Does the Community Reinvestment Act Cause Banks to Provide a Subsidy to Some Mortgage Borrowers?[1]

Glenn B. Canner, Elizabeth Laderman, Andreas Lehnert, and Wayne Passmore[2]

First Version: Sep. 2001
This Version: Sep. 2004

---

[1]The opinions, analysis and conclusions of this paper are solely ours and do not necessarily represent those of the Board of Governors of the Federal Reserve System, the Reserve Bank of San Francisco, or their staffs. We received valuable comments and suggestions from seminar participants at Columbia University, the Federal Reserve Board, and the Federal Reserve Bank of Chicago. We thank especially Mark Flannery and Greg Udell for specific comments, Paul Calem for allowing us to use his credit scoring model, and Gerhard Fries for computing statistics from the restricted-use SCF. We received excellent research assistance from Mary DiCarlantonio, Jamie Ingpen, Erin Klein, Judy Peng, and Guillermo Pinczuk. Any remaining errors are our own.

[2]Canner, Lehnert, and Passmore are at the Board of Governors of the Federal Reserve System in Washington DC, Laderman is at the Federal Reserve Bank of San Francisco. Corresponding author: Andreas Lehnert, Board of Governors of the Federal Reserve, Mail Stop 93, Washington DC 20551, Andreas.Lehnert@frb.gov, 202-452-3325.

Abstract

The Community Reinvestment Act (CRA) encourages banks and other depository institutions to lend to lower income households and to households of any income who live in a lower income area; non-depository lenders, such as independent mortgage bankers, are not directly affected by the CRA. Although the CRA applies to many types of loans (and banking services), we concentrate on mortgage lending; in particular, we test whether the CRA leads depository institutions to cross-subsidize from non-CRA-eligible borrowers to CRA-eligible borrowers. We construct a novel dataset containing mortgage terms and borrower characteristics; in addition, we use auxiliary data to bound the selection bias in our estimates. We find that the CRA leads to economically insignificant changes in mortgage rates for CRA-eligible mortgage borrowers. Our data also provides evidence that banks play a different role from mortgage bankers in mortgage markets.

Journal of Economic Literature classification numbers: G21, G28, H23, R21

Keywords: Community Reinvestment Act, mortgages, bank regulation

1. Introduction

The Community Reinvestment Act (CRA) requires the federal agencies that supervise commercial banks and savings associations to encourage such lenders "to help meet the credit needs of the local communities in which they are chartered, consistent with the safe and sound operation of such institutions." In practice, regulators judge banks primarily by the number and dollar amount of loans that they make to lower-income borrowers, or borrowers in lower-income neighborhoods. If banks satisfy their regulators by attracting such borrowers with lower interest rates than justified by objective risk measures or expected return, the CRA would lessen the profitability of the banking system.[3]

We test for such an effect of the CRA on interest rates using data on home purchase mortgages made over the period 1995-2000. We exploit variation in the institutions subject to the CRA and the types of borrowers (and neighborhoods) for whom banks receive CRA credit.[4] One of the main contributions of this paper is linking data on mortgage characteristics, including interest rates, with household-level data, including borrower income and geographic location. This allows us to test for a CRA effect on interest rates, a lacuna in the current literature.

Mortgages are made by a wide variety of financial institutions to many different borrower types. The greatest distinction among lenders is between depository institutions, such as commercial banks and savings associations, and non-depository institutions, such as independent mortgage companies.[5] Banks are subject to the CRA, while mortgage companies are not. Further, only loans to certain borrowers are eligible for CRA credit. We use this variation to test whether the CRA encourages banks to lower mortgage rates in order to attract CRA-eligible borrowers.

---

[3]For studies of the effects of the CRA on bank profitability see Canner and Passmore (1996), Avery, Bostic and Canner (2000,2002), Board of Governors of the Federal Reserve System (1993,2000), Meeker and Myers (1996), Johnson and Sarkar (1996) and Malmquist, Phillips-Patrick, and Rossi (1997). For a review of the theoretical arguments surrounding CRA, see Canner and Passmore (1995b). Our paper is similar in interest to Zinman (2002), who studies the small business aspect of CRA while we concentrate on the mortgage aspect, and to Bostic and Robinson (2003) who study the effect of CRA agreements on bank lending patterns.

[4]For a study of the extent of lending to lower-income borrowers by CRA-affected institutions compared to other institutions, see Canner and Passmore (1996).

[5]We refer to depository institutions as "banks" throughout this paper.

One might object that in a competitive mortgage market, interest rates that do not result in a competitive return to capital cannot be sustained. However, if one views the banking industry as imperfectly competitive, so that loan demand is not perfectly elastic at a uniform market-clearing rate, lower interest rates on certain types of loans can be sustained. The main effect of a binding CRA constraint would be to decrease the present discounted value of economic rents associated with a bank charter.

Simple comparisons of interest rates between CRA-eligible and non-CRA-eligible loans may be subject to selection bias because we cannot observe all of the relevant details for each borrower. In particular, we cannot observe the borrower's credit quality as proxied by a credit score. Thus, on a risk-adjusted basis, two apparently similar interest rates may actually be quite different. We use auxiliary data sources to characterize the distribution of credit scores conditional on observable characteristics (primarily income). We find that higher-income households have much lower variation in credit scores than do other households. By restricting the data to just higher-income households, we can bound the bias introduced as a result of selection.

We find that banks do charge CRA-eligible borrowers slightly lower interest rates. Our estimates of the magnitude of this subsidy range between 0 and 8 basis points; depending on specification, we are often not able to reject the hypothesis that there is no subsidy However, even our largest point estimate of the subsidy is not economically meaningful.

The plan of this paper is as follows: section 2 reviews the salient features of the CRA, section 3 describes in some detail the construction and nature of our data, section 4 presents our identification strategy and section 5 presents our results. Section 6 briefly concludes. Details on the construction of our dataset are in an appendix.

2. The Community Reinvestment Act

The CRA was enacted in 1977 and is intended to encourage commercial banks and savings associations to help meet the credit needs of the local communities in which they are chartered. In adopting the CRA, the Congress reaffirmed the principle that depository institutions have an obligation under their charters to serve “the convenience and needs” of their communities by extending credit to all parts of those communities.[6]

The CRA is directed primarily at four federal supervisory agencies--the Board of Governors of the Federal Reserve System, the Comptroller of the Currency, the Federal Deposit Insurance Corporation, and the Office of Thrift Supervision. The Act calls upon these agencies to (1) use their supervisory authority to encourage each financial institution to help meet local credit needs in a manner consistent with safe and sound operation, (2) assess an institution's record of meeting the credit needs of its entire community, including lower-income neighborhoods, and (3) consider the institution's CRA performance when assessing an application for a charter, deposit insurance, branch or other deposit facility, office relocation, merger, or acquisition.

To enforce the CRA, the regulatory agencies conduct periodic CRA examinations of commercial banks and savings associations and, as required by the statute, evaluate CRA performance during the application process for bank acquisitions, mergers and other actions. The vagueness of the affirmative responsibility placed on lenders by the Congress has made it more difficult for regulators to determine compliance with the CRA. Most institutions receive a rating of satisfactory or better on their CRA performance, and few institutions have had their applications for mergers or acquisitions denied. The CRA has, however, prompted institutions to undertake specific actions to enhance their CRA performance before and during the application process; for an overview of these programs and their profitability, see Avery, Bostic, and Canner (2000).

In 1995, the agencies began implementing a revised CRA regulation that, for larger institutions, uses three distinct performance-based measures: a lending test, an investment test and a service test.[7] These tests combine the judgment of CRA examiners with quantitative measures of performance, such as the ratio of mortgages extended to lower-income borrowers to all mortgages and the ratio of mortgages extended in lower-income neighborhoods to all mortgages. When adopting the new regulation, the agencies noted that the examination process is inherently subjective and requires that performance be measured within the context of (1) a community's

[6]For an overview of the history of the CRA, see Garwood and Smith (1993).

[7]In this paper, we focus exclusively on the lending test portion of CRA regulation (see 12 CFR 228, regulation BB Community Reinvestment"; section 28 contains the lending test).

credit needs and (2) the capability of the lender. These two standards are referred to as the "performance context".

