Appendix A
Description of the Data, Data Issues & Complications

In this Appendix we describe in detail the various data sets that were used in the analysis, and why. We also discuss several data issues and problems that required special attention. These include a comparison of the GSE versus HMDA data, a discussion of lenders whose data needed unique treatment, and an explanation of the matching procedures used to combine data sets. The latter includes HUD's list of B & C and manufactured housing lenders and their ID numbers.

Description of the Data

Wherever possible, data were collected for each of the years 1992-1996. By looking at trends over a five year period, it is much easier to assess whether GSEs (and CRA lenders) were "leading the market" or simply following it[1]. Data were collected from several sources.

HMDA LOAN APPLICATION REGISTERS AND TRANSMITTAL SHEETS. Starting in 1990, most lenders were required to provide information on every home mortgage application they received. The information included the type of loan (conventional, FHA or VA), the requested amount, the final disposition of the application (e.g., approved, denied, withdrawn, not accepted), the census tract in which the desired property was located, the income, race and gender of the applicant(s), and the ultimate purchaser of the loan (e.g. not sold, sold to Fannie Mae or Freddie Mac). The HMDA transmittal sheets (one record per lender per year) indicate the lender's name, address, and parent company (if any).

GSE DATA. The GSEs have recently begun providing HUD with loan-level data on each of their mortgage transactions since the beginning of 1993. This includes information on demographic characteristics of both the borrower and the census tract the property is located in. In many ways, the GSE data provide a more accurate description of GSE purchases than the HMDA data do. For example, as various authors have pointed out (Canner and colleagues, 1996), loans made late in the year are especially likely to have their ultimate purchaser misclassified by HMDA. As we discuss later in this Appendix, key features of the way the GSE data sets are constructed greatly limit their usefulness for the sort of regional analysis undertaken here. *We therefore primarily relied on the HMDA data*, and where possible used the GSE data to double-check the accuracy of our results. We also extracted from the 1996 GSE data a list of census tracts defined as "targeted" under the Final Rule[2]. Further, we computed from the GSE data the percent of all GSE purchases in a census tract that were from first-time homebuyers, on the rationale that the higher this percentage was, the more aggressive the GSEs were being in helping needy markets.

---

[1] One minor problem with the longitudinal approach is that the boundaries of some Indiana MSAs were expanded after 1993. This accounts for some of the increase in the number of loans made in MSAs between 1993 and 1994 but appears to have no other meaningful impact on our results.

[2] HUD's underserved area definition was introduced in 1996. Prior to that, the GSE data sets did not include information on targeted tracts; and the HMDA data has never had this information. We therefore extracted the list of targeted tracts from the 1996 GSE data and matched them up with the 1992-1996 HMDA data and the GSE data from earlier years. For about 2 percent of the HMDA loan applications, we were unable to determine whether the tract was targeted or not; we coded these tracts as missing on the targeted tract variable. Using the guidelines given in the Final Rule, we also tried to compute directly whether a tract was "targeted" or not. We found that our algorithm gave very close, but not identical, classifications to those contained in the 1996 GSE data.

CENSUS TRACT DATA. The HMDA data include key information on census tracts, making it possible to determine whether a neighborhood is low-income or minority. The St. Joseph County data also includes other information that was gathered directly from published census reports for the county.

MANUFACTURED HOUSING AND B&C LOANS. There is an ongoing debate about whether manufactured housing and B&C (below-investment-grade, or subprime) loans should be included in analyses. These are generally higher-risk, higher interest loans that the GSEs will not buy. Various researchers (Lind 1996a, 1996b; Bunce and Scheessele, 1996) have therefore attempted to exclude such loans in their analysis. However, given the increasing importance of these loans to underserved markets, we were not sure it was right to exclude them; simply because GSEs won't buy such loans doesn't mean that they couldn't. Indeed, in October 1997 Freddie Mac announced plans to move into the subprime market. Using a list of subprime lenders provided to us by HUD, we originally planned to include subprime loans throughout our analysis and apply appropriate controls for them. However, it quickly became apparent to us that this would greatly complicate the analysis and make a fair evaluation of GSEs and CRA much more difficult. We therefore decided to leave subprime loans out of our main analysis, and instead include a chapter where we examined them separately. As Chapter 4 shows, subprime lending has risen dramatically in Indiana during the 1990s, and any analysis that does not somehow take this into account has the potential to be highly misleading.

