NBER WORKING PAPER SERIES

ASSET QUALITY MISREPRESENTATION BY FINANCIAL INTERMEDIARIES: EVIDENCE FROM RMBS MARKET

Tomasz Piskorski
Amit Seru
James Witkin

Working Paper 18843
http://www.nber.org/papers/w18843

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2013

We are grateful to Equifax and BlackBox Logic for their data which were invaluable for the analysis in this paper. We thank Charlie Calomiris, John Cochrane, Gene Fama, Chris Mayer, Lasse Pedersen, Amir Sufi, and Luigi Zingales as well as seminar participants at Columbia Business School for valuable comments. We thank Ing-Haw Cheng and Andrew Ellul for graciously sharing their data on incentive compensation and the risk management index. All errors are our own. Piskorski thanks the Paul Milstein Center for Real Estate at Columbia Business School for financial support. Seru thanks the Initiative on Global Markets at Booth for financial support. Contact Authors: Tomasz Piskorski (tp2252@columbia.edu) and Amit Seru (amit.seru@chicagobooth.edu) The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Asset Quality Misrepresentation by Financial Intermediaries: Evidence from RMBS Market
Tomasz Piskorski, Amit Seru, and James Witkin
NBER Working Paper No. 18843
February 2013
JEL No. G14,G21,G24,G28,K22

ABSTRACT

We contend that buyers received false information about the true quality of assets in contractual disclosures by intermediaries during the sale of mortgages in the $2 trillion non-agency market. We construct two measures of misrepresentation of asset quality – misreported occupancy status of borrower and misreported second liens – by comparing the characteristics of mortgages disclosed to the investors at the time of sale with actual characteristics of these loans at that time that are available in a dataset matched by a credit bureau. About one out of every ten loans has one of these misrepresentations. These misrepresentations are not likely to be an artifact of matching error between datasets that contain actual characteristics and those that are reported to investors. At least part of this misrepresentation likely occurs within the boundaries of the financial industry (i.e., not by borrowers). The propensity of intermediaries to sell misrepresented loans increased as the housing market boomed. These misrepresentations are costly for investors, as ex post delinquencies of such loans are more than 60% higher when compared with otherwise similar loans. Lenders seem to be partly aware of this risk, charging a higher interest rate on misrepresented loans relative to otherwise similar loans, but the interest rate markup on misrepresented loans does not fully reflect their higher default risk. Using measures of pricing used in the literature, we find no evidence that these misrepresentations were priced in the securities at their issuance. A significant degree of misrepresentation exists across all reputable intermediaries involved in sale of mortgages. The propensity to misrepresent seems to be largely unrelated to measures of incentives for top management, to quality of risk management inside these firms or to regulatory environment in a region. Misrepresentations on just two relatively easy-to-quantify dimensions of asset quality could result in forced repurchases of mortgages by intermediaries up to $160 billion.

Tomasz Piskorski
Columbia Business School
3022 Broadway, Uris Hall 810
New York, NY 10027
tp2252@columbia.edu

Amit Seru
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
amit.seru@chicagobooth.edu

James Witkin
Columbia Business School
jlw2188@columbia.edu

Section I: Introduction

Market rules and regulations that require disclosure of information and prohibit misleading statements on the financial products being manufactured by intermediaries play an important role in the functioning of capital markets (Akerlof, 1970). However, the nature of intermediation has changed dramatically over the past decade–with the introduction of more agents in the supply chain of credit (Loutskina and Strahan, 2009; Keys et al. forthcoming; Nadauld and Sherlund, 2013) --- potentially weakening the ability of existing market arrangements and regulatory oversight in ensuring truthful disclosure of asset quality. This concern has gained momentum in the aftermath of the crisis, with a precipitous decline in the value of supposedly safe securities as well as large investor losses (Acharya, Schnabl, and Suarez, forthcoming).[1] This paper adds to this debate, and more generally to one on the design of market rules and future financial regulation, by quantifying the extent to which buyers may have received false information about the true quality of these assets by sellers of these securities and by exploring factors that may have moderated such behavior by financial intermediaries.

Our study focuses on misrepresentation of asset-quality that consist of securities collateralized by residential mortgages that are originated without government guarantees (non-agency RMBS) -- representing a $2 trillion market in 2007.[2] These misrepresentations are not instances of the usual asymmetric information problem in which the buyers know less than the seller. Rather, we contend that they are instances where, in the process of contractual disclosure by the sellers, buyers received false information on the characteristics of assets.[3]We identify misrepresentations by comparing the characteristics of mortgages in the pool that were *disclosed* to the investors at the time of sale with *actual* characteristics of these loans at the same time and show that such misrepresentations constitute a significant proportion of the loans. We investigate if lenders and investors were aware of such behavior and assess characteristics of intermediaries involved in the sale of assets, such as incentives for managers, that could have moderated this behavior. In doing so we demonstrate the limits of existing market and regulatory arrangements in preventing such behavior in the capital market.

The RMBS securitization process involves aggregating mortgages into loan trusts, either through direct origination or indirect acquisition, and using their underlying cash flows to issue securities. The sale of these securities is organized by underwriters who, as part of this process, must collect and verify information regarding the quality of the underlying collateral backing these securities.

[1] Critics of imposing more regulation argue that reputational concerns of large, well-established financial intermediaries would prevent such violations of investors' rights. In contrast, proponents of increased regulation argue that intermediaries were able to exploit investors despite their reputation (and existing regulation).
[2] See Agarwal and Ho (2007), and Keys et al. (2012), and the Federal Reserve Board March 2009 Statistical Release (http://www.federalreserve.gov/econresdata/releases/mortoutstand/mortoutstand20090331.htm)
[3] Akerlof (1970) also considers settings where quality of goods can be potentially misrepresented by sellers.

The underwriters in this market are large, reputable financial intermediaries who are considered to be more sophisticated than the buyers in this market, which are typically institutional investors such as pension funds, mutual funds, and insurance companies. Thus, given the lighter regulatory oversight relative to public capital markets that are open to retail investors, this market serves as a good laboratory to study whether existing market mechanisms based on implicit and explicit contracting are sufficient to safeguard investors' rights.

In general, several factors make it difficult to perform a systematic study of the nature and determinants of asset misrepresentation. For instance, most studies of corporate fraud define fraud as events in which the rights of the shareholders of a firm are violated by its managers. However, such violations of investors' rights involve firms with varied production technologies and in different time periods, making such comparisons difficult. Moreover, most of these studies focus only on the cases where fraudulent behavior was ex-post discovered. We focus on the RMBS market and study the disclosure activity related to sale of assets by intermediaries, which circumvents some of these challenges. Since residential mortgages are fairly standardized and have a simple structure, comparisons of disclosure activity are relatively easier. In addition, detailed histories of such activity by financial institutions involved in sale of RMBS are available, allowing us to study asset misrepresentation by these firms over time as well as in the cross-section. Finally, our methodology will identify asset misrepresentation regardless of whether or not it is ex-post discovered by the investors.

We start our analysis by combining two sources of data, loan-level data on mortgages from BlackBox with data on consumer credit files from Equifax, to construct two measures of misrepresentation regarding the quality of mortgages backing the RMBS pools. The mortgage-level data include characteristics of loans that were *disclosed* to the investors at the time of asset sale. The consumer credit data, which were not available to investors at the asset sale date, contains the *actual* characteristics of loans at the same time. The matching between the two datasets is done by the credit bureau.

Using the combined dataset we identify two, relatively easy-to-quantify, dimensions of asset quality misrepresentation by intermediaries during the sale of mortgages. The first misrepresentation concerns loans that are reported as being collateralized by owner-occupied properties when in fact these properties were owned by borrowers with a different primary residence (e.g., a property acquired as an investment or as a second home). The second form of misrepresentation concerns loans that are reported as having no other lien when in fact the properties backing the first (senior) mortgage were also financed with a simultaneously originated closed-end second (junior) mortgage. The advantage of looking at this second type of misrepresentation is that it clearly indicates that the distortion of information occurred within the boundaries of the financial industry, as some institutions (e.g., a lender financing a second lien loan) had to be aware of the presence of such higher liens. Our hope in both instances is to

directly identify asset misrepresentations relative to sellers' contractual disclosure. This differentiates our study from the literature that infers the decisions of agents in the supply chain of credit (e.g., screening by lenders) based on outcome variables such as delinquencies (e.g., Keys et al. 2010).

A high-quality match between the dataset that contains actual loan characteristics and that containing characteristics reported to investors is critical for constructing measures of misrepresentation. We conduct extensive analysis that suggests that the misrepresentations we identify are not likely to be an artifact of matching error between the datasets that were merged by the credit bureau. In particular, we follow a conservative approach and conduct our analysis only on loans that were matched across the datasets with highest confidence level by the credit bureau. We confirm the high quality of this merge when we conduct an independent analysis of various data fields and find that, for the vast majority of cases, these fields are the same across the two databases. More importantly, as we will demonstrate, the misrepresented loans we identify strongly predict ex post delinquencies, further confirming our assertion that these loans misreport meaningful information about asset quality. This is not the pattern we find for a placebo test where we use few incidences of likely data errors across datasets to construct a measure of pseudo misrepresentation.[4] Finally, we cross-validate the merge quality using an internal database from a large subprime lender. We find that almost all loans we identified as having second liens that are not disclosed to investors do indeed have such liens reported in the bank's internal data.

Using our measures we find a significant degree of misrepresentation of collateral quality across non-agency RMBS pools. More than 6% of mortgage loans reported for owner-occupied properties were given to borrowers with a different primary residence, while more than 7% of loans (13.6% of loans using a broader definition) stating that a junior lien is not present actually had such a second lien. Alternatively put, more than 27% of loans obtained by non-owner occupants misreported their true purpose and more than 15% of loans with closed-end second liens incorrectly reported no presence of such liens. The propensity of banks to sell loans that misrepresented asset quality increased as the housing market boomed. This pattern is consistent with the model of Povel, Singh, and Winton (2007). Overall, more than 9% of loans had one of these misrepresentations in our data. Note, however, that because we look only at two types of misrepresentations, this number likely constitutes a conservative, lower-bound estimate of the fraction of misrepresented loans.

[4] The placebo test uses the notion that incorrectly merged records should not have such a strong relationship with the subsequent adverse performance of loans. In particular, we focused on the few records in our database for which the loan balance of the first mortgage does not exactly match across the two databases. We find that the subsequent performance of such loans in terms of their default risk is similar to that of loans with perfectly matching balances across the two databases (see Appendix A for more details).

We find that these misrepresentations have significant economic consequences. In particular, loans with misrepresented borrower occupancy status have about a 9.4% higher likelihood of default (90 days past due on payments during the first two years since origination), compared with loans with similar characteristics and where the property was truthfully reported as being the primary residence of the borrower. This implies a more than 60% higher default rate relative to the mean default rate of owner-occupants during our sample period. Similarly, loans with a misrepresented higher lien -- which we find are likely to be fully documented loans -- have about a 10.1% higher likelihood of default compared with loans with similar characteristics and no higher lien. This is again a large increase, about 70%, relative to the mean default rate of loans without higher liens. Thus, our results indicate the same pattern for the two measures, which confirms that the misrepresentations we identified capture economically meaningful information about asset quality. Because of their substantially worse performance, misrepresented loans account for more than 15% of mortgages that defaulted in our sample, a higher share than their proportion in the overall sample (about 10%).

Next, we investigate if misrepresentations on the two dimensions, given that they are associated with higher mortgage defaults, appear to be priced in by lenders and investors. To do this we first assess whether lenders charged higher interest rates for loans with misrepresented quality. We find that mortgages with misrepresented owner occupancy status are charged interest rates that are higher when compared with loans with similar characteristics and where the property was truthfully reported as being the primary residence of the borrower. Similarly, interest rates on loans with misrepresented second liens are generally somewhat higher when compared with loans with similar characteristics and no second lien. Given the increased defaults of these misrepresented loans, this suggests that lenders were partly aware of the higher risk of these loans. Strikingly, however, we find that the interest rate markups on the misrepresented loans are much smaller relative to loans where the property was truthfully disclosed as not being primary residence of the borrower and as having a higher lien. This suggests that relative to prevailing interest pricing of that time, interest rates on misrepresented mortgages did not fully reflect their higher default risk.

We also examine whether pools with a higher incidence of misrepresented assets were perceived to be of lower quality by investors. For that purpose, we investigate the relation between misreporting in a pool and the measures of pricing used in the literature, i.e., imputed average yield of the pool as well as the subordination level that protects its AAA-rated tranches.[5] Using these measures, we find little evidence that investors were compensated for a greater risk of securities involving a higher share of misreported assets. These results suggest that RMBS

[5] As is the case in the literature, we do not have access to data on actual prices paid by the investors at the time of pool issuance (see Keys et al. 2012). Instead, consistent with literature as well as industry practice, we use proxies of prices such as average balance-weighted coupon. These proxies are the same as those used by prior work to investigate whether risk of collateral was reflected in investor prices (e.g., Faltin-Traeger, Johnson, and Mayer 2010; Demiroglu and James, 2012; He, Qian, and Strahan, forthcoming).

investors had to bear a higher risk than they might have perceived based on the contractual disclosure. As a result, investors could argue that the ex ante value of the securities with misrepresented assets that were sold to them was less than the price paid, and truthful disclosure of the characteristics of the assets could have prevented some of their losses. Assuming that our estimates are broadly applicable to the entire stock of outstanding non-agency securitized loans just prior to the crisis, enforcement of contractual guarantees by investors in response to these misrepresentations could result in forced repurchases of mortgages with combined balance of up to $160 billion.[6]

We also investigate the variation in asset misrepresentation across underwriters. We demonstrate that there is substantial heterogeneity in the extent of these misrepresentations across underwriters. The propensity to misrepresent is largely unrelated to measures of incentives for top management and to quality of risk management inside these firms. While misrepresentations appear to be less prevalent among commercial banks, underwriters with more RMBS experience, and underwriters with more high-powered incentives given to its top management and better internal management, none of these associations is statistically significant. Importantly, a significant degree of misrepresentation exists *across all* reputable intermediaries in our sample. Finally, we also find no relation between share of misrepresented loans and leniency of regulatory environment in a region.

Our earlier analysis suggests that lenders were partly aware of the risk of misrepresented loans since it was reflected in the interest rates charged on these loans. In the last part of the paper we examine where in the supply chain of credit--i.e., at the borrower, lender, or underwriter level--these misrepresentations likely took place. This analysis requires significantly richer micro data than are available to us, but we are able to provide some evidence that suggests that part of the misrepresentation occurs at the level of the financial institution. In particular, we use an internal database of a large subprime lender in which we observe the data that were available to the lender as well as data that were disclosed to the investors. To the extent that practices in this subprime lender are representative, our findings suggest that misrepresentation concerning owner-occupancy status was made early in the origination process, possibly by the borrower or broker originating a loan on behalf of the lending institution. In contrast, the lender was aware of the presence of second liens, and hence their misreporting likely occurs later in the supply chain.[7] This last result is also consistent with our earlier evidence that loans misrepresented on the dimension of higher liens are more likely to be fully documented.

Our findings contribute to the debate on the recent crisis. We provide evidence showing that

[6] Asset misrepresentations by intermediaries could induce other costs beyond those directly incurred by the investors. For instance, these misrepresentations could have also led to misallocation of scarce investor capital among the proposed real investment projects in the economy.

[7] Our review of mortgage deeds records from the FirstAmerican database indicates that both first-lien and closely situated second-lien mortgages on a given property were commonly financed by the same lending institution.

investors bought assets that not only proved to be ex post risky but may also have been, at least in part, ex ante misrepresented by financial intermediaries. More broadly, our results suggest that the current market arrangements -- based on reputational concerns and explicit incentives -- may have been insufficient to prevent and eliminate misrepresentations of asset quality in a large capital market. These findings suggest that a critical inspection of the protection of investors in other capital markets, especially those with more passive investors like the high-grade investment debt market, may be warranted. Our findings resonate well with studies that suggest that the existing regulatory framework may have also been insufficient in preventing such behavior by various actors in the supply chain of credit (see Keys et al. 2009; Keys et al. forthcoming, and Agarwal et al. 2012).

Our work most directly relates to the recent empirical literature on the housing and financial crisis.[8] In this literature, our research on the extent and consequences of loan misrepresentation in the mortgage securitization market is closest to studies that attempt to infer decisions of borrowers and financial institutions (e.g., lenders) using data based on outcome variables such as delinquencies (see Keys, Mukherjee, Seru, and Vig (2010), Ben-David (2011)). It is also related to Jiang, Nelson, and Vytlacil (2012) that provides evidence of misrepresentation of income for low documentation borrowers, and Garmaise (2012) that provides evidence that suggests that borrowers overstated their assets and that loans granted to such borrowers had higher default rates. In contrast to these papers, we attempt to directly identify misrepresentation of asset quality rather than infer them from outcomes, and we do so across the main players in the market.