The CRA legislation places a heavy emphasis on the analysis of the geographic distribution of an institution's lending across its entire community. The current CRA regulation implements this legislative intent by classifying neighborhoods in a lender's CRA assessment area as low-, moderate-, middle-, or higher-income. A lower-income area is defined as an area where the median family income is less than 50 percent of the median family income for the broader area (such as a metropolitan statistical area or MSA). In a moderate income area, the median family income is at least 50 percent and less than 80 percent of that for the broader area. In a middle-income area, the percentages range from at least 80 percent up to 120 percent. And in a higher-income area, the percentage is at least 120 percent. These income definitions divide the population and the number of census tracts into groups of unequal size, with far fewer people, owner-occupied homes, and census tracts in the lower-income groups.[8] We refer to neighborhoods (or borrowers) with less than 80 percent of the MSA median family income as lower-income neighborhoods (or borrowers).

The current CRA regulation also extends the evaluation of a bank's lending to encompass the distribution of loans across low-, moderate-, middle-, and higher-income borrowers, where the income categories follow the same groupings as neighborhoods but rely on the individual's income relative to the MSA median family income. Thus, while continuing to place a heavy emphasis on the geographic distribution of an institution's lending, the agencies also favorably consider loans made to lower-income individuals.

One result of this dual approach (considering both the borrower's own income, as well as the income of the neighborhood in which the borrower seeks to purchase a home) is that loans to higher-income borrowers are favorably considered by regulators for CRA purposes only if the borrower is purchasing a home in a lower-income neighborhood. Indeed, loans to higher-income borrowers not purchasing homes in lower-income neighborhoods might be seen as counting against the lender for CRA purposes, because regulators use the ratio of loans to lower-income borrowers

[8]See Canner and Passmore (1995b).

relative to the total to gauge a lender's CRA performance. We exploit this fact in our empirical work below.

CRA examinations consider a broad range of loan products, including all types of residential, consumer, and business loans. Our paper is focused only on home purchase lending, an important component of the lending test, because the data available pursuant to the Home Mortgage Disclosure Act (HMDA) allow the empirical investigation of the nature and extent of this type of lending by the mortgage industry to different neighborhoods and different borrowers in all MSAs.[9]

3. Data

3a. Construction of dataset

It is difficult to come by data to empirically test the effects of CRA.[10] In this paper, we merge three separate datasets in order to create a single dataset that can be used to test the interest rate effects of the CRA. We begin with the loan-level microdata collected under the Home Mortgage Disclosure Act (HMDA). These records are filed for every mortgage application received by mortgage lenders covered by the law, and constitute close to a census of mortgage activity.[11] We merge these records with the Federal Housing Finance Board's Monthly Interest Rate Survey, or MIRS. This survey is restricted to a sample of lenders, who are asked to report terms and conditions on all single-family, purchase-money, fully amortized, conventional loans closed in the last five business days of the month. Finally, we use data on mortgages with private mortgage insurance (PMI)

---

[9]The HMDA requires financial institutions with offices in metropolitan areas to provide information on the geographic location of the properties related to the home loans they originate or buy. HMDA also requires lenders to disclose information on the disposition of home loan applications, the date of the loan, and on the race or national origin, gender, and annual income of loan applications or borrowers.

[10]However, there are a few studies: Avery, Bostic and Canner (2000, 2002), Board of Governors of the Federal Reserve System (1993, 2000), Canner and Passmore (1996), Canner, Passmore and Surette (1996), Evanoff and Segal (1996,1997)CPS96b, Evanoff and Segal (1996, 1997) and Harvey, Collins, Nigro, and Robinson (2001).

[11]Federal Reserve estimates suggest that HMDA covers between 82 and 85 percent of all mortgage lending.

collected by the Federal Financial Institutions Examination Council (FFIEC). These data follow the same format as the HMDA data.

The HMDA data are the main public source of mortgage borrower information, but lack pricing information (interest rates, points, and PMI) and the value of the property securing the mortgage. HMDA provides information on residential mortgages, including their type: (conventional or government-backed), purpose of the loan (home purchase, home improvement, or refinancing), and the amount of the loan. In addition, HMDA includes some information about borrowers (including income, race, ethnicity, and gender), as well as the lender type (commercial bank, savings association or mortgage bank), and location of the property securing the loan (state, MSA, county and census tract). With the property location, the characteristics of the neighborhood the property is located in can be determined from the 1990 Census of Population and Housing.

The MIRS collects pricing information on mortgages, including the contract interest rate, points, the effective interest rate and term to maturity, but only covers conventional home purchase loans.[12] It also includes the loan amount, property value and loan-to-value ratio for the mortgage and the location of the property by zip code.

Finally, the FFIEC data on private mortgage insurance (PMI) identifies the lender, location of the property by census tract, loan amount and characteristics of the borrower (e.g. race, ethnicity, gender and income) for each conventional mortgage backed by PMI.

We briefly describe in this section the procedure used to match records from these three separate datasets. A more complete description of the procedure can be found in the appendix, where we also present some statistics that allow an evaluation of the quality of the match.

There is no unique borrower identifier (such as borrower name, social security number or street address) in any of these databases nor is there a unique lender identifier across all three databases, so we "statistically" matched the loan records on conventional home purchase mortgages across the three databases using the following procedure: first, we use the date the mortgage was closed from each database to determine the month of mortgage origination. Second, we converted the geographic property identifiers in HMDA into zip codes and grouped the HMDA

[12]The effective interest rate accounts for both the contract interest rate, and points and fees paid by the borrower. As such the effective rate is a more accurate representation of the mortgage rate than the contract rate.

and FFIEC PMI data records by zip codes. Only loans located in MSAs were considered in our analysis.

At this point, we had a set of mortgage loans from each database for each month of the years 1995 to 2000 grouped by zip code. For each of these month-location groups, we matched records in HMDA and MIRS using the loan amount. Records with the least difference in loan amount were matched and kept unless the absolute difference was greater than $2,000; in this case, no match was made. The result was a set of statistically matched borrower/loan records that contained both the HMDA and MIRS information.

The PMI information was added to these HMDA/MIRS matched records using another statistical match. For the records in a particular month-location group, we matched the PMI to the HMDA/MIRS record based on loan amount (using the $2,000 absolute difference criterion again), and borrower race and ethnicity (which had to be exact matches).

Although the resulting data contains information on adjustable-rate mortgages, 15-year fixed-rate mortgages and 30-year fixed-rate mortgages, we eliminated all observations that did not relate to a 30-year fixed-rate mortgage. Such mortgages made up about 90 percent of our data, and by concentrating solely on a single financial product, we could compare interest rate spreads directly, without having to control for amortization length or the fixed/floating spread. See the appendix for evidence that mortgage type and CRA eligibility are not related.

Our empirical analysis features a full set of lender and MSA fixed effects; if a particular lender or MSA appears only seldom in our sample, the fixed effect may completely absorb any identification provided by those observations. Further, we want to concentrate on active lenders, as opposed to lenders who make few loans (of the type we study) per year. Thus we impose two conditions on each observation for it to be included in our final dataset. First, we require that each observation come from an institution that accounts for at least 0.1 percent of the records in our original database. Second, we require that each observation come from an MSA that also accounts for at least 0.1 percent of the records in our original database.[13]

---

[13]Further, the survey conducted by the FHLB to construct the MIRS excludes very small lenders; for this reason, excluding less-active lenders from our dataset probably purges some spurious matches.

The original dataset contained information on 314,009 conventional home purchase mortgages made by 2,685 different lenders in 304 different MSAs; about 76 percent of the mortgages in the original dataset were made by commercial banks, saving associations and their mortgage lending affiliates. After applying our two conditions, the final dataset comprised 250,593 mortgages made by 84 different lenders operating in 144 different MSAs; about 77 percent of the mortgages in this final dataset were made by banks.[14]

3b. Sample statistics

Our measure of the loan interest rate is the spread, in percentage points, between the effective rate on the loan and the average prevailing 10-year Treasury rate in the month in which the loan was made. Figure 1 compares the time series of this spread measure in our data to the Freddie Mac conforming mortgage index; also, the figure compares the number of observations in our data (per month) to the Mortgage Banker's Association purchase index. The mean spread in our data tracks the spread from Freddie Mac's primary mortgage market survey quite well, with a correlation of about 90 percent. The mean spread in our data, however, consistently exceeds Freddie Mac's published average, probably because our data include non-conforming and high LTV mortgages and because we use the effective rate on the mortgage while the index tracks the contract rate and points separately. The monthly pattern of observations in our data closely mirrors the pattern in the Mortgage Banker's Association (MBA) purchase index, with a correlation of about 90 percent between the two series.