LENDING INSTITUTION DATA. Information on lender characteristics comes from several sources, both local and national. In earlier studies (Williams and Nesiba, 1997) characteristics of lenders in St. Joseph County were hand-coded[3]. Because such coding is very tedious, we developed means for automating much of this process on a state and national level. Using the HMDA data, we found that it is fairly easy to determine whether a lender is a commercial bank, credit union, etc. Different types of institutions report to different agencies[4]. The HMDA data includes the agency to whom a lender reports. However, this information is not quite sufficient, since non-independent mortgage companies (e.g. mortgage banking subsidiaries) also report to some of these agencies. Fortunately, mortgage companies can be distinguished from other types of lenders by using the lender code variable contained in the HMDA Expected Reporter Panel, which is available at additional cost for the years 1994-1996[5].

Other measures of institutional characteristics are more problematic. While lenders report their parent institution's name and address in the HMDA data, the measure seems to be riddled with errors. It may be that many lenders do not understand the question or do not know the answer.

---

[3] HMDA lists the location of the parent institution, making it possible to code whether the bank was locally headquartered or not. We used Moody's Bank and Finance Manual and McFadden's American Financial Directory to look up the amount of assets held by lenders. Simply by looking at the lender's name or by drawing on other knowledge we had, we could determine whether an institution was a commercial bank, credit union, savings and loan, or mortgage company.

[4] The OCC, the FRS, and the FDIC all deal with different kinds of banks, S&Ls report to OTS, credit unions report to the NCUA, while independent mortgage companies report to HUD.

[5] We also found that an alternative coding scheme, which uses the agency code and looks for key terms in the lender's name, produces nearly identical results. Hence, where possible, we used the Expected Reporter Panel to code lenders; when that was not possible (primarily for lenders in 1992-1993 who had "disappeared" by 1994) we used our alternative scheme.

For example, hundreds of lenders claim that their parent institution is the Federal agency they report to (e.g. FRB, OTS).

We therefore came up with an alternative procedure that coded lenders by the location of their headquarters and branch offices in Indiana. The web page of the Federal Reserve System's National Information Center (http://www.ffiec.gov/nic/) contains detailed information on the ownership and organizational structure of many lenders. We looked up lenders who were active in Indiana during the years 1995-1996. If we failed to find the lender on the NIC pages, we did statewide phone directory searches to see if the lender had any offices in Indiana. We then coded each lender as either (a) having its headquarters in Indiana, (b) having branches in Indiana but headquarters elsewhere, or (c) having no branches that we could identify in Indiana.[6]

In addition, the HMDA Expected Reporter Panel contains information on the assets of the lending institution. This measure seems fairly reliable (e.g. it gives similar numbers to what we got for most lenders in St. Joseph County) and FRB officials have told us they believe this information to be of high quality (with the main problem being that numbers are sometimes a year or so out of date). Unfortunately, the ERP is not available before 1994. Thus, lenders in 1992 and 1993 were assigned the assets from the earliest year in which they appeared in the ERP. However, if a lender was not active nationwide between 1994-1996 we were not able to code their assets for earlier years. Hence, while we have some asset data for 1992-1993, the data for 1994-1996 are much more complete and reliable. We coded lenders as (a) small – assets of $100 million or less (b) medium – assets of $100 million to $1 billion, and (c) large – assets greater than $1 billion.

MEDIAN INCOME DATA. The official HMDA reports (and this study) use the MSA median family income when classifying applicants as low income, moderate income, etc. These numbers are based on HUD estimates that change yearly. This information can be obtained for free from the FFIEC. In addition, information for selected years is available on the world wide web at http://www.ffiec.gov/hmda/hardcopy.htm and at http://www.huduser.org/data/factors.html.

ALTERNATIVE LENDING PROGRAMS. Some special programs aimed at low income and minority borrowers are not reflected in the HMDA data. For example, during 1994-1996 the Community Homebuyer's Corporation (CHC) made 102 loans in St. Joseph County. The CHC pools money from area lenders with block grant support from the government to provide loans that make

---

[6] There are numerous limitations to the measure we constructed. For one thing, location of headquarters is not the same as location of ownership; for example, Norwest Bank of Indiana lists its headquarters as being in Fort Wayne, IN although it is part of the much larger chain of lenders owned by Norwest nationwide. Still, given that much of the concern over increasing bank concentration has been over the possible loss of sensitivity to local needs caused by decision making power concentrated far away (Campen, 1993), we thought a lender with a local headquarters might be better than a lender without one. We also thought that a lender who apparently had no physical presence at all in the state (perhaps doing business by phone, mail, or through mortgage brokers that acted on behalf of several lenders) might be different than one that was physically present. There are also concerns about the quality of the measure's coding. In between the time lenders reported to HMDA and when we looked them up, there could have been changes in branches and headquarters. We made no attempt at all to look up lenders who were active in 1992-1994 but disappeared after that. Such lenders are likely to have gone out of business, changed their names, or merged with others. We coded such lenders as not being active after 1994, with their headquarters and branches unknown. Hence, while we think our branch and location data are fairly good (but not perfect) for 1995-1996, for earlier years the data are much more suspect.

