

The Prevalence and Impact of Misstated Incomes on Mortgage Loan Applications

McKinley L. Blackburn, University of South Carolina
Todd Vermilyea, Federal Reserve Bank of Philadelphia

Current Abstract

Misstatement of income on mortgage loan applications (the “liar-loan” problem) is thought to have been a contributor to the boom and bust of mortgage markets. We provide nationwide measurements that reflect the degree to which incomes on mid-2000 home-purchase mortgage loan applications were overstated relative to the actual incomes of mortgage applicants. Our results suggest a substantial degree of income overstatement in 2005, one consistent with the average mortgage application overstating income by 20-25 percent. We find the tendency to misstate income was influenced by securitization markets. We find no evidence that income overstatement played a role in subsequent foreclosures.

Anticipated Major Revisions to the Paper

1. The paper compares of income reports for new mortgage holders from two different sources – income reported on mortgage applications (from the HMDA) and income reported in a national survey (the American Housing Survey). The idea is to consider the importance of the liar loan problem in mortgage loans by examining whether the difference in these two income reports increased during the mid-2000s housing boom. The approach requires matching across the HMDA and AHS data, and the current version does that matching on the basis of time and MSA (using averages within MSAs for each quarter to make these comparisons). We have decided to redo the matching to take into account the reported loan amount – that is, for each owner of a newly purchased home in the AHS, we attempt to find a match (in practice, we find a set of matches) based on time, MSA, and the reported loan amount. This should provide a more accurate assessment of the extent of income misstatement.
2. The liar loan problem should only be relevant for a fraction of the mortgage loans made in the mid 2000s. This suggests that the distribution of income differences (between the HMDA and AHS) should become more varied and skewed in the mid 2000s, compared to earlier years. Given the revised matching procedure discussed above, we should be able to estimate quantile regressions to see if the impact was largely concentrated at the upper end of the distribution.
3. It has been suggested by Shiller that the housing boom was associated with the most rapid house price increases occurring at the lower end of the house price distribution, fueled by the subprime mortgage boom. We will now be able to look for corroborating evidence, as we can examine whether the mid-2000s income differences vary with the size of the loan amount.

4. At the time the current version of the paper was written, the latest data we could get from the AHS ended in 2005 (there was no 2006 survey). We now have the AHS from 2007, so we are able to extend the analysis to include some loans made at the end of 2006 and the beginning of 2007.

5. We may also change the variable we examine when we look for an impact on loan performance. We currently use a foreclosure measure, but given many of the unique responses of mortgage lenders to troubled mortgages (with loan terms renegotiated rather than foreclosure being pursued) it may be appropriate to use a measure of delinquencies in payment rather than foreclosures.

I'll append the current version of the paper, as the first three sections of the paper should be quite similar, and the basic empirical approach would be unchanged other than the more detailed level for the matching.

The Prevalence and Impact of Misstated Incomes on Mortgage Loan Applications*

McKinley L. Blackburn
Department of Economics
University of South Carolina
Columbia, SC 29208
(803) 777-4931
blackbrn@gmail.com

Todd Vermilyea
Federal Reserve Bank of Philadelphia
Ten Independence Mall
Philadelphia, PA 19106
(215) 574-4125
todd.vermilyea@phil.frb.org

September 2009

JEL Codes: G01, G21

Keywords: mortgage lending; liar loans; housing crisis

Abstract

Misstatement of income on mortgage loan applications (the “liar-loan” problem) is thought to have been a contributor to the boom and bust of mortgage markets. We provide nationwide measurements that reflect the degree to which incomes on mid-2000 home-purchase mortgage loan applications were overstated relative to the actual incomes of mortgage applicants. Our results suggest a substantial degree of income overstatement in 2005, one consistent with the average mortgage application overstating income by 20-25 percent. We find the tendency to misstate income was influenced by securitization markets. We find no evidence that income overstatement played a role in subsequent foreclosures.

*Helpful comments were provided by seminar participants at the Federal Reserve Bank of Philadelphia. We thank Andrew Kish for outstanding research assistance. The views expressed in this paper do not necessarily represent those of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

Try these handy steps to get SISA findings ...

(3) If you do not get [approval], try resubmitting with slightly higher income. Inch it up \$500 to see if you can get the findings you want. Do the same for assets.

It's super easy! Give it a try!

Internal Memo Circulated at JPMorgan Chase, as reported in *The Oregonian*

I. Introduction

After a number of years of rapid house-price appreciation, the third quarter of 2006 saw the beginning of a major decline in prices, suggesting that the mid-2000s housing boom may have been fed by a speculative bubble.¹ Accompanying the fall in house prices has been an increase in foreclosure rates. A major part of this boom-and-bust episode seems to have been over-lending to individuals unprepared to make payments without substantial increases in home equity. With hindsight, both lenders and borrowers entered into contracts that now seem excessively risky.

One of the explanations that is commonly offered for the increased rates of lending in the mid 2000s is a lack of diligence in documenting income on mortgage loan applications by lending institutions. The beginning quote is from a memo from that time, and refers to attempts by loan officers to get Chase's automated underwriting software to approve "stated income/stated assets" applications, applications that allowed income and assets to be stated by the applicant without verification. The mid-2000s saw an increased use of "low-doc" or "no-doc" lending in which the traditional verification processes regarding income sources were no longer part of the loan application process. Historically, these type loans were marketed to a small number of high-income individuals who may have been self-employed or had highly variable

¹ The second quarter of 2006 was the peak for the S&P/Case-Shiller national house price index. This price index had fallen 20 percent from its peak by the third quarter of 2008.

income, but over the 2000s this characterization appears to have changed dramatically. The prevalence of this type of lending is now thought to have given scope for applicants (and their loan officers or brokers) to massage income levels on applications so as to meet standards required in underwriting software. While stories of this type of activity have been noted, no academic study has clearly documented the prevalence or importance of income misstatement in the boom period for conventional mortgage lending.

Our study provides nationwide measurements that reflect the degree to which incomes in mid-2000 home-purchase mortgage loan applications were overstated relative to the true incomes of mortgage applicants. We do so by comparing reports on incomes of mortgage applicants from two different data sources. One data source – the Home Mortgage Disclosure Act data – allows us to measure incomes as reported on actual home mortgage applications. The second data source – the American Housing Survey – provides measures of incomes of new-home purchasers. Data from both sources are examined from the period 1992-2005. Our findings suggest that, while reports of income between these two sources do differ in any year, this difference is stable over time. The primary exception is for 2005, in which there was an increase in the reported incomes on mortgage loan applications relative to those reported in the housing survey.

We are able to construct measures of the degree of income overstatement across MSAs. In what follows, we examine potential MSA-level lending characteristics that might have helped contribute to income overstatement. We also develop simple models of foreclosure rates during 2007, in which income misstatement is allowed to be a potential explanation. While we do find a simple correlation between income overstatement and higher foreclosure rates, this correlation does not hold up when other factors are incorporated.

II. Mortgage Loans in the 2000s

A. Subprime and Alt-A Loans

Prior to the housing boom, the mortgage market was dominated by “conforming” home-purchase loans that met certain credit, income, and loan-limit guidelines. One advantage of conforming loans is the ability for resale to one of the government-sponsored enterprises (GSEs) – namely, Fannie Mae and Freddie Mac. An important component of meeting the GSE guidelines was documentation on the applicant’s employment, income, and debts. As a result, before 2000 the large majority of home-purchase mortgage loans were “full-doc” loans with a thorough investigation of the applicant’s debt and income situation. A small minority of loans were “low-doc” or “no-doc,” with the usual explanation for these type loans being that the lender was reasonably assured of the borrower’s capacity to repay the loan without this documentation.

