

Syracuse University

SURFACE at Syracuse University

Center for Policy Research

Maxwell School of Citizenship and Public
Affairs

11-1995

Estimating the Volume of Multifamily Mortgage Originations By Commercial Banks Using the Survey of Mortgage Lending Activity and the Home Mortgage Disclosure Act Data

Amy D. Crews
Syracuse University

Robert M. Dunskey
Syracuse University

James R. Follain
Syracuse University

Follow this and additional works at: <https://surface.syr.edu/cpr>



Part of the [Economic Policy Commons](#), [Economics Commons](#), and the [Public Policy Commons](#)

Recommended Citation

Crews, Amy D.; Dunskey, Robert M.; and Follain, James R., "Estimating the Volume of Multifamily Mortgage Originations By Commercial Banks Using the Survey of Mortgage Lending Activity and the Home Mortgage Disclosure Act Data" (1995). *Center for Policy Research*. 458.
<https://surface.syr.edu/cpr/458>

This Working Paper is brought to you for free and open access by the Maxwell School of Citizenship and Public Affairs at SURFACE at Syracuse University. It has been accepted for inclusion in Center for Policy Research by an authorized administrator of SURFACE at Syracuse University. For more information, please contact surface@syr.edu.

**Metropolitan Studies Program Series
Occasional Paper No. 172**

**ESTIMATING THE VOLUME OF MULTIFAMILY MORTGAGE
ORIGINATIONS BY COMMERCIAL BANKS USING THE
SURVEY OF MORTGAGE LENDING ACTIVITY AND
THE HOME MORTGAGE DISCLOSURE ACT DATA**

**Amy D. Crews, Robert M. Dunskey,
and James R. Follain**

**Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
Syracuse, New York 13244-1090**

November 1995

\$5.00

This paper is based upon research partially supported by the U.S. Department of Housing and Urban Development. Helpful comments were received from Dixie Blackley, Harold Bunce, John Dickie, John Gardner, William Reeder, and Randy Scheessele. Remaining errors are our own.

CENTER FOR POLICY RESEARCH - SPRING 1996

Timothy M. Smeeding
Professor of Economics & Public Administration

Director

Richard V. Burkhauser
Professor of Economics

Margaret M. Austin

John Yinger
Professor of Public Administration
& Economics

Associate Director,
Aging Studies Program

Associate Director,
Budget and Administration

Associate Director,
Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Amy Crews	Economics	Bernard Jump	Public Administration
William Duncombe	Public Administration	Duke Kao	Economics
Thomas Dunn	Economics	Marcia Meyers	Public Administration
James Follain	Economics	Jerry Miner	Economics
Vernon Greene	Public Administration	Jan Ondrich	Economics
David Greytak	Economics	John Palmer	Public Administration
Barbara Grosh	Public Administration	Michael Wasylenko	Economics
Christine Himes	Sociology	Douglas Wolf	Public Administration
Douglas Holtz-Eakin	Economics	Assata Zerai	Sociology

RESEARCH ASSOCIATES

Debra Dwyer	Economics
James McNally	Demography

GRADUATE ASSOCIATES

Hakan Aykan	Public Administration	Suzanne McCoskey	Economics
Debra Bailey	Economics	Michael McLeod	Public Administration
Lloyd Blanchard	Public Administration	Meeae Park	Social Science
Meg Bower	Public Administration	John Phillips	Economics
Don Bruce	Economics	John Poupore	Economics
Shannon Felt	Economics	Mark Robbins	Public Administration
Tess Heintze	Public Administration	Katherin Ross	Social Science
Chin Hu	Sociology	Alex Striker	Economics
Christopher Ingram	Public Administration	Robert Weathers	Economics
Carol Jenkins	Public Administration	David Wittenburg	Economics
Rick Joy	Economics	Wilson Wong	Public Administration
Anna Lukemeyer	Public Administration		

STAFF

Martha W. Bonney	Publications and Events Coordinator	Alex Mintskovsky	Programmer
Barbara Butrica	Data Manager	Inge O'Connor	Administrative Assistant, Luxembourg Income Study
Karin D'Agostino	Computer Consultant	Mary Santy	Receptionist
Julie DeVincenzo	Librarian/Office Coordinator	Ann Wicks	Administrative Secretary
Esther Gray	Administrative Secretary	Jodi Woodson	Secretary
Gina Husak	Secretary to the Director	Sheng Zhu	Manager - Computing Services
Detlef Jurkat	Translator		

Abstract

Two public data sets on multifamily mortgage originations are used to resolve the \$15 billion discrepancy between the published estimates of the size of the multifamily lending market covered by commercial banks. The data are from the Survey of Mortgage Lending Activity and the Home Mortgage Disclosure Act. The analyses show the primary sources of the differences in the estimates are differences in the populations covered, nonreporting biases, and the methods used to expand the reported values to aggregate values. The 1993 multifamily mortgage originations volume by commercial banks is estimated to be about \$7-8 billion.

Introduction

With strong incentives to change government as we know it, the agencies responsible for creating, implementing, and monitoring housing policies are under ever greater pressure to streamline and improve their operations. In the area of low and moderate income housing in particular, the federal government has undergone a major policy shift away from the direct provision of housing into one of creating incentives for the private and nonprofit provision of these services. However, a necessary ingredient in forming effective policies for achieving specific goals is high quality information about the structure of the markets in question.

One market of particular interest currently is the multifamily mortgage market. One aspect that makes this market difficult to study is that the market is comprised of many lenders of widely varied types. Further complicating this issue are the enormous structural changes in the banking industry that have occurred in the past few years.

The data sources for analyzing this market also vary widely in their quality, but unfortunately, the number of sources is very few. The Survey of Mortgage Lending Activity (SMLA) is the oldest of the surveys covering this market. It has high quality information on some aspects of the multifamily mortgage market, such as the lending activity by thrifts, but also has some severe weaknesses. Unfortunately, the weakest part of the SMLA multifamily data is that on commercial and mutual savings banks, which comprise a large segment of this market.

A second source of data on this market is the Home Mortgage Disclosure Act (HMDA) data, which has (theoretically) universal coverage of all lenders covered by the Act. Although this data has been collected since 1976, its use in estimating multifamily mortgage activity has occurred only recently. The HMDA data also has some weaknesses, and like the SMLA data, one of these weaknesses is the data on the multifamily lending activity by commercial banks.

The purpose of this study is to evaluate the data sources available in search of an answer to one question in which HUD and other agencies are interested: what was the total volume of multifamily housing loan originations by commercial banks in 1993? This report focuses on the SMLA and HMDA data. A third data set, the Residential Finance Survey (RFS), is analyzed in the paper by Dunskey, Follain, and Ondrich (1995), and will not be considered here.