home ownership more affordable to low income persons. While the CHC makes relatively few loans, the vast majority of these (90 percent) go to underserved markets. Since the CHC is a nonprofit entity, its loans are not reported to HMDA; and since most of the lenders who back CHC are subject to CRA, exclusion of these loans runs the risk of understating the true impact of CRA in St. Joseph County. The CHC has graciously provided us with HMDA-style information on its lending, which we incorporate in our analysis of St. Joseph County.[7]

HISTORICAL EVENTS IN ST. JOSEPH COUNTY. As residents of St. Joseph County, we are familiar with important events during the 1990s that may have affected area lending. We know which lenders have engaged in mergers. We also know which institutions have entered into CRA agreements with community organizations and which ones were asked to do so but refused.

Advantages of a multi-level/multi data source approach. We see several advantages to combining a county case study approach with a statewide analysis using several data sets:

- As noted above, both Indiana and St. Joseph County share many similarities with the nation as a whole. Indeed, they may be more representative of the entire country than many of the large urban centers previous studies have focused on. At the same time, there is considerable diversity across Indiana MSAs, allowing us to examine the determinants of home mortgage lending in a variety of settings. Studying multiple MSAs also reduces the risk that idiosyncratic or atypical factors are responsible for the results.

- All key concepts can be operationalized in both the state and county data.

- Using the HMDA data, we can determine characteristics of the loans made by primary market lenders and sold (or not sold) to GSEs; and whether an institution sells any of its loans to GSEs.

- Using the HMDA data and the Expected Reporter Panel, we can also determine whether the lender is subject to CRA; the legal structure of the lender (commercial bank, mortgage company, S&L, or credit union); and the assets of the lender

- Using the HMDA data in conjunction with the GSE data, we can further determine the percentage of GSE loans in an area involving first-time home buyers

- These data sets allow us to distinguish between Fannie Mae and Freddie Mac, and to examine changes in primary and secondary market loan activity across time.

---

[7] Time constraints kept us from following through on our original plan to incorporate data from the Indiana Housing Finance Authority (IHFA), although we eventually hope to do so. Between 1989 and 1996, IHFA financed the purchase, development or rehabilitation of 40,000 affordable homes. IHFA estimates that during this period it helped more than 125,000 Indiana residents, or about 2 percent of the state's population. IHFA generally works through "participating lenders;" hence, most of the loans it helps make are probably "hidden" among HMDA records. An important implication of this is that the effects of other influences on lending to underserved markets (e.g., CRA, GSEs) are potentially distorted. For example, if lenders subject to CRA are coincidentally more likely to participate in IHFA's programs, CRA may appear to have more impact than it really does. (Of course, it could also be the case that CRA obligations make lenders more willing to cooperate with IHFA.) Similarly, GSEs may look "better" than they really are if they are buying loans that never would have been made without IHFA support. In general, we think that future studies need to look more carefully at the role that government and alternative lending programs are having on home mortgage lending.

- Other data sets also play key roles in both the state and county analysis. The Subprime & Manufactured housing lender list helps identify loans that that do not meet GSE underwriting standards, and the median income data helps classify the income status of applicants.

- All of the above measures are, of course, also available for the St. Joseph County study. In addition, locally we can use more powerful operationalizations of concepts than we can with the state and national data alone. For example, at the state level, we can measure whether the institution was subject to CRA; at the county level we can also measure whether the institution was pressured to sign a CRA agreement or was involved in merger activity that may have made it more susceptible to CRA pressure. We also have the data we collected on our own that describes local lending activity not reflected in HMDA.

- In the county case study, we can better deal with flaws and limitations in the data. We have come across instances where the HMDA data were either wrong or misleading. We discuss these problems in more detail shortly.

- Most primary market studies have been done on large urban areas, many of which have long histories of racial conflict and discrimination. The situation in other types of areas is unknown. To the extent possible, we can replicate and extend previous studies to see whether similar results can be found in a moderate-sized urban area. Conversely, most studies of GSEs have focused on the entire country. Several authors (e.g. Bunce and Scheessele 1996) have cited the need for more disaggregated analyses both at the individual lender level and at the regional and metropolitan area level. A study such as this can help to meet that need.

- Finally, by starting with a more manageable area, the analytic techniques and methods we develop for St. Joseph County and Indiana can serve as a template for a larger national study.