During the run-up in house prices, the mortgage market saw an important weakening in the dominance of loans that met conforming guidelines. Non-conforming loans – consisting of jumbo, subprime, and Alt-A loans -- all became more prevalent as the decade continued. Borrowers for jumbo loans typically meet the “prime” standards for being purchased by the GSEs, but the purchase value of the property exceeds the limit imposed by the GSEs (this limit was \$417,000 in 2006). Although this limit was increased in the mid 2000s, it did not keep pace with house price appreciation in many markets, and so increasingly limited the ability of loans to meet conformability standards in high-price markets. More important were the increases in subprime and Alt-A loans. As shown in Table 1, subprime loans grew from roughly 9 percent of mortgage loan value in 2001 to 24 percent in 2006, while Alt-A loans grew from 3 percent in 2001 to 16 percent in 2006. The rate of increase in the importance of this type lending was

largely concentrated in the 2004-2005 period, with Alt-A loans in particular increasing six-fold in their importance over this two-year period.

Both subprime and Alt-A loans fail to meet the traditional conforming standards of the GSEs, though for different reasons. Subprime loans are typically targeted towards borrowers with poor credit histories. However, the 2000s saw an increased use of subprime mortgages to finance borrowers with somewhat better credit scores than in the past, but who were attempting to finance purchases that would leave the mortgage with a high loan-to-value ratio, or a high debt-to-income ratio (see Foote, et al., 2008). The desire to avoid full-documentation requirements was also noted as an increasingly common motivator for subprime mortgages in this decade, although this desire was perhaps a greater motivation for the growth in Alt-A loans. Borrowers on Alt-A loans typically have good credit histories (though they may still be less than perfect), but typically desire nontraditional loan or underwriting terms. Novel payment structures – such as interest-only or negatively amortizing payments – were common for Alt-A loans, and this characteristic combined with less-than-full-doc requirements likely allowed the purchases of homes by owners that would not have occurred under conforming standards. As reported in Ashcraft and Scheurmann (2008), 65 percent of Alt-A loans were less than full-doc in 2001, with this percentage growing to 84 percent by 2006. By comparison, only 28 percent of subprime loans were less than full-doc in 2001, this percentage increasing to 42 percent by 2006.

The distinction between subprime and Alt-A mortgages is not an official designation. By the mid-2000s, the large majority of mortgage loans originated were packaged with other loans for sale to private investors in securitized form. For prime loans, this had been the case for many years, with most of the mortgage-backed securities issued by the GSEs. Private issuers of securitized loans were less common for subprime loans and, in particular, for Alt-A loans in the

early part of the 2000s. However, by 2005, 74 percent of subprime loan origination value was securitized, and 87 percent of Alt-A loans (by comparison, 82 percent of prime loans were securitized). As argued by Ashcraft and Schuermann (2008), several sources of friction potentially arise between financial market participants in the securitization of nonconforming loans, leading to a substantial disconnect between the motives of the borrower and originator and the desires of the investor who ends up holding the security. This disconnect was perhaps enhanced by the growing tendency for less-than-full-doc loans to go through mortgage brokers rather than retail lenders (Green, 2008). As a result, by 2005 investors were holding securities that they may have mistakenly felt were almost risk-free, with the underlying assets consisting of poorly underwritten nonprime loans that were at substantial risk if housing prices were to fall.²

Shiller (2008) places much of the blame for the mid-2000s housing bubble on subprime lending, citing a high growth in house prices at the lower end of the house-price distribution in San Francisco in the mid-2000s as supporting evidence. Along this line, Table 2 reports percentage changes in the S&P/Case-Shiller house price index for large MSAs (for which indices are provided) over the 2000-2005 period, where the changes are broken down by whether the house sold was originally in the bottom tier (bottom third), middle tier, or upper tier for that MSA. With only two exceptions (Las Vegas and Phoenix), growth in the bottom tier was more rapid than growth in the middle or top tiers. And in many cases – notably Boston, New York, San Diego, and San Francisco – the growth at the bottom tier was almost twice as fast as at the top. Shiller argues that this would be expected if the growth in subprime loans was a major contributing factor to the housing boom. It might also suggest that the growth in less-than-full-

² Geradi et al. (2008) provide evidence that market analysts in the mid 2000s appreciated the consequences of a nationwide reversal of house price appreciation, but rated the likelihood of this occurrence as very low.

doc loans could have been concentrated in individuals at the bottom end of the income and house-price distribution.

B. Performance of Loans Made During the Housing Bubble

Former Federal Reserve Bank Governor Randall Kroszner has pointed to the prevalence of “stated-income” loans as a “clear culprit” in the rise in mortgage problems.³ Mayer et al. (2009) note that the growth in no-doc and low-doc loans was indicative of the slackening in underwriting standards, and that default rates increases have been particularly high for these type loans. Sanders (2008) provides additional evidence that banks were reporting a weakening of underwriting standards in the 2004-2006 period. He also notes that delinquency rates for subprime mortgages were falling in the 2000s until 2006 (delinquency rates were steady for prime loans). Green (2008) finds falling delinquency rates from 2001 to 2005 among both low-doc and full-doc subprime ARM loans (for loans with a set of fixed credit characteristics). However, he notes that the delinquency rates started to rise with the 2005 vintage, though only for the low-doc subprime loans. It would seem that this fall in delinquencies before 2005 could have contributed to a growing confidence in the subprime part of the market, and hence an increased desire to make loans to borrowers that in an earlier period might have been considered too risky. Contributing to this growing confidence was the rapid house price appreciation occurring in the mid-2000s, as an expected increase in house prices could make attractive a loan that would be considered unprofitable in a period of stable expected prices. The appearance of exotic mortgage products – for example, Alt-A loans that allowed interest-only (or even less-

³ In remarks to the National Association of Hispanic Real Estate Professionals Legislative Conference in March 2008, Governor Kroszner noted that “When we looked closely at why so many borrowers had mortgages that they struggled to repay so soon after taking out the loan, the prevalence of ‘stated-income’ lending was a clear culprit. Substantial anecdotal evidence indicates that failing to verify income invited fraud. Moreover, when we looked at the loan-level data we saw a clear correlation between ‘low-doc’ or ‘no-doc’ lending and performance problems, particularly early payment defaults.”

than-interest) payments – became increasingly common in the mortgage market, again indicative of a belief that house price appreciation could decrease the loan-to-value ratio without there being direct contributions to principal through loan payments.⁴

In explaining the rise in mortgage default rates since 2006, Haughwout et al. (2008) point to changes in the economy as the most important factor, with falls in house prices as the dominant explanation.⁵ However, a reduction in underwriting standards also appears to have played a role. Foote et al. (2008) note that, although FICO scores were actually rising among subprime borrowers in the early-to-mid 2000s, on net the creditworthiness of these loans was still falling due to decreases in requirements for loan-to-value, debt-to-income, and documentation status. They argue that a combination of forces may have been important to the severity of the decline – the reason why the turnaround in house prices had such a large effect on mortgage default was that the 2005 vintage of subprime loans had been underwritten with reduced standards (such as high loan-to-value ratios) that put these loans at substantially higher risk if the house-price boom were to unravel.

In the mid-2000s, it became common knowledge in the mortgage industry that “stated-income” and other less-than-full-doc loans opened up the possibility of a substantial misstatement of income, and the term “liar loan” was adopted by those analysts that were concerned about this possibility (see, for example, Harney, 2005). This overstatement of incomes may have led to loans being made to borrowers that would not have been considered creditworthy if true income had been utilized. If this were the case then, the resulting fall in

⁴ A belief of many borrowers in the subprime market may have been that an increase in the house price would enable a refinance at a lower rate, as the loan-to-value ratio would be increased at this time. Mortgage payments that might have been unsustainable for the full 30 or more years of a loan would then be refinanced to a lower, sustainable payment.