We find that the SMLA is weak on multifamily mortgage origination activity by commercial banks. The worst fault is that the sample contains observations on only 48 banks. Only 28 banks report any multifamily originations in 1993. Using the SMLA, we estimate the total volume of multifamily mortgage originations in 1993 by commercial and mutual savings banks to be \$20.50 billion. This compares with SMLA's own estimate of \$19.93 billion. However, because of the way the SMLA data is used to estimate the size of the multifamily mortgage market, it is not known whether the SMLA estimate is too high or too low.

We also find the HMDA data to be weak on multifamily originations activity by commercial banks. However unlike the SMLA, the HMDA data has a large number of banks and a wide distribution of banks by size. Using the HMDA data, we estimate the total volume of multifamily mortgage originations by commercial banks to be \$4.84 billion under the assumption that HMDA contains the universe of these banks. We conclude, however, that HMDA does not contain the full universe of commercial banks that write multifamily originations, and thus estimate the volume of multifamily loans based on HMDA data to be approximately \$8.48 billion.

Based only on the HMDA sample of multifamily lenders, HMDA data misses 5 of the 10 largest banks (those with assets exceeding \$36 billion) and 468 of the largest 1000 banks (those with assets exceeding \$300 million). Although some of the missing banks are no doubt due to some banks not writing multifamily loans, the SMLA and HMDA data that we analyzed cover two (nearly) mutually-exclusive subsets of multifamily lenders.

The next section discusses the strengths and weaknesses of the SMLA and HMDA data we examined. In Section III, estimates of 1993 multifamily mortgage originations are given for commercial banks, and a comparison of the banks in the samples is made with aggregate data on the commercial banking industry. We provide suggestions for how the use of existing data to estimate originations volume may be improved for existing data in Section IV and how the data collection may be improved in Section V. Section VI concludes.

The Data

This section describes the SMLA and HMDA data sets. The real and potential strengths and weaknesses of both are discussed, and a brief statistical description of the samples is given.

The Survey of Mortgage Lending Activity

The Survey of Mortgage Lending Activity is a monthly survey of lending institutions. Some of the data contained in the survey is collected through other agencies, such as universal data on thrifts which is collected by the OTS, and the rest is collected from a sample of institutions.

The SMLA was carefully analyzed by ICF, Inc. in 1994. Their report provides a detailed analysis of the quality of the data, so we will only briefly summarize their conclusions here.

ICF believes the data on commercial and mutual savings banks is by far the weakest part of the survey. The primary reason for their conclusion is that the stratified random sample on which the survey is based has not been redrawn in over 20 years. At the inception of the survey, the sample of banks was likely very representative of commercial banking industry. However, with the dramatic changes in the banking system as a whole, this current sample is most certainly not representative of the commercial banking industry. For example, the survey only contains banks that have been in business in some form over the full 20-year period since the sample was

drawn. In the event of a merger involving a bank in the sample, the resulting bank is retained in the sample only if it is designated as the “parent bank” by the FDIC.

The 1993 survey has 537 observations on 48 commercial and mutual savings banks, of which 41 report in all 12 months.¹ Multifamily mortgage originations are reported by 171 observations representing 28 banks. These originations total \$1.85 billion. We were not provided with any indicators separating the commercial banks from the mutual savings banks. Therefore, hereafter, the SMLA banks are simply referred to as commercial banks. The sample moments of these data are presented in the Appendix to this paper.

Inclusion in the SMLA is voluntary for commercial banks, however banks are removed from the survey for systematic nonresponse. This system further compounds the selection bias caused by the longevity requirements, although we are unable to determine from the SMLA how reporting banks differ from non-SMLA banks.

There is some concern over the quality of the reports made by banks that do participate in the sample. For example, some banks may participate in community development consortia which provide money for multifamily housing and other projects. Since a bank participating in such a consortia would not originate a loan made through the consortia, some under-reporting of originations may occur.² This problem may also affect the HMDA data.

Some of the reported originations in the SMLA are quite small; of the 171 observations with multifamily originations activity, 12 report total monthly originations below \$85,000, and 5 are below \$50,000. We do not know the circumstances determining these loans, however, these outliers are unlikely to have a large effect on the resulting estimates, which are based on annual total originations in this paper.³

The ICF report suggests a second type of respondent error where some banks may report loans originated through their mortgage banking subsidiaries. Although the directions for filling

in the survey questionnaire specifically state that the information in the survey should pertain only to that particular bank's activities, the ICF report suggests that some banks may report these loans anyway. Thus it is possible that some double counting of loans may be occurring. Our attempts to discover the frequencies of these occurrences lead us to believe that this rarely, if ever, occurs since the mortgage companies are managed independently from their parent companies, as are other subsidiaries. Moreover, the SMLA values are cross checked against FDIC reports which cover only the banks' activities, and HUD staff find the SMLA to be accurate in this respect.

The Home Mortgage Disclosure Act Data

The HMDA data has been collected since 1976. However, the reporting requirements under the Act were greatly revamped in 1988 and 1989. These changes expanded the scope of the institutions that are covered under HMDA and the detail of the information provided by the reporting institutions. Reporting by HMDA covered institutions is required and, therefore, should produce a sample equivalent to the universe of these lenders. However, HMDA does not cover all lenders. Traditionally, HMDA required reporting by all depository institutions with more than \$10 million in assets and a home office or branch in an MSA. The 1988 and 1989 amendments expanded this coverage to nondepository institutions.⁴

The HMDA data suffers from problems of its own, and thus is not a perfect alternative to the SMLA. First, some institutions are missed due to confusion over who has to report. This was a particularly common problem with mortgage bankers who were brought into the HMDA reporting circle in 1988.⁵ Also, as banks merge with one another or just simply grow, their exemption status may change. Thus, some banks may fail to report if they do not realize that HMDA now applies to them.

Second, the HMDA was intended to gather information about *home* lending, which is more often thought of as single-family housing. Banks that do little of this type of lending directly

may, however, participate in the multifamily housing mortgage lending market, which may be considered commercial lending by some banks. Some multifamily housing loans may also be missed in HMDA because of the way some banks separate their single and multifamily lending operations.

A final way in which HMDA underreports originations is that some loans purchased by Fannie Mae are being missed. This problem is analyzed more carefully by Crews, Dunsky, and Follain (1995), who show that up to 50 percent of Fannie purchases are not reported in HMDA. That analysis shows that only a small number of Fannie purchases are from commercial banks; most of the loans are purchased from mortgage bankers. Nonetheless, the volume of Fannie loans missed by HMDA is nontrivial at \$2.5 billion.

In the HMDA data we analyzed, there are 37,007 total observations of which 23,502 are conventional mortgage originations.⁶ Of those observations, 9090 correspond to commercial banks and 190 belong to mortgage companies that are subsidiaries of commercial banks, which are hereafter omitted.⁷ The commercial bank originations correspond to 1,531 commercial lenders and total \$4.84 billion.⁸ Adding the restrictions that the originations correspond to properties in an MSA and that the lender's assets exceed \$1 million reduces the sample to 1,418 lenders and \$4.11 billion in originations. This restricted sample allows us to augment the HMDA data with census tract data. The sample statistics for these data are given in the appendix to this report.