GSE versus HMDA data

Both the HMDA and GSE public use data sets provide information on GSE lending activities. Even though the GSE data sets have many advantages, we ultimately decided to rely primarily on the HMDA data. The GSE data were then used to check whether the HMDA results were plausible. There were several reasons for this.

First and foremost, it would have been impossible to test our hypotheses using only the GSE data. A key component of our argument is that GSEs can both be a cause of primary market activity and a reflection of it. An improvement (or decline) in GSE performance could simply reflect changes in the markets from which GSEs buy their loans. We therefore need to look at all home mortgage lending, not just the loans bought by GSEs, to assess how GSEs are doing. Indeed, as our main analysis shows, the improvements that GSEs made in the 1990s can be misleading if one does not take into account that similar improvements occurred throughout the home mortgage markets.

Ideally, then, we would have liked to use both data sets together, perhaps substituting records from the GSE data sets for the corresponding records from HMDA. Unfortunately, several characteristics of the GSE data sets greatly limited their usefulness for us. For proprietary reasons, the GSE data are divided into three unlinkable data sets. Key information appears in

one data set but not another, or else is missing altogether. Only the census tract data file makes it possible to select loans from Indiana; the other two data sets lack state identifiers. Unfortunately, unlike some of the other GSE data sets, the census tract file does not indicate whether (a) the loan was for home purchase or refinance, (b) whether the loan was bought in the current year or not, or (c) whether the loan was conventional or FHA-insured. It is therefore impossible, with the GSE data, to make what we considered to be the reasonable and appropriate sample selection we used in our analysis. While we appreciate the proprietary concerns of the GSEs, we hope that these sorts of crippling limitations can be reduced in future releases of the data. Indeed, given that similar information can be obtained from the HMDA data it is not clear to us why the GSE data sets have to be so limited.

The more detailed information in the HMDA data sets is of little value, however, unless it is also accurate. A major advantage of the GSE data sets is that their coverage of GSE loans is more complete. Bunce and Scheessele (1996) found that, nationwide, because of reporting errors and other problems, HMDA reports include only 75 to 85 percent of GSE purchases in metropolitan areas. Our own analysis of Indiana confirms that as much as 1/3 to 1/2 of all GSE purchases are not reflected in the HMDA data.

A critical question, then, is whether the GSE purchases not included in HMDA are missing more or less at random, or whether there are systematic biases in the missing data. If GSE loans are randomly missing, then GSE/Non-GSE comparisons will be more or less correct. If, however, there are systematic biases in the exclusions, then GSE performance will appear to be better or worse than it really is. For example, if the missing GSE loans are largely going to underserved markets, then the HMDA data will understate how well GSEs are serving those markets. If, on the other hand, the missing loans are all from markets that are served, then HMDA will exaggerate how well the GSEs are doing.

Fortunately, Bunce and Scheessele (1996) found that, nationwide, both GSE-based and HMDA-based reports of lending to underserved markets gave similar results. To confirm that this is also true in our sample of Indiana, we compared, as closely as possible, the underserved market measures that exist in both the GSE and HMDA data sets. Our procedure was as follows:

- From the GSE data sets for 1993-1996, we selected all loans from Indiana MSAs. We did not employ any of the other sample selection criteria used in our main analysis (e.g. home purchase conventional loans) since the GSE census tract data set does not include the information needed to make these selections.

- From the HMDA data sets for 1993-1996, we selected all originated loans from Indiana MSAs that were coded as being sold to Fannie Mae or Freddie Mac. Again, we did not employ our other sample restrictions, so this is a much broader sample than our main analysis uses.

Even though we tried to make our selections from both data sets as comparable as possible, there is still one important respect in which they differ. Unlike HMDA, the GSE data can include

loans that were originated in a prior year[8]. Hence, even if the HMDA data were completely
unbiased, the figures from HMDA and GSE could differ somewhat; in particular, we might
reasonably expect that the GSE figures for a given year would be somewhere in between the
HMDA figures for that year and the previous year. This is because the HMDA data only include
loans from one year while the GSE data includes loans from two or more. Hence, some year–to-
year small differences between GSE and HMDA are to be expected, but over the long run the
two should give more-or-less comparable results if both are unbiased.