Table 1 shows sample statistics of selected variables from our dataset. As discussed in section 2, banks can count loans towards their CRA goals if the borrower's income is relatively low, if the borrower is purchasing a property in a census tract with a relatively low median family income, or both. As shown in table 1, banks are more likely than transaction lenders to make CRA-eligible loans; 29.5 percent of loans made at banks are CRA eligible, compared to 22.6 percent at transaction lenders. Most of this difference appears to stem from the fact that lower-

[14]For the purposes of CRA examination, banking institutions can choose to include the mortgage lending of their mortgage affiliates.

income borrowers are likelier to use banks, where 25.4 percent of loans are to lower-income borrowers compared to 18.8 percent of loans at transaction lenders. Loans at transaction lenders are slightly larger, more likely to be sold to Fannie Mae or Freddie Mac, and less likely to carry PMI.

Table 2 shows the spread conditional on borrower's income status and lender type. Lower-income borrowers at banks have spreads about 12 basis points below lower-income borrowers at transaction lenders. Also, among all borrowers at banks, those with lower incomes have spreads about 10 basis points less than borrowers with moderate or high incomes. Although these differences are far from statistically significant, they at least suggest that banks might cut interest rates to attract CRA-eligible borrowers. Of course, this simple table does not control for a host of covariates (such as loan size, loan-to-value ratio, or PMI) or the selection problem that we discussed earlier.

4. Identification Strategy

4a. Identification Problem

In order to best identify whether the CRA results in lower mortgage rates for some borrowers, lenders would ideally be randomly subject to the CRA, borrowers would be randomly eligible for CRA credit, and all relevant borrower-level variables would be observed. In such a world, borrowers who were CRA-eligible and got loans from CRA-affected institutions would be "treated". We could identify a subsidy by comparing the mortgage rates paid by treated borrowers to the mortgage rates paid by otherwise-identical untreated borrowers.

Notice that even in this ideal world, a CRA subsidy would lead CRA-eligible borrowers to prefer CRA-affected institutions. This selection problem would be overcome by assuming that all variables influencing the selection decision be observable by the econometrician.

Given such an ideal setup, one would regress mortgage interest rate spreads on variables related to the borrower's riskiness, the prevailing cost of funds, date-, MSA- and institution-level fixed effects and a treatment indicator variable. If the estimated effect of CRA eligibility at CRA-affected institutions (the coefficient on the treatment variable) were negative in this regression, we would conclude that the CRA was causing affected lenders to attract CRA-eligible borrowers with lower mortgage rates than the lenders would normally charge such borrowers, i.e. that the CRA caused banks to subsidize mortgages to eligible borrowers.

Our actual data deviate from this ideal in two related ways. First, only banks are subject to the CRA. Banks, unlike transaction lenders, may have an ongoing relationship with a borrower and thus may use soft information in pricing the loan, or bundle the loan with other depository services, particularly in the cases of wealthy borrowers or borrowers who are CRA-eligible. (Hence the commonly-used term "relationship lenders" for banks.) Mortgage bankers (or "transaction lenders") are more commonly seen as making commodity loans, that is, loans not tied to any other relationship with the borrower and often have no other interaction with the borrower than originating (and then selling) their mortgage. Thus, mortgage interest rates may differ between banks and transaction lenders even without the CRA .

Second, we face a more standard selection problem. We do not observe all relevant information used by lenders in determining mortgage interest rates. In particular, we do not observe a borrower's credit history. Thus, two borrowers could appear identical in our dataset, and be charged the same interest rate, but if the borrower at the bank had a poor credit history, the bank would in effect be charging a lower risk-adjusted interest rate than the transaction lender.

We use two strategies to identify the effect of the CRA. First, in our basic empirical specification we interact lender type with borrower-level characteristics, including whether or not the borrower is CRA-eligible. As a result, identification of the CRA treatment effect comes from comparing borrowers within banks. Observations from transaction lenders are used to identify other parameters, such as time and MSA fixed-effects.

Second, we deal with the problem of unobserved borrower heterogeneity by restricting the borrower population to those with relatively little credit score variation (discussed below).

4b. Controlling for Endogenous Credit Quality

As we have noted, we do not observe borrower credit quality, leading to potential selection bias in the estimated effect of CRA on mortgage rates. However, our dataset includes several important variables related to borrower credit risk. These are: loan size, borrower income, house value, the presence of PMI, whether and how the loan was securitized and several neighborhood characteristics.[15]

In addition to these variables, lenders also observe the borrower's credit history, usually summarized as a credit score. Credit scores reflect a borrower's repayment history, other debts, and credit utilization.[16] Lenders are likely to charge otherwise-identical borrowers different mortgage rates if their credit scores vary widely.

If borrowers that look the same in our data have systemically lower credit scores at banks than at transaction lenders we will understate the true magnitude of any CRA-related subsidy.[17] Although we do not observe credit scores in our data, we can statistically match borrowers from our dataset with borrowers in the Survey of Consumer Finances (SCF), where can use imputed credit scores. We can then measure the residual variation in credit scores once we control for the variables we observe.

More specifically, we use the credit-scoring algorithm developed by Barakova, Bostic, Calem, and Wachter (2003), which approximates the credit scoring model used by a large credit reporting agency.[18] Because the SCF has high-quality data on income, assets and credit accounts, including information on recent payment performance, these imputed credit scores are likely to be highly accurate reflections of the credit scores actually assigned to households in the SCF sample.

Table 3 shows the imputed credit scores for five income categories (corresponding to the absolute income thresholds used below) and, separately, for six loan-to-value categories

[15]Note that we transform several of these variables, so that we work with the loan-to-value ratio, for example, as opposed to loan size and house value separately.

[16]Repayment history includes current and past bankruptcies, serious delinquencies, and collection agency accounts; other debts include consumer credit, auto loans, and student loans; and credit utilization is defined as the ratio of outstanding balances to available credit limits.

[17]More specifically, if higher-income borrowers buying houses in lower-income neighborhoods using loans from banks have lower credit scores than higher-income borrowers buying houses in higher-income neighborhoods, we would understate the true subsidy.

[18]This work would not have been possible without the active assistance of Federal Reserve Board staff. We are grateful to Paul Calem for making his SCF credit scoring algorithm available to us. Also, Gerhard Fries kindly produced a variety of statistics using the non-public-use SCF microdata for us.

(corresponding to the LTV categories used in our regressions). The sample of households in the SCF was restricted to those who had taken out a new mortgage within the 24 months before they were surveyed. The mortgages in our primary data, of course, are all newly originated. Tightening this window to 12 months (which might better match the primary data) has no significant effect other than to decrease the sample size.

It is interesting to note that, while mean credit scores drop with increasing loan-to-value ratios, ranges do not tighten. This is probably because high LTVs are associated both with loans made to financially troubled borrowers and with loans made to high quality borrowers purchasing extremely expensive properties.

However, the most important features of table 3 are the statistics conditional on annual income. Among the group with annual incomes of $120,000 or more, the amount of absolute variation in credit scores is much smaller than for any other group. Further, the tenth percentile credit score is 680, a relatively high value. We use these income thresholds in section 5b to restrict the sample. Table 3 demonstrates that borrowers with high incomes have much less variation in their credit scores than borrowers in other categories. In addition, higher-income borrowers are much less likely to have credit scores considered subprime.[19] The standard deviation of scores is lower for this group; more importantly, though, so is the interdecile range. This bounds the amount of variation in credit scores, in turn limiting the role that selection could play in our results.

4c. Empirical Specification

Our identification strategies suggest both empirical specifications and restrictions on the data. Because banks may price borrower risk differently than transaction lenders, we interact lender type with borrower-specific variables (include CRA eligibility). In order to limit the role played by unobserved credit score variation, we limit our dataset to higher-income borrowers only. Thus, our specification will identify a CRA subsidy by comparing the interest rate paid by higher-income borrowers at banks who buy a house in a lower-income neighborhood (CRA-eligible) to the interest

[19]Although the credit score cutoff used to categorize a borrower as "subprime" is subject to some variation among lenders, values around 620-660 are typically cited.

rate paid by higher-income borrowers at banks who do not buy a house in a lower-income area (not CRA-eligible).