Estimating Multifamily Originations

This section is devoted to both replicating the HUD estimates of multifamily mortgage originations by commercial banks and extending the analysis to test the sensitivity of the estimates to the method used.

Replicating HUD's Estimates Using the SMLA

HUD estimates that multifamily mortgage originations by commercial banks and mutual savings banks amounted to \$19.93 billion in 1993 (HUD 1995a). These estimates are obtained by taking the reported total multifamily mortgage originations for each bank and scaling that total by an expansion factor. The scaled values are then totaled for the national estimate for banks of these types.

The expansion factors are determined by the share of total real estate assets held by reporting SMLA banks relative to that held by all banks in a similar stratum. Specifically, the total multifamily loan originations values for each stratum are multiplied by the ratio of total real estate assets held by all banks in the stratum to the assets of SMLA banks. Thus, if SMLA banks account for one quarter of all assets in the stratum of largest banks, the SMLA estimate of loan originations for this stratum is four times the sample total. The total real estate assets for each stratum are obtained using the Call Reports data maintained by the Federal Reserve.

Our estimates from following this technique are provided in Table 1. We did not have access to the Call Report data, so the estimates in the table are based upon the total real estate assets for commercial banks in 1993 reported in the *Federal Reserve Bulletin* (1994b).⁹ No distinction is made for bank size since we do not observe total assets, nor are banks broken down by type. The total real estate assets of all commercial banks in 1993 is \$916.8 billion as reported in the *Bulletin*.

For the banks that reported multifamily originations activity in the SMLA sample, the total multifamily originations are \$1.85 billion. Total real estate assets for each bank are calculated by taking the average total ending balance in real estate for each month that the bank reported in the sample. Two estimates of total assets are provided in the table, one for the sample of all multifamily mortgage originations (171 observations), and the second for all monthly multifamily

mortgage originations that exceed \$50,000. In both cases, total real estate assets represented by these banks are approximately \$83 billion.¹⁰

The expansion factors are created by dividing the Federal Reserve's (1994) value for total real estate assets by the sample totals. These factors are 11.07 and 11.08 for the unrestricted and restricted samples, respectively. The resulting multifamily mortgage originations estimate in both cases is \$20.50 billion. This value is slightly higher than the estimate provided in the HUD (1995a) report (\$19.93 billion), and the discrepancy is due almost certainly to the different measures of total real estate assets. For example, our measure did not take into account different sizes of banks which is done in the HUD estimate. Furthermore, we did not separate commercial banks from mutual savings banks.

Estimating Multifamily Loan Originations using the SMLA: Regression Analysis

The use of expansion factors based on assets is sensible only if there is a strong correlation between the bank's originations activity and its size. Therefore, we examine the relationship between these two variables in a simple regression framework. Table 2 reports the estimates for two models using the SMLA.

In the first model total multifamily originations are regressed on a constant and a cubic polynomial in assets. The second model regresses the ratio of multifamily originations to assets on a cubic polynomial in assets. The two models suggest a highly nonlinear relationship with assets, and, while the use of expansion factors based on real estate assets is supported, a more sophisticated weighting system would reduce bias in the SMLA estimate of multifamily mortgage originations volume.

For heuristic purposes, Figures 1 and 2 show the plots of total originations and originations to asset share against real estate assets for the SMLA banks. The small sample sizes

allow the regression estimates to be dominated by outliers. However, we can see in both pictures that the correlation with real estate assets is quite strong.

Replicating HUD's Estimates Using the HMDA Data

The HMDA data offers a larger sample, and while not without problems of its own, it is an alternative to the SMLA. We begin again by trying to replicate the estimates given in HUD (1995b). The sample is restricted to observations that reported loan agency codes consistent with commercial banks (reporting to either Office of the Comptroller of the Currency, the Federal Reserve System, or the FDIC, and not a mortgage subsidiary of a bank). Observations are further restricted to include only banks located in MSAs with assets in excess of \$1 million. The resulting sample includes 1,418 banks. The restricted sample is required for matching the HMDA data with census tract data.¹¹

Table 3 reports the estimates for average loan amounts, average loan originations by institutions, average assets of these institutions, total multifamily originations and total assets by all members in the two groups. The estimate of total originations by HMDA commercial banks is \$4.84 billion. For reference, total originations by commercial banks located in an MSA and having more than \$1 million in assets are \$4.11 billion. Both totals are approximately \$3 billion below the RFS estimate with HMDA (Gardner) restrictions given in Crews, Dunsky, and Follain (1995).

Both HMDA estimates of total multifamily originations are calculated under the assumption that the HMDA sample contains the universe of commercial banks. However, with only 1,531 lenders in the multifamily lender sample, this assumption cannot be true. Table 4 shows the distribution of the HMDA banks we analyzed by asset size. The size classifications are based on the Federal Reserve definition given in the *Federal Reserve Bulletin* (1994a). They are: (1) largest banks, corresponding to the 10 largest commercial banks, which have more than \$36

billion in assets; (2) large banks, corresponding to the next 90 largest banks, which have more than \$6 billion in assets; (3) medium banks, corresponding to the next 900 largest banks, which have more than \$300 million in assets; and (4) small banks, corresponding to banks not in the largest 1000, which have assets less than \$300 million.

The sample commercial banks has 5 banks in the largest class and 62 in the large class. In all, 468 of the 1000 largest commercial banks are missing from the multifamily sample because the bank does not lend in the multifamily market, the bank does not report is multifamily loans, or the bank does not report to HMDA.¹² Therefore, the estimate of multifamily originations should take account of the fact that the sample is not a universe. If the same methodology used in the SMLA estimate is applied here using total assets, the HMDA estimate of multifamily originations volume is \$8.48 billion. This expanded estimate implicitly assumes that all commercial banks write multifamily loans, but not all banks report in the HMDA data. This estimate is close to the estimate obtained using the HMDA (Gardner) restrictions on the RFS data.

Examining the Relationship Between Bank Size and Originations Activity Using HMDA Data

Based on the results from the SMLA regression estimates, an interesting question worth exploring in HMDA that could not be examined in SMLA is the relationship between bank size, total assets, and multifamily originations for commercial banks. Table 4 also provides the results from this comparison. For each group, average assets, average total multifamily originations, and multifamily originations as a percent of total assets are calculated for the HMDA sample. For comparison, the percent of assets held in multifamily loans for commercial banks reported in the *Federal Reserve Bulletin* (1994a) are given in the last column. Although the Fed value is a ratio of two stocks while the HMDA value is the ratio of a flow to a stock, the values are surprisingly

close. Furthermore, these values suggest that the relationship of multifamily originations with assets is nonlinear.

Estimating Multifamily Loan Originations Using the HMDA: Regression Analysis

In this exercise, we make use of the merged HMDA-census tract data on originations for commercial banks in MSAs, which was provided by HUD. This richer data gives some indication where banks are making their loans. Two regressions are estimated; one for the total multifamily loan volume of the reporting institution and the other for the share of originations activity to total assets by the lender. These estimates are reported in Tables 5 and 6.