Table A-1 gives the results of these comparisons. Several points stand out:

- For both GSEs together over the combined four-year period, the GSE and HMDA data sets
  give very similar estimates of GSE lending to Final Rule Underserved Markets. HMDA
  reports that 20.3% of all GSE loans during this time went to Final Rule Underserved
  Markets; the GSE data reports a virtually identical 20.4%. For specific types of underserved
  markets, HMDA gives a slightly lower figure for lending to very low-income borrowers and
  a slightly higher number for low income borrowers in low-income neighborhoods. For
  targeted areas and minority tracts, the GSE and HMDA numbers are again virtually identical.
  Only with lending to blacks is there a clear systematic bias, as the HMDA data consistently
  gives higher figures than the GSE data does.

- Between GSEs, there is less consistency. For Fannie Mae, the HMDA data report that 21.1%
  of its loans went to Final Rule underserved markets, while the GSE data reports a somewhat
  higher 21.9%. A closer examination of the data reveals, however, that in three out of four
  years, HMDA reports higher numbers than does GSE; but in 1995 GSE reports almost 5
  percentage points more underserved market loans (29.5% v. 24.8%) than does HMDA.
  When underserved markets are looked at in more detail, it is apparent that the major reason
  the 1995 GSE/HMDA gap exists is because GSE says Fannie made 20.4% of its loans to
  very low income borrowers, while HMDA reports only 11.5%. However, this figure of
  20.4% seems quite extraordinary: it is three times as high as Fannie's 1993 tally, twice as
  high as Fannie's 1996 figure, and about double what Freddie Mac did in the same year. All
  of this suggests that the 20.4% may be inaccurate, misleading or at least atypical. Perhaps in
  1995 Fannie bought many very low-income loans that had been made in previous years.

- For Freddie Mac, if the HMDA data are biased, they are biased in Freddie's favor. The
  HMDA data give modestly higher figures than does GSE for Freddie Mac lending to very
  low-income borrowers and low income borrowers in low-income neighborhoods. However,
  for targeted areas, blacks, and minority neighborhoods, the GSE and HMDA figures for
  Freddie Mac are virtually identical.

---

[8] Actually, one of the problems with the HMDA data is that it can include loans made in earlier years – but if you try
to use them you run the risk of double-counting the same loans. It is possible for one lender to make a loan, sell it to
someone else, and then that lender sells the loan to Fannie Mae or Freddie Mac (either in the same year or a later
one). In HMDA, the first lender will be recorded as originating the loan and then selling it to another. The second
lender will be recorded as purchasing the loan and then selling it to a GSE. We include only loan originations in this
analysis. This avoids the very serious problem of double counting the same loans, but also opens the possibility of
creating greater discrepancies between HMDA and GSE data reports.

Table A-1: Comparison of GSE and HMDA Data Sets, Indiana MSAs, 1993-1996

% of Loans going to:	GSE	Source	1993	1994	1995	1996	TOTAL
Final Rule	Fannie	HMDA Data	16.6%	25.9%	24.8%	22.6%	21.1%
Underserved Markets		GSE Data	15.8%	24.4%	29.5%	22.3%	21.9%
	Freddie	HMDA Data	15.0%	22.3%	22.7%	22.0%	19.1%
		GSE Data	14.7%	20.7%	21.7%	20.6%	18.4%
	Both	HMDA Data	15.9%	24.3%	24.0%	22.3%	**20.3%**
		GSE Data	15.3%	22.7%	26.7%	21.6%	**20.4%**
Very low	Fannie	HMDA Data	7.2%	14.6%	11.5%	10.1%	9.9%
income borrowers		GSE Data	6.8%	12.3%	20.4%	10.8%	11.6%
	Freddie	HMDA Data	6.8%	11.6%	10.3%	9.9%	8.9%
		GSE Data	6.2%	9.1%	9.3%	9.2%	8.0%
	Both	HMDA Data	7.0%	13.3%	11.1%	10.0%	**9.5%**
		GSE Data	6.5%	10.9%	16.3%	10.1%	**10.1%**
Low-income borrowers	Fannie	HMDA Data	2.2%	4.9%	5.1%	3.9%	3.6%
in low-income areas		GSE Data	1.9%	4.1%	4.4%	3.6%	3.3%
	Freddie	HMDA Data	2.1%	3.2%	3.7%	3.4%	2.8%
		GSE Data	1.7%	2.8%	2.9%	3.1%	2.4%
	Both	HMDA Data	2.2%	4.2%	4.6%	3.7%	**3.3%**
		GSE Data	1.8%	3.5%	3.9%	3.4%	**2.9%**
Targeted Areas	Fannie	HMDA Data	11.4%	16.7%	18.4%	16.5%	14.8%
		GSE Data	11.3%	16.7%	16.6%	16.0%	14.6%
	Freddie	HMDA Data	10.1%	14.3%	16.6%	15.4%	13.1%
		GSE Data	10.7%	14.5%	16.0%	14.4%	13.2%
	Both	HMDA Data	10.9%	15.7%	17.8%	16.0%	**14.1%**
		GSE Data	11.1%	15.7%	16.4%	15.4%	**14.0%**
Blacks	Fannie	HMDA Data	2.1%	4.0%	4.1%	3.2%	3.1%
		GSE Data	1.6%	3.3%	3.4%	2.8%	2.6%
	Freddie	HMDA Data	1.2%	2.2%	2.7%	2.4%	1.9%
		GSE Data	1.1%	1.9%	2.4%	2.2%	1.8%
	Both	HMDA Data	1.7%	3.2%	3.6%	2.9%	**2.6%**
		GSE Data	1.4%	2.7%	3.0%	2.6%	**2.2%**
Tracts > 30% minority	Fannie	HMDA Data	3.6%	4.7%	5.4%	4.4%	4.3%
		GSE Data	3.4%	4.7%	4.8%	4.2%	4.1%
	Freddie	HMDA Data	2.2%	3.0%	2.9%	3.1%	2.7%
		GSE Data	2.2%	3.1%	2.8%	3.3%	2.7%
	Both	HMDA Data	3.0%	4.0%	4.6%	3.9%	**3.6%**
		GSE Data	2.8%	4.0%	4.1%	3.9%	**3.6%**