⁵ Haughwout et al. only attempt to explain rising default rates within subprime, or within Alt-A loans. Of course, one factor in the increase in default rates is the shift in lending towards more nonprime loans, which have always had higher default rates (see Sanders, 2008).

house prices left these applications at particular risk, and so the rate of foreclosures should be expected to be particularly high where liar loans were more prevalent. An alternative view is that participants in mortgage loan markets were aware of the liar loan problem and took this into account in their lending decisions, implying that an increase in liar loans need not have led to higher foreclosure rates.

An open question is why the mortgage industry may have condoned this increase in less-than-full-doc lending. The evidence seems to suggest lending institutions were weakening their underwriting standards in many dimensions of mortgage lending (for example, credit score, loan to value requirements, debt-to-income requirements) at the same time. It may have been that many lending institutions desired to weaken standards to an even greater degree than their official policies seemed to indicate. Rather than officially set a less-restrictive debt-to-income threshold (or abolish this requirement altogether), lenders might prefer to maintain official standards while increasing their use of stated income loans. In packaging these loans to sell to investors, it may have looked better if mortgages met normal underwriting standards, even if this appearance was based on unsubstantiated characteristics.

No academic study has previously examined system-wide evidence for the U.S. on the existence and magnitude of potential income overstatement in home mortgage loans. There have been a small number of previous compilations of limited evidence on income misstatement, but these generally use small or select samples. An interesting example is cited by Gimein (2008), referring to Steven Krystofiak's testimony to the Federal Reserve that 60 percent of 100 stated-income loans that he examined appeared to overstate income by at least 50 percent. In the following, we combine evidence from two large nationwide datasets to assess the relevance of income overstatement on mortgage loans in the mid-2000s. We also examine whether markets

that seem to exhibit a tendency for income overstatement also had higher foreclosure rates two years later. To our knowledge, ours is the first attempt to use U.S. data on incomes across MSAs to measure the importance and impact of potential income overstatement in this period.

III. Data Sources

A. Incomes from the HMDA and AHS

As our desire is to measure the extent to which reported incomes reflected true incomes among loan borrowers, we require a data set for reported incomes on mortgage applications. A nationally representative census for such applications is available from the Home Mortgage Disclosure Act (HMDA) data.⁶ All mortgage lending institutions with offices in metropolitan areas are required to report data on home-mortgage applications, and the coverage of applications is almost universal. While the HMDA requirement has existed since 1975, the level of information required was substantially expanded starting with the 1990 HMDA. Beginning with that year, lending institutions were required to provide application-level data on applicant income, loan request, sex and ethnicity of the applicant(s), property location, and the decision outcome on the application. While the data are not suitable for reliably assessing the creditworthiness of applicants – for instance, no information is collected on the property value or the applicant’s debt or credit history – it is ideal for our purpose of measuring incomes of originated loan applications across metropolitan statistical areas (MSAs).

We primarily use the HMDA data to construct measures of average incomes as stated on loan applications across homeowners with newly-purchased homes within specific MSAs. In so doing, we restrict our sample of HMDA applications to those that were conventional mortgages

⁶ Studies that have examined recent problems in the mortgage market have primarily relied on individual loan data available from FirstAmerican CoreLogic’s Loan Performance data, which carries extensive information on loan characteristics and outcomes for securitized loans. However, this is a select data set (only securitized loans are included), and so is not as useful as the HMDA data for our purposes.

that were: (1) originated; (2) home-purchase in a 1-to-4 family dwelling; and (3) owner-occupied. A small number of loans report income as zero, so we treat this as a missing value and exclude these applications in our calculations.⁷ In cleaning the data, we also excluded applications with requested loan amounts of zero, or action dates that were either before the application date, or more than a year after that date. The enhanced data reporting requirements in HMDA started in 1990, but as lending institutions were just getting used to the requirements the data tend to be poor at first. Therefore, we begin our analysis with applications originated in 1992.

Our comparison data set for incomes of new homeowners is constructed from the American Housing Survey. The AHS is conducted by the U.S. Bureau of the Census, and data are available for every year from 1992 to 2005 (except 2000). The data are collected at a household level, and the primary purpose is to obtain information on the quality of housing. The AHS does not allow us to match individual households to their loan application data in the HMDA, so we instead compare reported income averages in the AHS and the HMDA measured at the MSA level. The AHS data actually go back to 1973, but as income only began to be reported in the HMDA data after the 1990 revisions, we restrict our analysis of the AHS to the surveys beginning in 1992.

The AHS actually consists of two different survey mechanisms. National surveys incorporate a nationwide sampling scheme (of roughly 55,000 households), while metro surveys focus on about 14 MSAs in a given year (with at least 3,200 households sampled within each MSA in a particular year). In recent years, national surveys have been conducted in odd-numbered years and metro surveys in even-numbered, though neither survey was conducted in

⁷ We also top code incomes at an upper limit of \$10 million dollars, as a similar top code is imposed on the comparison data in the AHS.

2000 or 2006. Also, metro surveys were conducted every year before 1996 in our sample period. Both surveys are conducted in a 3-to-7 month period in the middle of the year. The MSAs included in the metro survey are cycled from a list of 47 large MSAs, with some slight variation in this list over time. As a result, we can calculate average incomes for a sample of MSAs, but the particular MSAs in the sample will vary over time, with the number of MSAs represented considerably higher when using the AHS from an odd-numbered year. As the final year in the data – 2005 – was odd-numbered, we have a reasonably-sized sample of MSAs to examine in the year in which we anticipate income misstatement to be most evident.

Incomes in the HMDA data are for owners in newly-purchased homes, so we also identify homeowners in the AHS that moved with a new mortgage in the last year. We further restrict AHS respondents to those with conventional mortgages living in a one-unit building or mobile home.⁸ This leaves us with a total of 10,755 AHS household-level observations from 1992 to 2005 to use in calculating AHS specific means. The Census Bureau provides sampling weights for households that allow for MSA-level estimates, and we use these weights in calculating our AHS income means.⁹ We should note that in many cases there are only a handful of observations from an MSA in a given quarter to use in constructing the AHS means (there are obviously always a correspondingly large number of HMDA observations for an MSA). As a result, we use the number of observations from the AHS as a weight in all regressions in which the AHS mean is a dependent variable. The lack of a large AHS sample within each MSA does not pose a serious problem for our regressions that use the AHS mean as part of the dependent

⁸ The similar restriction in the HMDA was to individuals in dwellings with less than five families, which is not perfect. However, we do not anticipate this slight difference will bias comparisons over time.

⁹ In a limited number of cases we have observations from both the national and metro surveys for a given MSA in a given year. In these cases, we calculate weighted means for the MSA in each of the surveys, and then form a weighted mean of those means based on the number of underlying households for that MSA in each of the surveys.

variable, as this random error in the dependent variable should serve to decrease the precision of the regression estimates but not bias them in any sense.

A key component of matching means from the AHS and HMDA data is matching the date of the mortgage origination. The HMDA data provide the date of mortgage approval, which we use to construct quarterly/MSA level HMDA averages. The AHS provides month of the move, and variation in the time of the survey and the date of the move generally allows us to find observations in all quarters (though some quarters end up being more heavily sampled than others). We then match AHS average income for moves reported in a given quarter to HMDA average income on applications originated in that quarter. The AHS income is measured at the time of the survey while the HMDA income is measured at the time of the loan application, so we use the nationwide rate of personal income growth between the move and survey date to index AHS incomes to the move date. This leaves a small discrepancy between the application date in the HMDA and the move date in the AHS, although this difference is on average likely to be small. All incomes are adjusted to be stated in the value of the dollar in 2005, using the personal consumption expenditure deflator. In the end, we are able to construct 1,827 quarterly MSA-level observations in which comparisons can be made between the HMDA and the AHS, with 127 unique MSAs represented in the sample.