Our analysis is also closely related to recent research by Haughwout et al. (2011) who employ credit bureau data to identify properties acquired by real estate investors. Like our paper, Haughwout et al. also identify some proportion of such buyers as misreporting their intentions to occupy a property. Their focus is on explaining the role such buyers played in the recent boom and bust in the housing market.[9] In contrast, our focus is on identifying the consequences of such misrepresentations on additional defaults and on understanding whether these misrepresentations were priced by investors and lenders as well as which underwriter characteristics relate to such behavior.

Our work is also related to a large theoretical and empirical literature in accounting and finance that investigates various aspects of corporate fraud (e.g., Richardson et al. 2002; Burns and Kedia 2006; Kedia and Philippon 2009; Povel, Singh, and Winton, 2007; Efendi et al. 2007;

[8] See Mian and Sufi 2009, forthcoming; Mayer et al. 2009; Loutskina and Strahan, 2009 and 2010; Campbell et al., forthcoming; Keys, Mukherjee, Seru, and Vig 2010; Keys, Seru, and Vig, forthcoming; Rajan, Seru, and Vig, forthcoming; Piskorski, Seru, and Vig 2010; Melzer 2010; Berndt, Hollifield, and Sandås, 2010; Mian and Sufi 2011; Agarwal et al. 2011; Demyanyk and Van Hemert 2011; Demiroglu and James (2012), Nadauld and Sherlund, 2013; Purnanandam, 2012; Acharya, Schnabl, and Suarez, forthcoming and He, Qian, and Strahan, forthcoming.

[9] See also related research by Chinco and Mayer (2012) who use deeds records to identify distant real estate investors and study their role in fueling the recent boom in house prices.

Dyck et al., forthcoming; Wang, Winton, and Yu 2010; Wang forthcoming). It is also connected to the work in law and economics that focuses on securities fraud litigation (e.g., Choi et al., 2009 Griffin et al. 2001). Finally, our analysis is also related to recent studies that explore the determinants of risk taking undertaken by financial intermediaries (e.g., Cheng et al., 2010 and Ellul and Yerramilli, forthcoming). We contribute to these areas by providing systematic evidence on the characteristics of asset misrepresentation -- in the cross-section of financial intermediaries and across time -- in a large debt market.

Section II: Information Disclosure on Collateral Quality of RMBS

The vast majority of mortgages originated in the United States are not held by the banks that originated them but are instead securitized and sold as securities to investors. In this paper we focus on residential mortgage-backed securities (RMBS) that are collateralized by mortgage loans that are originated without government guarantees (non-agency RMBS). This sector was a significant portion of the overall mortgage market and reached more than $2 trillion of securities outstanding (Keys, Piskorski, Seru, and Vig, forthcoming).

In our analysis we are concerned about the misrepresentation of collateral backing RMBS. When a pool of mortgages is sold to investors, the underlying characteristics and quality of these mortgages are disclosed in the prospectus of the pool in the section related to "representations and warranties." This disclosure usually contains information about risk-relevant variables, such as loan-to-value ratios of loans in the pool, their interest rates, borrowers' credit scores, and the occupancy status and location of the properties backing the mortgages in the pool. Our analysis focuses on two misrepresentations that our data enable us to identify. The first misrepresentation concerns the occupancy status of the property backing the loan because, as we explain later, such status can impact how much risk is associated with the mortgage. In particular, we are interested in identifying loans reported as being collateralized by owner-occupied properties when in fact these properties were owned by borrowers with a different primary residence (e.g., a property acquired as an investment or as a second home). We now provide an example of a contractual disclosure that refers to asset quality on this dimension -- namely, the prospectus supplement for Series 2006-FF15 states:

> *The prospectus supplement will disclose the aggregate principal balance of Mortgage Loans secured by Mortgaged Properties that are owner-occupied.*[10]

The second case of asset-quality misreporting we consider concerns loans that are reported as having no associated higher liens when in fact the properties backing the first (senior) mortgage were also financed with a simultaneously originated second (junior) mortgage. Prospectus

[10] The prospectus supplements will also usually describe the frequency of loans for second homes or investor properties. For example, the prospectus supplement for the Wells Fargo Mortgage Backed Securities Series 2007–8 states, "Approximately 0.02% (by aggregate unpaid principal balance as of the cut-off date) of the mortgage loans in the mortgage pool are expected to be secured by investor properties."

supplements commonly make statements regarding the total value of all liens on the collateralized property. As we explain, omitting information on the junior mortgages on a property will also understate the true risk associated with the pool collateral. Again, it is useful to provide an example of such contractual disclosure. The prospectus supplement of WFMBS Series 2007-7 states:

> *For purposes of Appendix A, the "Combined Loan-to-Value Ratio" or "CLTV" is the ratio, expressed as a percentage, of (i) the principal amount of the Mortgage Loan at origination plus (a) any junior mortgage encumbering the related Mortgaged Property originated by the Sponsor or of which the Sponsor has knowledge at the time of the origination of the Mortgage Loan or (b) the total available amount of any home equity line of credit originated by the Sponsor or of which the Sponsor has knowledge at the time of the origination of the Mortgage Loan, over (ii) the lesser of (a) the appraised value of the related Mortgaged Property at origination or (b) the sales price for such property.*[11]

As we explain in detail in Section IV, in order to identify such a misrepresentation we use loan-level data concerning the characteristics of mortgages in the pool that were *disclosed* to the investors at the time when the pool was sold. This data contains the most detailed information that is contractually provided to investors by asset sellers and pertains to the quality of loans that back the mortgaged-backed securities. Then, using another proprietary matched dataset that contains information on the *actual* characteristics of these loans, we verify whether the information disclosed to the investors was accurate. Thus, it is important to keep in mind that our analysis is from the perspective of an investor who used detailed loan level information available from the trustee at the time of sale of the pool. We discuss how this assumption could impact our inferences in Section VIII.

The information contractually disclosed to investors allows them to assess the risk of the security. Previous research has shown that mortgages with a higher loan amount relative to property value and loans backed by non-owner-occupied properties are often associated with greater default risk (see Mayer et al. 2009). Thus, both of these misrepresentations could imply that RMBS investors had to bear a higher risk than they might have perceived based on the contractual disclosure. As a result, investors could argue that the ex ante value of the securities with misrepresented collateral that was sold to them was less than the price paid, and full disclosure of the true characteristics of the collateral could have prevented their loss.

Misrepresentation can also have consequences for the party securitizing the mortgages (i.e., the intermediary who does the underwriting of the pool) since such parties are often contractually

[11] Similarly, the prospectus supplement for First Franklin's Series 2006-FF15 reports that "original full Combined Loan-to-Value Ratio reflects the original Loan-to-Value Ratio, including any subordinate liens, whether or not such subordinate liens are owned by the Trust Fund."

liable to guarantee that the underlying collateral in a pool is accurately represented. Of particular relevance are contractual obligations that force the lender or underwriter to repurchase the loan from the securitization trust. For instance, the prospectus supplement for Series OOMC 2005-HE6 of Asset Backed Pass-through Certificates states:

> *If the seller or the originator fails to cure a material breach of its representations and warranties with respect to any mortgage loan in a timely manner, then the seller or the originator would be required to repurchase or substitute the defective mortgage loan.*[12]

It is worth noting that though such repurchase clauses were common in the market during 2005 and 2006, actual enforcement of these contracts was not as strict (see Piskorski et al. 2010). However, the enforcement of these contractual agreements has become more important since the crisis and has been the center of many recent lawsuits filed by investors.

Section III: Data

Our primary dataset links two databases that allow us to construct our measures of asset collateral misrepresentation: (i) loan-level mortgage data collected by BlackBox Logic and (ii) borrower-level credit report information collected by Equifax.

BlackBox is a private company that provides a comprehensive, dynamic dataset with information about twenty-one million privately securitized Subprime, Alt-A, and Prime loans originated after 1999. These loans account for about 90% of all privately securitized mortgages from that period. The BlackBox data, which are obtained from mortgage servicers and securitization trustees, include static information taken at the time of origination, such as the mortgage date and amount, borrower FICO credit score,[13] servicer name, interest rate, term, interest rate type, CLTV, and borrower occupancy status. The BlackBox data also include dynamic data on monthly payments, mortgage balances, and delinquency status. Importantly, this database collects information from trustees of mortgage pools and contains information concerning characteristics of mortgages in the pool that were *disclosed* to the investors at the time when the pool was sold.

The other dataset that we use is from Equifax -- a credit reporting agency that provides monthly data on borrowers' *actual* current credit scores, payments and balances on mortgage and installment debt, and balances and credit utilization for revolving debt (such as credit cards and home-equity lines of credit [HELOCs]). Equifax has recently linked their credit information data

[12] Similarly, the prospectus supplement for Series 2006-FF15 of the First Franklin Mortgage Loan Trust states: "Following the discovery of a breach of any representation or warranty that materially and adversely affects the value of a mortgage loan, or receipt of notice of that breach, the mortgage loan seller or the sponsor will be required to either (1) cure that breach, (2) repurchase the affected mortgage loan from the trust fund or (3) in certain circumstances, substitute another mortgage loan."

[13] The FICO is the most widely known type of consumer credit score, developed by FICO, previously known as Fair Isaac Corporation. The FICO scores range from 350 to 850, with a higher score indicating a more creditworthy borrower. See Keys et al. (2010) for more discussion.

to the BlackBox data using a proprietary match algorithm that merges on more than twenty-five variables (see Appendix A for more details on our data). Consequently, we use the information from Equifax to verify whether information disclosed to the investors -- captured through the BlackBox sample -- was misrepresented.

Overall, we impose three restrictions on the merged BlackBox-Equifax data. First, we limit our dataset to loans originated from 2005 to 2007, years for which we have full data from Equifax. Second, we retain only first liens, identified as loans with the following characteristics in the BlackBox dataset: either (i) a lien type of "first" or (ii) if the lien type is "unknown," an origination LTV of at least fifty-five. The second restriction allows us to include loans of unknown lien type in our sample that are not likely to be subordinate liens. Third, we restrict the sample to include only loans that have the highest Equifax merge-confidence level (Equifax reports a merge-confidence level that ranges from low to high confidence). As we discuss in the next section this step is important because without Equifax data correctly matched with BlackBox data, we would not be able to plausibly construct our misrepresentation measures. After imposing these restrictions, we obtain a base sample that includes approximately 1.9 million loans. As we show in the Appendix A, imposing these restrictions leads to a sample with slightly higher-quality observables than in the overall data.

Section IV: Constructing Measures of Loan Misrepresentation

We construct two measures of asset-quality misrepresentation in RMBS pools, the first of which examines whether a given loan reports the occupancy status of the borrower correctly. In particular, we are interested in examining whether the borrower is reported as being an owner-occupant when in fact the property is not the borrower's primary residence. To construct this measure, we start with our base sample and restrict the sample to include only loans that are reported as owner-occupied in the BlackBox dataset at the time of origination. This gives us a sample of 1,563,223 loans.

Our measure of owner-occupancy misrepresentation is constructed by taking advantage of the mailing address zip code reported to Equifax. The zip code reported to a credit bureau such as Equifax should generally match the property address zip code reported to the trustee of the RMBS for owner-occupied homes (available in the BlackBox dataset). To allow for reporting delays, we follow a very conservative approach and compare the Equifax zip code reported each month over the first year of the loan's life to the property zip code reported to BlackBox at origination. If *none* of these twelve Equifax zip codes match with the BlackBox dataset, we classify this loan's borrower as a misreported non-owner-occupant.

The second measure of asset-quality misrepresentation concerns identifying whether loans that were reported to RMBS investors as having no associated second lien at the time of sale of the

pool did in fact have a second lien. Among other characteristics, the loan-level information disclosed to investors at the time of purchase consists of the combined loan-to-value (CLTV) ratio for all liens on the property at the time of loan origination. If the lender was unable to ascertain this value or was not willing to disclose this information, the loan is usually given a missing CLTV. While it is possible that some of these loans with missing CLTV values were set to be missing due to asset misrepresentation, it is hard to know this with certainty. Therefore, we again take a conservative approach and do not consider these loans to have been misrepresented on this dimension. Instead, we identify loans with no explicitly reported second liens by restricting our sample to loans with non-missing values of CLTV, where the reported CLTV is within 1% of the loan's LTV.[14] For these loans we are confident that the securitized first mortgage (senior loan) was reported to the trustee of RMBS as having only one lien on the property. This yields a sample of 854,959 loans that report no presence of second liens.

We construct our second measure of asset misrepresentation based on the observation that the credit bureau data (Equifax) includes information about other mortgages held by each borrower. We can therefore examine whether loans that are reported to RMBS investors as having no simultaneous second liens do in fact have a second lien reported in the credit bureau data.[15] As in the case of our first measure, because we focus on data with the highest merge quality, we are confident that such instances do in fact reveal second liens that were misreported to RMBS investors.

We must make a judgment call in identifying second liens as either simultaneously or subsequently originated. In this paper, we classify simultaneous second liens as those that appear in the credit bureau data with an origination date within 45 days of that of the first lien. We classify a loan as having a misreported second lien if the loan is reported to investors as having no second lien but does show presence of such a lien in the credit bureau data within this time window. The 45-day window we use allows for reporting delays or potentially imputed origination dates in the Blackbox data. It is also quite unlikely that a borrower would obtain a subsequent second lien on a mortgage within this time period without the lender of the first lien having that information when reporting to the trustees. Regardless, we have experimented with this window and the results reported later are not sensitive to changes in the length of this window.

[14] We have also verified that loans reported to investors as having second liens do indeed have such mortgages reported in the credit bureau data. In fact about 90% of loans that report CLTV greater than LTV has junior loan present in the Equifax database. The remainder likely represents the cases where lenders have not promptly reported their records to the credit bureau data, which we confirmed with the data vendor. If anything, this makes our measure of misrepresentation a conservative estimate since we may not detect some misreported second liens.

[15] We take a conservative approach and throughout most of our analysis we do not consider loans as being misrepresented if reported CLTV by the trustee does not include HELOC balance in its calculation. Despite the fact that prospectuses often include such loans in definition of CLTV (see example in Section II), one could argue that such loans are harder to categorize as being misrepresented mortgage and instead could be perceived as revolving debt similar to credit card debt. The level of misrepresentation is much higher if we consider HELOCs as second liens (see Section V.A) and our other results are robust to such extension of our misrepresentation measure.

IV.A Verifying Quality of Misrepresentation Measures

We need to discuss two issues related to the quality of misrepresentation measures that we have constructed. First, both measures rely on Equifax data being correctly matched with BlackBox data in order to plausibly construct our misrepresentation measures. Thus, we need to ensure that the match between these two datasets, done by the credit bureau, is of high quality. Second, even if the match between the two databases is perfect, only our second measure (misreported second lien) allows us to directly identify asset misrepresentation. For the first measure, which is based on occupancy status, the misrepresentation is inferred. The reason is that we do not observe occupancy status in the Equifax database. Instead, we infer whether or not a property is owner-occupied based on the match between the mailing zip code reported in the credit bureau data and the property zip code reported to the trustee of the RMBS. Thus, we need to verify that the measure based on occupancy status is not mainly reflecting errors made in this inference.

To address the first issue, we restrict our sample to those loans that have the highest merge-confidence level assigned by the credit bureau. We then conduct an independent analysis of the merge quality of this data, checking fields such as dynamic payment history, origination balances, and origination dates. Regardless of whether or not the loans are misrepresented based on our measures, the vast majority of other fields are the same across the two databases (see Data Appendix for more details).[16]

Next, we note that available evidence indicates that loans backed by non-owner-occupied properties and loans with second lines have higher delinquency rates (see Mayer et al. 2009). Hence, if our measures correctly identify the two misrepresentations, we would expect such loans to have a *higher* default pattern compared with loans that are not misrepresented. In fact, as we show in Section V.C, we find that ex post delinquencies of such loans are more than 60% higher when compared with otherwise similar loans. It is worth noting here that we also conducted a placebo test using the notion that incorrectly merged records should not have such a strong relationship with the subsequent adverse performance of loans. In particular, as discussed in Appendix A, we focused on the few records in our database for which the loan balance of the first mortgage does not exactly match across the two databases. The balance of the securitized first mortgage is unlikely to be misreported to investors because servicers verify and report on a monthly basis the outstanding loan amount and payments to the securitization trust; hence, such records may indicate incorrectly merged loans across the two databases. We find that subsequent performance of such loans in terms of their default risk is similar in economic terms to loans with a perfectly matching balance across the two databases. This evidence provides

[16] For example, for the entire sample, the match between payment histories increase monotonically with Equifax's confidence measure, with only 0.3% of the highest confidence loans having different BlackBox and Equifax payment statuses.

further support that our methodology allows us to identify *actual* misrepresentation of asset quality instead of reflecting few incorrectly merged records.