The simple models shown in each table regress the dependent variable on a constant, assets, assets squared, and assets cubed. The second model is a regression on these variables and a the average characteristics of the census tracts to which these loans apply (see Table A2 for these characteristics). From Table 5, multifamily loan originations are nonlinear in total assets. In Table 6, the ratio of multifamily originations to assets is not constant with respect to bank size. Overall, about 10 percent of the total lender multifamily originations are explained by the bank's size and the characteristics of the areas served.

As with the SMLA data, we plotted the relationship between asset size and multifamily originations. Using the same banks size classifications as in Table 4, these relationships are shown in Figure 3. Small banks do very little multifamily originations volume, and volume by banks does not vary with the bank's asset size. Medium and large banks do a large volume of originations, and when compared with the originations volume by the largest banks, the relationship appears to be nonlinear in asset size, first increasing then decreasing.

Other Data Analysis

In an earlier preliminary report to HUD we had proposed using the Call Reports data in a selection model with the SMLA data to improve our estimates of the appropriate expansion

factors. However, the small sample size in the SMLA and time constraints made this approach infeasible. Since we already had the HMDA data, we attempted to do the same type of analysis by matching the SMLA with HMDA. This attempt was unsuccessful because there was only one match between the two data sets. The same match was performed by PD&R at HUD, and they also achieved only one match.¹³

This exercise provides strong evidence that the HMDA data on multifamily mortgage originations is not a universe, and moreover, that the failure of our sample to show a universe of commercial banks is not due to a large number of commercial banks that do not write multifamily mortgages. Since the SMLA banks combined show originations volume of \$1.85 billion and the HMDA commercial banks combined report originations of \$4.84 billion, the HMDA is missing some very important lenders. The simple total for multifamily originations volume combining the two samples is \$6.69 billion using all HMDA banks.¹⁴ The RFS estimate for the HMDA type originations is approximately \$7 billion. (See Crews, Dunskey, and Follain, 1995.) This number appears to be a reasonable estimate of total multifamily lending volume by commercial banks in 1993.

A Proposal for Improving the Estimates of Multifamily Originations Activity of Commercial Banks Using Existing HMDA Data

Even though the HMDA is presumed to be universally reported for all banks covered by the Act, many banks are missing from the sample. In this section we propose a method for using the HMDA data, supplemented with the Call Reports data, to obtain a more complete estimate of multifamily originations using the existing data. The benefit of using this model, originally proposed in an earlier report by our team for correcting bias in the SMLA estimates, is that it would not require the collection of additional data yet would correct the estimates for the biases

caused by the systematic (either by the HMDA reporting criteria or noncompliance) exclusion of certain banks.

This model has the basic form of

$$MFOs_i = f(\mathbf{X}_i | \text{sample selection rule})$$

where \mathbf{X}_i is a set of bank characteristics such as real estate holdings, assets, years in operation, etc. The sample selection rule is what determines which banks are in the sample.

A common model of this type is the Heckit model, or Heckman 2-step regression model.¹⁵ In these types of models, the first step involves estimating the probability that the bank's MFOs are observed (i.e., that the bank is in the HMDA multifamily sample). This estimated probability is then used to estimate the bank's MFOs as a function of bank characteristics in a basic regression framework.

In our model, which assumes the errors are distributed normally, the probability that a bank is chosen to be in the HMDA sample (without distinction between the bank's choice of whether to comply or the exclusion by reporting mandates) is a function of the bank's characteristics w_i . However, the actual selection index value, z_i^* , on which the decision of whether to include the banks is made, is not observed. Instead, the outcome of the decision is observed; the bank is either in or not in the sample. Under the assumption that $u_i \sim N(0,1)$ and letting $\Phi(\bullet)$ denote the normal cumulative distribution function, the selection mechanism has the following mathematical form,

$$z_i^* = \gamma' \mathbf{w}_i + u_i,$$

$$z_i = 1 \quad \text{if } z_i^* > 0,$$

$$z_i = 0 \quad \text{if } z_i^* \leq 0,$$

$$\text{prob}(z_i = 1) = \Phi(\gamma' \mathbf{w}_i),$$

$$\text{prob}(z_i = 0) = 1 - \Phi(\gamma' \mathbf{w}_i).$$

The regression model of interest depends on the characteristics of the observed banks which in turn depend on the sample that was drawn. Therefore, the regression of bank characteristics on MFOs has the following form,

$$MFOs_i = \beta' \mathbf{x}_i + \epsilon_i, \quad \text{observed only if } z_i = 1,$$

$$(u_i, \epsilon_i) \sim \text{bivariate normal} [0, 0, 1, \sigma_\epsilon, \rho].$$

If the z_i and \mathbf{w}_i are observed for an entire sample of banks (which they would be for the universe in the FDIC Call Reports, for example), but MFOs are observed only when $z_i = 1$ (the HMDA multifamily lenders sample), then the expected value of the observed originations is given by

$$E(MFOs_i | z_i = 1) = \beta' \mathbf{x} + \rho \sigma_\epsilon \lambda(\gamma' \mathbf{w}),$$

where

$$\lambda_i = \frac{\phi(\gamma' \mathbf{w}_i)}{\Phi(\gamma' \mathbf{w}_i)}.$$

The standard normal density function is denoted by $\phi(\bullet)$, and, as before, $\Phi(\bullet)$ denotes the standard normal distribution function.

We are interested in getting the expected value of the total MFO market. Therefore, to get an unbiased estimate of the population values of β , the estimated regression equation should be in the form of

$$MFO_{S_i} = \beta'x_i + \beta_\lambda \lambda(\gamma'w_i) + v_i,$$

where

$$\beta_\lambda = -\rho\sigma_\epsilon \text{ and } v_i = \rho\sigma_\epsilon \lambda(\gamma'w_i) + \epsilon_i.$$

This MFO equation is estimated in two steps. First, the vector γ is estimated from a probit equation (z_i is the dependent variable) to determine the probability that a bank in the HMDA multifamily lender sample is selected from the universe of banks. In the second step, the MFO regression equation is estimated using the imputed values of $\lambda(\gamma'w_i)$ and the observed bank characteristics x_i .

We propose to use the HMDA data merged with the Call Reports data. Together, these data provide extensive data on the banks and their lending activity. The proposed model would correct the HMDA estimates for bias due to the sample design. Once the merged data is in hand, this method should be easy to implement since the procedure is part of many standard statistical packages.

One issue remains to be resolved in this model. To this point we have only suggested how the expected value of originations as a function of bank characteristics could be obtained. This, however, does not provide us with the estimate we desire, namely, total multifamily originations. If the relationship between originations and total assets (or total real estate assets) is linear, then the second stage regression is not necessary. Instead, the first stage probabilities could be used to create expansion factors for inflating the HMDA originations total. If, however, the relationship is nonlinear (as we found in our simple regression models for SMLA and HMDA data) then the second stage estimates would be required. The Heckit model provides unbiased estimates of the

β' vector. These estimated coefficients could then be applied to the data on commercial banks contained in the Call Reports data to estimate total multifamily originations.