There are reasons to anticipate that incomes reported in HMDA will differ from incomes in the AHS. Income in the HMDA is the total gross annual income that the lending institution relied upon in making their credit decision. Thus, the exact measure may vary from institution to institution, due to variations in policies and procedures for underwriting mortgages.¹⁰ The AHS provides several income definitions that could be used to measure gross annual income. Our

¹⁰ For example, one institution might include income from the rental of a room or garage in the new home, while another does not. Treatment of investment income might also vary across institutions.

results are based on an income definition that consists of the joint annual wage and salary income for a married couple (if the householder is a partner in a marriage) or the householder's wage and salary income for an unmarried householder.¹¹ We recognize that these measures do not represent a perfect comparison, and so anticipate a bias in using AHS income as a measure in what might be expected for HMDA income in any given MSA/quarter (even if there were no income misstatement in the HMDA). However, we also expect that this bias will tend to be constant over time.

Figure 1 presents a time-series plot of our basic averages from the two data sources. As average incomes do vary across MSAs, and the MSA composition of the sample does vary over time, we present means that remove MSA-related variation in incomes in each survey.¹² The lack of an AHS survey in 2000 leaves us unable to construct AHS measures for the first two quarters of 2000. In any given quarter, there is a clear tendency for the HMDA income to be higher than the AHS income. We suspect this tendency has to do with differences in income definitions across the surveys. Over time, this difference appears relatively stable; with perhaps some indication of tendency for the two income means to get closer together as move to the early 2000s. In fact, the AHS mean income is only slightly below the HMDA mean in the first quarter of 2003. However, after this quarter, means from the two surveys appear to diverge, and by 2005 the average difference in the reported means is on the order of \$35,000. While we will explore the statistical significance and possible explanation of this difference in the next section, we find the pattern interesting: the HMDA income is relatively constant in the 2003-2005 period, but

¹¹ Several other income definitions could be constructed – for example, wage and salary income of all individuals in the household, or total income of just the householder. Mortgage decisions are usually based on steady sources of income that might best be reflected in wage and salary income. In practice, our basic results are not sensitive to the particular definition of income we use.

¹² In particular, we estimate separate regressions for HMDA and AHS incomes using the MSA/quarter observations. These regressions include separate dummies for each quarter and each MSA in the data. We then predict income for each quarter for each survey, using a fixed MSA distribution across the two surveys.

there is a substantial decline in the AHS income – consistent with a shift of the distribution of actual incomes toward individuals with low incomes without any accompanying shift reflected in the distribution of incomes stated on mortgage applications.

B. Other Variables Used in the Analysis

In our statistical analyses, we make use of several MSA-level factors that might be associated with a tendency to income overstatement or the propensity for foreclosure. First, we construct a measure of subprime lending in the MSA in each quarter using the HMDA data. While HMDA does not identify any particular loan as subprime, we are able to use information on the lending institution as an indication of whether or not a loan was made by a lender that tends to specialize in subprime loans. Since 1993, the Department of Housing and Urban Development has maintained an annually updated list of lending institutions that tend to specialize in subprime and manufactured-home loans.¹³ Inclusion on the list is largely determined by the lending institution possessing certain characteristics in the HMDA data (low origination rates, low selling to GSEs, and a high refinance composition), although communication with the institution is also utilized in making this assessment. While this measure is not perfect -- counting prime loans by subprime specialists as subprime, while missing subprime loans made by other institutions -- it perhaps provides a reliable indicator of the prevalence of subprime loans in a given MSA. Mayer and Pence (2008) show that this measure performs very similarly to two other subprime measures they consider, when they estimate regressions that attempt to explain subprime prevalence across MSAs.¹⁴ Our particular measure

¹³ We use the 1993 list to assign subprime indicators for 1992. Also, beginning with 2004, lenders who specialize in manufactured-home were no longer included in the list, while some Alt-A lenders were included if their APR characteristics tended to be similar to subprime lenders.

¹⁴ These other measures are based either on Loan Performance data – which explicitly indicates subprime in the securitization process, but is then limited only to securitized loans – or interest rate data in the HMDA which itself may suffer from reporting problems (especially with teaser rates on many subprime loans). Neither measure would be available for the same length of time as the HUD measure, in any case.

is the percentage of originated loans that are from a subprime lender, using the sample restrictions on loans mentioned above.

Second, we also construct annual measures of house-price changes within MSAs. While data on median home prices is available from the National Association of Realtors, we do not use this measure as it has no control for changes in the quality of housing. We instead use the house-price index constructed by the Office of Federal Housing Enterprise Oversight (OFHEO), based on changes in house prices in repeat sales or refinances. In particular, we take four-quarter moving averages of the OFHEO index in any given MSA, and then calculate the percentage change in that moving average from one year prior.¹⁵ One limitation of the OFEHO index is that it is only constructed for houses in which mortgages are sold to either Fannie Mae or Freddie Mac (see Calhoun, 1996). The S&P/Case-Shiller index is constructed in a similar manner – in fact, OFHEO based their index on the methods developed by Case and Shiller. The S&P/Case-Shiller index is not restricted to GSE loans in its coverage, but it also has a much more limited offering for price indices at the MSA level.

The HMDA data also contain a variable that indicates whether the loan was sold by the end of the year in which the action was taken. We measure the percentage of originated loans in an MSA that were sold to a GSE, namely Fannie Mae, Freddie Mac, Ginnie Mae, or the FmHA. We also measure the percentage held by the institution. For a sale to be reported in the HMDA, the sale must occur before the end of the calendar year. Loans made near the end of the year will tend to have lower percentages sold within that same year, and we worry that we will miss the tendency for loans to be sold in the latter quarters. As a result, we calculate the percentages sold

¹⁵ For example, for the 3rd quarter of 2005, we calculate a simple moving average of the index in that quarter and the three previous quarters. A similar moving average is constructed for the 3rd quarter of 2004. Then, the house-price percentage change for the third quarter of 2005 is the percentage change in these two moving averages.

(or held) using loans with approval dates in the first quarter of the year, and use these measures for all quarters in that calendar year.

Descriptive statistics and estimates for the trends of these variables are reported in Table 3. The trend estimates are from a regression that also allows for MSA effects, and the regressions also include a dummy variable for the observation being from 2005 (similar to the regressions discussed in the next section). The average measure of annual house price appreciation in the data over this period was 5 percent with an upward trend over the period, but 2005 still experienced a 4 percentage-point higher rate of appreciation than would be predicted by this trend. While the percent held by the originating bank was falling over our sample period, the year 2005 did not represent an outlier from this trend. The reverse is the case for the percentage sold to GSEs, as this was rather stable over time, but dropped rather dramatically in 2005 – a 10 percentage-point drop which is more than 25 percent below the usual average for this variable. While merely suggestive, the GSE numbers are consistent with an increased tendency for lending institutions to fail to verify incomes fully, because there was no intention to resell the loans to entities where this failure would be an important issue. The percentage originated by subprime lenders was increasing over the 1991-2005 period, but was nonetheless considerably above trend in 2005 – the estimated deviation from trend is roughly 50 percent of the average value for this percentage over our sample period.

A final set of variables that we make use of are two measures of default and foreclosure activity in 2007 within an MSA. Our data on foreclosure is obtained from Realtytrac, which tabulates the number of homes in different foreclosure states. We consider two aggregated foreclosure states in our measure. Our primary focus is on a *pre-foreclosure* count, consisting of homes that have either been given a notice of default, or a *lis pendens*. In these cases, a

homeowner has been notified that there is a limited amount of time in which to make payment before eviction or a final lawsuit is initiated. In pre-foreclosure, the homeowner still has possession and the right to resell the home. We also estimate models for a *final-foreclosure* count that consists of homes in which the borrower has lost possession of the home, a notice of trustee's sale or foreclosure sale has been filed, or the property has become "real estate owned" by the lender. For both measures, we use a Census Bureau measure of the number of housing units in the MSA in 2007 to express the variable in rate form as a percentage of the number of total housing units that are in the indicated foreclosure state.