However, for completeness we also present results without these refinements. This also has the advantage of allowing us to present results from both forms of CRA eligibility, buying a home in a lower-income area or having a lower income. Our income restriction rules out the latter eligibility channel. In all the specifications we present here, we include fixed effects for the institution (i.e. the particular bank or mortgage company that made the loan), the MSA, and the year.

In addition to the standard set of variables, there are four variables of particular interest. First is an indicator variable, *CRA_ELIG*, which is set to unity if the loan is eligible for CRA credit. This variable comprises the union of two other indicator variables, *LOWMOD* and *IRLT80* . The first of these, *LOWMOD*, indicates whether the borrower's neighborhood is lower income, which makes the loan eligible for CRA credit, regardless of the borrower's own income. The second of these, *IRLT80*, indicates whether the borrower's own income is below 80 percent of the median income of the MSA, in which case the loan also is eligible for CRA credit, regardless of the borrower's neighborhood. Note that *CRA_ELIG* does not depend on the lender type, only the borrower's characteristics.

The final variable of interest is *REL_LEND*. This is an indicator variable set to unity if the lender is subject to the CRA. In practice, of course, *REL_LEND*=1 for loans made by banks and *REL_LEND*=0 for loans made by transaction lenders.

We now are ready to describe our specifications. In all of our regressions, the dependent variable will be the spread, in percentage points, between the effective rate on the loan and the average prevailing 10-year Treasury rate in the month in which the loan was made. We label the specifications (*A*) and (*B*):

$$(A) \qquad S_i = X_i B + \alpha CRA\_ELIG_i + \lambda REL\_LEND_i + \alpha_\lambda CRA\_ELIG_i \times REL\_LEND_i + u_i$$

$$(B) \qquad S_i = W_i B^w + Z_i B^z + \alpha CRA\_ELIG_i + REL\_LEND_i \times [\lambda + W_i B_\lambda^w + \alpha_\lambda CRA\_ELIG_i] + u_i \quad .$$

Here, *X* is the set of control variables other than those designating CRA eligibility or lender type. These variables are partitioned into *W*, designating borrower-level variables associated with risk, and other variables, Z, such as institution and MSA dummies.

In both specifications, the estimated value of α_λ gives the estimated CRA subsidy (if negative). Specification (B) interacts borrower-level risk characteristics with lender type. As we have discussed, this narrows identification to a comparison of loans made within banks and not between banks and transaction lenders.

One immediate generalization that we apply to equations (*A*) and (*B*) is to split the definition of CRA eligibility into its parts. Thus, for example, we would estimate the coefficients of an alternative version of equation (*A*), denoted equation (*A'*):

$$(A') \qquad S_i = X_i B + \alpha^{(1)} LOWMOD_i + \alpha^{(2)} IRLT80_i + \lambda REL\_LEND_i$$
$$+\alpha_\lambda CRA\_ELIG_i \times REL\_LEND_i + u_i$$

In the same way, we would estimate the extra coefficients from the alternate specifications of equation (*B*) with expanded definitions of CRA eligibility; we denote this specifications as (*B'*).

5. Results

Before turning to the regression results, it is instructive to consider some conditional sample means. Consider again table 2, which gives the means and standard deviations of the spreads on the mortgages in our final dataset conditional on lender type and borrower income category. Notice that transaction lenders charge lower-income borrowers a higher spread than they charge other borrower types, while relationship lenders charge lower-income borrowers a lower spread than they do other borrowers. For other borrower income groups, though, relationship lenders charge a slightly higher spread than do transaction lenders. Without any further investigation, one might conclude that relationship lenders are increasing spreads on non CRA-eligible borrowers slightly in order to offer CRA-eligible borrowers attractive, low, spreads.

We present a complete set of regression coefficients for our four specifications (equations (*A*) and (*B*) with both definitions of CRA eligibility) using data from all borrowers in table 4. Table 5 presents estimates of the parameters of interest under both specifications and with different income thresholds. (Complete regression results are available upon request.)

5a. Results From All Borrower Types

Estimated regression coefficients from specifications (A), (A'), (B) and (B') are shown in table 4. In these regressions we used all available data, including data from lower-income borrowers. The estimated CRA subsidy in these regressions can be found using the estimated interaction coefficients α_λ, $\alpha_\lambda^{(1)}$ and $\alpha_\lambda^{(2)}$.

Under specification (A) and (A'), which does not allow banks and transaction lenders to price risk differently, the estimated CRA subsidy is 7.2 basis points using a unitary definition of CRA eligibility. Separating CRA eligibility into its parts (in specification A'), gives a subsidy estimate of 6.2 basis points for buyers in lower-income areas and 7.8 basis points for lower-income borrowers.[20]

Specifications (B) and (B'), which interact borrower risk variables with lender type, however, decrease the estimated subsidy by at least half in absolute value. Using the unitary definition of CRA eligibility, the estimated interaction effect drops from 7.2 basis points to 1.8 basis points. The estimated subsidy to buyers in lower-income neighborhoods drops from 6.2 basis points to 3.6 basis points, and the subsidy to lower-income borrowers drop even more, from 7.8 basis points to 2.2 basis points.

5b. Results Using Absolute Income Thresholds

Table 3 allows us to quantify the amount of possible unobserved credit score variation in the sample if we restrict attention to borrowers of a particular income threshold. However, because we concentrate on relatively higher-income borrowers, none will be eligible for CRA credit purely on the basis of the income test alone (in other words, none will have incomes below 80 percent of the MSA median income). As a result, we can estimate specifications (A') and (B') only, where *IRLT80=0* always. Identification of a CRA subsidy will come from borrowers that purchase homes in lower income areas (*LOWMOD=1*). We estimated our model after restricting the dataset to only those borrowers with annual incomes above \$60,000, \$80,000, \$100,00 or \$120,000, as suggested by table 3.

[20]Very few borrowers are *both* lower-income *and* buying in a lower-income area, so we do not estimate a separate interaction term, nor do we consider summing the two coefficients.

Table 5 presents sample statistics for each of the restricted samples. Notice that as borrower income grows, the average loan-to-value and loan-to-income ratios fall slightly, while the loan spread rises slightly.

If banks subsidized loans to CRA-eligible borrowers, we would expect CRA-eligible borrowers to be more likely to use banks. Table 5 gives the fraction of borrowers using banks and the fraction who are CRA eligible, the *product* of these two fractions, and the fraction who are *jointly* CRA-eligible and who use banks. If CRA-eligible borrowers are more likely to use banks, we would expect the joint fraction to exceed the product fraction. From table 5, we do in fact see that CRA-eligible borrowers are likelier to use banks than chance alone would dictate. However, this effect is small. Among borrowers with annual incomes above $100,000, for example, 2.98 percent are CRA eligible and 74.19 percent use banks, giving a product fraction of 2.21 percent. The joint fraction is very slightly higher, at 2.29 percent. Note that this borrower preference drops at higher incomes, suggesting that selection (if present at all) is decreasing.

By eliminating lower-income borrowers from the sample, we will identify a subsidy by comparing the interest rates paid by higher-income CRA-eligible borrowers at banks to the interest rates paid by otherwise identical non CRA-eligible borrowers at banks. Thus, the number of CRA-eligible borrowers at banks is crucially important. As shown in table 5, as we raise the income threshold, the number of CRA-eligible borrowers at banks drops. At the lowest threshold ($60,000) the sample contains about 3,400 CRA-eligible borrowers at banks. At the highest threshold ($120,000) the sample contains only 366 such households. Thus, although we may be controlling for borrower heterogeneity by increasing the threshold, we do so at the cost of statistical precision.

In table 6 we present the estimated subsidy under each income threshold for both of our specifications. In addition, table 6 shows the standard error of the estimate and the p-value from a test of the hypothesis that the subsidy is zero. As we discussed, the number of observations used to identify the subsidy at the highest income threshold is quite small, so those results, as expected, have large standard errors.

For both specifications, the estimated subsidy rises as the income threshold rises (again, ignoring the highest income threshold). This suggests, consistent with our other results, that selection bias is present. Under specification (A’) the maximum subsidy is 7.2 basis points, with a p-value of 1.69 percent. Under specification (B') the maximum subsidy is 5.2 basis points, with a p-value of 8.70 percent.

5c. Mortgage Pricing Differences Between Banks and Transaction Lenders

The estimated subsidy depends on which specification we use; the estimated subsidies are uniformly smaller under specification (B) and (B') than under specification (A) and (A'). In this section, we argue that the data support specification (B) over specification (A).