- There is also some fluctuation across years. For example, in 1994, the HMDA data says 24.3% of both GSE's loans went to Final Rule Underserved Markets, while the GSE data only reports 22.7%. However, in 1995, the HMDA figure drops to 24% while the GSE data claims 26.7%. Most of the small year-to-year differences between HMDA and GSE pretty much offset each other over the four year period, and could well be due to the fact that the GSE data sets include loans from earlier years while HMDA does not.

In conclusion, despite their differences, the HMDA and GSE data generally give very similar estimates of GSE activity in underserved markets. In most cases, differences between the data sets are small and tend to offset each other across time, and indeed we would expect such patterns given that the GSE data include loans from more than one year. In the one instance where there is a very sharp difference between the GSE and HMDA data – Fannie Mae loans to very low-income borrowers in 1995 – it is the figure from the GSE data that appears more problematic.

Further, if there are any biases in the HMDA data, they generally seem to work in the GSEs favor. As noted before, in three out of four years, HMDA reports slightly better underserved market lending by Fannie Mae than does GSE; and for Freddie Mac, the HMDA estimates of lending to very low income borrowers and low income borrowers in low income areas are always slightly ahead of what GSE claims. Further, for both Fannie and Freddie, the HMDA data always report slightly more loans to blacks than GSE does. As Berkovic and Zorn note, lenders tend to disproportionately report their "good" loans in HMDA; it appears that this bias benefits the GSEs as well.

Ideally, we would like HMDA to provide perfect coverage of GSE purchases; and barring that, we would like a GSE data set that was not so crippled by the proprietary restrictions placed on it. Given that neither of these is currently possible, and given that it is impossible to test our ideas with the GSE data alone, we think our current strategy of relying primarily on the HMDA data and double-checking it against the GSE data is the best possible course.

Special Problems with Lender Data

One of the advantages of working within a smaller geographic area is that problematic cases in the data can be more easily identified and dealt with. We encountered three lenders that required special handling on our part.

Trustcorp Mortgage. According to official HMDA statistics, in St. Joseph County during the 1990s there was a dramatic increase in lending to heavily minority neighborhoods. Census tracts which were more than 50% minority went from receiving about 1% of all mortgage loans in 1992, to about 7% in 1996. Further, even though there was a dramatic increase in lending to minority neighborhoods, very few of those loans were reported as going to minority individuals. We considered this highly suspicious. A closer examination revealed that all of these loans were going to a single census tract, #20; in fact the tract was supposedly receiving more loans per year than it had houses. This was obviously an error, and since tract #20 is one of the poorest and most heavily minority tracts in the area, lending statistics for the entire county were being wildly distorted.

We discovered that a single lender was making all of the reported loans: Trustcorp Mortgage. When we pointed this out to Trustcorp, it investigated the matter and discovered that a programming error had caused most of its loans since 1994 to be reported as going to Tract #20. Trustcorp graciously provided us with a corrected data set for the three years. The final data set Trustcorp provided us for 1996 still included an implausibly large number of loans for Tract #20. Those records were discarded from our analysis, while the rest of the corrected records were substituted for the original Trustcorp reports.