IV. Empirical Results on the Extent of Income Misstatement

A. Empirical Model and Estimation Approach

For a varying sample of MSAs over time, we have observations on the average income of new home buyers from two different surveys. Our expectation is that data from the HMDA source will tend to show an overstatement of income for new homeowners relative to data from the AHS source, especially in the mid-2000s period in which nonprime became more prevalent. In this section, we develop an empirical model that allows us to test if the difference in repeated incomes from the two surveys increased in the final part of the 1992-2005 period.

As discussed earlier, there are reasons to anticipate that average incomes between the two surveys will differ for a given MSA even without the increase in income overstatement generally thought to have occurred in the mid 2000s. As such, we want to develop a statistical model that allows for a difference in income (between the surveys) that can vary both across MSAs at a point in time, and within an MSA over time. To do so, we assume that average income (y_{smt}) of new homeowners from a particular data source income is determined by:

$$y_{smt} = \tau_{sm} + \eta_{st} + \beta'x_{smt} + \varepsilon_{smt} \quad (1)$$

where s is a subscript indicator for the survey (AHS or HMDA), m indicates the MSA, and t indicates the quarter. The vector x_{smt} consists of measured aspects of the new homeowner population, in MSA m at time t , that are characteristic of new home buyers in survey s . The equation allows the MSA (τ_{sm}) and time (η_{st}) effects to vary between the surveys, but assumes that the coefficients on x are the same in the two surveys.¹⁶

Our intention is to characterize the difference in reports between the two surveys, so our sample construction is dictated by the ability to measure income for a particular MSA in a particular quarter using the AHS. With this sample, we can construct a matching sample of similar observations from the HMDA data. Given these observations, we difference equation (1) across the two surveys (HMDA values minus AHS values) to provide

$$\Delta y_{mt} = \Delta \tau_m + \Delta \eta_t + \beta' \Delta x_{mt} + \Delta \varepsilon_{mt} . \quad (2)$$

In this equation, the difference in average incomes across surveys is the dependent variable. Our focus is on the values for $\Delta \eta_t$, which reflect differences across the surveys in time effects for income. While the equation could be estimated with a complete set of quarter effects, this does not provide a frame of reference with which to argue that the last quarters of the sample differed in this income difference compared to the early part of the sample. So we instead restrict the temporal path of these effects by initially assuming a possible linear trend for this survey difference, but one that allows deviations from trend in certain quarters. In particular, we assume

$$\Delta \eta_t = \gamma_0 + \gamma_1 t + \gamma_2' QS_t \quad (3)$$

where QS is a vector of dummies for “suspect” quarters that might differ from the early years, and γ_2 measures deviations from the trend in those suspect quarters. One limitation of this

¹⁶ This is relaxed if the demographic measure is constructed differently in the two surveys.

approach is the necessity of assuming a functional form for the trend, so we do experiment with generalizing this assumption in our analysis.

Equation (3) can be substituted into the one above to provide our estimating equation

$$\Delta y_{mt} = \Delta \tau_m + \gamma_0 + \gamma_1 t + \gamma_2' QS_t + \beta' \Delta x_{mt} + \Delta \varepsilon_{mt} .$$

This linear equation can be estimated by a fixed-effects type estimator (incorporating MSA effects), so as to identify the primary parameter of interest γ_2 . The error term in the equation ($\Delta \varepsilon_{mt}$) is likely to be correlated over time for observations from the same MSA, but this potential correlation is allowed for by using standard errors that are robust to arbitrary correlation at the MSA level. In practice, we estimate these equations weighting by the number of households used in constructing the AHS income average, as MSAs with more households in the AHS should have a more accurately measured dependent variable and thus a smaller variance for the error term.

B. Results

The results from estimating our income-difference model are reported in Table 4, employing a variety of specifications that explore the importance of the linear trend assumption as well as the incorporation of demographic controls. Our basic finding -- 2005 is a year with a dramatic overstatement of income in the HMDA data -- is robust to this variation in model specification.

Specification (1) of Table 4 estimates a very simple model that omits the trend and demographic controls (though still including MSA and quarter-of-year effects). The “suspect” quarter dummies are represented by quarter dummies specific to the first three quarters of 2005 (the last 3 quarters in our data). A test that all three of these coefficients are equal to zero is easily rejected (with p-value of 0.003), while both the first and third quarter effects are

individually statistically significant. The estimated effects are quite sizeable – for instance, the estimated coefficient of 12.4 for the first quarter of 2005 implies that the HMDA/AHS difference in income increases by \$12,400 relative to the average difference in the years 1991-2004. This increase is roughly a 20 percent overstatement of the average reported AHS income in those years.

When the trend and demographic controls are added to the equation (in specification 2), the estimated effects for 2005 continue to be strongly statistically significant as a group. The estimated trend effect is negative, leading to the estimated quarter effects for 2005 – which are deviations from this trend – increasing in value. However, the trend effect itself is not statistically significant. The estimated demographic controls are as expected, reflecting lower average incomes for blacks and Hispanics relative to whites (although the estimated negative effect for Asian is somewhat surprising). The controls for marital status differ across the surveys, as we do not have direct measures of marital status in the HMDA data, and can only use the percent of applications for which there was “no coapplicant” as a rough proxy (which is generally not statistically significant). For the AHS, we can measure both marital status and the presence of one adult in the household, and the estimated coefficient on the percent unmarried in the AHS has the expected positive sign (suggesting lower AHS incomes for those unmarried). Even though no trend in the difference was suggested by specification (2), we still explored whether this may be masking some nonlinear trend pattern by including trend squared in the regression (see specification 3), with no suggestion of a trend nor impact on the estimated 2005 effects.¹⁷

¹⁷ We also experimented with spline functions for the trend, again finding no suggestion of any trend in the difference.

The estimated effects for the three 2005 quarter dummies in specification (2) are quite similar in magnitude, and a test that they are all equal has a p-value of 0.60. In specification (4), we collapse the three 2005 quarter dummies into a single 2005 year dummy, providing an estimated income overstatement of roughly \$15,000 for that year. To explore the extent to which 2005 is a special year, we then added dummies for 2004 (in specification 5) and then a dummy for the period beginning in 2001. Both of these specifications continue to suggest that there was something special about 2005 relative to the years preceding. The estimated effect for 2004 is also positive, though smaller in value than the 2005 effect as well as being statistically insignificant. Adding a dummy for the post-2000 period does not change the conclusions. In later specifications, we will use a single 2005 dummy rather than separate quarter effects as this facilitates our use of interactive specifications.

The time series for average incomes presented in Figure 1 suggested that the major change in the 2005 numbers had to do with average income in the AHS falling, with a potential rise in HMDA incomes being less important. To pinpoint this more specifically, we also estimated separate (but similar) regressions in which either the AHS income level, or the HMDA income level, were used as the dependent variable. With HMDA income, we obtain a positive estimated effect (standard error) for the 2005 dummy of 2.1 (1.1), suggesting some upward deviation trend for that year. With AHS income as the dependent variable, the estimated 2005 dummy effect is equal to -15.2 (3.8). As suggested by Figure 1, the increased disparity in incomes between the HMDA and the AHS is largely driven by a statistically significant fall in the actual incomes of buyers of newly-purchased home. This suggests that the “liar loan” phenomenon was less one of average homebuyers overstating their incomes so as to be able to afford a better home, and more related to lower-income individuals overstating incomes to

become eligible for loans by appearing more like the usual homebuyer population. This result is also consistent with the argument of Shiller (2008) that the housing boom was predominately a low-end boom.