Next, recall that if our sample consists of correctly merged records, we can directly identify whether or not a given loan has an undisclosed second lien. As we discuss in Section VII, we are able to cross-validate the merge quality using an internal database from a large subprime lender in which almost all loans identified by us as having misreported junior liens do indeed have such liens reported in the bank's internal data.[17]

Finally, turning to the second issue, we verify that our method of inferring whether or not a property is owner-occupied is reasonable. In particular, we find that our method of inference matches well with the status reported to the trustee of the RMBS. Among loans reported as being for non-owner-occupied properties, the majority (about two-thirds) have a mailing address zip code of the borrower in the Equifax database that does not match with the property address in the BlackBox database.[18] Similarly, a majority (about 70%) of reported owner-occupants residing in the same zip code have only one first mortgage on the credit file from Equifax. Moreover, our measure of owner-occupancy misreporting also correlates strongly with borrowers that acquired multiple properties and hence are likely to be real estate investors. In particular, a majority (more than 60%) of reported owner-occupants whose property zip code does not match with their credit file zip code do have multiple first mortgages on their credit file, indicating that they acquired multiple properties.[19]

Section V: Empirical Facts on Misrepresentation

In Section V.A, we present descriptive statistics. Section V.B and Section V.C present the empirical results related to the two measures of asset misrepresentation.

V. A Descriptive Statistics

Panel A of Table 1 displays summary statistics for loans reported as for owner-occupied properties. As we observe, 6.43% of mortgages reported as for owner-occupied properties are

[17] We find that more than 93% of such loans have second liens recorded in the bank's internal database. This exercise confirms the quality of merged data we use and shows that almost all loans identified by us as being misrepresented on the dimension of second lien are correctly assigned into this category. The remainder, less than 7%, of loans that do not have a second lien recorded in the New Century database are not necessarily incorrect merges because they may simply represent the cases where New Century was not aware of the simultaneous second-lien mortgage originated by another lending institution.

[18] This indicates that about a third of properties whose owner-occupancy status is reported to a trustee of the RMBS as being investor-owned have owners residing in the same zip code as the property. This likely reflects the fact that a non-negligible fraction of investors purchase properties in their zip code of residence – a fact we corroborated to be potentially true based on our conversations with several industry practitioners.

[19] This is also true for collateral with reported non-owner-occupants. In particular, more than 80% of such borrowers have multiple first mortgages, based on the credit file.

misrepresented based on our method. In other words, these records have a mailing address zip code of the borrower in the Equifax database that does not match with the property address in the BlackBox database, suggesting that the borrower was not a primary resident of these properties. Alternatively put, more than 27% of loans obtained by non-owner occupants misreported their true purpose.

Panel B of Table 1 reports similar statistics for loans that report no presence of second liens to RMBS investors (e.g., CLTV and LTV ratios are the same in the BlackBox data). As we observe, we find that about 7.1% of mortgages in this sample are misrepresented as these loans do have associated second liens in the credit bureau data. This implies that more than 15% of loans with closed-end second liens in our sample incorrectly reported that such loans were not present. It is worth noting that if we treat HELOCs as second liens, the overall level of misreported second liens is almost twice as large. In particular, 13.6% of loans that report that second lien is not present do actually have closed-end second mortgage or a HELOC with positive outstanding balance originated at the same time as a senior mortgage.

One could contemplate that possibility of asset misrepresentation and more broadly a lower quality of the reported data may have been indicated to the investors through low or no documentation status of loans. However, as Table 1 shows, we find significant extent of misrepresentation even when we focus on fully documented loans. About 4.7% of fully documented loans reported as for owner-occupied properties were given to borrowers with a different primary residence, while about 7.9% of fully documented loans stating that a junior lien is not present actually had such a second lien. Interestingly, while the extent of misrepresentation of owner-occupancy status is somewhat lower for fully documented loans, it is actually higher for misrepresentation of second-lien.

Overall, our measures imply that about 9.1% of loans have either misrepresentation (i.e., misreported owner-occupancy status or misreported second lien). This number would be about 12.2% were we to include HELOCs when inferring misreported second liens.

In Figures 1A and 1B, we display the geographical concentration of loan misrepresentations. Figure 1A presents the fraction of loans with misrepresented owner-occupancy status across the United States in our data. As can be observed from this figure, misrepresentations of owner-occupancy status are most pronounced in coastal and West Coast markets such as Florida, California, Arizona, and Nevada (with more than 7% of loans being misreported, according to our measure), which were booming real estate markets leading up to the crisis. This is consistent with the significant activity of non-owner-occupant borrowers documented in these regional markets prior to the recent housing crisis (see Chinco and Mayer 2012). Our measure indicates that a significant portion of the purchases made by non-owner-occupants in these regions was financed with mortgages that misrepresented their true purpose. For example, in our data about

22.5% of loans in Florida were reported to be non-owner-occupant loans. At the same time, according to our measure, about 9.5% of mortgaged properties in Florida that were reported as owner-occupied were actually non-owner-occupied properties. This implies that about a third of loans obtained by non-owner occupants in Florida misreported their true purpose. In contrast, Figure 1A shows that rural states such as Nebraska and South Dakota have some of the lowest fractions of loans that misrepresent the owner-occupancy status (2.5% and 2.8%, respectively). This lower incidence of misrepresentation is most likely tied to the less pronounced presence of non-owner-occupant borrowers in these regional markets.

Figure 1B presents the corresponding results for the fraction of loans that our measure indicates were financed with misrepresented second liens. We note that the geographical distribution of this misrepresentation is somewhat different from the owner-occupancy status results displayed in Figure 1A. In particular, West Coast states and coastal states that have a pronounced presence of owner-occupancy misreporting show a relatively lower incidence of misreported second liens. In contrast, states such as Texas and some of the Midwestern states show a more pronounced incidence of misreported second liens (more than 12% of loans erroneously report the absence of a second lien based on our measure).

To investigate this further, we examine how the extent of misrepresentations across MSAs (Metropolitan Statistical Areas) covaries with house price movement and volume of loan origination. Specifically, we regress the mean value of given misrepresentation at the MSA level against annualized HPI growth from 2001 to 2004 at the MSA level, annualized HPI growth from 2005 to 2006 at the MSA level, and a variable accounting for the MSA loan volume. This last variable, *Share of 2005–2007 Loan Originations*, is defined as the number of loans originated in the MSA from 2005 to 2007 divided by the total number of loans originated over this time period, using the overall BlackBox loan sample. Table 2 reports the results.

Consistent with our findings in Figure 1A, we observe in Column (1) of Table 2 that misrepresentations regarding owner-occupancy status were most prevalent in regions with large pre-crisis HPI growth, which typifies states such as Arizona, California, Florida, and Nevada. Similarly, the misreported second lien results in Column (2) corroborate our state-level findings from Figure 1B, as these misrepresentations are prevalent in regions that experienced less home price appreciation, such as Texas and the Midwestern states. Both misrepresentations are correlated with larger loan volumes, suggesting a connection between increased lending activity and asset misrepresentations.

V.B Relation of Misrepresentation with Observables and Its Evolution over Time

In this section we investigate the degree to which asset misrepresentation varies with observables and over time. We begin by estimating three loan-level regressions of the following form:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (1)$$

In these specifications the dependent variable is a dummy variable, *Misreported Non-Owner-Occupant*, that takes a value of 1 if the owner-occupancy status of a loan i is misrepresented to be owner-occupied when it is truly non-owner-occupied as identified by our method, and is 0 otherwise. Because we are interested in assessing how asset misrepresentation is related to the observable characteristics of loans, we include a vector X_i which, depending on the specification, consists of a set of loan-level observable characteristics, such as origination interest rate, credit score, and LTV ratio.

We present the results in Table 3. Column (1) of Panel A presents the results of these regressions with basic controls.[20] The second column adds a control for variation in merge confidence. This accounts for heterogeneity in the confidence with which we are able to infer whether or not there was asset misrepresentation using our method. Finally, the third column also includes half-year fixed effects capturing the loan origination date. This allows us to track the evolution of asset misrepresentation over time.

As is evident, all coefficients reported on observables are stable across these three specifications and are statistically significant. These results suggest that loans with asset misrepresentation on owner-occupancy status, as identified by our method, are more likely to have higher loan interest rates and CLTV ratios. Moreover, these loans often tend to be of an option adjustable-rate mortgage (ARM)[21] type and have low or no additional documentation provided. We note that the relation between observables and asset misrepresentation also seems economically meaningful. For instance, the probability of asset misrepresentation based on the owner-occupancy status of a loan increases by more than 7% (more than 100% relative to the mean) if the loan contract is an option ARM. Likewise, the low-documentation status is associated with a nearly 3% increase in the probability of asset misrepresentation. Broadly, the features that are positively related to asset misrepresentation on the dimension of occupancy status are typically associated with purchase mortgages of the Alt-A type of credit risk.[22]

[20] Throughout the paper we estimate our specifications using the OLS despite the binary nature of several of the dependent variables. Our OLS specification with flexible controls to capture nonlinearity allows us to estimate our coefficients consistently even with multiple fixed effects (Dinardo and Johnston 1996). As we illustrate in Appendix B, our inferences are very similar using a non-linear specification instead (probit).

[21] Unlike more traditional fixed-rate mortgages (FRMs) or adjustable-rate mortgages (ARMs), an option ARM is an adjustable rate mortgage that lets borrowers pay only the interest portion of the debt or even less than that, while the loan balance can grow above the amount initially borrowed up to a certain limit. See Piskorski and Tchistyi (2010), Amromin et al. (2011), and Keys, Piskorski, Seru, and Vig (forthcoming) for more discussion of these contracts.

[22] The Alt-A risk category consist of loans that, for various reasons (e.g., low documentation), are considered riskier by loan originators than prime loans but less risky than subprime mortgages.

Although the addition of origination cohort controls does not seriously impact the sign or significance of the regression coefficients, the fixed effects themselves illustrate an interesting trend. In Column (1) of Table 4, we report these coefficients relative to the mean asset misrepresentation in 2005 (the omitted category in these regressions). As is evident, loans with asset misrepresentation regarding their owner-occupancy status became more prevalent in the run-up to the crisis. Note that we start in 2005 because we do not have reliable credit bureau data prior to this period. The average propensity of asset misrepresentation peaks in the first half of 2007 and then drops sharply thereafter. This could reflect the fact that, with the contraction of the non-agency securitization market, scrutiny in the private market was elevated around this time period and credit standards tightened.

We now follow the same empirical approach and use our second measure of asset misrepresentation, which identifies loans that state that only a single lien is present while in fact there is an associated second junior lien. Specifically, we estimate specifications of form (1) where now the dependent variable in these regressions is a dummy, *Misreported Second,* that takes a value of 1 if the loan states that it has no simultaneous second lien but our method infers that there is one, and is 0 otherwise. The results are displayed in Panel B of Table 3. The specifications follow the same order as Panel A.

The results indicate that loans that are classified as having misrepresented asset quality on the dimension of second liens tend to have lower interest rates, smaller origination balances, and lower CLTV ratios. In contrast to what we found in Panel A, the coefficients are negative and significant for low- and no-documentation loans and for option ARMs. These features broadly depict patterns that are expected of typical loans *with* second liens. These findings suggest that the lenders may have been aware of the presence of second liens -- given that they were more likely to collect information on the assets, income, and employment of borrowers when granting the first mortgage -- and that this information was indeed misrepresented to the investors of RMBS. We investigate this hypothesis further in Section VIII.

Similar to the inferences in Panel A, the cohort fixed effects do not change the qualitative or quantitative relation between asset misrepresentation and loan observables, but they again paint a consistent picture. In particular, in Column (2) of Table 4, we again find a growing pattern of asset misrepresentation on the dimension of misreported second liens through the boom until about the first half of 2006. Thereafter, the prevalence decreased with time, with small negative fixed effects beginning in the second half of 2006. Overall, these time series patterns are consistent with the model of Povel, Singh, and Winton (2007) who predict misrepresentations are most likely to occur in relatively good times.

V.C Loan misrepresentation and ex post defaults

V.C.1 Mortgage Debt

Do our measures of asset misrepresentation imply that RMBS investors bought assets that were of a higher risk than they could have estimated using the information disclosed at the time of purchase? Evaluating this issue requires us to assess whether these asset misrepresentations did indeed imply that the quality of these loans was worse than what the investors of the RMBS might have perceived given the disclosed information.

To make this assessment, we examine how the measures of asset misrepresentation are related to subsequent defaults on mortgages, accounting for the observables that were disclosed to investors at the time of RMBS issuance. To that end, we estimate various specifications investigating the impact of the two types of asset misrepresentation on the mortgage's ex post default likelihood and report the results in Panels A and B of Table 5.

We start by examining the relation between default likelihood and asset misrepresentation based on the owner-occupancy status of the loans by estimating the following loan-level specification:

$$Y_i = \alpha + \beta X_i + \gamma \delta_i + \varepsilon_i \qquad (2)$$

The dependent variable in these regressions is a dummy that takes a value of 1 if the mortgage goes 90 days past due on payments during the first two years since origination and is 0 otherwise. The right-hand variables include a dummy (δ_i), *Misreported Non-Owner-Occupant,* which takes value of 1 if the loan *i* has misrepresentation on owner-occupancy as per our method and is zero otherwise. The key coefficient of interest is γ, which captures the estimated relation between owner-occupancy misrepresentation and the subsequent loan performance. As in specification (1), vector X_i consists of a set of loan-level observable characteristics, such as origination interest rate, credit score, and LTV ratio.

Panel A of Table 5 presents the results. The first column estimates this specification for the entire sample and includes origination cohort fixed effects along with other controls. As is evident, we find that the *Misreported Non-Owner-Occupant* dummy is positive and statistically significant. This implies that, conditional on all the relevant observables that are commonly used in the literature to predict delinquency (e.g., see Piskorski, Seru, and Vig 2010; Keys, Mukherjee, Seru, and Vig 2010), asset misrepresentation on the dimension of owner-occupancy positively predicts the subsequent delinquency of the loan. This estimate is economically significant as well. In particular, ceteris paribus, a loan with misrepresented borrower's occupancy status have about a 9.4% higher likelihood of default (90 days past due on payments during the first two years since origination), compared with loans with similar characteristics and where the property was

truthfully reported as being the primary residence of the borrower. This implies more than 60% higher default rate relative to the mean default rate of owner-occupants during our sample period.

The remaining columns of Panel A of Table 5 re-estimate the specification in Column (1) for different origination cohorts. As before, we report the results for 2005 as a whole, rather than in half-yearly cohorts as in the subsequent years due to our concerns about the data quality in the earlier part of 2005. However, the qualitative nature of the results is similar to what is reported in the table. As is evident, the positive correlation between the *Misreported Non-Owner-Occupant* dummy and the delinquency of the loan is stable across time. In terms of economic magnitudes, the effect is the smallest in 2005. However, even in 2005, the estimate suggests an increase in absolute delinquency by 4.04% if the asset quality of a loan is misrepresented on the dimension of owner-occupancy (a relative increase of more than 56% over the mean default rate of 7.15%).

We also find that the absolute magnitude of the coefficients on the *Misreported Non-Owner-Occupant* variable grows with time. This is consistent with the notion that loans originated in 2006 and 2007 experienced larger and more immediate house price declines -- thereby increasing the chance of default, especially for riskier loans such as misreported non-owner-occupant loans.

We repeat the same analysis using the misrepresentation of second lien status. For that purpose, we estimate the specification (2) where now the right-hand variable (δ_i), *Misreported Second Lien*, takes value of 1 if the loan *i* has misrepresented second lien and is zero otherwise. The results are presented in Panel B of Table 5. Similar to our findings in Panel A, we find that the coefficient on *Misreported Second* is positive and significant. This suggests that, that asset-quality misrepresentation on the dimension of second liens also strongly predicts the likelihood of a loan becoming delinquent, even after conditioning on all the relevant observables. The economic magnitudes are also large. In particular, loans with misrepresented higher lien have about 70% higher likelihood of default in relative terms compared with loans with similar characteristics and no higher lien (a 10.2% absolute increase over a mean delinquency rate of 14.6% for this sample).