Specification (B) interacts borrower risk characteristics with lender type; thus, if banks price borrower risk differently than transaction lenders, the coefficients relating to borrower risk will be improperly estimated in specification (A). A simplified version of specification (B), ignoring CRA eligibility for a moment, is:

$$S = aR + u$$

A simplified version of specification (B) is:

$$S = bR + \lambda(c + dR) + u.$$

Here R is a scalar measure of borrower risk, λ is an indicator variable for using a bank, u is the error term and $a, b, c,$ and d are estimated coefficients. From the estimated coefficients on loan-to-value ratio shown in columns (A) and (A') in table 4, we see that a is positive but small in absolute value because loans with higher LTVs carry slightly higher interest rates. The coefficients on the same variables in columns (B) and (B') show the pattern $b>0$, $c>0$, $d<0$ and $(b+d)>0$. In other words, loans with higher LTVs carry much higher interest rates, except at banks, where they carry only slightly higher rates. At the same time, all loans at banks carry higher interest rates, with c estimated to be about 35 basis points.

Thus, banks and transaction lenders do price borrower risk characteristics differently. Among borrowers with low LTVs (that is, safe loans), borrowers pay a premium at a bank over a transaction lender. For borrowers with high LTVs (that is, high-risk loans), borrowers at banks get a significant discount.

Low-risk borrowers may be willing to pay a small premium to deal with a bank for, essentially, concierge services. These borrowers are presumably higher-income or high-wealth households with associated high opportunity cost of time. If relationship lenders are able to process loan documentation with a minimum amount of borrower effort, borrowers with high opportunity costs of time may be willing to pay for this service.

The difference in loan pricing among high-risk borrowers may be due to superior information gathering technology available to banks, or because the loan is bundled as part of a larger set of services not available to transaction lenders.

Because banks and transaction lenders price borrower risk differently, estimated coefficients from equations (A) and (A') will suffer from specification bias. Indeed, the coefficients on LTV in columns (A) and (A') of table 4 show a nonsensical pattern. For this reason, the subsidy estimates from specifications (B) and (B') are probably better.

6. Conclusion

In this study, we test whether lenders affected by the CRA cut interest rates to attract CRA-eligible borrowers. The analysis relies on a novel dataset that contained mortgage interest rates, geographic location and borrower- and lender-level variables. In addition, we used an auxiliary dataset containing the estimated credit scores of individuals.

We confronted two serious identification problems. First, banks (relationship lenders) and independent mortgage companies (transaction lenders) may price borrower risk differently; however, banks are the only institutions affected by the CRA. Second, our data do not contain the borrower's credit history, which is an important risk variable used in setting interest rates.

We approached the first problem by limiting our test to comparing CRA-eligible borrowers and non-CRA eligible borrowers within banks. Our results indicate that banks and transaction lenders do price risk differently. In particular, low-risk borrowers appear to pay a premium in order to borrow from banks, while high-risk borrowers actually may pay substantially higher interest rates at transaction lenders. These findings are consistent with the theory that banks and transaction lenders have access to different underlying technologies.

We bounded the effect of unobserved credit history by using data on credit score variation by income. By limiting the dataset to higher-income borrowers only, we showed that the amount of credit score variation was sharply limited.

Limiting our dataset to higher-income borrowers forces us to use the geographic definition of CRA eligibility, because no borrower with a high absolute income also has a low enough income to quality for CRA credit. Thus, our preferred test for CRA-induced price effects uses only data from higher-income borrowers at banks; we compare rates paid by those borrowers who buy homes in lower-income neighborhoods to those borrowers who buy homes in higher-income neighborhoods. However, it is precisely among these higher-income, high-credit score borrowers that we would expect competition for CRA-eligible borrowers to be fiercest.

In no case did we find that CRA-affected institutions cut interest rates by an economically meaningful amount in order to attract CRA-eligible borrowers. Indeed, after limiting our dataset to higher-income borrowers and allowing banks and transaction lenders to price risk differently, the largest subsidy estimate was 5.2 basis points. Even in this case, the p-value for the test of the hypothesis that the true subsidy is zero was 8.7 percent.

Our findings are consistent with the view that the CRA does not cause banks to extend mortgage loans with substantially lower mortgage rates to attract CRA-eligible borrowers. However, it may still be the case that the CRA forced banks to institute special lending programs or otherwise pay a fixed investment in order to make loans to CRA-eligible borrowers. Beyond measuring the effects on mortgage rates this paper does not address the overall costs or benefits of the CRA.

References

Avery, R., R. Bostic, and G. B. Canner (2002). Assessing the CRA's necessity and efficiency. Manuscript, Federal Reserve Board, Washington DC.

Avery, R., R. Bostic, and G. B. Canner (2000). CRA special lending programs. *Federal Reserve Bulletin 86*(11), 711–31.

Barakova, I., R. Bostic, P. Calem, and S. Wachter (2003). Does credit quality matter for homeownership? Manuscript, Federal Reserve Board, Washington DC.

Board of Governors of the Federal Reserve System (1993). *Report to the Congress on Community Development Lending by Depository Institutions.* Washington, DC: Federal Reserve Board.

Board of Governors of the Federal Reserve System (2000). *Report on the Performance and Profitability of CRA-Related Lending.* Washington, DC: Federal Reserve Board.

Bostic, R. and B. L. Robinson (2003). Do CRA agreements influence lending patterns? *Real Estate Economics 31*(1), 23–51.

Canner, G. and W. Passmore (1995a). Credit risk and the provision of mortgages to lower-income and minority homebuyers. *Federal Reserve Bulletin 81*(11), 989–1016.

Canner, G. and W. Passmore (1995b). Home purchase lending in low-income neighborhoods and to low-income borrowers. *Federal Reserve Bulletin 81*(2), 71–103.

Canner, G. and W. Passmore (1996). The relative profitability of commercial banks active in lending in lower-income neighborhoods and to lower-income borrowers. In *Proceedings of the 32nd Annual Conference on Bank Structure and Competition.* Federal Reserve Bank of Chicago.

Canner, G., W. Passmore, and B. Surrette (1996). Distribution of credit risk among providers of mortgages to lower-income and minority homebuyers. *Federal Reserve Bulletin 82*(12), 1077–1101.

Evanoff, D.D. and L. M. Segal (1996). CRA and fair lending regulations: Resulting trends in mortgage lending. *Federal Reserve Bank of Chicago Economic Perspectives 20*(6), 19–46.

Evanoff, D.D. and L. M. Segal (1997). Strategic responses to bank regulation: Evidence from the HMDA data. *Journal of Financial Services Research 11*(1-2), 69–93.

Garwood, G. L. and D. S. Smith (1993). The community reinvestment act: Evolution and current issues. *Federal Reserve Bulletin 79*(4), 251–67.

Harvey, K. D., M. C. Collins, P. Nigro, and B. Robinson (2001). Disparities in mortgage lending, bank performance, economic influence, and regulatory oversight. *Journal of Real Estate Finance and Economics 23*, 379–410.

Johnson, S. A. and S. K. Sarkar (1996). The valuation effects of the 1977 community reinvestment act and its enforcement. *Journal of Banking and Finance 20*(5), 783–803.

Malmquist, D., F. Phillips-Patrick, and C. Rossi (1997). The economics of low-income mortgage lending. *Journal of Financial Services Research 11*(1-2), 169–88.

Meeker, L. and F. Myers (1996). Community Reinvestment Act lending: Is it profitable? In *Financial Industry Perspectives 1996*, pp. 13–35, Kansas City, MO: Federal Reserve Bank of Kansas City.

Zinman, J. (2002). The causes and real effects of credit constraints: Evidence from the community reinvestment act. Manuscript, M.I.T. Department of Economics, Cambridge MA.

