed	Customer Receives Fixed
Average	-0.3	2.4***	-0.4###
Standard deviation	2.9	2.8	2.2
Number of observations	77	19	10

Source: Authors'calculations, based on data from Barclays, Bloomberg, and MarkitSERV.

Notes: The table reports statistics for inflation swap rate differentials according to the direction and counterparties of a trade. The rate differential is calculated as the transaction rate from MarkitSERV minus the average quoted rate from Barclays and Bloomberg for the same tenor and day and is measured in basis points. The sample includes new trades only and excludes forward transactions. Statistical significance is determined from Wald tests using heteroskedasticity-consistent (White) standard errors.

*A mean for a group of customer transactions is statistically different from the mean for interdealer transactions at the 10 percent level.
**A mean for a group of customer transactions is statistically different from the mean for interdealer transactions at the 5 percent level.
***A mean for a group of customer transactions is statistically different from the mean for interdealer transactions at the 1 percent level.
#The means for the groups of customer transactions are statistically different from one another at the 10 percent level.
##The means for the groups of customer transactions are statistically different from one another at the 5 percent level.
###The means for the groups of customer transactions are statistically different from one another at the 1 percent level.

5. Conclusion

Our analysis of a novel transaction data set uncovers relatively few trades—just over two per day –in the U.S. zero-coupon inflation swap market. Trade sizes, however, are large, averaging almost $30 million. Sizes are generally larger for new trades, especially if they are bulk and allocated across subaccounts, and tend to decrease with contract tenor.

We also identify concentrations of activity—with 45 percent of trades at the ten-year tenor, and 36 percent of all trades (and 48 percent of new ones) for a notional amount of $25 million. Over half the trades (54 percent) are between G14 dealers, 39 percent are between G14 dealers and other market participants, and 7 percent are between other market participants. We identify just eighteen market participants during our study's sample period, made up of nine G14 dealers and nine other market participants.

[20] While dealers report that spreads vary by tenor, and they likely vary by other attributes of a trade, such as trade size, our small sample of customer-dealer trades limits our ability to examine how bid-ask spreads vary with contract terms.

Despite the low level of activity in this over-the-counter market, we find that transaction prices are quite close to widely available end-of-day quoted prices. The differential between transaction prices and end-of-day quoted prices tends to decrease with tenor and increase with trade size and for customer trades. By comparing trades for which customers pay fixed with trades for which they receive fixed, we are able to infer a realized bid-ask spread for customers of 3 basis points, which is consistent with the quoted bid-ask spreads reported by dealers.

In sum, the U.S. inflation swap market appears reasonably liquid and transparent despite the market's over-the-counter nature and modest activity. This likely reflects the fact that the market is part of a larger market for transferring inflation risk that includes TIPS and nominal Treasury securities. As a result, inflation swap positions can be hedged quickly and with low transaction costs using other instruments, and prices of these other instruments can be used to efficiently price inflation swaps, despite modest swap activity.

An earlier version of this article appeared as an appendix to "An Analysis of OTC Interest Rate Derivatives Transactions: Implications for Public Reporting," by Michael Fleming, John Jackson, Ada Li, Asani Sarkar, and Patricia Zobel, Federal Reserve Bank of New York Staff Reports*, no. 557, March 2012.*

Appendix: Additional Information on Our Measure of Inflation Swap Activity

We note in the "Data" section that our data set covers less than 100 percent of activity in the U.S. zero-coupon inflation swap market. Additional factors relevant to the activity covered by our data set and to the measurement of a trade are as follows:

- Our data set is limited to "price-forming" transactions—defined as trades representing new activity—and excludes "non-price-forming" transactions, such as those related to portfolio compression. Fleming et al. (2012) show that the number and volume of non-price-forming trades in the interest rate derivatives market exceed the number and volume of price-forming trades.
- Our data are aggregated to the execution level, and not examined at the allocated level, so that a trade executed by a money manager on behalf of five accounts is counted once. As noted in the "Data" section, 17 of our trades are allocated, with an average of 6.9 allocations per primary (or bulk) trade.
- There appear to be some "spread" trades in our data set, in which a dealer buys an inflation swap of one tenor and sells a swap of another tenor. Such spread trades appear in the MarkitSERV database as two separate transactions, even though they might be thought of as a single transaction.[21]
- It appears that most assigned trades are executed as part of larger transactions. On June 29, 2010, for example, five ten-year swaps of varying sizes—all with a June 4, 2010, start date—were traded from a customer to a dealer and submitted to MarkitSERV within a three-minute period. Overall, the thirty-five assigned trades in our data set occurred with just six unique combinations of counterparties, trade dates, and start dates.

[21] In the six instances of such apparent spread trades, the submission times for the two sides of the trade differ by only one to five minutes. Moreover, in all six instances, the trade size for the longer tenor is for a round amount (for example, $25 million) and the trade size for the shorter tenor is for a larger and nonround amount (for example, $42.25 million), suggesting that the shorter tenor side may be duration-matched to the longer tenor side.

References

Campbell, J. Y., R. Shiller, and L. Viceira. 2009. "Understanding Inflation-Indexed Bond Markets." Brookings Papers on Economic Activity 40, spring: 79-120.

Christensen, J. H. E., and J. M. Gillan. 2011. "Could the U.S. Treasury Benefit from Issuing More TIPS?" Federal Reserve Bank of San Francisco Working Paper no. 2011-16, June.

Fleckenstein, M., F. A. Longstaff, and H. Lustig. Forthcoming. "The TIPS-Treasury Bond Puzzle." Journal of Finance.

Fleming, M. J. 2003. "Measuring Treasury Market Liquidity." Federal Reserve Bank of New York Economic Policy Review 9, no. 3 (September): 83-108.

Fleming, M., J. Jackson, A. Li, A. Sarkar, and P. Zobel. 2012. "An Analysis of OTC Interest Rate Derivatives Transactions: Implications for Public Reporting." Federal Reserve Bank of New York Staff Reports, no. 557, March.

Fleming, M. J., and N. Krishnan. 2012. "The Microstructure of the TIPS Market." Federal Reserve Bank of New York Economic Policy Review 18, no. 1 (March): 27-45.

Haubrich, J. G., G. Pennacchi, and P. Ritchken. 2011. "Inflation Expectations, Real Rates, and Risk Premia: Evidence from Inflation Swaps." Federal Reserve Bank of Cleveland Working Paper no. 11-07, March.

Hinnerich, M. 2008. "Inflation-Indexed Swaps and Swaptions." Journal of Banking and Finance 32, no. 11 (November): 2293-2306.

Jarrow, R., and Y. Yildirim. 2003. "Pricing Treasury Inflation-Protected Securities and Related Derivatives Using an HJM Model." Journal of Financial and Quantitative Analysis 38, no. 2 (June): 409-30.

Krishnamurthy, A., and A. Vissing-Jorgensen. 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." Brookings Papers on Economic Activity 43, fall: 215-87.

Lucca, D., and E. Schaumburg. 2011. "What to Make of Market Measures of Inflation Expectations?" Federal Reserve Bank of New York Liberty Street Economics blog post, August 15.

Mercurio, F. 2005. "Pricing Inflation-Indexed Derivatives." Quantitative Finance 5, no. 3: 289-302.

Rodrigues, A. P., M. Steinberg, and L. Madar. 2009. "The Impact of News on the Term Structure of Breakeven Inflation." Unpublished paper, Federal Reserve Bank of New York, September.

Federal Reserve Bank of New York
Economic Policy Review

Economic Policy Review articles may be reproduced for educational or training purposes, provided they are reprinted in full; include credit to the author(s), the publication, and the Bank; and include the publication's disclaimer.

Library of Congress Card Number: 77-646559

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*