As other columns in Panel B suggest, the relation between the misrepresentation on the dimension of second lien and the ex post delinquency of the loan is persistent and significant across time. Moreover, as was the case with the first measure of asset quality misrepresentation, the absolute effect of misreported second liens on subsequent default rates grows over time through mid-2007: from about a 7% higher default rate for loans originated in 2005 to about a 16.7% higher default rate for loans originated in the first half of 2007. We do observe smaller effects for loans originated in the second half of 2007 (about a 6.7% increase in default rates for loans with misreported second lien), but we note that there are very few such loans in our sample. So far, we have analyzed how the performance of loans misrepresented on the dimension of owner-occupancy (*Misreported Non-Owner-Occupant*) compares with that of loans for the

properties that are indeed the primary residence of the borrowers. We now conduct analysis where we contrast how the performance of mortgages misrepresented on the dimension of owner-occupancy compares with loans where the occupancy status is *identical* but was accurately reported to the investors.

In Panel C of Table 5, we examine whether loans made to misreported non-owner-occupants performed differently relative to loans that truthfully reported the non-owner-occupancy status. For this purpose, we estimate regressions similar to specification (2) where the dependent variable is the same delinquency dummy variable used in Panel A and Panel B. There are two variables of interest in these regressions: the previously defined *Misreported Non-Owner-Occupant* dummy as well as the additional *Reported Non-Owner-Occupant* dummy, which takes a value of 1 if the loan is reported as secured by either a second home or investor property, and is 0 otherwise. For these regressions we expand the earlier sample of loans with reported owner-occupied status to also include loans backed by properties that are reported as either second homes or investor properties. The difference in the specifications results from whether we also include half-year vintage effects (Column 2), state fixed effects (Column 3), and whether we cluster standard errors at state level (Column 4).

Panel C of Table 5 shows that *Reported Non-Owner-Occupant* has a positive and significant coefficient in the default regression, implying an about 3.5% absolute increase in the default rate for these loans (about 23% relative increase in default rate compared with loans with owner-occupied purpose). This result is in line with the literature that finds that loans with non-owner-occupancy status are more likely to default when controlling for other observable risk characteristics (see Mayer et al. 2009 and Haughwout et al. 2011). The reason is that such borrowers are likely to default more quickly on homes that are not their primary residence, for reasons such as weaker neighborhood and social ties.

More interestingly, *Misreported Non-Owner-Occupant* has a positive and much larger coefficient in the default regression, implying an approximately 9.5% absolute increase in the default rate for these loans (about a 64% increase in the default rate relative to loans for owner-occupied properties). These results show that loans with misreported non-owner-occupancy status of the borrower perform worse *not only* with respect to loans of owner-occupants but *also* relative to loans of non-owner-occupants that are correctly reported as such to investors. Thus, misrepresentation on the dimension of owner-occupancy is indicative of borrowers with significantly lower quality unobservable characteristics even compared with truthfully reported non-owner-occupants.

Next we conduct a similar exercise with the second measure of misrepresentation. In particular, we examine the relationship between a loan's delinquency and whether this loan has a truthfully reported second lien, a misreported second lien, or no second lien. Importantly, given the non-

linearity in relation of debt and defaults, we need to control for not only the presence of a misreported second lien but also its size. For loans with reported second lien we already know the CLTV ratio. For an misreported second lien, we construct a new measure of CLTV by using the original balance of the first mortgage reported to BlackBox, the second lien balance reported to Equifax, and the original home value (as imputed from the LTV ratio). We then allow for the differential impact of higher level of debt on defaults by constructing four dummy interaction variables. First, we create *Misreported Second X CLTV >= 100,* which takes a value of 1 if the loan has a misreported second lien and a CLTV greater than or equal to 100, and is 0 otherwise. Similarly, we construct *Misreported Second X CLTV < 100,* which takes the value of 1 if the loan has a misreported second lien and a CLTV less than 100, and is 0 otherwise. We then construct the corresponding measures, *Reported Second X CLTV >= 100* and *Reported Second X CLTV < 100,* for loans with a truthfully reported second lien and CLTV greater and less than 100, respectively.

We now repeat the analysis of Panel B with these four interaction terms as our variables of interest. Panel D of Table 5 reports our results. We observe that senior mortgages with second liens are much more likely to default than those without such liens, and the effect is nearly twice as large when the CLTV ratio is greater than or equal to 100. Moreover, unlike in the case of misrepresentation of owner-occupancy status, we observe similar default patterns for loans with second liens, regardless of whether or not the presence of second lien was reported to the investors.

Finally, we also investigate how the impact of borrower occupancy status on default varies over time. For that purpose, we use monthly data to estimate the transition probability that a given loan goes 90 days past due on its payments in its first two years since origination using the sample of loans from Panel C of Table 5. We include a set of fixed effects representing the first eight quarters of a loan history, which allows us to estimate the dynamic pattern of defaults. Other controls include the variables used in Panel C of Table 5, as well as *Reported Non-Owner-Occupant* and *Misreported Non-Owner-Occupant* interacted with quarterly time dummies. Panel A of Figure 2 plots the resulting cumulative delinquency rates over time for properties with different borrower occupancy status, holding all other observables at the overall sample mean. We observe that regardless of loan age, misreported non-owner-occupants consistently default at significantly higher levels on their mortgages than truthfully reported non-owner-occupants, who in turn default at a higher rate than truthfully reported owner-occupants.

Next, we repeat the same analysis for the impact of junior lien status on defaults over time using the sample of loans from Panel D of Table 5. Controls include those used in Panel D of Table 5, as well as quarterly time dummies and *Reported Second* and *Misreported Second* interacted with quarterly time dummies. Panel B of Figure 2 plots cumulative delinquency rate over time by junior lien status, holding all other observables at the overall sample mean. We observe that

loans with either a misreported or a reported second lien consistently default at significantly higher levels than loans that truthfully reported as having no junior liens. Moreover, consistent with our prior results, loans with second liens have similar default patterns, regardless of whether or not the presence of a second lien is correctly reported to investors.

V.C.2 Other Forms of Debt

Next, we consider whether the two identified misrepresentations are related to how the borrower performs on other credit products. The motivation for this analysis is to assess if borrowers with misrepresented loans differ in terms of their inherent riskiness. We hope to make this assessment by examining the probability of default on other forms of debt.

We utilize the borrower credit card payment histories provided by Equifax and create a dummy variable that takes a value of 1 if the borrower becomes 60 days past due on any credit card account in the first two years since mortgage origination. We run a series of regressions with this measure as the dependent variable and a vector of controls that includes credit card utilization, available credit, monthly credit card payments, unpaid credit card balance, FICO, and merge confidence. We define credit card utilization as the borrower's unpaid credit card balance divided by the total credit card limit, effectively measuring how credit-constrained a borrower is. We require that the borrower be current on all credit card accounts in the first month that we observe the mortgage in the data in order to make meaningful inferences.

Table 6A shows the results of three credit card default regressions that include *Misreported Non-Owner-Occupant* and *Reported Non-Owner-Occupant* as dependent variables. The excluded category represents loans for which the borrower has been truthfully reported as an owner-occupant based on our measure. The first column includes the above controls and origination vintage fixed effects, the second column adds state fixed effects, and the third column incorporates mortgage-based controls such as the CLTV ratio, loan balance, and borrower documentation status. We find that an owner-occupancy misrepresentation increases the likelihood of credit card default by about 7.6%. At the same time, truthfully reported non-owner-occupants have a likelihood of default on credit cards similar to truthfully reported owner-occupants. Throughout, the other covariates behave as we would expect, with lower FICO scores and higher credit card utilization being correlated with higher credit card default rates. This reinforces the conclusion that borrowers with misrepresented owner-occupancy status are a distinct group of riskier borrowers.

Next, we repeat this analysis for reported and misreported second liens. Again we use the four interaction terms employed in Table 4D. Panel B of Table 6 presents the results. Although the controls behave as in Panel A, we observe that presence of a second lien is negatively correlated with credit card default. The absolute effect ranges from –0.70% to –3.11%, a relative decrease

in the default probability of about 2% to 8%. This may be consistent with the notion that mortgages with second liens are more likely to be fully documented and therefore tend to be selected by better-quality borrowers.[23] As in Table 4D, we do not observe a consistent, significant difference between loans with misreported and truthfully reported second lien.

This evidence, taken together with our findings from the previous section, supports the view that there is significant variation among borrowers that have misrepresentations on their mortgages, depending on which misrepresentation is identified. It appears that misrepresentations on the dimension of owner-occupancy are indicative of borrowers with significantly worse unobservable characteristics even compared with those who are truthfully reported as being non-owner-occupants. The case of misrepresentation of second liens is more subtle. We find that while mortgages with misreported second liens perform significantly worse than those that truthfully reported no presence of second liens, the borrowers associated with these loans appear to perform similarly on mortgages and credit cards as compared to borrowers with mortgages for which the presence of such liens is correctly reported to investors.

Section VI: Were Misrepresentations Priced?

We now investigate whether misrepresentations on the two dimensions -- given that loans with such misrepresentations were more likely to default ex post -- were priced by lenders and investors. To shed light on this question, we first assess whether lenders charged higher interest rates for loans that were misrepresented. Second, we examine whether pools with a higher incidence of misrepresented collateral were perceived to be of lower quality by examining if such pools had higher subordination levels protecting the safest AAA-tranches and higher imputed yields compensating RMBS investors for the greater risk of these securities.

VI.A Misrepresentations and Mortgage Interest Rates

We begin by investigating whether lenders charged higher interest rates for loans with misrepresented collateral, given that such loans were more likely to default ex post. In Panel A of Table 7, we first examine whether lenders recognized that loans with misrepresented borrower occupancy status were riskier and thus charged a higher interest rate on such loans. In particular, we estimate a regression similar to specification (2) using interest rate as the dependent variable. In these regressions the two variables of interest are the previously defined *Misreported Non-Owner-Occupant* and *Reported Non-Owner-Occupant* dummy variables. The other covariates used in the regression include the common variables that have been shown to be related to risk of

[23] Because banks often keep second liens on their own balance sheets, obtaining a second lien may be evidence of a generally less risky borrower (see Lee et al., forthcoming). Our *Misreported Second* variable thus contains information related to overall borrower credit quality as well as the misrepresentation itself. In examining credit card defaults, it appears that the credit quality effect dominates, and hence the misrepresentation is correlated with a lower level of default.

the loan such as FICO, LTV, whether the loan is an ARM, and so on. These coefficients are unreported for brevity. The difference across specifications results from whether we also include half-year origination (Column 2), state fixed effects (Column 3), and whether we cluster standard errors at state level (Column 4).

As illustrated in Panel A of Table 7, we find that *Reported Non-Owner-Occupant* has a positive and significant coefficient. This result is in line with the literature that finds that loans with non-owner-occupancy status are perceived to be riskier than those for owner-occupied properties. The more interesting result, however, is that the coefficient on *Misreported Non-Owner-Occupant* is also positive and significant. This suggests that asset misrepresentations on this margin were, to some extent, captured in interest rates charged by lenders.[24] However, we note that the increase in interest rates for misreported non-owner-occupants is smaller than for truthfully reported ones (by around 13 basis points) -- even though our earlier analysis shows that loans with misreported non-owner-occupants also default *significantly more* than loans of truthfully reported non-owner-occupants.

Next we repeat this analysis for reported and misreported second liens. For that purpose we estimate similar specifications for mortgage interest rate, replacing *Misreported Non-Owner-Occupant* and *Reported Non-Owner-Occupant* with the four interaction terms as used in our earlier analysis. Panel B of Table 7 presents the results. As expected, loans with a second lien and a CLTV below 100 consistently receive lower interest rates than those with a CLTV of 100 or greater. When the presence of second lien is truthfully reported, the borrower receives a higher interest rate, with increases ranging from 12 to 44 basis points, depending on CLTV and other controls. More important, senior mortgages with misreported second liens carry, on average, only slightly higher interest rates compared with loans which truthfully reported not having second liens (the interest rate difference ranges from close to 0 to about 10 basis points) -- even though our earlier analysis shows that such loans default at a much *higher* rate than loans without second liens.

To further investigate the relationship between occupancy status and origination interest rates, we plot the error terms for predicted interest rates using a sample and specification similar to Column (3) of Panel A of Table 7, with the *Misreported Non-Owner-Occupant* and *Reported Non-Owner-Occupant* variables removed. Panel C of Figure 2 shows the kernel density distributions for these error terms, defined as the difference between the actual origination interest rate and the predicted value, with loans grouped by actual occupancy status. Consistent with our previous results, we observe that loans that truthfully reported as having non-owner-

[24] While not reported, the other covariates behaved as expected. Indicators associated with higher risk, such as high CLTV and low or no documentation, were correlated with higher interest rates, whereas high FICO scores and balances led to lower interest rates. Regardless of the controls and type of specifications, the qualitative and quantitative nature of results reported in the table remained the same.

occupants receive higher interest rates (larger error terms) than those of misreported non-owner-occupants, who in turn have interest rates that are higher than loans that truthfully have owner-occupants.

We next repeat this analysis for reported and misreported second liens, now predicting interest rates using a sample and specification similar to that of Column (3) of Panel B of Table 7, with the second lien dummies and CLTV interaction terms removed. Panel D of Figure 2 shows the kernel density plots for the error terms of these predicted rates, with loans grouped by their second lien status. Again, consistent with our previous results, we observe that loans with truthfully reported second liens receive higher interest rates than loans that truthfully report as not having any second lien. Importantly, loans with misreported second liens have origination interest rates that are only slightly higher than those that truthfully report as not having a second lien.

Overall, the above evidence suggests that lenders were partly aware of the higher risk of misrepresented loans as they charged higher interest rates on these loans. However, the interest rate markups on misrepresented loans are much smaller relative to loans where the property was truthfully disclosed as not being primary residence of the borrower and as having a higher lien. Therefore, relative to prevailing interest pricing, interest rates on misrepresented mortgages did not fully reflect their higher default risk.

VI.B Misrepresentations and Pool Yields and Subordination Levels

We now turn to the second empirical test in this section, in which we assess whether these misrepresentations were reflected in the prices that RMBS investors paid for the securities. In other words, we examine if the RMBS pricing reflected the lower quality of deals that had a larger share of misrepresented loans.

We note that investigating this notion is difficult for several reasons. First, we do not have access to actual prices paid by the investors at the time of pool issuance. This is consistent with the existing literature that uses model-generated prices rather than actual prices (e.g., prices generated by Bloomberg or Intex) or indirect proxies such as subordination rates. Second, even if we were to observe the pool tranche prices, we would need a structural model to assess whether the variation in these prices is sufficient to compensate the investors for additional default risk. Constructing such a structural model is difficult, as it requires, among other challenges, modeling expectations of market participants while accounting for the complex structure of many pools with multiple tranches and rules governing the distribution of cash flows.

Nevertheless, to shed some light on this question we start by investigating the statistical relationship between the imputed average yield spreads on each pool and our measures of

collateral misreporting. To determine yield spread, we take the balance-weighted average coupon for the pool from ABSNet and subtract the yield of a 10-year U.S. Treasury bond at the time of securitization, as reported by the Federal Reserve. This is arguably a crude approximation of the actual yields received by investors on MBS pools, but due to a lack of better data, such measures have been used in a number of recent studies (e.g., Faltin-Traeger, Johnson, and Mayer 2010; Demiroglu and James, 2012; He, Qian, and Strahan, forthcoming).

To run our pool-level analysis, we first restrict our sample to pools with at least 25% of loans for which we can compute whether the loan was misrepresented or not (given our high quality match restrictions). We have experimented with this threshold and found our qualitative inferences similar with different thresholds. We then remove pools that do not have reliable coupon data on ABSNet and are left with a sample of 353 pools containing approximately 698,000 mortgages. In the following specifications, we consider three variables of interest separately: the percentage of loans with misreported non-owner-occupant borrowers, the fraction of loans with misreported second lien, and the percentage of mortgages with either a misreported non-owner-occupant borrower or misreported second lien. To calculate these measures, we divide the number of loans in the pool with the particular misrepresentation by the number of loans in the pool for which we could potentially identify such a misrepresentation. Other variables, such as CLTV ratios and FICO scores, are taken as an average across all loans in the pool.