Table 1: Sample statistics by lender type

Variable	All	Lender Type Bank/Relationship	Transaction
Mortgage Spread[a]	1.85	1.85	1.87
	(0.51)	(0.52)	(0.49)
Loan Amount[b]	126.43	124.22	134.01
	(66.09)	(65.13)	(68.72)
Loan-to-value ratio (*v*)	82.05	81.93	82
	(14.51)	(14.78)	(13.54)
Loan-to-income ratio (*l*)	2.15	2.14	2.06
	(1.24)	(0.87)	(2.06)
		—Percent—	
FANFRED	70.5	67.1	82.4
SOLDOTH	11.8	11.5	12.7
CONFORMSIZE	93.8	94.3	91.9
PMI	36.3	39.2	26.2
PCT_VAC	3.1	3.3	2.5
CRA_ELIG	28.0	29.5	22.6
LOWMOD	8.1	8.6	6.1
IRLT80	23.9	25.4	18.8
Observations	250,593	193,827	56,766

Note. Table gives means and standard deviations (in parentheses) of selected variables for the indicated subsets of the data. "Banks" are defined as commercial banks and savings institutions; "transaction" lenders are defined as independent mortgage bankers.

a Spread over 10-year constant-maturity 10-year Treasury yield, in percent
b In thousands of real 1996 dollars.

Table 2. Mortgage interest rate spreads

	Lender Type		
	Bank	Transaction	All
	Lower-Income Borrowers		
	Income < 80 % of MSA Median		
Spread Mean	1.78	1.90	1.80
Spread Std. Dev.	0.60	0.51	0.58
Observations	49,174	10,700	59,874
	Medium-Income Borrowers		
	(80 % ≤Income≤120% of MSA Median)		
Spread Mean	1.86	1.86	1.86
Spread Std. Dev.	0.49	0.49	0.49
Observations	55,198	16,086	71,284
	Higher-Income Borrowers		
	(Income > 120% of MSA Median)		
Spread Mean	1.85	1.87	1.85
Spread Std. Dev.	0.49	0.49	0.49
Observations	89,455	29,980	119,435
	All Income Categories		
Spread Mean	1.85	1.87	1.85
Spread Std. Dev.	0.52	0.49	0.51
Observations	193,827	56,766	250,593

Note. Table gives mortgage interest rate spread (to the 10 year constant maturity Treasury rate) by lender and borrower type.

Table 3: Imputed Credit Scores By Income and LTV

	Mean	Std. Dev.	Percentile 10th	50th	90th	Range
Income Category (thousands of 1996 dollars)						
Y<60	680	70	570	690	760	170
60≤Y<80	710	50	630	710	770	140
80≤Y<100	700	60	590	710	770	180
100≤Y<120	700	60	600	710	760	160
120≤Y	730	40	680	740	770	90
Loan-to-Value Category (percent)						
LTV<75	710	60	630	720	770	140
75≤LTV<81	680	70	570	690	760	190
81≤LTV<91	680	60	580	690	750	170
91≤LTV<96	670	60	570	680	760	190
96≤LTV<99	650	60	570	670	740	170
99≤LTV	670	80	560	710	750	190

Note. Table gives imputed credit score statistics from the 1989, 1992, 1995, 1998, and 2001 Surveys of Consumer Finance; higher values indicate better credit. To be included in the sample, households must have gotten a mortgage within 24 months of the survey date. Conditional statistics are given for five income and six loan-to-value categories. The column marked “range” gives the 90th percentile less the 10th percentile for the indicated category.

Table 4: Complete Regression Results, All Borrowers

Variable	Specification A	A'	B	B'
Relationship Lender: REL_LEND λ	0.0850	0.0880	0.3470	0.3490
	0.0040	0.0040	0.0140	0.0140
CRA Eligible: CRA_ELIG α	0.1080		0.0760	
	0.0040		0.0040	
Interaction term: CRA_ELIG times REL_LEND α_{λ}	-0.0720		-0.0180	
	0.0050		0.0050	
CRA Eligible: LOWMOD α^{1}		0.0630		0.0450
		0.0080		0.0080
CRA Eligible: IRLT80 α^{2}		0.1140		0.0810
		0.0050		0.0050
Interaction term: LOWMOD times REL_LEND $\alpha^{(1)}_{\lambda}$		-0.0620		-0.0360
		0.0080		0.0080
Interaction term: IRLT80 times REL_LEND $\alpha^{(2)}_{\lambda}$		-0.0780		-0.0220
		0.0050		0.0060
Risk Variables				
Loan-to-value ratio LTV				
75<=LTV<81	-0.0240	-0.0230	-0.0030	-0.0020
	0.0020	0.0020	0.0050	0.0050
81<=LTV<91	0.0300	0.0310	0.0730	0.0750
	0.0030	0.0030	0.0060	0.0060
91<=LTV<96	0.0570	0.0580	0.1650	0.1670
	0.0030	0.0030	0.0060	0.0060
96<=LTV<99	0.0430	0.0450	0.2650	0.2640
	0.0040	0.0040	0.0110	0.0110
LTV>=99	-0.0460	-0.0430	1.2400	1.2400
	0.0080	0.0080	0.0260	0.0260
Loan-to-income ratio LTY				
LTY	-0.0680	-0.0690	-0.0320	-0.0340
	0.0010	0.0010	0.0020	0.0020
$LTY^2/100$	0.4700	0.4700	0.2100	0.2200
	0.0100	0.0100	0.0100	0.0100
FANFRED	-0.0080	-0.0080	-0.1380	-0.1370
	0.0030	0.0030	0.0080	0.0080
SOLDOTH	-0.0920	-0.0920	0.0340	0.0340
	0.0040	0.0040	0.0100	0.0100

CONFORMSIZE	-0.1790	-0.1790	0.0000	-0.0010
	0.0040	0.0040	0.0090	0.0090
PMI	0.0100	0.0100	0.0150	0.0150
	0.0020	0.0020	0.0040	0.0040
PCT_VAC	0.0130	0.0160	0.0340	0.0350
	0.0050	0.0050	0.0110	0.0110
Risk Variable Interaction Terms				
Loan-to-value ratio LTV				
75<=LTV<81			-0.0220	-0.0230
			0.0060	0.0060
81<=LTV<91			-0.0520	-0.0520
			0.0070	0.0070
91<=LTV<96			-0.1390	-0.1390
			0.0060	0.0060
96<=LTV<99			-0.2610	-0.2580
			0.0120	0.0120
LTV>=99			-1.4040	-1.4020
			0.0270	0.0270
Loan-to-income ratio LTY				
LTY			-0.0590	-0.0590
			0.0020	0.0020
$LTY^2/100$			0.8100	0.8100
			0.0300	0.0300
FANFRED			0.1430	0.1420
			0.0090	0.0090
SOLDOTH			-0.1670	-0.1680
			0.0110	0.0110
CONFORMSIZE			-0.1940	-0.1930
			0.0100	0.0100
PMI			0.0020	-0.0020
			0.0050	0.0050
PCT_VAC			-0.0240	-0.0210
			0.0120	0.0120
Other Control Variables				
Intercept	2.4400	2.4400	2.2400	2.2400
	0.0100	0.0100	0.0100	0.0100
CLOANS	-0.0740	-0.0770	-0.0710	-0.0740
	0.0070	0.0070	0.0070	0.0070
LENDERS	0.1710	0.1820	0.1570	0.1680
	0.0240	0.0240	0.0240	0.0240
Year Dummies				
YR95	-0.5060	-0.5070	-0.5110	-0.5120
	0.0050	0.0050	0.0050	0.0050
YR96	-0.6610	-0.6620	-0.6610	-0.6630
	0.0050	0.0050	0.0050	0.0050
YR97	-0.6650	-0.6650	-0.6690	-0.6700
	0.0050	0.0050	0.0050	0.0050
YR98	-0.3680	-0.3680	-0.3740	-0.3750
	0.0040	0.0040	0.0040	0.0040
YR99	-0.3660	-0.3660	-0.3670	-0.3670
	0.0030	0.0030	0.0030	0.0030

POSTRUSSIA	0.2380	0.2380	0.2380	0.2380
	0.0020	0.0020	0.0020	0.0020
F-Value	101.5700	101.5700	95.7800	95.9600
Pr > F	<0.001	<0.001	<0.001	<0.001
R^2	0.3300	0.3300	0.3400	0.3400

Note. Table gives regression coefficients and standard errors from OLS regressions of the mortgage interest rate spread against the specified variables; see discussion in section 4 of the text and equations (A) and (B) for more details. All regressions contain a complete set of MSA and lending institution dummies (not shown); the *F*-test statistic of the hypothesis that all dummies were jointly equal to zero is shown.