C. MSA Characteristics and Income Overstatement.

While the estimated models in Table 4 clearly suggest that was tendency to income overstatement on loan applications in 2005, it does not help identify the particular MSAs in which this overstatement might have been more prevalent. As discussed earlier, the tendency to overstate incomes is often thought to have been related to whether or not the MSA market was a “hot market” at that time, as characterized by the tendency for house prices to increase in that market. The characterization of an MSA as a fast-appreciating market was made by examining the percentage change in the OFHEO house price index. Given the higher underwriting and documentation requirements, we expect loans sold to GSEs to display a lower tendency for income overstatement. A similar expectation holds for loans held by the bank, which we anticipate for asymmetric information reasons will be the ones the bank find less risky than the market does. Income overstatement may have been particularly symptomatic among subprime lenders in 2005, so we also examine whether income overstatement is a function of the percentage of loans made by subprime lenders.¹⁸

We also estimated models that allow us to study whether income overstatement was a more severe problem in some markets than others related to these market characteristics. These models were generalizations of the previous models to incorporate interactions of the market characteristics with the 2005 dummy. The first specification of Table 5 repeats specification (4) of Table 4 for the smaller sample available for these regressions, with little consequence for the

¹⁸ Ideally, we would also include a measure for the percentage of loans made by Alt-A lenders. Unfortunately, a list of Alt-A lenders is not generally available for the years under consideration.

estimated 2005 effect.¹⁹ Specification (2) simply adds the non-interacted version of the market characteristics – in fact, the evidence does not suggest that any of these characteristics have a long-standing relationship with the income difference between the HMDA and the AHS. This could be indicative of the fact that over the large majority of our sample period, this income difference does not reflect a tendency to intentionally misstate incomes in the HMDA. Specification (3) adds the house-price appreciation interaction (removing the other market characteristics), and while the sign of the coefficient is as expected – income overstatement in 2005 appears more severe in markets with higher appreciation – the estimated effect is not statistically significant. If we remove this variable and add in the purchase variables (specification 4), we find statistically significant evidence that income overstatement in 2005 was less of a problem in markets where the percentage resold to GSEs was higher. On the other hand, there is no suggestion that income overstatement was any less of a problem among loans held by the originating bank compared to loans sold to private institutions. Specification (5) incorporates the subprime lending interaction (removing the other market characteristics) and the estimated coefficient (which just misses statistical significance at the 5 percent level) is consistent with income overstatement being more prevalent when more lending is by subprime lenders.

The basic results for all three characteristics are as we would expect, though the statistical confidence is weak for the house-price measure. In specification (6), we incorporate all of these factors to see the robustness of these findings. Evidence for income overstatement in 2005 being a function of rapid house-price appreciation is further weakened in this specification, as the coefficient estimate actually becomes negative (though still statistically insignificant). The

¹⁹ We lose 165 quarter/MSA observations because of difficulties in matching MSA definitions in the AHS/HMDA data and the OFHEO house-price data.

estimated subprime effect remains positive, but it is considerably smaller and statistically insignificant. The one robust finding is for the GSE interaction, which continues to suggest that income overstatement was less characteristic of markets where a large share of loans was resold to GSEs. The estimated effect is also considerably large. For instance, the almost 10 percentage-point fall below trend in 2005 for this variable (see Table 3) is estimated to have led to an increase in the average income difference by roughly \$18,500 – more than fully explaining the estimated increase in overstatement of \$14,000 for 2005 we identified in specification (1) of the table. By comparison, the estimated effect of the increase above trend in subprime lending in 2005 would be to raise income overstatement by less than \$700, suggesting that any effect was only minor by comparison.

V. Impact of Income Misstatement on Loan Performance

Our regressions in section IV suggested that income overstatement was a characteristic of the general housing market in 2005, with income overstatement particularly prevalent in markets with a low tendency to engage in traditional reselling of mortgages to GSEs. The occurrence of income misrepresentation leads to the natural question of whether this has had apparent effects on subsequent performance of loans originated in these markets. As noted earlier, loan performance has seriously deteriorated in the years following 2005, so we wish to examine the role that income overstatement and other measures of market characteristics may have played in this deterioration.

We do not have performance measures in either the AHS or HMDA data, and obviously cannot tie performance to income overstatement at the loan level. But we can construct average measures of income overstatement in 2005 by MSA using these data. In particular, we estimate a generalization of our basic income-difference model (specification 4 of Table 4) that includes a

full set of interactions of the 2005 dummy with the MSA dummies, thereby allowing us to identify an average income difference in 2005 for each MSA. The primary loan performance measure we use is the “pre-foreclosure rate” (described earlier in section III), which can be measured for 68 of the 72 MSAs included in the AHS sample for 2005. In 2007, the average value for the pre-foreclosure rate was 1.1 percent, ranging from 0.002 percent to 6.9 percent across the MSAs in our sample.

Several recent papers have estimated models for foreclosure rates, with many focused on foreclosures in the recent period of housing-market decline (see Immergluck and Smith, 2005; Grover et al., 2008; Haughwout et al., 2008; Richter, 2008; Sherlund, 2008). Two basic characteristics of a market are focused on in these models as potential explanation for an increase in foreclosure rates: one, a fall in house prices can lead to an increase in negative-equity situations for borrowers; or, two, economic conditions can worsen leading to problems in borrowers making regular mortgage payments. The basic creditworthiness of the mortgagers in a market can also be important. Bad loans typically go bad in the first few years of the mortgage, so we analyze the extent to which pre-foreclosure in 2007 depends on initial conditions in 2005, and changes in those conditions from 2005 to 2007. Changes in house prices are measured using percentage changes in the annual average of the OFHEO house-price index for the MSA. The MSA-level unemployment rate is incorporated as a measure of economic conditions, both for 2007 and 2005. While we do not have extensive measures of creditworthiness of new mortgage holders in the MSA, we do incorporate the subprime percentage variable for 2005 as a “quality of borrower” measure. The percentage of loans sold to a GSE is another potential quality measure, as is the MSA-specific income overstatement variable discussed above. As a

foreclosure rate is a probability that tends to be close to zero, a logistic functional form was used in estimation. In particular, we assume the true foreclosure rate (r) follows the form

$$r = \frac{e^{\beta'x_i}}{1 + e^{\beta'x_i}}$$

and estimate this model with a generalized-linear-model estimator assuming a binomial likelihood function.²⁰

Estimates of several specifications of this model using the pre-foreclosure rate as the dependent variable are reported in Table 6. A simple model in which pre-foreclosure rates are determined as a function of the measure of income overstatement in 2005 does suggest some connection, as the coefficient estimate in specification (1) is positive and statistically significant. The size of the estimated effect is reasonably large, as a one standard deviation increase in the overstatement measure would be predicted to increase foreclosure rates by roughly 35 percent.²¹ Controls for economic conditions and changes in house prices (specification 2) lessen the estimated size of this impact, and cause it to be statistically insignificant. Specifications that also incorporate controls for the type of loan (specification 4) also leave the estimated effect of income overstatement on foreclosure rates as quite small and statistically insignificant. There are two interpretations that might be given to this finding. One is that the tendency to overstate incomes in certain MSAs in 2005 did not ultimately affect loan performance, because at the time loan decisions were being made the mortgage industry fully appreciated that this was occurring. Alternatively, there may be a problem with measurement error in our income misstatement

²⁰ This follows the suggestion of Papke and Wooldridge (1996) in estimating fractional-response models. The advantage of this approach over the usual estimator that uses OLS with the log-odds ratio for r as the dependent variable is that it avoids potential retransformation problems in interpreting the coefficient estimates.