In Tables 8A through 8C, we run a series of pool-level regressions of the following form:

$$Y_i = \alpha + \beta X_i + \gamma PM_i + \varepsilon_i \qquad (3)$$

in which the dependent variable is the pool's average yield spread as described above. The variable of interest, PM_i, is the percentage of misrepresented loans in a pool i. Table 8A shows the results using the percentage of loans in a pool with misreported owner-occupancy while Table 8B reports the results using the percentage of loans with misreported second liens. Table 8C shows the results using the percentage of loans having either misrepresentation. The vector of control variables X include the pool-level means of common risk factors used in the regressions in Table 2, such as mean FICO of the pool, average CLTV ratio, and the fraction of loans with low documentation. Column (1) contains our base specification, Column (2) adds vintage fixed effects, Column (3) includes the pool overcollateralization percentage, and fixed effects for the largest six underwriters are added in Column (4).[25]

We find limited evidence that misrepresentations were priced in by investors, as only misreported non-owner-occupants led to statistically significantly higher yield spreads, and even this significance was only at a 90% level. More importantly, the estimated effect is economically

[25] We do not include full underwriter fixed effects due to the small sample size, but specifications with complete underwriter fixed effects, grouping the smallest underwriters as one entity, yield similar results.

small. For instance, a one-standard-deviation increase in the pool of misreported non-owner-occupants from the mean level of misrepresentation would be associated with only about a 4-basis-point increase in the yield spread over 10-year Treasury bonds. For other measures, these estimated effects are not only economically very small but also statistically insignificant.[26]

Next we also investigate whether pools with a higher share of misrepresentations have higher subordination levels protecting AAA-rated tranches. Tables 9A through 9C run corresponding regressions to those described above, now with the dependent variable being the lowest subordination level in the pool given to a AAA-rated tranche. The control variables in Columns (1) and (2) are defined identically as before. Column (3) includes the average initial interest rate of loans in the pool, and Columns (4) and (5) add overcollateralization percentages and fixed effects for the six largest underwriters, respectively. Again, we see little evidence that investors were compensated for the greater risk of collateral involving a higher share of misreported non-owner-occupants and junior liens. As before, not only are these estimated effects economically very small but they are statistically insignificant as well.

Overall, we view these results as providing little evidence that misrepresentations were reflected in the initial prices of the RMBS securities. It is important to note that we do find evidence that *reported* levels of average CLTV and fraction of non-owner occupied loans had a meaningful relation with our price measures. For example, controlling for other observable pool characteristics, a 10 percent absolute increase in reported average CLTV of the pool is associated with about 0.31% increase in the imputed yield spread (more than 38% increase relative to the mean spread) and about 1.7% increase in subordination level (more than 15% increase relative to the mean subordination level). This suggests that had misrepresented loans been correctly reported to investors, there would have been a meaningful impact of this disclosure on the prices of these securities at their issuance.

VI.C Institutional Evidence for Misrepresentation and Mispricing of MBS

We now discuss some evidence that is consistent with the broad, systematic nature of the findings uncovered by our study. Recently, there has been an increasing number of lawsuits filed against the originators, servicers, and underwriters of RMBS that attribute the deterioration of the performance of these securities to poor underwriting standards and misstated representations and warranties (see Appendix C for a list of such lawsuits). Although representations and warranties are also provided by the mortgage lenders, investors are holding the underwriters of

[26] For example, a one-standard-deviation increase in a share of *Misreported Second* in a pool at the mean level of such misrepresentation would be associated with only about a 1.3-basis-point increase in yield.

the securities accountable, claiming that these parties neglected their responsibility to verify the stated quality of the collateral.[27]

While the specific claims vary by lawsuit, investors generally charge that loans in the securitized pools frequently and materially failed to satisfy the representations and warranties made in security prospectuses. Though these lawsuits do not present any clear systematic empirical evidence, among the most frequently cited violations of representations and warranties are that combined loan-to-value (CLTV) ratios were higher than stated, homes identified as owner-occupied were either non-owner occupied or second homes, property values were improperly inflated, and unreported second liens secured by the property were present.[28] For example, in a representative suit filed by the Federal Home Loan Bank of San Francisco (FHLB) against eight security dealers (Credit Suisse, Deutsche Bank, JPMorgan, Bear Stearns, RBS, Morgan Stanley, Bank of America, and Countrywide Financial), FHLB claims that the security prospectuses misled RMBS investors and did not accurately depict the underlying loans.[29] Specifically, in regard to owner-occupancy, the suit alleges:

> *These statements were materially untrue or misleading because (i) the stated number of mortgage loans that were secured by primary residences was higher than the actual number of loans in that category; (ii) the stated number of mortgage loans that were secured by second homes was lower than the actual number of loans in that category; (iii) the stated number of mortgage loans that were secured by investment properties was lower than the actual number of loans in that category; or (iv) the Defendants omitted to state that the occupancy status of a significant number of the properties that secured the mortgage loans in the collateral pool was misstated because of fraud.*

Similarly, in regard to statements on the underlying mortgages' LTV and CLTV ratios, the FHLB charges:

> *These statements were materially untrue or misleading because (i) the stated LTVs and CLTVs of a significant number of those mortgage loans were lower than the actual LTVs or CLTVs; (ii) the Defendants omitted to state that the appraisals of a significant number of the properties that secured the mortgage loans in the collateral pools were biased upward, so that stated LTVs and CLTVs based on those appraisals were lower than the true LTVs and CLTVs of those mortgage loans; or (iii) the stated CLTVs did not reflect second*

[27] Underwriters often argue that they reported information that was provided to them by borrowers and primary mortgage lenders and point to such qualifying clauses in the prospectuses. On the other hand, investors in these securities contend that underwriters are *also* responsible for verifying that such information was correct instead of merely reporting it. We discuss this issue in conclusion.

[28] See Chris Gamaitoni, Jason Stewart, and Mike Turner, "Mortgage Repurchases Part II: Private Label RBMS Investors Take Aim—Quantifying the Risks." Compass Point Research & Trading, LLC, 17 August 2010.

[29] *FHLB of San Francisco vs. Credit Suisse Securities LLC et al.*, Superior Court of California, County of San Francisco, No. CGC.10.497840.

mortgages on a significant number of the properties that secured the mortgage loans in the collateral pools.[30]

Two issues emerge from this discussion. First, while the lawsuits alleging misrepresentations of the type we analyzed are becoming increasingly common and are often anecdotally based, the broad flavor of these assertions corroborates the large-scale systematic findings in this paper.[31] Second, in most of these cases, RMBS investors argue that these misrepresentations were not reflected in the initial prices of securities, which is also consistent with our analysis in this section.

Section VII: Asset Misrepresentation, Underwriters and Regulation

Having established that the asset misrepresentations we focus on imposed potentially significant risk on the investors of RMBS, we now demonstrate that there is substantial heterogeneity in the extent of these misrepresentations across underwriters. We then try to assess whether the degree of asset misrepresentation at the underwriter level varies with underwriter-specific factors related to their business model and executive compensation structure. Finally, we end by assessing if regional level variation in misrepresentation is related to regulatory environment in that region.
We start by visually demonstrating the presence of heterogeneity across underwriters in terms of their propensity to misrepresent the asset quality on the dimension of misreported borrower occupancy status. Specifically, we use a specification similar to Panel A of Table 2 with *Misreported Non-Owner-Occupant* as the dependent variable and with underwriter fixed effects in addition to the control variables. In Figure 3, we plot the estimated level of underwriter misrepresentation, calculated using the underwriter fixed effects that are obtained from this regression and fixing other controls at their means, along with the 95% confidence interval (omitted category is Credit Suisse, with 4.43% of loans in pools it underwrote identified as having misrepresented the owner-occupancy status of the borrower).

As can be seen, there is substantial heterogeneity across underwriters on the dimension of misreported borrower occupancy status. For instance, loans in pools underwritten by Lehman Brothers and Washington Mutual are about 4% more likely in absolute terms to contain misreported non-owner-occupant properties as compared with loans in pools underwritten by

[30] These claims are very representative of other cases. See, for example, *Assured Guaranty Municipal Corp. et al. v. DLJ Mortgage Capital Inc. et al*., Supreme Court of New York, County of New York, No. 652837-2011, *Federal Housing Financing Agency et al. v. UBS Americas Inc. et al.*, U.S. District Court for S.D.N.Y., No. 2011-cv-5201; No. 2011-cv-02341; *Metropolitan Life Insurance Co. v. Morgan Stanley*, Supreme Court of New York, County of New York, No. 651360/2012.

[31] These do not seem to be frivolous lawsuits with only a small chance of success. For example, in 2011 Bank of America proposed $8.5 billion settlement with twenty-two large investors to resolve litigation concerning violations of representations and warranties in more than $200 billion worth of outstanding MBS originally issued by Countrywide Financial, which Bank of America purchased in 2008. See "In the matter of the application of The Bank of New York Mellon," New York State Supreme Court, No. 651786/2011.

Credit Suisse. In contrast, Bank of America and HSBC underwrite pools that contain a similar proportion of misrepresented collateral on the dimension of owner-occupancy as Credit Suisse. Finally, recall that the mean level of misreported non-owner-occupants across loans in our sample is 6.43%. Comparing the magnitudes of the fixed effects reported in Figure 3 against the sample mean (displayed in a dashed line in Figure 3) also shows that the variation in misrepresentation of collateral across underwriters is economically significant. An important point to note is that although there is substantial heterogeneity across underwriters, a significant degree of misrepresentation exists across all underwriters, which includes the most reputable financial institutions.

Next, we repeat the analysis above using the underreporting of second liens as the dimension of misrepresentation. In particular, we use a similar specification to Panel B of Table 2 with *Misreported Second* as the dependent variable and with underwriter fixed effects in addition to the control variables. In Figure 4, we plot the estimated level of underwriter misrepresentation, calculated as above, along with the 95% confidence interval (the omitted category is Credit Suisse, with 3.37% of loans in pools it underwrote having misreported second liens.)

As is evident from Figure 4, there is significant variation among underwriters' propensity to securitize loans with misreported second lien. For instance, loans in pools underwritten by Lehman Brothers were more than 13% more likely in absolute terms to contain misreported second liens relative to loans in pools that were underwritten by Credit Suisse. Similarly, Merrill Lynch, Deutsche Bank, Nomura, and UBS underwrote pools that contain a similar proportion of misrepresented collateral on the dimension of reported second liens as Credit Suisse. Again, note that the mean level of misreported second liens in our sample is 7.13%. Comparing the magnitudes of fixed effects shown in Figure 4 with this mean (displayed in a dashed line in Figure 4) also shows that the variation in misrepresentation of collateral across underwriters is economically significant.

Finally, we study misrepresentation across both dimensions simultaneously. Specifically, we construct a variable, *Either Misreported*, which takes the value of 1 if the loan has a misreported second lien, a misreported non-owner-occupant borrower, or both. We then regress this variable against the controls used in Panel B of Table 2, as well as underwriter fixed effects, using the sample of loans that are reported as having owner-occupied properties and/or have no reported second lien. In Figure 5, we plot the level of underwriter misrepresentation, calculated as above, along with the 95% confidence interval (the omitted category is Credit Suisse, with 5.69% of loans in pools it underwrote having a misrepresentation on either of the dimensions).

Consistent with the previous results, we observe a sizeable degree of heterogeneity in the share of misrepresented loans across underwriters with all underwriters having violation levels significantly greater than 0. For loans in pools underwritten by Countrywide and Lehman

Brothers, there was a greater than 10% chance that one or both of the occupancy and second lien statuses were not correctly reported to investors. More than one in seven loans in pools underwritten by Lehman Brothers contained at least one misrepresentation. Conversely, underwriters such as Bank of America and Morgan Stanley underwrote pools with a roughly similar level of misrepresentations as those underwritten by Credit Suisse.

In the next set of analysis, we assess how the likelihood of misrepresentation on the dimensions we have been analyzing varies with underwriter characteristics. We focus on several measures that relate to the nature of the underwriter's business, such as the relative importance of the mortgage underwriting business for the underwriter, the number of years of experience in underwriting subprime deals, and whether the underwriter is a commercial bank. In addition, we focus on the internal organization of underwriters, focusing on the nature of compensation contracts and the importance a given firm places on its risk management. Both these factors have been shown in the previous literature to impact the risk-taking and screening decisions of banks (e.g., Cheng et al. 2010; Ellul and Yerramilli , forthcoming; Keys, Mukherjee, Seru, and Vig 2009). In particular, to capture the nature of compensation we construct the average ratio of cash bonus to salary of its top executives as in Cheng et al. 2010. In addition, we use the Ellul and Yerramilli index of the importance of risk management inside the organization. The details on construction of all these measures and the data sources are provided in the Data Appendix.

The specifications and control variables used are similar to those used in Table 2. Column (1) of Table 10 reports the results with *Misreported Non-Owner-Occupant* as the dependent variable, Column (2) reports the results using the *Misreported Second* misrepresentation, and Column (3) shows the results where the dependent variable takes the value of 1 if a loan has any of these two misrepresentations, and is 0 otherwise. The standard errors are clustered at the underwriter level. As we observe from Table 10, there is a very weak relation between the degree of asset misrepresentation at the underwriter level and underwriter-specific factors. We do find some evidence that the propensity to misrepresent is moderated for underwriters with more diversified business outside of the RMBS market. However, the effect is only marginally statistically significant and only for the misreported non-owner-occupant misrepresentation. Asset misrepresentation is less prevalent among commercial banks, underwriters with more RMBS experience, and underwriters with more high-powered incentives given to its top management and better internal risk management. However, none of these effects is statistically significant.

We also investigate how the level of misrepresentation aggregated at the state level relates to regulatory environment in that state. We obtain estimates of state level misrepresentation by estimating state fixed effects from the specification of likelihood of either misrepresentation that we used in Table 3. In tests, unreported for brevity, we explore how these state fixed effects relate to empirical measures of regulatory environment. For this purpose we employed the state regulator leniency index of Agarwal et al. (2012). We find limited evidence that states with more

softer banking regulators had higher share of misrepresented loans. In particular, while the association is positive, it is not statistically significant (coefficient is 1.20; t-statistic of 0.29). This result is robust to exploring different specifications and alternative measures of state level regulatory environment.

Section VIII: Discussion

The results in our paper indicate the presence of sizeable asset misrepresentations even among the most reputable underwriters. Thus, the existing market arrangements, including the reputational concerns and high-powered explicit incentives that are common in these institutions, seemed to have had limited ability to eliminate such behavior. In addition, this behavior also escaped regulators who were in charge of safeguarding rights of investors. The latter finding may suggest that the current level of protection of investors in capital markets, especially those with more passive investors, like the high-grade investment debt market, may be inadequate. Given that the problem we document is pervasive, any solution needs to impact the entire industry (e.g., culture). What this solution should be, or the exact form it should take, is a fruitful area for further research.

VIII.A Where does misrepresentation occur in the supply chain of credit?

Our earlier analysis suggests that lenders were partly aware of the higher risk of misrepresented loans since they charged higher interest rates on these loans. Our data do not allow us to investigate where in the supply chain of credit (i.e., borrower, lender, or underwriter) the misrepresentation took place. This is a difficult issue to analyze because it requires knowing about the entire information set of various agents responsible in the chain of credit. However, we can shed some light on this issue by providing limited evidence based on bank-level data from a subprime mortgage lender (New Century).

The advantage of this dataset is that it contains the loan characteristics that were recorded by the lender at loan origination. This allows us to assess whether loans backed by misreported non-owner-occupied properties were correctly recognized as for non-owner-occupied properties by loan officers and brokers of New Century. Similarly, we can assess whether New Century was aware of the presence of second liens for loans for which such presence was not reported to RMBS investors.[32] Of all loans in this sample that we identified as having misreported non-owner-occupied status, none was reported as being for non-owner-occupied properties in the

[32] To conduct this analysis, we merge the New Century data with the BlackBox-Equifax dataset. This matching is done using "Loan-id" variable that are identical in the two databases (i.e., there is *no error in matching* between these datasets). After merging, we have a sample of 3,160 loans that were reported to investors as being for owner-occupied properties. Of these, our method identifies 148 loans (4.7%) as misrepresented non-owner-occupants. In addition, 10,924 loans report no second liens to investors. Of these, our method identifies 1,279 loans (11.71%) as having a misreported second lien (Table 11B-C presents summary statistics for these loans).

New Century database. This evidence suggests that the misrepresentation concerning owner-occupancy status was made early in the origination process, possibly by the borrower or broker originating the loan on behalf of New Century. In fact, virtually all loans that misrepresented borrower occupancy status were originated by the brokers (more than 99%). This is a much larger proportion relative to the overall percentage of broker originated loans by New Century (about 65%). This misrepresentation likely reflects screening practices on the side of the lender that were not corrected by the underwriter.

In contrast, of all mortgages identified as having misreported second lien status to investors, 93.3% had a second lien reported in the New Century database. This confirms that the lenders were often aware of the presence of second liens, and hence their underreporting occurs later in the process of intermediation. This result is also consistent with our earlier findings that these loans are more likely to be fully documented.[33] Importantly, in both Panels B and C of Table 11, we find that similar to our overall sample, misreported loans in the matched New Century data default at a much higher rate relative to similar loans. This is the case despite misreported loans being of better quality on observables (such as FICO).

Overall, we find that misrepresentation concerning owner-occupancy status was made early in the origination process, possibly by the borrower or broker originating a loan on behalf of the lending institution. In contrast, the lender was aware of the presence of second liens, and hence their misreporting likely occurs later in the supply chain. One must interpret these findings with caution, however, since asserting the same holds for all the lenders requires one to assume that lending practices in New Century were representative of the industry.