Table 5. Sample statistics by income category

	Real income threshold (thousands)			
Variable	60	80	100	120
Spread: Mean	1.8669	1.8772	1.8978	1.9103
Spread: Std. dev.	0.4865	0.4881	0.4864	0.4826
LTV (ν): Mean	81.38	79.90	78.49	77.43
LTV (ν): Std. dev.	13.71	13.79	13.86	13.72
LTY (l): Mean	1.8656	1.7272	1.6041	1.4980
LTY (l): Std. dev.	0.6425	0.6254	0.6272	0.6336
		—Percent—		
LOWMOD	4.00	3.38	2.98	2.78
Relationship lender	74.93	74.43	74.19	74.03
LOWMOD at Relationship[a] (*joint)*	3.15	2.65	2.29	2.13
LOWMOD × Relationship[b] (*product*)	3.00	2.52	2.21	2.06
Sample size	114,256	60,349	32,180	17,805

Note. Table gives sample statistics for selected variables when the sample is restricted to include only households with annual incomes greater than or equal to the indicated levels (in thousands of real 1996 dollars). With the sample restricted so that lower-income borrowers are excluded, borrowers are only eligible for CRA credit if they buy homes in lower-income areas (LOWMOD); as always, only relationship lenders are affected by the CRA.

a Percent of sample buying homes in lower-income neighborhoods (LOWMOD) who actually get loans at relationship lenders; that is, the fraction of CRA-eligible borrowers using CRA-affected institutions.

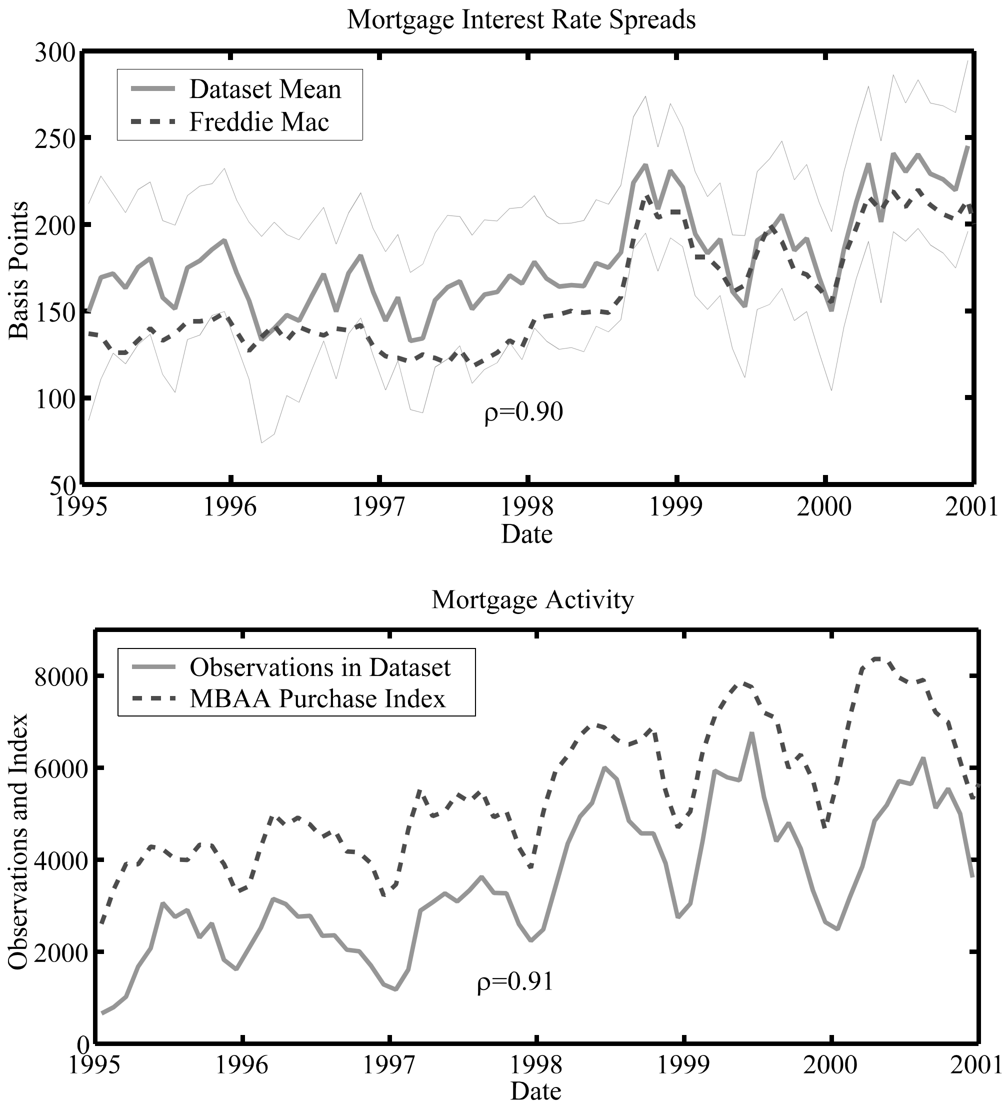
b Percent of sample buying homes in lower-income neighborhoods (LOWMOD) *times* the percent of sample who get loans from relationship lenders.

Table 6: Subsidy estimates from the restricted samples

	Specification					
	(A')			(B')		
	Coeff.	Std. Err.	p-value	Coeff.	Std. Err.	p-value
Income ≥$60,000	-0.0321	0.0145	0.0274	-0.0086	0.0146	0.5577
Income ≥$80,000	-0.0542	0.0214	0.0114	-0.0327	0.0215	0.1270
Income ≥$100,000	-0.0722	0.0302	0.0169	-0.0521	0.0304	0.0870
Income ≥$120,000	-0.0394	0.0413	0.3395	-0.0098	0.0416	0.8128

Note. Table gives the coefficient estimates and standard errors on the interaction term of CRA-eligible borrowers are CRA-affected institutions when the sample is restricted to contain only borrowers with annual incomes above the indicated thresholds. Borrowers can only be CRA eligible by buying homes in lower-income neighborhoods (LOWMOD=1). The specification marked (A') contains a full set of borrower risk controls, while the specification marked (B') interacts these borrower risk variables with lender type. Full regression results are available by request.

Figure 1: Data Validation Evidence



A Constructing the Base Dataset

Our final dataset, as discussed in the main text, contains information on individual mortgage loans, borrowers and lenders drawn from three primary datasets: HMDA, MIRS and a PMI dataset. Further, we dropped all loans that were not 30-year and fixed-rate, and those loans made by lenders or in MSAs for whom we had relatively few observations. In this section we describe in some detail the precise nature of the statistical matching procedure used to construct the base dataset, and the effect of our restrictions on the final dataset.

A.1 The Statistical Match

Our first major step was to match the loan application register (the "LAR") from the Home Mortgage Disclosure Act records to the records from the Federal Housing Finance Board's (FHFB) Monthly Interest Rate Survey (the "MIRS"). Our second major step was to match these records to the Private Mortgage Insurer (the "PMI") records.

Timing of the LAR-MIRS match

Ideally, both the LAR and the MIRS would contain specific loan identifiers (such as borrower name or property street address) that would allow precise matches. Such identifiers are not present in either data file. Similarly, it would be desirable if both files contained precise dates with the same date concept (that is, the same definition of the day on which the mortgage is complete). Instead, MIRS dates are monthly, and the LAR and MIRS feature slightly different date concepts. The MIRS records contain a "cycle date" variable (coded, e.g., as 9601, 9602, etc for January 1996, February 1996 and so on) rather than an "action date" variable as in the LAR records.[1] Information from the institution conducting the MIRS (the FHFB) led us to believe that observations are drawn mostly from the end of the cycle date month (the official documentation says the loans must close in the last five working days of the month). Thus, we matched MIRS records from one cycle date to LAR records from the middle of the cycle date month to the middle of the next month.

Summary of procedure

Our procedure can be divided into several parts. First, we prepared the LAR records, then we prepared the MIRS records, then we matched these two datasets. After that, we matched to the PMI records.

[1] In the case of loan originations, the action date is the closing date of the loan. The date gives the day and month.

In producing the LAR dataset to prepare to match, we extracted only those HMDA LAR records that:

1. Were originations
2. Were for home purchases
3. Were conventional
4. Had a valid geocode (state, county and census tract)
5. Had a valid MSA (no "non-MSA" records)

We then calculated the cycle date for the LAR records (using the month and day of the LAR action date) and determined the ZIP codes for the LAR records. Note that LAR records may sometimes be matched to more than one ZIP code.

We then turned to the MIRS dataset. We discarded only those MIRS records with invalid or missing MSA codes.