Suggestions for Improving the Collection of Multifamily Originations Data

The SMLA clearly needs major revisions in its sampling scheme for commercial and mutual savings banks, and HUD has already issued a Request for Proposals for the new SMLA. The survey itself is rather straightforward in its requests for filers, yet two problems arise which are simple to correct. First, the survey should be expanded slightly to include some demographic information on the lending institutions. For example, total assets should be recorded for all reporting banks. This information is currently available by supplementing the SMLA data with data from the Federal Reserve's Call Reports. Adding the question in the survey eliminates the need for matching observations with the call reports for the estimates in question. The second problem is one of response compliance with the survey. There is no particular need to require monthly reporting, which places a heavy burden on the reporting institutions. Therefore, to increase compliance, banks could be sampled quarterly or annually for this information.

The sample could be redesigned with the intention of either supplementing the information in HMDA or as an alternative to HMDA, depending on the needs of HUD and other researchers, and funding availability. To increase total coverage of banks and keep collection costs to a minimum, the SMLA could be designed to catch the banks missed by HMDA. For example, based on Figure 3 and Table 4, the SMLA could sample the 468 medium, large, and largest banks that are not part of the HMDA multifamily sample. By merging the new data with HMDA reports the combined sample would contain the universe of the largest 1,000 banks plus

approximately 1,000 small banks, and would almost certainly capture the majority of commercial bank multifamily originations.

The information provided in the SMLA for other lending institutions is of fairly high quality with the exception of the mortgage bankers. However, the data collection for these institutions is being modified with the assistance of the Mortgage Bankers Association. Thus, the criticisms we have of the SMLA collection of information on commercial banks should not be interpreted to suggest that the entire survey is of low quality.

Although HMDA contains a better distribution of banks across the size classes, many large and medium banks are being missed. In our analysis, we were not able to determine the reasons for why these banks are being missed, but likely reasons are that banks may not be aware that they are covered by HMDA reporting requirements or the banks are legitimately not required to report. In the former case, attempts should be made to get the word out to these institutions, such as the efforts being made in collecting the mortgage bank information. In the latter case, the solution is not so easy to implement and a simpler alternative to changing the HMDA reporting requirements (which come from Congress) would be to use the redesign of the SMLA to fill in the weak areas in HMDA.

Conclusions

We find the Survey of Mortgage Lending Activity and the Home Mortgage Disclosure Act data both have weaknesses when used to estimate the volume of multifamily mortgage originations by commercial banks. First, the SMLA sample covers only a very small number of lenders, and while the volume of mortgage originations does seem to be highly correlated with total real estate assets, the estimate of industry total originations volume using expansion factors is suspect. Our SMLA originations estimate for commercial banks is \$20.5 billion using the

expansion factor methodology. The SMLA banks in our sample report \$1.85 billion in originations.

Second, the HMDA data is not a universe of all commercial banks. This conclusion is supported by two facts. First, our sample contained only 1,531 lenders, or just over 10 percent of all commercial banks. While some lenders would necessarily not be part of our sample if they do not write any multifamily mortgages, this number still seems very low. The second fact is that when the SMLA sample and the full HMDA multifamily sample were matched by lender identification numbers, only one match occurred. Thus, there are at least 27 banks that write multifamily mortgages that HMDA missed.

The HMDA based estimate of total multifamily originations is \$4.84 billion for all banks that reported multifamily mortgages, and \$4.11 billion for MSA banks with more than \$1 million in assets. If an expansion factor method is used with the total HMDA multifamily lender sample, the estimate for total multifamily originations becomes \$8.48 billion.

Our best guess of the volume of multifamily lending activity by commercial banks in 1993 is \$7 to \$8 billion, which is supported by the estimates using the RFS data in the companion Crews, Dunskey, and Follain (1995) report, HMDA expansion factor estimates, and simple addition of the reported originations in HMDA and the SMLA. Combining the data in the HMDA and SMLA with the call reports data would further strengthen the available information and improve the estimates given in this report. In the future, the commercial bank portion of the SMLA should be designed as a supplement to the HMDA as a relatively inexpensive way to greatly improve the collection of data on multifamily mortgage originations.

Endnotes

1. The banks which do not report in all 12 periods by identification number are: 40100 (did not report in August), 50020 (reported January through April only), 91563 (did not report August through December), 110010 (did not report in November or December), 172660 (did not report April through December), 172940 (reported in January only), 212090 (did not report in October, November, or December).
2. This point was raised in the ICF report. We were unable to determine the frequency of these types of loans, however, the discrepancy between the HMDA and Fannie Mae data, found by Crews, Dunsky, and Follain (1995), suggests that these loans may be a nonnegligible part of the multifamily mortgage originations volume.
3. The banks with low reported total monthly originations below \$85,000 by identification number are: 81010 (month reported = March, total monthly MFO = \$74,000), 901540 (June, \$21,000), 120540 (March, \$46,000), 132810 (November, \$16,000), 171490 (March, \$57,000), 171560 (January, \$2,000; November, \$50,000; and December, \$50,000), 171940 (January, \$30,000; and February, \$85,000), 196070 (November, \$40,000), and 212090 (May, \$74,000).
4. See *Federal Reserve Bulletin* (1994c) for more information on the HMDA and the data collection.
5. Through the efforts of the Mortgage Bankers Association, compliance with the HMDA is expected to increase in the future.
6. Observations where “type” = 1 and “action” = 1.
7. Commercial banks are observations reporting “code” = 1, 2, or 3 and “ocode” = 0. Commercial banks’ mortgage subsidiaries are observations reporting “code” = 1, 2, or 3 and “ocode” = 1.
8. These values were checked against those obtained by HUD staff for accuracy.
9. Arrangements with Policy Development and Research at HUD had been made for us to receive the Call Reports data, however, in light of the problems outlined above in the SMLA and the need for a timely report, this extension was terminated.
10. The estimates are not very sensitive to the choice of banks. The same estimation was performed using the full SMLA sample, including those banks that do not write multifamily loans, and the total assets increase to \$88.9 billion with a resulting originations estimate of \$19.06 billion.
11. Two observations with very low reported assets are eliminated by the \$1 million lower limit.
12. This issue is given further treatment below in Subsection F.

13. This matched bank was First Waco National Bank.
14. For these calculations, the one matched bank was double counted. Leaving the second contribution in simply accounts for omitted institutions.
15. See Greene (1993) for more about this type of model.