VIII.B How large are the effects?

Our estimates can be used to quantify the cost of these misrepresentations to RMBS investors. We note that, based on our measures, about 9.1% of loans have some misrepresentation (a misreported owner-occupancy status or misreported second lien). Moreover we find that this estimate applies to about 85% of loans in our data since loans that truthfully report that they are non-owner occupied, have second liens, or for which investor status and CLTV are missing are not part of our analysis. Hence, assuming that our estimates are broadly applicable to the entire stock of outstanding non-agency securitized loans just prior to the subprime crisis (more than $2 trillion), such misrepresentations could impact mortgages with combined outstanding balance of up to $160 billion. This simple calculation implies that enforcement of representations and warranties by investors in response to these misrepresentations could result in significant costs to

[33] This exercise also cross-validates our methodology as almost all loans identified by us as having misreported second liens do indeed have associated second liens in the internal New Century database.

intermediaries.[34] Of course, not all of the misrepresented loans may eventually default. However, given the significantly higher default rates of these mortgages, such repurchases could still result in a substantial loss to intermediaries.[35]

Note, however, that since we focus only on two easy to quantify cases of asset misrepresentation, this estimate is likely a lower bound on the actual degree of misrepresentation and consequently the amount of assets facing possibility of forced repurchases. Likewise, asset misrepresentations by intermediaries could also induce other losses beyond those directly incurred by the investors -- for instance, by disrupting the allocation of capital by such intermediaries in the economy. Finally, such misrepresentations, by overstating the true quality of assets, might have had a meaningful impact on fueling the demand for these securities.[36] Quantifying such effects is a fruitful area of future research.

VIII.C Caveats

A few caveats about our analysis are worth discussing. First, note that our inference is based on the trustee reports data as provided by BlackBox. While the trustee reports are the primary source of data used in this market to assess the quality of collateral backing RMBS, one could argue that an average investor may have been unsophisticated and relied on more basic data to make assessment about the collateral quality backing the RMBS. Our analysis is based on an investor who would use the detailed loan level information that would be available from the trustee at the time of sale of the pool. Thus, the extent of misrepresentations for unsophisticated investors would have been even higher. Alternatively, one could argue that investors may have relied on other sources of data including potentially informative RMBS ratings, not captured in our database that could allow them to understand the true characteristics of these assets.[37] Though this scenario is possible, it seems unlikely as we find little evidence that misrepresentations we identify were priced by investors in the securities at their issuance.

Second, it is possible that our inferences are confounded due to matching error between the datasets that contains actual loan characteristics and that containing characteristics reported to

[34] We note that it is possible that not all misrepresentations we identify would result in legal liability of some intermediaries given the specific nature of their contracts with investors and their interpretation by the courts.

[35] For example, assume that over the course of the crisis about a third of non-agency securitized loans will be foreclosed and that, in line with our estimates, this rate would be 60% higher for misrepresented loans. This would result in foreclosure of misrepresented mortgages with about $85 billion of outstanding principal. Assuming that the recovery rate on these loans in the event of foreclosure would be equal to about 25% of outstanding loan balance, capturing declines in house prices and some deadweight costs of foreclosures, the forced repurchases (at par) of these loans would result in the overall net loss to financial intermediaries of more than $60 billion.

[36] We also note that ex post, once investors become aware of asset misrepresentation, a large degree of such activity can lead to market shutdowns, potentially limiting sources of financing to valuable investment projects. See Akerlof (1970) and Akerlof and Romer (1993) for a discussion.

[37] See Rajan et al (2012) for evidence that rating models, because they ignore the changed nature of intermediation, may have been generating systematic errors well before the crisis.

investors is critical for constructing measures of misrepresentation. As discussed in detail, our extensive analysis suggests that the misrepresentations we identify are not likely to be an artifact of matching error between the datasets that were merged by a credit bureau.

We note that our inferences are based on a sample of loans that does not include all non-agency securitized mortgages. While the sample used in our study is sizeable (roughly 2 million loans) and is quite comparable to the overall sample of mortgages (see Appendix A for details), it is possible that the intensity of asset misrepresentation -- for instance, across states or across underwriters -- in the population could differ from the one we document. However, our analysis does establish that a significant level of asset quality misrepresentation already exists when examining just two dimensions of asset quality.

It is possible that underwriters may have not known about some aspects of the misrepresentation that occurred at the level of the originator. Related, one could argue that contractual disclosures may have included statements that limit the responsibility of underwriters regarding the quality of disclosed information. For example, in the case of owner-occupancy status, underwriters could argue that they were merely reporting the intent to occupy the property as disclosed by the borrower. In fact, some pool prospectus documents state that borrower's disclosure could serve as the basis for the owner-occupancy status of a loan and that there may be some leeway in the classification of the borrower as owner-occupant (e.g., some investor properties could be classified as owner-occupied properties if borrowers reside in them for considerable part of the year). Such arguments are harder to apply to our second measure of misrepresentation (misreported second liens) since, as argued earlier, it involves information distortion within the boundaries of the financial industry. Regardless, by classifying misrepresentation to an underwriter, we are implicitly taking a stand that these firms, as part of the process of sale, should have collected and verified information regarding the quality of the underlying collateral backing these securities.

Finally our findings that the propensity to misrepresent loans is pervasive among reputable firms and is largely unrelated to underwriter level measures such as incentives for top management or quality of risk management inside these firms should not be taken to imply that the quality of lending and intermediation is unrelated to these factors. We explored the relation between the extent of misrepresentation and the equilibrium level of these factors in the data. It is, of course, possible that the variation in these factors across the firms in our data may be insufficient to identify such a relationship. Moreover, while we have focused on one specific aspect of lending and intermediation, there remain several other functions, some of which have been shown by prior literature to be related to these factors (e.g., see Keys et al. 2009 and Elul et al. forthcoming for screening during origination by lenders and for intermediary risk taking, respectively). Why would some activities of intermediaries respond to governance mechanisms while leaving others relatively unaffected? We think answering this requires understanding the internal organization

and governance of financial intermediaries. In particular, as we showed, intermediaries with misrepresented loans would face some ex-post liability due to contractual representations and warranties. However, employees inside firms who have control over decisions, such as whether to misrepresent, may not directly face this liability. Consequently, as is emphasized in Akerlof and Romer (1993), some employees in control might take actions that transfer wealth from the firm to themselves. Understanding whether, and the extent to which, liability at the intermediary level impacts actions of various decision makers inside the firm requires more inquiry.

References

Acharya, Viral, Philipp Schnabl, and Gustavo Suarez, forthcoming Securitization without risk transfer, *Journal of Financial Economics*.

Agarwal, Sumit, and Calvin T. Ho, 2007, Comparing the prime and subprime mortgage markets, *Chicago Fed Letter*, Number 241.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff, 2011, The role of securitization in mortgage renegotiation, *Journal of Financial Economics* 102, 559-578.

Agarwal, Sumit, David Lucca, Amit Seru, and Francesco Trebbi, 2012, Inconsistent Regulators: Evidence from Banking, *Working paper*

Akerlof, George A., 1970, The Market for ”lemons”: Quality uncertainty and the market mechanism, *Quarterly Journal of Economics* 84, 488–500.

Akerlof, George A., and Paul M. Romer. 1993, Looting: The economic underworld of bankruptcy for profit. *Brookings Papers on Economic Activity* 2, 1-74.

Amromin, Gene, Jennifer Huang, Clemens Sialm, and Edward Zhong, 2011, Complex mortgages, NBER Working Paper 17315.

Ben-David, Itzhak, 2011, Financial constraints and inflated home prices during the real-estate boom, *American Economic Journal: Applied Economics* 3, 55-78.

Berndt, Antje, Burton Hollifield, and Patrik Sandås, 2010, The role of mortgage brokers in the subprime crisis, NBER WP 16175.

Burns, Natasha, and Simi Kedia, 2006, The impact of performance-based compensation on misreporting, *Journal of Financial Economics* 79, 35-67.

Campbell, John Y., Stefano Giglio, and Parag Pathak, forthcoming, Forced sales and house prices, *American Economic Review*.

Cheng, Ing-Haw, Harrison Hong, and Jose Scheinkman, 2010, Yesterday's heroes: Compensation and creative risk-taking, working paper.

Chinco, Alexander, and Chris Mayer, 2012, Distant speculators and asset bubbles in the housing market, working paper.

Choi, Stephen J., Karen K. Nelson, and A. C. Pritchard, 2009, The screening effect of the securities litigation reform act, *Journal of Empirical Legal Studies* 6, 35-68.

Demiroglu, C., and Christopher C. James, 2012, How important is having skin in the game? Originator-sponsor affiliation and losses on Mortgage-Backed Securities, *Review of Financial Studies* 25, 3217-3258.

Demyanyk, Yuliya, and Otto Van Hemert, 2011, Understanding the subprime mortgage crisis, *Review of Financial Studies* 24, 1848-1880.

Dyck, Alexander, Adair Morse, and Luigi Zingales, forthcoming, Who blows the whistle on corporate fraud? *Journal of Finance*.

Efendi, Jap, Anup Srivastava, and Edward Swanson, 2007, Why do corporate managers misstate financial statements? The role of in-the-money options and other incentives, *Journal of Financial Economics* 85, 667-708.

Ellul, Andrew, and Vijay Yerramilli, forthcoming, Stronger risk controls, lower risk: Evidence from US bank holding companies, *Journal of Finance*.

Faltin-Traeger, Olivier, Kathleen Johnson, and Chris Mayer, 2010, Issuer credit quality and the price of asset-backed securities, *American Economic Review Papers and Proceedings* 100, 501-505.

Garmaise, Mark, 2012, Borrower misrepresentation and loan performance, working paper.

Griffin, Paula, Joseph Grundfest, and Micael Perino, 2001, Stock price response to news of securities fraud litigation: Market efficiency and the slow diffusion of costly information, *Stanford Law and Economics Olin Working Paper* 208.

He, Jie, Jun 'QJ' Qian, and Philip E. Strahan, forthcoming, Are all ratings created equal? The impact of issuer size on the pricing of mortgage-backed securities, *Journal of Finance*.

Haughwout, Andrew, Donghoon Lee, Joseph Tracy, and Wilbert van der Klaauw, 2011, Real estate investors, the leverage cycle, and the housing market crisis, *Federal Reserve Bank of New York Staff Reports,* 514.

Jiang, W., Ashlyn A. Nelson, and Edward Vytlacil, forthcoming, . Liar's loan? Effects of origination channel and information falsification on mortgage delinquency, *Review of Economics and Statistics*.

Kedia, Simi, and Thomas Phillipon, 2009, The economics of fraudulent accounting, *Review of Financial Studies* 22, 2169-2199.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2009, Financial regulation and securitization: Evidence from subprime loans, *Journal of Monetary Economics* 56, 700-720.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2010, Did securitization lead to lax screening: Evidence from subprime loans, *Quarterly Journal of Economics* 125, 307-362.

Keys, Benjamin J., Tomasz Piskorski, Amit Seru, and Vikrant Vig, forthcoming, Mortgage financing in the housing boom and bust, in *Housing and the Financial Crisis*, Edward Glaeser and Todd Sinai, editors, NBER and University of Chicago Press.

Keys, Benjamin J., Amit Seru, and Vikrant Vig, forthcoming, Lender screening and role of securitization: Evidence from prime and subprime mortgage markets, *Review of Financial Studies*.

Lee, Donghoon, Christopher Mayer, Joseph Tracy, forthcoming, A new look at second liens, in the *Housing and the Financial Crisis*, Edward Glaeser and Todd Sinai, editors, NBER and University of Chicago Press.

Loutskina, Elena, and Philip E. Strahan, 2009, Securitization and the declining impact of bank finance on loan supply: Evidence from mortgage acceptance rates, *Journal of Finance* 64, 861-889.

Loutskina, Elena, and Philip E. Strahan, 2010, Informed and uninformed investment in housing: The downside of diversification, forthcoming in the *Review of Financial Studies.*

Mayer, Christopher, Karen Pence, and Shane Sherlund, 2009, The rise in mortgage defaults, *Journal of Economic Perspectives* 23, 27-50.

Melzer, Brian, 2010, Debt overhang: Reduced investment by homeowners with negative equity, working paper.

Mian, Atif, and Amir Sufi, 2009, The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis, *Quarterly Journal of Economics* 124, 1449-1496.

Mian, Atif, and Amir Sufi, 2011, House prices, home equity-based borrowing, and the U.S. household leverage crisis, *American Economic Review* 101, 2132-56.

Mian, Atif, Amir Sufi, and Francesco Trebbi, 2010, Foreclosures, house prices, and the real economy, working paper.

Nadauld, Taylor D., and Shane M. Sherlund, 2013, The impact of securitization on the expansion of subprime credit, *Journal of Financial Economics* 107, 454-476.

Povel, Paul, Rajdeep Singh, and Andrew Winton, 2007, Booms, busts, and fraud, *Review of Financial Studies* 20, 1219-1254.

Piskorski, Tomasz, and Alexei Tchistyi, 2010, Optimal mortgage design, *Review of Financial Studies* 23, 3098-3140.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig, 2010, Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis, *Journal of Financial Economics* 97, 369-397.

Rajan, Uday, Amit Seru, and Vikrant Vig, V., 2012, The failure of models that predict failure: Distance, incentives and defaults, *Journal of Financial Economics*, forthcoming.

Purnanandam, A., forthcoming, Originate-to-distribute model and the subprime mortgage crisis. *Review of Financial Studies*.

Richardson, Scott, Irem Tuna, and Min Wu, 2002, Predicting earnings management: The case of earnings restatements, working paper.

Wang, Tracy Yue, forthcoming, Corporate securities fraud: Insights from a new empirical framework, *Journal of Law, Economics and Organization*.

Wang, Tracy Yue, Andrew Winton, and Xiaoyun Yu, 2010, Corporate fraud and business conditions: Evidence from IPOs, *Journal of Finance* 65, 2255-2292.

Table 1: Descriptive Statistics and Percent of Misrepresented Loans

The table presents summary statistics of key variables for mortgages reported as owner-occupied (Panel A) and loans reported as having no second liens (Panel B) to the RMBS trustee (loans reported as such in the BlackBox dataset). The sample consists of these loans merged with high confidence level with the credit bureau data. *Interest Rate* is the loan interest rate at origination in percentage terms. *FICO* is the variable capturing the borrower's FICO credit score at loan origination. *Balance* is the initial loan balance (in thousands of dollars). *CLTV* is the loan's origination combined loan-to-value ratio in percentage terms. *No Cash Out Refi* and *Cash Out Refi* are dummies that take a value of 1 if the loan purpose was a no cash out refinancing or cash out refinancing, respectively, and are 0 otherwise. *Low or No Doc.* is a dummy that takes a value of 1 if the loan was originated with no or limited documentation, and is 0 otherwise. *ARM* and *Option ARM* are dummies that take a value of 1 if the loan type was an ARM or option ARM, respectively, and are 0 otherwise.