1st Source Bank. 1st Source Bank is one of the largest lenders in the South Bend and the neighboring Elkhart-Goshen MSAs. In 1992 and 1993, HMDA shows 1st Source selling a large portion of its loans to Fannie Mae. However, after 1993, the number of loans reported as sold to Fannie Mae plummeted to zero while loans "sold to others" substantially increased. We thought it would be important to find out why one of the largest lenders in the area had made such a dramatic change. We learned, however, that when 1st Source acquired Trustcorp in the mid-1990s, it adopted a policy of selling its loans to Trustcorp, which then bundled them together and sold them to Fannie Mae. We therefore recoded 1st Source loans that were listed as "Sold to others" as "Sold to Fannie Mae." Statewide, this had virtually no effect on our results, changing most of the key statistics by .1 or .2 percentage points at most. Within the South Bend and Elkhart-Goshen MSAs the effect was somewhat larger, changing some statistics by 2 to 3 percentage points. In particular, the differences between loans bought by GSEs and the loans not bought by GSEs are 2 to 3 percentage points smaller in the county when this change is not made.

Bank of America, FSB. As noted before, HUD generously provided us with a list of subprime and manufactured housing lenders, as well as their ID numbers. Among these was Bank of America, FSB. We noticed, however, that in 1995, Bank of America FSB was not being coded as a subprime lender in our data, even though it was in 1996. Further examination revealed that the lender changed its ID number between 1995 and 1996, perhaps because of some sort of corporate restructuring (BA-FSB of Oregon, was replaced by BA-FSB of California). To make sure that this really was the same S&L, we compared lending patterns across years and found that the "old" BA-FSB behaved pretty much the same as the new one, e.g. made very many of its loans to underserved markets and had exceptionally high denial rates. We therefore added the old id number for BA-FSB to our list of subprime lenders and treated it accordingly. Failure to make this change would not have dramatically affected our results; the main effect would have been to create a curious spike in underserved market loans made by S&Ls in 1995.

Matching and combining data sets

The use of multiple data sets requires that information from different sources be combined somehow. Fortunately, this is not that difficult (although the number and size of the data sets make it a time-consuming process). Matching generally involves the following information and data.

1. Lender ID codes. In HMDA, the respondent ID and the Agency code uniquely identify each lender. With this information, one can match the HMDA loan application records (detailed information on each loan application), transmittal sheets (one record for each lender, including name and address) and Expected Reporter panels (which gives additional information about legal structure of the lender and the lender's assets). The HUD list of subprime lenders also includes

lender ID codes. HMDA records for those lenders were excluded from the analysis. The complete list of lenders and their ID numbers appears in Table A-2. As noted earlier, subprime lenders actually fall into two categories, B & C lenders and manufactured housing lenders.

2. Area ID codes – state, MSA, county and census tract. HMDA has all of these, making it possible to match HMDA data with area-specific information from other sources. With the GSE census tract file, one can compute the level and type of GSE activity in a census tract (e.g. number and percent of GSE loans that went to first-time homebuyers), and then merge that information with HMDA. HUD and HMDA make available annual estimates of median family income by MSA; this too can be merged with HMDA, making it possible to compute whether an applicant should be coded as low income or not.