**Table 1. Estimates of Multifamily Mortgage Originations
Using the SMLA and Federal Reserve Bulletin ^a
(billions of dollars)**

Federal Reserve Bulletin (October 1994)		
Real Estate Assets of Commercial Banks in 1993	916.8	
Total Assets of Commercial Banks in 1993	3,598.6	
	All Banks Reporting MFOs > \$0	All Banks Reporting MFOs > \$50,000
SMLA Banks Reporting Multifamily Activity		
Total Multifamily Originations	1.85	1.85
Total Real Estate Assets ^b	82.81	82.77
Number of Banks	28	26
Fed Real Estate Assets to SMLA Real Estate Assets Ratio	11.07	11.08
Estimated Multifamily Mortgage Originations ^c	20.48	20.50
HUD's Estimated Multifamily Mortgage Originations Based on SMLA^d		
Commercial Banks	18.82	
Mutual Savings Banks	1.11	

^a See Table A1 for sample statistics.

^bTotal Assets is the sum of average total ending balances by bank.

^cCalculated as the product of the Fed to SMLA real estate assets ratio and total multifamily originations.

^dData from U.S. Department of Housing and Urban Development (1995a).

Sources: Authors' calculations, U.S. Department of Housing and Urban Development (1995a), and the *Federal Reserve Bulletin* (October 1994).

Table 2. Total Multifamily Loan Originations and MFO/Assets Share as a Function of Assets Using 1993 SMLA Data^a

Variable	Dependent Variable			
	Total MFOs (in millions)		MFO/Assets Share	
	Coefficient	t-statistic	Coefficient	t-statistic
Intercept	6.5400	0.603	0.0743	2.784
Total Assets (in millions)	0.0057	0.620	-3.85E-05	-1.689
Total Assets ² (in millions ²)	3.53E-06	3.837	3.69E-09	-1.588
Total Assets ³ (in millions ³)	-7.66E-11	-4.470	-6.48E-14	-1.536
Coefficient of Variation and Number of Observations	R ² = 0.974	n = 26	R ² = 0.117	n = 26

^aAssets are average end balance reported real estate loans held by bank.
Source: Authors' calculations.

**Table 3. 1993 Conventional Multifamily Loan Originations
Based on Home Mortgage Disclosure Act Data
(dollars)**

	Sample	
	All HMDA Commercial Banks ^a	All HMDA Commercial Banks in MSA ^b
Average Loan Amount (in thousands)	532.2	526.5
Average Total Multifamily Originations (in thousands)	3,160	2,928
Average Assets (in billions)	1.342	1.410
Total Assets (in billions)	2,054.60	1,999.80
Total Multifamily Originations (in billions)	4.838	4.113
Number of Lenders	1,531	1,418

^aOriginations accruing to commercial banks (code=1, 2, or 3, ocode=0) make up unrestricted sample, which corresponds to sample used in HUD's calculations.

^bThe restricted sample eliminates observations with MSA = 0000 and assets < \$1 million. See Table A2 for the sample moments.

Source: Authors' calculations.

Table 4. 1993 Conventional Multifamily Loan Originations Based on HMDA Data for Commercial Banks by Institution Size

	Bank Class ^a			
	Largest	Large	Medium	Small
Defined by Rank	10 Largest	Next 90 Largest	Next 900 Largest	Not in 1,000 Largest
Defined by Asset Size	More than \$36 Billion	Between \$6 and \$36 Billion	Between \$300 Million and \$6 Billion	Less than \$300 Million
Number of Lenders	5	62	465	999
Average Assets (millions of dollars)	105,716	13,292	1,270	112
Average Total Multifamily Originations (thousands of dollars)	22,025	21,863	5,262	926
Multifamily Originations as a Percent of Total Assets	0.45	0.58	1.09	1.33
Commercial Banks Multifamily Loans/Assets ^b (percent)	0.53	0.71	1.07	0.84

^aInstitution size determined by total assets, and categories are the same as those used in *Federal Reserve Bulletin* (June 1994) for Commercial Banks. U.S. Commercial Banks have \$3,598.6 billion in total assets (*Federal Reserve Bulletin*, October 1994).

^bThe *Federal Reserve Bulletin* reports share of assets held in multifamily residential loans, not originations to assets, for domestic commercial banks and nondeposit trusts.

Source: Authors' calculations, *Federal Reserve Bulletin* (June 1994), *Federal Reserve Bulletin* (October 1994).

Table 5. Multifamily Loan Originations to Total Assets as a Function of Assets and Characteristics of Areas Served Using 1993 HMDA Data

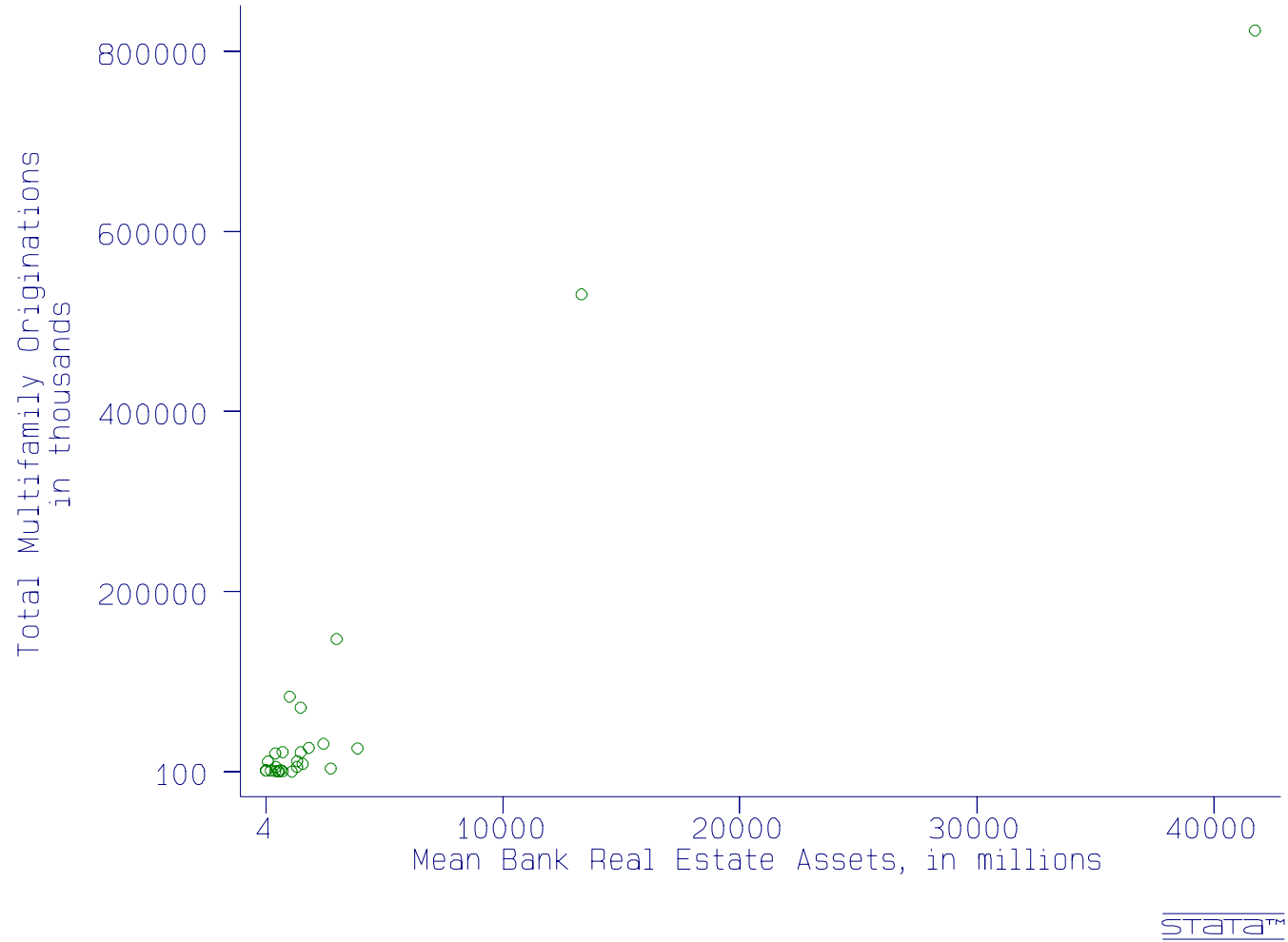
Variable	Regression Model 1		Regression Model 2	
	Coefficient	t-statistic	Coefficient	t-statistic
Intercept	1474.955	4.351	-3104.052	-1.434
Total Assets (in millions)	1.525	10.377	1.482	10.086
Total Assets ² (in millions ²)	-2.08E-05	-5.652	-2.01E-05	-5.476
Total Assets ³ (in millions ³)	7.55E-11	3.913	7.27E-11	3.777
Percent of Loans in Central City of MSA	●	●	1692.222	2.052
Average Income in MSAs Served	●	●	0.078	1.647
Average Minority Percentage in Areas Served	●	●	21.708	1.408
Percent of Areas Served with Decennial Median Incomes < 80 Percent of MSA Median	●	●	-1766.917	-1.765
Percent of Areas Served with Decennial Median Incomes > 120 Percent of MSA Median	●	●	2247.547	2.020
Coefficient of Variation and Number of Observations	R ² = 0.095	n = 1,418	R ² = 0.105	n = 1,418