Panel A: Sample of Loans Reported for Owner-Occupied Properties

	Mean	SD
Interest Rate	6.60	2.02
FICO	680.	73.20
Balance	324.6	248.3
CLTV	83.72	11.71
Purchase	0.49	0.50
No Cash Out Refi	0.13	0.34
Cash Out Refi	0.35	0.48
Low or No Doc.	0.46	0.49
ARM	0.52	0.49
Option ARM	0.10	0.30
Number of Loans	1,563,223	
Percent of Loans for Misreported Non-Owner-Occupants	6.42	
Percent of Fully Documented Loans for Misreported Non-Owner-Occupants	4.79	

Panel B: Sample of Loans Reported as Having No Second Liens

	Mean	SD
Interest Rate	6.45	2.42
FICO	663.5	76.90
Balance	293.1	238.9
CLTV	78.93	9.95
Purchase	0.36	0.48
No Cash Out Refi	0.12	0.32
Cash Out Refi	0.50	0.50
Low or No Doc.	0.51	0.50
ARM	0.47	0.49
Option ARM	0.16	0.37
Number of Loans	854,959	
Percent of Loans with Misreported Second Liens	7.13	
Percent of Fully Documented Loans with Misreported Second Liens	7.93	
Percent of Loans with Misreported Second Liens (including HELOCs)	13.64	

Table 2: Misrepresentations by Home Price Changes and Loan Volume

The table presents the OLS estimates from regressions where the dependent variable is the mean of the given asset misrepresentation at the MSA level. Column (1) shows the results for misreported owner-occupants, Column (2) shows the results for misreported second liens, and Column (3) shows the results for either misrepresentation. *Annualized HPI Growth, '01–'04* and *Annualized HPI Growth, '05–'06* are variables representing the annualized MSA-level HPI growth from 2001 to 2004 and 2005 to 2006, respectively, computed using FHFA's HPI measure. *Share of 2005–2007 Loan Originations* is a variable defined as the number of loans originated in the MSA from 2005 to 2007 divided by the total number of loans originated over this time period, using the overall BlackBox loan sample. The estimates show the effect of a one standard deviation change and are in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1) Misreported Non-Owner-Occupant	(2) Misreported Second Lien	(3) Either Misreported
Annualized HPI Growth, '01–'04	0.869***	-1.432***	0.0575
	(7.79)	(-7.72)	(0.43)
Annualized HPI Growth, '05–'06	0.486***	-0.241	0.323**
	(4.61)	(-1.37)	(2.54)
Share of 2005–2007 Loan Originations	0.365***	0.628***	0.509***
	(3.62)	(3.75)	(4.20)
Number of MSAs	381	381	381
Percent Misrepresented	4.604	7.513	7.929
R-squared	0.364	0.188	0.0929

Table 3: Misrepresentations and Loan Characteristics

The table presents the OLS estimates from regressions where dependent variable takes a value of 1 if the loan has the given misrepresentation, and is 0 otherwise. Panel A shows the results with *Misreported Non-Owner-Occupant* as the dependent variable, while Panel B shows results with *Misreported Second Lien* as the dependent variable. Column (2) includes controls for the level of merge confidence expressed by Equifax. Column (3) includes loan origination half-year fixed effects with 2005 omitted. The estimates are in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Misreported Non-Owner-Occupant

	(1)	(2)	(3)
Interest Rate	0.821***	0.812***	0.786***
	(55.80)	(55.22)	(49.00)
FICO	0.00922***	0.00907***	0.00902***
	(24.97)	(24.57)	(24.27)
Balance	0.00250***	0.00262***	0.00256***
	(28.60)	(30.00)	(28.73)
CLTV	0.0286***	0.0274***	0.0267***
	(14.27)	(13.70)	(13.34)
No Cash Out Refi	-5.497***	-5.499***	-5.536***
	(-89.01)	(-89.12)	(-89.40)
Cash Out Refi	-6.214***	-6.266***	-6.286***
	(-125.34)	(-126.46)	(-126.74)
Low or No Doc.	2.778***	2.824***	2.821***
	(65.74)	(66.86)	(66.70)
ARM	2.279***	2.153***	2.173***
	(52.14)	(49.22)	(49.41)
Option ARM	7.295***	7.229***	7.126***
	(79.65)	(78.99)	(74.90)
Confidence Match Controls	No	Yes	Yes
Half-Year Origination Cohorts	No	No	Yes
Number of Loans	1,563,223	1,563,223	1,563,223
Percent Misrepresented	6.426	6.426	6.426
R-squared	0.0283	0.0300	0.0301

Table 3 (Continued):

Panel B: Misreported Second Lien

	(1)	(2)	(3)
Interest Rate	-0.624***	-0.619***	-0.548***
	(-30.89)	(-30.65)	(-25.13)
FICO	0.00418***	0.00449***	0.00523***
	(8.94)	(9.60)	(11.11)
Balance	-0.00804***	-0.00810***	-0.00773***
	(-65.46)	(-65.92)	(-62.17)
CLTV	-0.163***	-0.163***	-0.161***
	(-53.35)	(-53.48)	(-52.84)
No Cash Out Refi	-8.181***	-8.202***	-7.990***
	(-89.21)	(-89.45)	(-86.84)
Cash Out Refi	-12.01***	-12.03***	-11.93***
	(-186.03)	(-186.38)	(-184.70)
Low or No Doc.	-1.487***	-1.501***	-1.504***
	(-25.08)	(-25.33)	(-25.35)
ARM	6.355***	6.406***	6.397***
	(99.27)	(99.97)	(99.92)
Option ARM	-2.825***	-2.797***	-2.630***
	(-23.57)	(-23.34)	(-21.09)
Confidence Match Controls	No	Yes	Yes
Half-Year Origination Cohorts	No	No	Yes
Number of Loans	854,959	854,959	854,959
Percent Misrepresented	7.131	7.131	7.131
R-squared	0.0643	0.0646	0.0668

Table 4: Time-Series Evolution of Misrepresentations

The table presents the OLS coefficient estimates for the half-year origination vintage fixed effects included in Column (3) of Table 3. The excluded categories are loans originated in 2005. The dependent variable takes a value of 1 if the loan has the given misrepresentation, and is 0 otherwise. Column (1) shows the results with *Misreported Non-Owner-Occupant* as the dependent variable, while Column (2) shows results with *Misreported Second Lien* as the dependent variable. The estimates are in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1) Misreported Non-Owner-Occupant	(2) Misreported Second Lien
Originated in 2006H1	0.0126	2.213***
	(0.24)	(29.72)
Originated in 2006H2	0.183***	-0.444***
	(3.21)	(-5.58)
Originated in 2007H1	0.456***	-1.820***
	(6.86)	(-18.62)
Originated in 2007H2	-0.773***	-3.993***
	(-4.58)	(-15.23)
Number of Loans	1,563,223	854,959
2005 Mean Percent Misrepresented	5.90	7.73

Table 5: Misrepresentations and Loan Default

The table presents the OLS estimates from regressions where the dependent variable takes a value of 1 if the mortgage ever defaults (ever goes 90 days past due on payments) in the first two years since origination, and is 0 otherwise. Panel A shows the results with *Misreported Non-Owner-Occupant* as a control variable. Panel B shows results with *Misreported Second Lien* as a control variable. Panel C shows the results with both *Misreported Non-Owner-Occupant* and *Reported Non-Owner-Occupant* as control variables, where the excluded category is loans truthfully reported as for owner-occupied properties. Panel D shows the results with *Misreported Second* x *CLTV >= 100, Misreported Second* x *CLTV < 100, Reported Second* x *CLTV >= 100,* and *Reported Second* x *CLTV < 100* as control variables, where the excluded category are loans that truthfully reported no second liens. The *CLTV* term in these interactions takes a value of 1 if the loan has a CLTV ratio in the appropriate range, and is 0 otherwise. "Other Controls" include origination variables used in Table 3 such as FICO, interest rates, and LTV ratios. Squared and cubed terms for FICO and CLTV ratios are also included to account for potential nonlinear effects. The estimates are in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Mortgage Default and Misreported Non-Owner-Occupant

	All Years	2005	2006H1	2006H2	2007H1	2007H2
Misreported Non-Owner-Occupant	9.35***	4.04***	10.13***	15.58***	14.00***	14.38***
	(85.49)	(32.73)	(40.66)	(51.35)	(36.35)	(15.25)
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes
Half-Year Origination Cohorts	Yes	No	No	No	No	No
Number of Loans	1,563,223	743,827	327,128	282,904	187,352	22,012
Percent Default	15.28	7.150	17.02	26.51	27.36	17.11
Percent Misrepresented	6.426	5.904	6.903	7.167	6.535	6.528
R-squared	0.157	0.0727	0.108	0.149	0.184	0.214

Panel B: Mortgage Default and Misreported Second Lien

	All Years	2005	2006H1	2006H2	2007H1	2007H2
Misreported Second Lien	10.15***	7.04***	13.40***	16.14***	16.72***	6.65**
	(70.38)	(46.29)	(42.46)	(34.49)	(22.17)	(2.36)
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes
Half-Year Origination Cohorts	Yes	No	No	No	No	No
Number of Loans	854,959	445,432	165,183	146,601	88,207	9,536
Percent Default	14.58	7.427	16.00	24.28	30.79	24.55
Percent Misrepresented	7.131	7.730	9.011	5.535	3.788	1.992
R-squared	0.160	0.0755	0.126	0.151	0.183	0.207

Table 5 (Continued):

Panel C: Mortgage Default and Misreported versus Reported Non-Owner Occupants

	(1)	(2)	(3)	(4)
Misreported Non-Owner-Occupant	10.72***	9.992***	9.48***	9.48***
	(96.16)	(92.23)	(87.74)	(13.39)
Reported Non-Owner-Occupant	4.02***	3.57***	3.45***	3.45***
	(53.29)	(48.64)	(46.88)	(5.60)
Other Controls	Yes	Yes	Yes	Yes
Half-Year Origination Cohorts	No	Yes	Yes	Yes
State Fixed Effects	No	No	Yes	Yes
SEs Clustered by State	No	No	No	Yes
Number of Loans	1,827,497	1,827,497	1,827,497	1,827,497
Percent Default	14.89	14.89	14.89	14.89
Percent Misrepresented	6.424	6.424	6.424	6.424
R-Squared	0.103	0.153	0.161	0.161

Panel D: Mortgage Default and Misreported versus Reported Second Lien

	(1)	(2)	(3)	(4)
Misreported Second x CLTV < 100	5.075***	5.725***	6.404***	6.404***
	(17.68)	(20.51)	(23.07)	(13.23)
Reported Second x CLTV < 100	8.627***	6.252***	6.636***	6.636***
	(70.28)	(52.22)	(55.65)	(7.09)
Misreported Second x CLTV >= 100	9.618***	10.19***	10.97***	10.97***
	(48.16)	(52.44)	(56.73)	(7.94)
Reported Second x CLTV >= 100	15.50***	13.26***	14.02***	14.02***
	(139.00)	(121.55)	(128.55)	(6.08)
Other Controls	Yes	Yes	Yes	Yes
Half-Year Origination Cohorts	No	Yes	Yes	Yes
State Fixed Effects	No	No	Yes	Yes
SEs Clustered by State	No	No	No	Yes
Number of Loans	1,109,250	1,109,250	1,109,250	1,109,250
Percent Default	17.10	17.10	17.10	17.10
Percent Misrepresented	5.212	5.212	5.212	5.212
R-Squared	0.113	0.162	0.172	0.172

Table 6: Misrepresentations and Defaults on Credit Cards

The table presents the OLS estimates from regressions where the dependent variable takes a value of 1 if the borrower ever defaults on credit card debt (goes 60 days past due on payments) during the first two years since mortgage origination, and is 0 otherwise. Panel A shows the results with *Misreported Non-Owner-Occupant* and *Reported Non-Owner-Occupant* as control variables, while Panel B displays results with *Misreported Second* x *CLTV >= 100, Misreported Second* x *CLTV < 100, Reported Second* x *CLTV >= 100,* and *Reported Second* x *CLTV < 100* as control variables. Column (1) includes FICO and controls related to credit card debt, such as utilization and monthly payments. Column (2) incorporates state fixed effects. Column (3) includes the full mortgage controls, such as loan balance and interest rate, used in Table 3. Column (4) clusters standard errors by state. The estimates are in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Credit Card Default and Misreported versus Reported Non-Owner-Occupants

	(1)	(2)	(3)	(4)
Misreported Non-Owner-Occupant	7.636***	7.305***	5.959***	5.959***
	(49.02)	(46.98)	(38.51)	(18.68)
Reported Non-Owner-Occupant	0.433***	1.144***	-1.024***	-1.024***
	(4.25)	(11.12)	(-9.69)	(-2.81)
Other Controls	Yes	Yes	Yes	Yes
State Fixed Effects	No	Yes	Yes	Yes
Mortgage Controls	No	No	Yes	Yes
SEs Clustered by State	No	No	No	Yes
Number of Loans	1,709,386	1,709,386	1,709,386	1,709,386
Percent Default	37.28	37.28	37.28	37.28
Percent Misrepresented	6.580	6.580	6.580	6.580
R-Squared	0.0760	0.0831	0.112	0.112

Table 6 (Continued):

Panel B: Credit Card Default and Misreported versus Reported Second Lien

	(1)	(2)	(3)	(4)
Misreported Second x CLTV < 100	-3.105***	-2.640***	-1.814***	-1.814***
	(-8.09)	(-6.91)	(-4.77)	(-3.07)
Reported Second x CLTV < 100	-2.759***	-2.462***	-1.601***	-1.601***
	(-17.63)	(-15.76)	(-9.84)	(-3.83)
Misreported Second x CLTV >= 100	-2.462***	-1.155***	-2.123***	-2.123***
	(-9.37)	(-4.41)	(-7.83)	(-3.20)
Reported Second x CLTV >= 100	-2.023***	-0.696***	-0.854***	-0.854
	(-16.27)	(-5.56)	(-5.63)	(-1.40)
Other Controls	Yes	Yes	Yes	Yes
State Fixed Effects	No	Yes	Yes	Yes
Mortgage Controls	No	No	Yes	Yes
SEs Clustered by State	No	No	No	Yes
Number of Loans	1,023,725	1,023,725	1,023,725	1,023,725
Percent CC 60 DPD	38.36	38.36	38.36	38.36
Percent Misrepresented	5.233	5.233	5.233	5.233
R-Squared	0.0769	0.0859	0.109	0.109

Table 7: Misrepresentation and Mortgage Interest Rates at Origination

The table presents the OLS estimates from regressions where the dependent variable is the mortgage's interest rate at origination. Panel A shows the results with *Misreported Non-Owner-Occupant* and *Reported Non-Owner-Occupant* as control variables. Panel B shows results with *Misreported Second* x *CLTV >= 100*, *Misreported Second* x *CLTV < 100*, *Reported Second* x *CLTV >= 100*, and *Reported Second* x *CLTV < 100* as control variables. "Other Controls" include origination variables used in Table 3, such as FICO, interest rates, and CLTV ratios. Column (1) includes only these variables. Column (2) incorporates half-year origination fixed effects. Column (3) also includes property state fixed effects. Column (4) clusters standard errors by state. The estimates are in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Interest Rates at Origination and Misreported versus Reported Non-Owner-Occupants

	(1)	(2)	(3)	(4)
Misreported Non-Owner-Occupant	0.258***	0.212***	0.229***	0.229***
	(60.45)	(54.38)	(59.01)	(41.94)
Reported Non-Owner-Occupant	0.393***	0.365***	0.351***	0.351***
	(136.13)	(137.82)	(132.93)	(17.30)
Other Controls	Yes	Yes	Yes	Yes
Half-Year Origination Cohorts	No	Yes	Yes	Yes
State Fixed Effects	No	No	Yes	Yes
SEs Clustered by State	No	No	No	Yes
Number of Loans	1,827,497	1,827,497	1,827,497	1,827,497
Mean Interest Rate	6.563	6.563	6.563	6.563
Percent Misrepresented	6.424	6.424	6.424	6.424
R-Squared	0.611	0.674	0.680	0.680

Table 7 (Continued):

Panel B: Interest Rates at Origination and Misreported versus Reported Second Lien

	(1)	(2)	(3)	(4)
Misreported Second x CLTV < 100	-0.0281***	-0.00237	-0.00358	-0.00358
	(-2.62)	(-0.24)	(-0.36)	(-0.11)
Reported Second x CLTV < 100	0.267***	0.121***	0.124***	0.124***
	(58.01)	(28.32)	(29.12)	(3.16)
Misreported Second x CLTV >= 100	0.0721***	0.0956***	0.0882***	0.0882
	(9.64)	(13.80)	(12.80)	(1.04)
Reported Second x CLTV >= 100	0.441***	0.295***	0.290***	0.290**
	(105.74)	(75.92)	(74.52)	(2.32)
Other Controls	Yes	Yes	Yes	Yes
Half-Year Origination Cohorts	No	Yes	Yes	Yes
State Fixed Effects	No	No	Yes	Yes
SEs Clustered by State	No	No	No	Yes
Number of Loans	1109250	1109250	1109250	1109250
Mean Interest Rate	6.654	6.654	6.654	6.654
Percent Misrepresented	5.212	5.212	5.212	5.212
R-Squared	0.624	0.677	0.682	0.682

Table 8: Misrepresentations and Yields: Pool Level Regressions

The table presents the OLS estimates from regressions where the dependent variable is the pool's issuance yield spread (in percentage terms), defined as the weighted average issuance coupon for the pool minus the U.S. Treasury yield at issuance. Panel A shows the results with *Percent Misreported Non-Owner-Occupant* as a control variable. *Percent Misreported Non-Owner-Occupant* is defined as the number of loans in the pool identified as being backed by misreported non-owner-occupied properties divided by the number of loans in the pool reported as for owner-occupied properties. Panel B shows the results with *Percent Misreported Second* as a control variable. *Percent Misreported Second* is defined as the number of loans in the pool identified as having a misreported second lien divided by the number of loans in the pool that report that no second lien is present. Panel C shows the results with *Percent Either Misreported* as a control variable. *Percent Either Misreported* is defined as the number of loans in the pool identified as having either misrepresentation divided by the number of loans in the pool reported as for owner-occupied properties and/or reported as having no second lien. Other controls include the pool-level means of origination variables used in Table 3, such as FICO and CLTV ratios, although loan interest rates are excluded. Column (1) includes only these controls. Column (2) incorporates half-year vintage fixed effects. Column (3) includes controls for the overcollateralization of the pool. Column (4) includes underwriter fixed effects for the six underwriters with twenty or more pools in our sample. The estimates are in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Pool Yield Spreads and Percent Misreported Non-Owner-Occupants