Table A-2: List of B & C and Manufactured Housing Lenders*

ID Number	Lender Name	Type of Lender
95-4438859-7	AAMES CAPITAL CORPORATION	B & C
95-4601683-7	AAMES CAPITAL CORPORATION OF M	B & C
95-2622032-7	AAMES FUNDING CORPORATION	B & C
95-4362095-7	AAMES HOME LOAN	B & C
95-2591924-7	AAMES HOME LOAN OF AMERICA	B & C
88-0303373-7	AAMES HOME LOAN OF NEVADA	B & C
6502700005-7	ACCESS FINANCIAL LENDING CORP	B & C
13-3237773-7	ADVANTA MORTGAGE CORP. - NE	B & C
23-2159309-7	ADVANTA MORTGAGE CORP. MID-ATL	B & C
23-2532654-7	ADVANTA MORTGAGE CORP. USA	B & C
23-2434974-7	ADVANTA MORTGAGE CORP.-MIDWEST	B & C
7083400004-7	ALLIANCE MORTGAGE BANKING CORP	B & C
59-2645397-1	ALTEGRA CREDIT COMPANY	B & C
0541664826-7	APPROVED RESIDENTIAL MORTGAGE	B & C
0000765578-2	BANC ONE FINANCIAL SERVICES	B & C
0000012416-4	BANK OF AMERICA, FSB	Manufactured Housing
0000008939-4	BANK OF AMERICA, FSB	Manufactured Housing
88-0331093-7	BELGRAVIA FINANCIAL SERVICES	Manufactured Housing
7766600004-7	BENEFICIAL MORTGAGE CORPORATIO	B & C
22-2630964-7	CHAMPION MORTGAGE CO.	B & C
0001035698-2	CIT GROUP/CONSUMER FINANCE,INC	Manufactured Housing
0001999138-2	CIT GROUP/CONSUMER FINANCE,INC	Manufactured Housing
0001035401-2	CIT GROUP/SALES FINANCING DE	Manufactured Housing
7496500002-7	CITYSCAPE CORPORATION	B & C
52-0278530-7	COMMERCIAL CREDIT CONSU	B & C
52-0278534-7	COMMERCIAL CREDIT CONSU	B & C
52-0278491-7	COMMERCIAL CREDIT CORPO	B & C
52-0278514-7	COMMERCIAL CREDIT CORPO	B & C
52-0278518-7	COMMERCIAL CREDIT CORPO	B & C
52-1264637-7	COMMERCIAL CREDIT CORPO	B & C
52-1690525-7	COMMERCIAL CREDIT CORPO	B & C
51-0372905-7	COMMERCIAL CREDIT OF AL	B & C
52-0609364-7	COMMERCIAL CREDIT PLAN	B & C
52-0278529-7	COMMERCIAL CREDIT PLAN,	B & C
52-0799008-7	COMMERCIAL CREDIT LOANS, INC.	B & C
52-1494782-7	COMMERCIAL CREDIT LOANS, INC.	B & C
52-1008409-7	COMMERCIAL CREDIT PLAN INC.	B & C
52-0808447-7	COMMERCIAL CREDIT PLAN, INCOR	B & C
7511600000-7	CONTIMORTGAGE CORPORATION	B & C
7568500004-7	CUSTOM MORTGAGE INC.	B & C
7069000008-7	DELTA FUNDING CORPORATION	B & C
56-1977469-7	DEUTSCHE FINANCIAL CAPITAL	Manufactured Housing
54-1779092-7	DYNEX FINANCIAL, INC.	B & C
59-3324910-7	EMERGENT MORTGAGE CORP	B & C
0002036450-2	EQUICREDIT CORP OF AMER	B & C
6473009998-7	FIRST FRANKLIN FINANCIAL CORP.	B & C
0000022559-1	FIRST UNION HOME EQUITY BK NA	B & C
6500200040-7	FORD CONSUMER FINANCE CO.,INC.	B & C
0000025653-3	FREMONT INVESTMENT & LOAN	B & C
2294709990-7	GREEN TREE FINANCIAL	Manufactured Housing
7566600002-7	GREENTREE FINANCIAL CORP	B & C
7568300002-7	GREENTREE MORTGAGE COMPANY, LP	B & C
7053300004-7	IMPERIAL CREDIT INDUSTRIES INC	B & C
7883200007-7	INDEPENDENT NATIONAL MTG	B & C
0000007946-4	LIFE SAVINGS BANK, F.S.B.	B & C
7756600001-7	LONG BEACH MORTGAGE COMPANY	B & C
6480209999-7	MASTER FINANCIAL INC.	B & C
0002142959-2	NATIONSCRE HOME EQUITY SERVICE	B & C
7506600003-7	OAKWOOD ACCEPTANCE CORPORATION	Manufactured Housing
0000008327-4	OCEANMARK BANK, A FSB	B & C
33-0536622-1	OPTION ONE MORT CORP	B & C
0000027415-3	PACIFIC T&LC	B & C
6487409995-7	QUALITY MORTGAGE USA, INC.	B & C
23-2772890-7	RESIDENTIAL MONEY CENTERS, INC	B & C
0000020589-3	SANWA BK CA	B & C
7650700000-7	SAXON MORTGAGE, INC.	B & C
6469509992-7	SOUTH PACIFIC FINANCIAL CORP.	B & C
63-0570060-1	SOUTH TRUST MOBILE SERVICES	Manufactured Housing
7893400007-7	SOUTHERN PACIFIC FUNDING CORP.	B & C
4864400009-7	THE MONEY STORE	B & C
0951428083-7	TRANSAMERICA FINANCIAL SERV.	B & C
7261100005-7	UNICOR FUNDING, INC.	B & C
7434800003-7	UNITED COMPANIES FINANCIAL COR	B & C
4856500006-7	VANDERBILT MORTGAGE	Manufactured Housing
7751500009-7	WALSH SECURITIES	B & C
0000011905-4	WASHINGTON MUTUAL BANK FSB	Manufactured Housing
0458600405-7	WEYERHAESER MORTGAGE COMPANY	B & C

* All information was provided by HUD, except for the second listing for Bank of America-FSB, which was added by the researchers of this report.