²¹ An easy interpretation to coefficients in logit models can be given when the underlying probabilities are very small. The standard formula for a marginal effect in a logit model is $\partial p / \partial x_j = \beta_j p (1 - p)$, so a percentage impact on the probability is $(\partial p / p) / \partial x_j = \beta_j (1 - p)$. When p is very small, $(1-p)$ is close to 1 and the percentage effect becomes approximately β_j .

variable, as these measures at the MSA level are often based on a very small number of income values from the AHS for many MSAs.²² The fact that significant correlations can be uncovered in the simpler specifications of Table 6, however, suggests that this problem does not lead to the variable being completely meaningless in its representation.²³

We explored a number of extensions and alterations to the model to help in refining the nature of the results or considering its robustness.²⁴ As noted in section II, it is likely that a combination of factors might be necessary to cause particularly high foreclosure rates, so we included as independent variables interactions between house price appreciation and the change in unemployment, between appreciation and the subprime percentage, and between subprime and the change in unemployment. While the estimated coefficients were generally in the expected sign, they were never statistically significant. We included a measure of the state's level of predatory lending protection (taken from Bostic, et al, 2008), expecting that more stringent protection might be associated with less foreclosure. However, the coefficient estimate was small and statistically insignificant. Previous studies have often found that the minority percentages in the local market affect default rates, so we included the percentage black and percentage Hispanic in 2005 as additional controls. Neither variable was statistically significant, as long as the percentage of subprime loans was included in the equation. We changed our foreclosure rate measure so as to be expressed as a percentage of 2005 loans in the MSA, rather than the number

²² Given it is reasonable to argue that this measurement error follows the classical assumptions, our coefficient estimates in the income-overstatement equations in the previous section should still be unbiased (as the error is in the dependent variable in those equations). However, it could cause a bias toward zero when income misstatement is used as an independent variable, as it is in Table 5.

²³ We also tried including a dummy variable for overstatement, with the dummy equal to one if the MSA had an estimated income overstatement of \$12,000 or more (about half are in this category). Our hope was that this variable might be less susceptible to measurement error. These results also did not support that income overstatement was associated with higher foreclosure rates.

²⁴ For brevity, these results are not reported here, but are available from the authors upon request.

of total housing units.²⁵ Finally, we used our final foreclosure measure (discussed above) in place of the pre-foreclosure rate. In all of these re-estimations, the pattern for the coefficient estimate on income overstatement was similar – the estimate was small and statistically insignificant when other MSA characteristics are included in the regression.

VI. Conclusions

We examine MSA-level reports on average incomes of buyers of newly-purchased homes in two different data sets – one reflecting income on originated mortgage applications, the other census-surveyed income among newly-purchased-home buyers. We substantiate a considerable difference in income between the two sources at the height of the mid-2000s housing boom. Our results suggest a substantial degree of income overstatement on average in 2005, one consistent with the average mortgage application overstating income by 20-25 percent. This income overstatement is largely characterized by a drop in the *actual* incomes of recent buyers, rather than in an increase in the *reported* incomes on accepted home-mortgage applications. It appears that the “liar loan” behavior was largely associated with the overstating of incomes on applications of buyers who had actual incomes lower than those typically acceptable in the mortgage market. This finding is consistent with Shiller’s contention that the housing bubble was greatest in the lowest tier of the housing market. Our results suggest that banks may have increased their willingness to serve the population segment most likely to buy these houses without the appearance of having lowered standards to an undue degree.

We also examined aspects of the MSA-level mortgage market that tended to be associated with the tendency to overstate income. Our estimates suggest that it was the potential sale to a GSE – such as Fannie Mae or Freddie Mac – that represented the most important factor,

²⁵ Another possible measure would be foreclosures as percentage of total outstanding mortgages. We find the use of 2005 mortgages more appropriate, given that more recent vintage loans are those at most risk of default.

with MSAs in which a large number of mortgage loans were sold to a GSE having the lowest tendency to overstate income. The most likely explanation for this connection was that requirements for purchase by a GSE generally required full documentation of the borrower's financial situation.

Has the tendency to overstate incomes on mortgage applications led to higher foreclosure rates for those loans in subsequent years? While the nature of our data limit the extent to which we can examine this question, we do find a positive simple correlation between the degree of income overstatement in an MSA and the subsequent tendency for houses in that MSA to enter an initial foreclosure preceding in 2007. However, this correlation appears to be explained by other factors affecting foreclosure rates. In particular, we find that both economic conditions in the MSA, and the degree of subprime lending in 2005, are important correlates with the pre-foreclosure rate.

References

- Ashcraft, Adam B. and Til Schuermann. 2008. "Understanding the Securitization of Subprime Mortgage Credit." *Federal Reserve Bank of New York Staff Reports*, No. 318.
- Bostic, Raphael W., Kathleen C. Engel, Patricia A. McCoy, Anthony Pennington-Cross, and Susan M. Wachter. 2008. "State and Local Anti-Predatory Lending Laws: The Effect of Legal Enforcement Mechanisms." *Journal of Economics and Business*. 60(1): 47-66.
- Calhoun, Charles A. 1996. "OFHEO House Price Indices: HPI Technical Description." Office of Federal Housing Enterprise Oversight Working Paper.
- Case, Karl E., Robert J. Shiller and Allan N. Weiss. 1996. "Mortgage Default Risk and Real Estate Prices: The Use of Index-Based Futures and Options in Real Estate." *Journal of Housing Research*. 7(2): 243-258.
- Foote, Christopher L., Kristopher Gerardi, Lorenz Goette, and Paul S. Willen. 2008. "Just the Facts: An Initial Analysis of Subprime's Role in the Housing Crisis." *Journal of Housing Economics*. 17: 291-305.
- Gerardi, Kristopher, Andreas Lehnert, Shane M. Sherlund, and Paul Willen. 2008. "Making Sense of the Subprime Crisis." *Federal Reserve Bank of Boston Public Policy Discussion Paper*, No. 09-1.
- Gimein, Mark. 2008. "Inside the Liar's Loan: How the Mortgage Industry Nurtured Deceit." *Slate*. April 24.
- Green, Richard K. 2008. "Imperfect Information and the Housing Finance Crisis: A Descriptive Overview." *Journal of Housing Economics*. 17: 262-271.
- Grover, Michael, Laura Smith, and Richard M. Todd. 2008. "Targeting Foreclosure Interventions: An Analysis of Neighborhood Characteristics Associated with High Foreclosure Rates in Two Minnesota Counties." *Journal of Economics and Business*. 60: 91-109.
- Harney, Kenneth R. 2005. "Agency Sounds Warning on Stated-Income and Interest-Only Mortgages." *Realty Times*. January 10.
- Haughwout, Andrew, Richard Peach, and Joseph Tracy. 2008. "Juvenile Delinquent Mortgages: Bad Credit or Bad Economy?" *Journal of Urban Economics*. 64: 246-257.
- Immergluck, Dan and Geoff Smith. 2005. "Measuring the Effect of Subprime Lending on Neighborhood Foreclosures: Evidence from Chicago." *Urban Affairs Review*. 40: 362-289.
- Mayer, Christopher and Karen Pence. 2008. "Subprime Mortgages: What, Where, and to Whom?" NBER Working Paper no. 14083.

Mayer, Christopher, Karen Pence, and Shane M. Sherlund. 2009. "The Rise in Mortgage Defaults." *Journal of Economic Perspectives*. 23: 27-50.

Papke, Leslie E. and Jeffrey M. Wooldridge. 1996. "Econometric Models for Fractional Response Variables with an Application to 401(k) Plan Participation Rates." *Journal of Applied Econometrics*. 11: 619-632.

Richter, Francisca. 2008. "An Analysis of Foreclosure Rate Differentials in Soft Markets." Federal Reserve Bank of Cleveland Working Paper No. 08-11.

Sanders, Anthony. 2008. "The Subprime Crisis and its Role in the Financial Crisis." *Journal of Housing Economics*. 17: 254-261.

Sherlund, Shane M. 2008. "The Past, Present, and Future of Subprime Mortgages." FEDS Working Paper No. 2008-63, Federal Reserve Board, Washington, D.C.