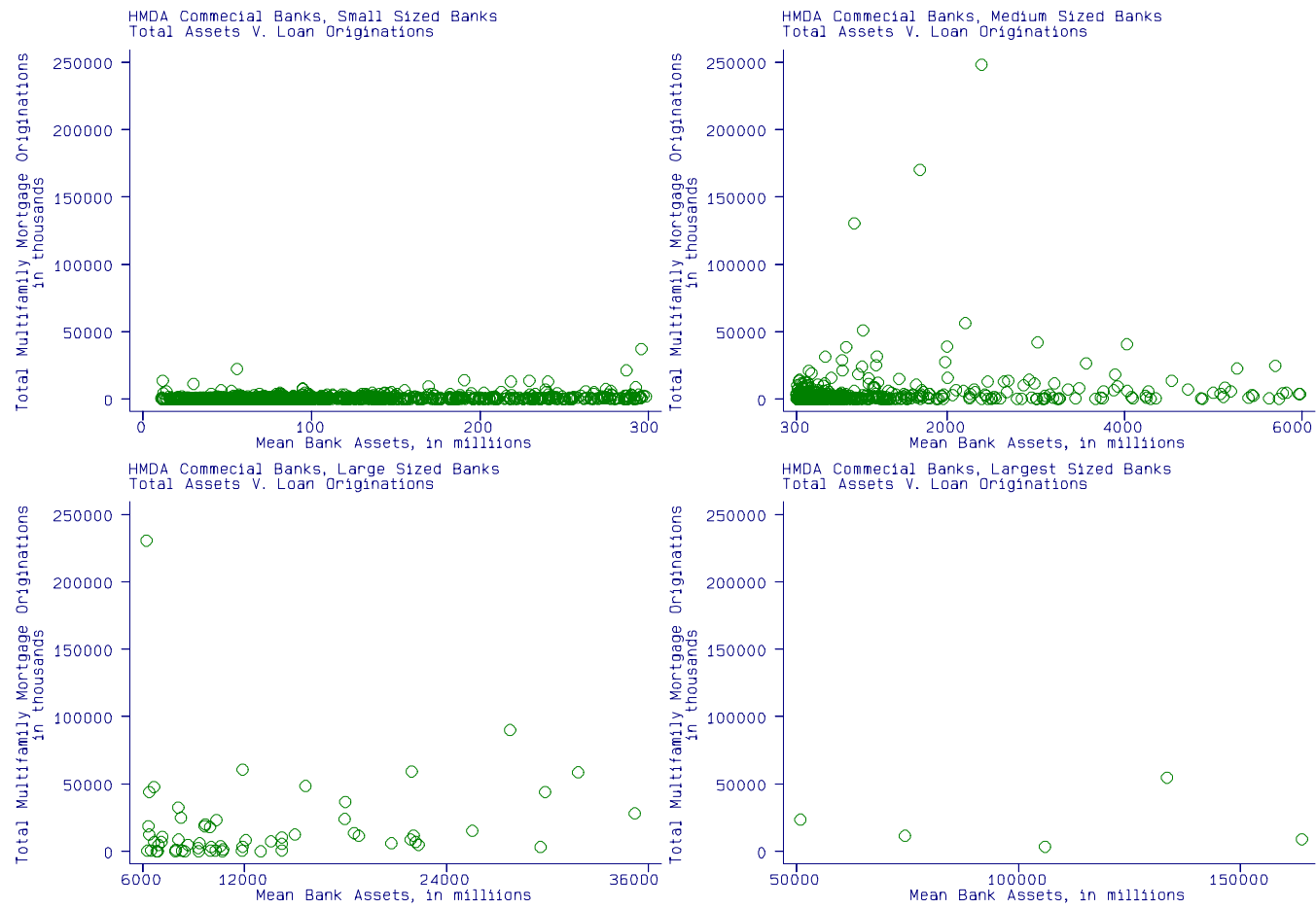
Source: Authors' calculations.

Table 6. Multifamily Loan Originations to Total Assets Share as a Function of Assets and Characteristics of Areas Served Using 1993 HMDA Data

Variable	Regression Model 1		Regression Model 2	
	Coefficient	t-statistic	Coefficient	t-statistic
Intercept	0.009	8.852	-0.002	-0.286
Total Assets (in millions)	-1.02E-06	-2.120	-1.08E-06	-2.229
Total Assets ² (in millions ²)	1.70E-11	1.409	1.77E-11	1.470
Total Assets ³ (in millions ³)	-6.96E-17	-1.101	-7.25E-17	-1.146
Percent of Loans in Central City of MSA	●	●	0.004	1.348
Average Income in MSAs Served	●	●	2.43E-07	1.549
Average Minority Percentage in Areas Served	●	●	4.38E-05	0.865
Percent of Areas Served with Decennial Median Incomes < 80 Percent of MSA Median	●	●	-0.004	-1.280
Percent of Areas Served with Decennial Median Incomes > 120 Percent of MSA Median	●	●	-1.20E-04	-0.033
Coefficient of Variation and Number of Observations	R ² = 0.004	n = 1,418	R ² = 0.007	n = 1,418

Source: Authors' calculations.