We then matched the two databases:

1. We sorted each database by MSA, state and cycle date to produce a set of potential matches ("couplets")
2. If the ZIP codes in a couplet were not an exact match, we discarded the couplet
3. For the remaining couplets, we calculated the difference in loan amounts
4. We then sorted the couplets into groups by MSA, state, and ZIP code (the "geocode") and the cycle date
5. In each geocode-month group, we chose the first couplet (that is, the couplet with the smallest difference in loan amount) as a *potential match couplet*
6. The next couplet in each group (the one with the second smallest loan amount difference) was examined to see if either side had previously been matched; if so it was discarded, if not, it was the next potential match couplet
7. All potential match couplets with loan amount differences greater than $2,000 were discarded.

At the end of this procedure we were left with a set of LAR-MIRS match couplets. In one sense, this database already contains almost all of the variables of interest. However, many of these mortgages will carry private mortgage insurance. PMI rates are (broadly speaking) set by state insurance commissions and PMI itself affects the effective interest rate on the mortgage because the payments are rolled into the mortgage's APR. About eight companies provide PMI insurance; these companies voluntarily submit a HMDA-like record to the FFIEC, which we refer to as the PMI database.[2] We augmented our LARS-MIRS couplet match dataset with a further statistical match against the PMI database (the lender name is suppressed in the PMI database by the PMI companies, so we cannot do an exact

[2]By 2000, a merger had reduced the number of PMI companies to seven.

match). Because the PMI database closely follows the HMDA database in form, we were able to match on several identifying criteria.

To prepare the PMI database for matching, we discarded all records without a valid geocode (MSA, state, county and census tract), that were attached to loans for purposes other than the purchase of a 1-4 family home, or were for an action other than loan origination. To match the resulting PMI database with our LARS-MIRS matched couplets we:

1. Sorted the PMI database by geocode (state, MSA, county and census tract) to produce the same geocode categories as in the LARS-MIRS database; this gave us our set of potential triplets
2. We then compared the "race or national origin" (RONO) codes from the LARS data and the PMI data; these had to match exactly, producing a winnowed set of potential triplets
3. For each of these potential triplets we calculate the absolute difference in three variables in the LARS and PMI data:
   (a) Loan amount
   (b) Borrower income
   (c) Action date
4. We discarded all potential triplets in which the loan amount or borrower income differed by more than $2,000 or in which the action date differed by more than 150 days
5. To produce the *A match*:

   We then sorted by geocode, loan amount difference, borrower income amount difference and action date difference; for each geocode group we picked the first triplet as an actual triplet, we then checked subsequent triplets to see if either side had already matched, if so then that triplet was discarded, if not, it became an actual triplet
6. To produce the *B match*:

   We then collected all unmatched couplets and PMI records and tried to match them again using slightly different criteria; in particular, we tightened the loan amount difference to $1,000 but widened the income amount difference to $10,000 (on the theory that differing definitions of "income" might be affecting our match).

In the end, we simply discarded the few extra matches generated under the "B" matching procedure.

A.2 Evaluating Match Quality

One might expect that the differences in loan amount would be zero; by allowing non-zero differences, we may be admitting spurious matches. However, both the PMI and the HMDA reporting guidelines call for loan amounts to be rounded to the nearest thousand; potentially, if the loan amount is close to the rounding midpoint, different institutions might report the loan amount differently. Further, the MIRS data come from a voluntary survey of institutions, who may truncate rather than round, or who may have a slightly different idea of what constitutes the loan balance, e.g. excluding certain fees rolled into the loan balance.

As mentioned in the text, we exclude all observations associated with smaller or inactive lending institutions. We have several reasons for doing so; in particular, because we include a full set of lender and MSA fixed effects, such observations will likely have little effect on our coefficient estimates anyway, and less-active lenders may be spreading their fixed costs over fewer loans. However, although we do not know the list of institutions polled by the FHFB in conducting the MIRS, we know that they are relatively large and well-established lenders. By excluding loans from smaller, less-active lenders, we are purging our dataset of spurious matches.

At each step in the matching procedure, we produced diagnostic variables that can be used to evaluate the quality of the match. In table A.1 we present summary statistics and the empirical CDF of the loan differences from each step of the match. As shown in the table, only about 36% of HMDA-PMI matches had a difference of exactly zero; however, 90% had a loan difference of $500 or less and only about four percent had a loan difference greater than $1,000 (the HMDA rounding point). We allowed the HMDA-PMI loan difference to be a little larger, because we were also matching on race or national original (which had to be an exact match) and borrower income (which generally had to be within $2,000 of each; see section A.1 for more information).

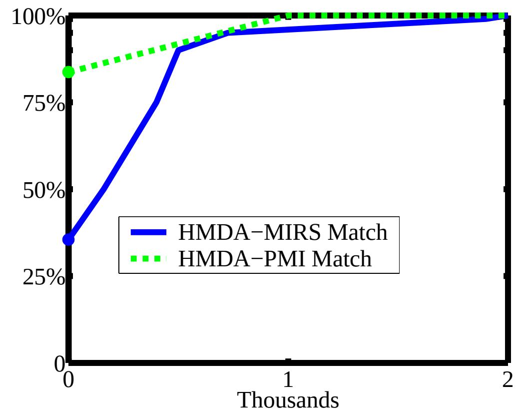
A.3 Other Types of Mortgages

In this paper we have exclusively used fixed-rate mortgages with a 30-year amortization; in reality, households use a variety of amortization horizons and rate structures. Among the most popular are 15-year fixed-rate mortgages and certain types of adjustable-rate mortgages (ARMs).

Thirty-year fixed-rate mortgages have the considerable advantage of being relatively homogeneous products; the main source of variation among them is the effective rate on the mortgage. However, if CRA-eligible borrowers are steered towards these other types of mortgages, we might be missing the relevant variation in mortgage terms. This is doubly true if one lender type rather than another favors ARMs or 15-year mortgages for lower-income borrowers.

Table A.1: Statistical match diagnostic: Loan amount differences

	HMDA Match	
	MIRS	PMI
	Thousands	
Mean	0.2363	0.1632
Median	0.1600	0
Std. Dev.	0.3200	0.3695
	Percent	
Diff= 0	36	84



NOTE. Table and figure give statistics on the distribution of the difference in loan amounts from the HMDA-MIRS and the HMDA-PMI match.

Because the FHFB ask about these alternate mortgage products, we can test this proposition directly. Although we exclude ARMs and 15-year mortgages from our final dataset, we can check their relative frequency in the base dataset by lender and borrower type. Table A.2 shows the percent of observations in our base dataset that are 30-year fixed-rate mortgages conditional on borrower type, neighborhood type and lender type. The table has several interesting features: First, in all combinations of borrower, neighborhood and lender type, at most 12% of mortgages are not 30-year FRMs. Second, transaction lenders are more likely than relationship lenders to use 30-year FRMs. Third, within each lender type, there is no appreciable difference in the treatment of CRA-eligible borrowers. Thus we are reassured that we are not missing an important feature of the treatment of CRA-eligible borrowers by studying 30-year FRMs exclusively.

Table A.2: Distribution of 30-year, fixed-rate mortgages in dataset

Income Class		Lender Type			
Borrower	Neighborhood	Transaction	Relationship	Total	
		—Percent—			
Lower	Lower	98.09	90.58	91.71	$\star$
Lower	Higher	97.79	90.04	91.52	$\star$
Middle	Lower	98.01	88.57	90.53	$\star$
Middle	Higher	97.67	89.35	91.26	
Higher	Lower	99.86	88.98	91.45	$\star$
Higher	Higher	97.65	88.87	91.06	
Total		97.73	89.31	91.23	
CRA-Eligible Borrowers		98.01	89.94		
Non CRA-Eligible Borrowers		97.65	89.05		

NOTE. Table gives percent of loans from base dataset that are 30-year, fixed-rate by borrower and neighborhood income class. Borrower and neighborhood income classes are defined relative to the MSA median income; lower income borrowers and neighborhoods are defined as having incomes less than 80% of the MSA's median income. Middle income borrowers have incomes between 80% and 120% of the MSA median and high-income borrowers have incomes greater than or equal to 120% of the MSA median. All lower-income borrowers, regardless of neighborhood and all borrowers who purchase homes in lower-income neighborhoods, regardless of income, are eligible for CRA credit.
$\star$: Loans eligible for CRA credit.