Table 1 Percentage of Origination Value from Mortgage-Loan Types, 2001-2006			
<i>Year</i>	<i>Subprime</i>	<i>Alt-A</i>	<i>Prime and Jumbo</i>
2001	9.0	2.8	88.2
2002	8.3	2.5	89.2
2003	8.9	2.3	88.8
2004	20.8	7.7	71.5
2005	22.7	13.8	63.5
2006	23.8	15.9	60.3
<i>Source: Ashcraft and Schuermann (2008)</i>			

Table 2 Percent Change in Home Prices within MSAs by Home Price Tier, Jan. 2000 to June 2005			
<i>MSA</i>	<i>Low Tier</i>	<i>Middle Tier</i>	<i>Upper Tier</i>
Atlanta	34.1%	25.8%	28.5%
Boston	114.9	86.2	67.7
Chicago	64.1	60.2	47.7
Cleveland	31.9	22.7	21.4
Denver	37.7	35.6	35.0
Las Vegas	118.4	124.4	117.0
Los Angeles	181.9	152.2	119.9
Miami	164.8	143.2	120.9
Minneapolis	80.6	63.5	59.1
New York	128.8	103.5	81.8
Phoenix	84.4	88.4	91.1
Portland	52.7	47.9	47.1
San Diego	188.1	149.5	120.7
San Francisco	156.5	115.2	82.5
Seattle	57.6	53.1	48.4
Tampa	122.2	98.8	91.1
Washington, D.C.	160.8	144.9	117.6
<i>Note:</i> Percent changes are based on the tier classifications available in the November 2008 S&P/Case-Shiller Tiered Price Indices. The three tiers separate houses in the area into three tiers (each with roughly one-third of homes sold) based on the starting price of the houses sold.			

Table 3 Estimated Models for Time Series Characteristics of MSA Lending-Market Variables				
<i>Independent Variables</i>	<i>Dependent Variable</i>			
	House Price Appreciation	Percent Held by Originating Bank	Percent Sold to Govt. Sponsored Entities	Percent Originated by Subprime Lenders
2005 dummy	4.21** (1.23)	-0.001 (0.007)	-0.097** (0.010)	0.032** (0.006)
Trend	0.129** (0.022)	-0.0035** (0.0002)	-0.0013 (0.003)	0.0029** (0.0001)
R ²	0.37	0.72	0.51	0.82
Mean (Std. Dev.)	4.95 (4.74)	0.295 (0.111)	0.377 (0.101)	0.066 (0.063)
<i>Note:</i> All models are estimated by OLS, with standard errors (in parentheses) robust to heteroskedasticity and any arbitrary correlation in the error terms over time within an MSA. A full set of MSA dummies and quarter of the year dummies are included in each specification. The sample includes 1,662 MSA/quarter observations.				

Table 4 Estimated Models of Income Differences between HMDA and the AHS						
<i>Independent Variable</i>	<i>Specification</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
2005-1 st Qtr.	12.4** (4.1)	14.9** (4.5)	14.6** (4.6)			
2005-2 nd Qtr.	6.6 (7.9)	12.3 (7.2)	12.0 (7.4)			
2005-3 rd Qtr.	20.6** (7.5)	24.3* (10.3)	24.0* (10.3)			
2005 dummy				14.8** (4.0)	18.0** (4.6)	17.9** (4.6)
2004 dummy					4.8 (2.9)	4.7 (2.8)
Post-2000 dummy						-1.1 (3.1)
Trend		-0.092 (0.109)	-0.136 (0.171)	-0.093 (0.109)	-0.042 (0.110)	-0.024 (0.134)
Trend Squared			0.001 (0.003)			
Black Difference		-18.9** (4.6)	-18.9** (4.6)	-18.9** (4.6)	-18.7** (4.6)	-18.7** (4.6)
Asian Difference		-13.6* (6.1)	-13.7* (6.1)	-13.5* (6.1)	-13.7* (6.1)	-13.7* (6.1)
Hispanic Difference		-24.6** (4.6)	-24.7** (4.6)	-24.7** (4.6)	-24.7** (4.6)	-24.7** (4.6)
Other-Race Difference		-9.5 (14.7)	-9.1 (14.6)	-9.3 (14.7)	-3.2 (14.3)	-2.3 (14.4)
No Coappl. (HMDA)		-25.8 (18.8)	-29.6 (19.3)	-25.6 (18.8)	-45.7 (23.5)	-44.0* (22.1)
Unmarried (AHS)		35.4** (3.5)	35.4** (3.5)	35.4** (3.5)	35.3** (3.5)	35.4** (3.5)
One Adult Hhhd. (AHS)		-5.7 (4.4)	-5.7 (4.4)	-5.7 (4.4)	-5.5 (4.4)	-5.6 (4.4)
P-value: No 2005 Effect	0.003	0.001	0.002	0.000	0.000	0.000
R ²	0.20	0.30	0.30	0.30	0.30	0.30
<i>Note:</i> All models are estimated by WLS, with the number of households in the AHS serving as weights. Standard errors (in parentheses) are robust to heteroskedasticity and any arbitrary correlation in the error terms over time within an MSA. A full set of MSA dummies and quarter of the year dummies are included in each specification. The dependent variable is expressed in thousands of year-2000 dollars, corrected for inflation using the personal consumption expenditure deflator. The sample includes 1,827 MSA/quarter observations. * indicates significance at the 0.05 level, ** at the 0.01 level						

Table 5 Estimated Models of HMDA-AHS Income Differences, with Interactions						
<i>Independent Variables</i>	<i>Specification</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
2005 dummy	14.0** (3.8)	16.4** (4.7)	10.4* (5.0)	-4.3 (10.8)	0.3 (8.4)	-0.9 (7.8)
Trend	-0.050 (0.122)	0.007 (0.132)	-0.033 (0.136)	0.067 (0.124)	-0.038 (0.135)	-0.015 (0.148)
House Price Appreciation		0.07 (0.20)	-0.047 (0.213)			0.047 (0.223)
Appreciation*2005			0.485 (0.473)			-0.888 (0.613)
Percentage Held		29.0 (19.1)		22.2 (15.7)		30.0 (19.5)
Held*2005				-37.9 (63.7)		-86.4 (80.9)
Percentage GSE		23.6 (16.7)		20.3 (14.8)		25.9 (16.9)
GSE*2005				-150.4* (71.9)		-191.4* (78.6)
Percentage Subprime		33.3 (32.0)			6.5 (26.8)	31.9 (32.6)
Subprime*2005					123.0 (72.8)	20.6 (91.2)
P-value: No 2005 Effects	0.000	0.000	0.004	0.001	0.001	0.001
R ²	0.30	0.30	0.30	0.30	0.30	0.30
<i>Note:</i> See the notes to Table 1. A full set of MSA dummies and quarter of the year dummies is included in each specification, along with the other variables in specification (4) of Table 4. The sample includes 1,662 MSA/quarter observations.						

Table 6 Estimated Fractional Response Models for Pre-Foreclosure Rates in 2007			
<i>Independent Variable</i>	<i>Specification</i>		
	(1)	(2)	(3)
Income Overstatement in 2005	0.0096* (0.0042)	0.0007 (0.0028)	0.0010 (0.0016)
Unemployment Rate in 2007		66.7 (39.7)	84.6** (24.8)
Unemployment Rate in 2005		-64.0 (37.5)	-93.8** (24.2)
Percent Change in OFHEO House Price Index, 2006 to 2007		-12.8** (3.9)	-0.7 (2.5)
Percent of Loans Subprime in 2005			8.9** (1.4)
Percent of Loans GSE in 2005			0.6 (1.6)
<i>Note:</i> Models are estimated as fractional response models with a logit functional form and a Bernoulli distributional assumption. Standard errors are robust to distributional assumptions and heteroskedasticity. The mean (standard deviation) of the dependent variable is 0.011 (0.015). The sample consists of 68 MSAs.			

Figure 1: Time Series of Corrected Means from HMDA and AHS