STATA™

Figure 3
HMDA Commercial Banks in MSAs by Asset Size: 1993

Appendix

Appendix Tables A1 and A2 contain the sample statistics from the SMLA and HMDA data we analyzed. In both tables, the first column contains the complete unrestricted sample, which in HMDA is the sample of commercial bank conventional originations. Subsequent columns contain the restricted samples we used in our estimation. The restrictions are described in the tables.

**Appendix Table A1. Sample Statistics for Commercial Bank Conventional
Multifamily Mortgage Originations 1993 SMLA
(millions of dollars)**

Variable Description	Unrestricted Data		Observations Reporting Total Monthly Originations > \$0		Observations Reporting Total Monthly Originations > \$50,000	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Monthly Beginning Balance in Real Estate Assets	1,854	6,243	2,965	8,035	3,191	8,306
Monthly Ending Balance in Real Estate Assets	1,853	6,218	2,957	8,004	3,183	8,273
Total Loan Originations, All Types	1,251	3,682	1,994	4,704	2,147	4,854
Monthly Beginning Balance in Multifamily Real Estate Assets	37.1	83.1	63.6	101	68.5	104
Monthly Ending Balance in Multifamily Real Estate Assets	37.0	82.0	63.3	99.9	68.2	102
Total Multifamily Mortgage Originations	38.6	140	66.1	180	71.2	186
Number of Observations	537		171		165	
Number of Lenders	48		28		26	

Source: Authors' calculations.

**Appendix Table A2. Sample Statistics for Commercial Bank Conventional
Multifamily Mortgage Originations 1993 HMDA
(millions of dollars)**

Variable Description	All HMDA Commercial Banks		Commercial Banks in MSAs with More than \$1 Million in Assets	
	Mean	Standard Deviation	Mean	Standard Deviation
Loan Amount (in thousands)	532	1,340	527	1,085
Lender Assets (in millions)	1,342	7,093	1,410	7,359
Loan in Central City of MSA (1=yes, 0=no)	0.804	0.557	0.614	0.418
Loan in Census Tract where Average Income Below 80 Percent of MSA Median (1=yes, 0=no)	0.439	0.398	0.346	0.384
Loan in Census Tract where Average Income Above 120 Percent of MSA Median (1=yes, 0=no)	0.155	0.283	0.181	0.306
MSA Median Income	41,705	6,787	41,704	6,792
Minority Percent in Census Tract	23.51	23.54	23.53	23.56
Census Tract Percentage of Decennial Median Family Income	97.29	47.33	95.16	30.22
Number of Loans		9,090		7,885
Number of Lenders		1,531		1,418

Source: Authors' calculations.

References

- Board of Governors of the Federal Reserve System. 1994a. *Federal Reserve Bulletin*, 80 (June): 483-507.
- _____. 1994b. *Federal Reserve Bulletin*, 80 (October): A18.
- _____. 1994c. *Federal Reserve Bulletin*, 80 (November): 859-881.
- Crews, Amy D., Robert M. Dunsky, and James R. Follain. 1995. "What We Know about Multifamily Mortgage Originations," Report submitted to the U.S. Department of Housing and Urban Development, Washington, DC: HUD, October.
- Dunsky, Robert M., James R. Follain, and Jan Ondrich. 1995. "An Alternative Methodology to Estimate the Volume of Multifamily Mortgage Originations," Report submitted to the U.S. Department of Housing and Urban Development, Washington, DC: HUD, October.
- Greene, William H. 1993. *Econometric Analysis*, Second ed. New York: Macmillan.
- ICF, Inc. 1994. "Evaluation of Design and Implementation of the Gross Flows Survey of Mortgage Lending Activity," Report submitted to U.S. Department of Housing and Urban Development, Washington, DC: HUD, July.
- U.S. Department of Housing and Urban Development, Office of Financial Management. 1995a. *Annual Gross Flows of Long-Term Mortgage Loans, 1993*, Washington, D.C.: HUD, April.
- _____. 1995b. "Multifamily Home Mortgage Originations," Task Order.

OCCASIONAL PAPER SERIES (\$5.00 each)

(complete list available upon request)

161.

Douglas Holtz-Eakin and Amy Ellen Schwartz

Infrastructure in a Structural model of Economic Growth,
June 1993, 27 pp.

162.

John Yinger

*Cash in Your Face: The Cost of Racial and Ethnic
Discrimination in Housing*, July 1993, 39 pp.

163.

William Duncombe, Jerry Miner and John Ruggiero

*Analysis of Scale Economies and Technical Efficiency in
New York Public Schools*, August 1993, 171 pp.

164.

Stephen L. Ross

*Economic Development Policy and Resident Welfare: An
Urban Economic Perspective*, September 1993, 38 pp.

165.

John Ruggiero

*Nonparametric Estimation of Cost Efficiency in the Public
Sector: With an Application to New York State School
Districts*, June 1994, 27 pp.

166.

James R. Follain and Orawin T. Velz

*Incorporating the Number of Existing Home Sales into a
Structural Model of the Market for Owner-Occupied
Housing*, August 1994, 36 pp.

167.

Jan Ondrich, Stephen L. Ross, and John Yinger

Measuring the Incidence of Discrimination, December 1994,
41 pp.

168.

Dixie M. Blackley and James R. Follain

*In Search of Empirical Evidence that Links Rent and User
Cost*, January 1995, 32 pp.

169.

**Robert Carroll, Douglas Holtz-Eakin, Mark Rider, and
Harvey S. Rosen**

Income Taxes and Entrepreneurs' Use of Labor, September
1995, 26 pp.

170.

Stephen Ross and John Yinger

*Sorting and Voting: A Review of the Literature on Urban
Public Finance*, September 1995, 78 pp.

171.

Robert M. Dunskey, James R. Follain, and Jan Ondrich

*Using Recurrence Probabilities to Estimate the Volume of
Multifamily Mortgage Originations*, November 1995, 21 pp.

172.

Amy D. Crews, Robert M. Dunskey and James R. Follain

*Estimating the Volume of Multifamily Mortgage Originations
by Commercial Banks Using the Survey of Mortgage
Lending Activity and the Home Mortgage Disclosure Act
Data*, November 1995, 34 pp.

Publications Officer
Center for Policy Research
The Maxwell School - Syracuse University
426 Eggers Hall
Syracuse, New York 13244-1090

Please send me _____ copies of Occasional Paper No. _____ at \$5.00 each.

NAME _____

ADDRESS _____

CITY, STATE, ZIP CODE _____

Postage is included. For air mail delivery to points outside the United States or Canada add \$5.00 per publication. Make checks payable to Center for Policy Research. Payment **MUST** accompany orders.