	(1)	(2)	(3)	(4)
Percent Misreported Non-Owner-Occupants	0.0211*	0.0209*	0.0204*	0.0189
	(1.81)	(1.78)	(1.74)	(1.57)
Half-Year FE	No	Yes	Yes	Yes
Overcollateralization	No	No	Yes	Yes
Top 6 Underwriter FE	No	No	No	Yes
Other Controls	Yes	Yes	Yes	Yes
Number of Deals	353	353	353	353
Mean Yield Spread	0.830	0.830	0.830	0.830
Percent Misrepresented	6.473	6.473	6.473	6.473
R-squared	0.268	0.268	0.273	0.281

Table 8 (Continued):

Panel B: Pool Yield Spreads and Percent Misreported Second Lien

	(1)	(2)	(3)	(4)
Percent Misreported Second	0.00342	0.00373	0.00416	0.00339
	(0.79)	(0.85)	(0.94)	(0.73)
Half-Year FE	No	Yes	Yes	Yes
Overcollateralization	No	No	Yes	Yes
Top 6 Underwriter FE	No	No	No	Yes
Other Controls	Yes	Yes	Yes	Yes
Number of Deals	353	353	353	353
Mean Yield Spread	0.830	0.830	0.830	0.830
Percent Misrepresented	4.191	4.191	4.191	4.191
R-squared	0.262	0.262	0.269	0.277

Panel C: Pool Yield Spreads and Either Misrepresentation

	(1)	(2)	(3)	(4)
Percent Either Misreported	0.00617	0.00638	0.00669	0.00589
	(1.39)	(1.43)	(1.49)	(1.26)
Half-Year FE	No	Yes	Yes	Yes
Overcollateralization	No	No	Yes	Yes
Top 6 Underwriter FE	No	No	No	Yes
Other Controls	Yes	Yes	Yes	Yes
Number of Deals	353	353	353	353
Mean Yield Spread	0.830	0.830	0.830	0.830
Percent Misrepresented	8.870	8.870	8.870	8.870
R-squared	0.265	0.265	0.272	0.279

Table 9: Misrepresentations and Subordination Level -- Pool Level Analysis

The table presents the OLS estimates from regressions where the dependent variable is the pool's AAA subordination level (in percentage terms), defined as the lowest subordination level given to a tranche in the pool that received a credit rating of AAA. All panels use the same sample of pools used in Table 8. Panel A shows the results with *Percent Misreported Non-Owner-Occupant* as a control variable. Panel B shows the results with *Percent Misreported Second* as a control variable. Panel C shows the results with *Percent Either Misreported* as a control variable. Other controls are defined as in Table 8. Column (1) includes only these controls. Column (2) incorporates half-year vintage fixed effects. Column (3) adds the pool-level average interest rate as a control variable. Column (4) includes controls for the overcollateralization of the pool. Column (5) includes underwriter fixed effects for the six underwriters with twenty or more pools in our sample. The estimates are in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Pool AAA Subordination and Percent Misreported Non-Owner-Occupants

	(1)	(2)	(3)	(4)	(5)
Percent Misreported Non-Owner-Occupant	-0.0512	-0.0401	-0.0597	-0.0305	-0.0225
	(-0.82)	(-0.64)	(-0.95)	(-0.51)	(-0.37)
Half-Year FE	No	Yes	Yes	Yes	Yes
Interest Rate	No	No	Yes	Yes	Yes
Overcollateralization	No	No	No	Yes	Yes
Top 6 Underwriter FE	No	No	No	No	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes
Number of Deals	353	353	353	353	353
Mean Subordination	8.444	8.444	8.444	8.444	8.444
Percent Misrepresented	6.473	6.473	6.473	6.473	6.473
R-squared	0.875	0.877	0.878	0.892	0.896

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9 (Continued):

Panel B: Pool AAA Subordination and Percent Misreported Second Lien

	(1)	(2)	(3)	(4)	(5)
Percent Misreported Second	0.0227	0.0157	0.0112	0.0313	0.0140
	(0.98)	(0.67)	(0.48)	(1.41)	(0.60)
Half-Year FE	No	Yes	Yes	Yes	Yes
Interest Rate	No	No	Yes	Yes	Yes
Overcollateralization	No	No	No	Yes	Yes
Top 6 Underwriter FE	No	No	No	No	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes
Number of Deals	353	353	353	353	353
Mean Subordination	8.444	8.444	8.444	8.444	8.444
Percent Misrepresented	4.191	4.191	4.191	4.191	4.191
R-squared	0.875	0.877	0.878	0.893	0.896

Panel C: Pool AAA Subordination and Either Misrepresentation

	(1)	(2)	(3)	(4)	(5)
Percent Either Misreported	0.0167	0.0120	0.00568	0.0270	0.0120
	(0.71)	(0.51)	(0.24)	(1.19)	(0.51)
Half-Year FE	No	Yes	Yes	Yes	Yes
Interest Rate	No	No	Yes	Yes	Yes
Overcollateralization	No	No	No	Yes	Yes
Top 6 Underwriter FE	No	No	No	No	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes
Number of Deals	353	353	353	353	353
Mean Subordination	8.444	8.444	8.444	8.444	8.444
Percent Misrepresented	8.870	8.870	8.870	8.870	8.870
R-squared	0.875	0.877	0.878	0.893	0.896

Table 10: Heterogeneity of Misrepresentation among Underwriters

The table presents the OLS estimates from regressions where the dependent variable takes a value of 1 if the loan has the given violation and is 0 otherwise. Column (1) shows the results with *Misreported Non-Owner-Occupant* as the dependent variable, Column (2) shows results with *Misreported Second* as the dependent variable, and Column (3) shows the results with misrepresentation on either dimension as the dependent variable (*Either Misreported*). *Underwriter RMBS Volume/Assets* is the volume of loans in our base sample underwritten in 2005 by the loan's deal underwriter divided by the underwriter's assets as of the end of 2005. *Underwriter Years in Subprime* is the number of years between the loan's securitization year and the first subprime deal underwritten by the deal's underwriter, excluding deals prior to 1999. *Commercial Bank Underwriter* is a dummy that takes a value of 1 if the underwriter is a commercial bank and is 0 otherwise. *RMI* is an index of the risk-management strength of an institution based on Elul et al. (2012). *Bonus to Salary Ratio* is the ratio of cash bonus to salary for the underwriter's five highest-paid executives. The description of these measures in discussed in Appendix A. "Other Controls" include origination variables used in Column (3) of Table 3, such as FICO, interest rate, and CLTV ratio. Standard errors are clustered by underwriter; the estimates are in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1) Misreported Non-Owner-Occupancy	(2) Misreported Second	(3) Either Misreported
Underwriter RMBS Volume/Assets	1.319*	0.546	1.528
	(1.75)	(0.21)	(1.20)
Underwriter Years in Subprime (Post-1999)	-0.0341	-0.299	-0.200
	(-0.22)	(-0.55)	(-0.93)
Commercial Bank Underwriter	-0.464	-0.0161	-0.110
	(-0.85)	(-0.01)	(-0.09)
RMI	-0.581	-0.434	-0.636
	(-0.72)	(-0.13)	(-0.35)
Bonus to Salary Ratio	-0.00423	-0.0126	-0.0141
	(-0.39)	(-0.24)	(-0.50)
Half-Year Origination Cohorts	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes
Number of Loans	1561625	853954	1700264
Percent Misrepresented	6.426	7.131	9.192
R-squared	0.0304	0.0672	0.0401

Table 11: Misrepresentations among New Century Loans

Panel A examines merged New Century loans identified as having a misrepresentation using the BlackBox-Equifax merged dataset and presents the percentage of these that have the same misrepresentation reported in the New Century dataset. Panels B and C of the table present summary statistics of key variables for New Century loans reported as for owner-occupants and reported as having no second lien to the RMBS trustees, respectively, that have been merged into the New Century dataset.

Panel A: Misrepresentations in New Century Data

	Percentage	Count
Misreported Non-Owner-Occupant Reported as Non-Owner-Occupant in New Century Dataset	0%	148
Misreported Second Lien Reported as Having a Second Lien in New Century Dataset	93.28%	1279

Panel B: Characteristics based on Owner-Occupancy Status in Merged Data

	Reported as Owner-Occupant		Identified by us as Having Misreported Non-Owner-Occupancy	
	Mean	SD	Mean	SD
Default within 2 Years	0.210	0.408	0.345	0.477
Interest Rate	8.222	1.140	8.236	1.269
FICO	617.7	58.12	656.3	67.32
Number of Loans	3,160		148	

Panel C: Characteristics based on Second Lien Status in Merged Data

	Reported as Having No Second Lien		Identified by us as Having Misreported Second Lien	
	Mean	SD	Mean	SD
Default within 2 Years	0.288	0.453	0.403	0.491
Interest Rate	8.409	1.356	7.642	1.082
FICO	594.9	55.65	645.2	41.69
Number of Loans	10,924		1,279	

Figure 1: Misrepresentations by State

Figures 1A and 1B provide mean levels of misrepresentation at the state level. Panel A displays these statistics for *Misreported Non-Owner-Occupant* and Panel B displays these statistics for *Misreported Second Lien* measure.

Figure 1A: Fraction of Loans Reported as Owner-Occupied that Misreport Occupancy Status

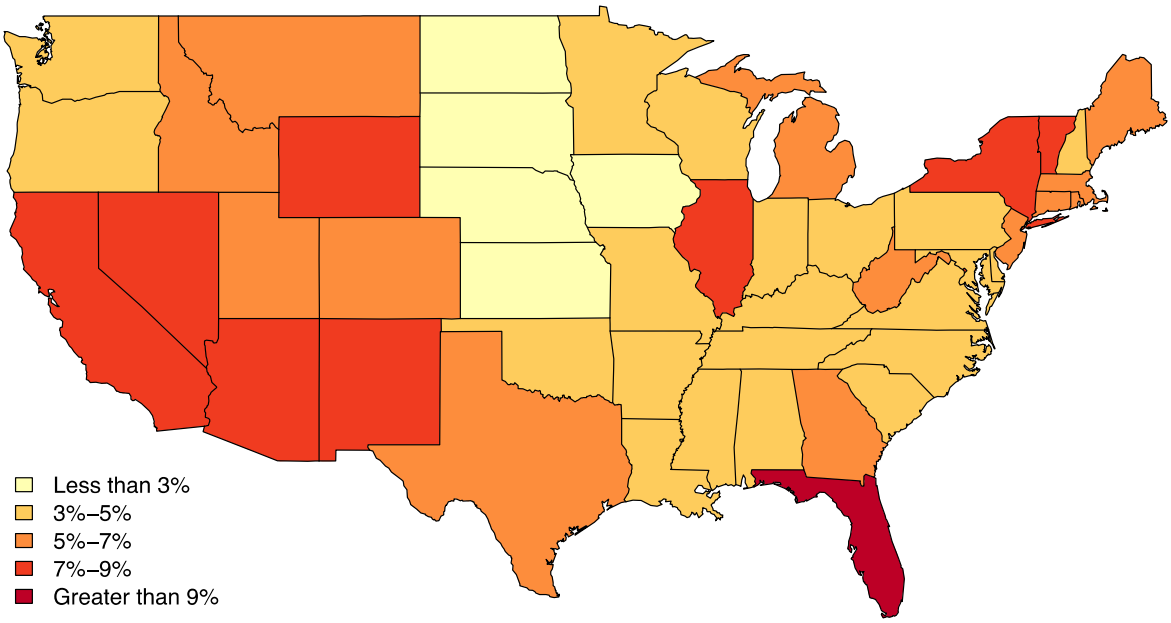


Figure 1B: Fraction of Loans Reported as Having No Second Lien that Misreport Second Lien

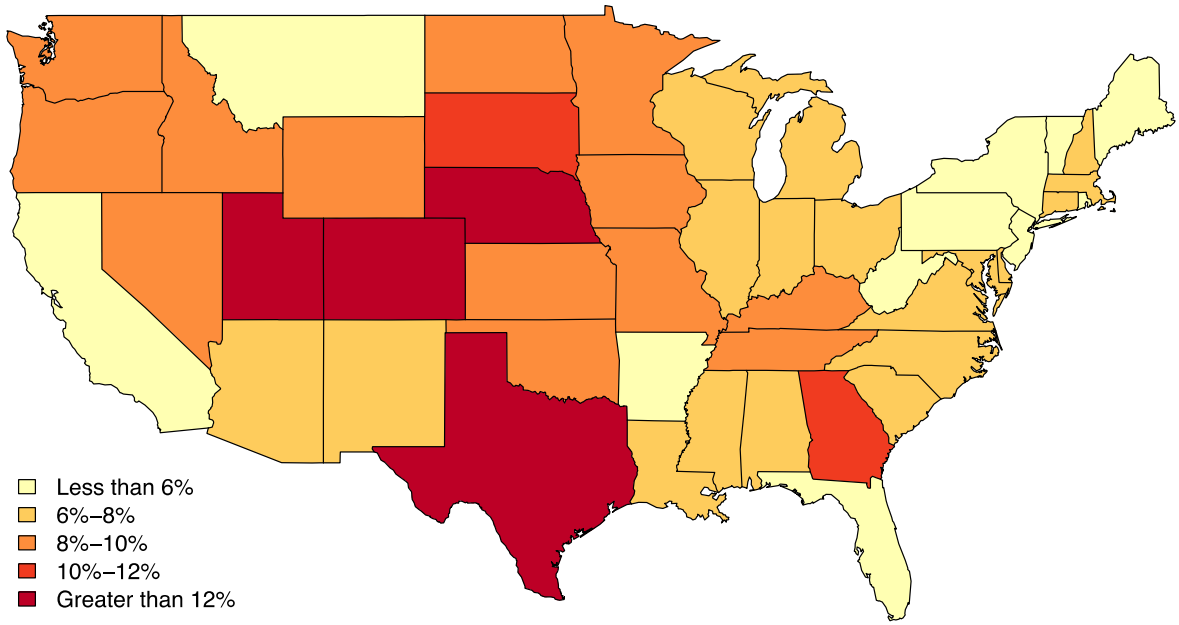
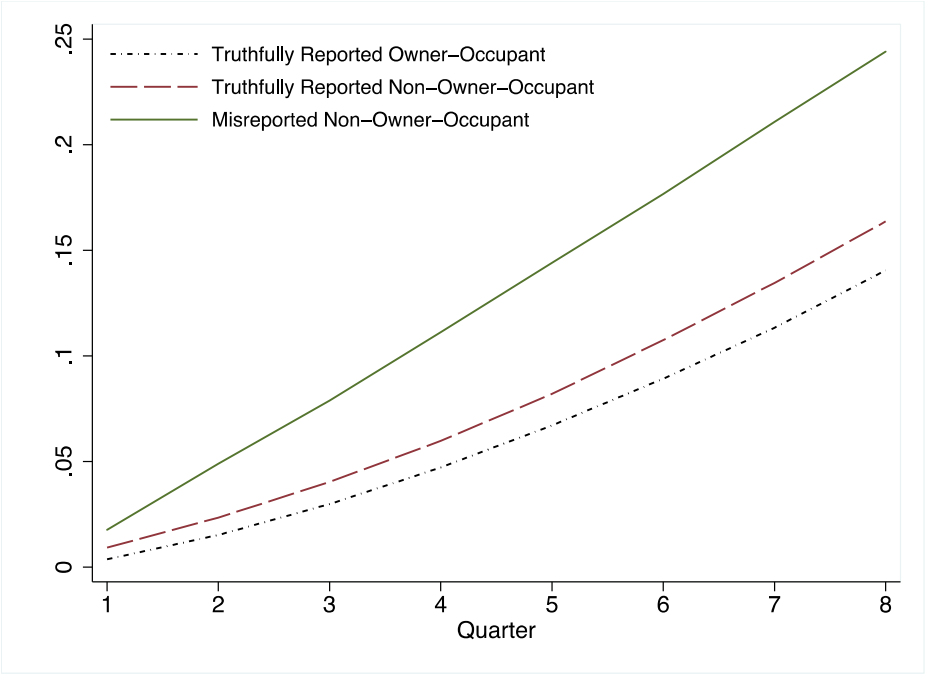


Figure 2: Cumulative Default Rates and Interest Rates

Panels A and B of the figure show the estimated cumulative default rates of a mortgage in the first two years since origination for different subsamples of the overall dataset, holding constant at their overall sample mean all other observables used in Table 2, such as FICO and interest rate. In both panels, the *y*-axis is the cumulative percentage of loans to have gone 90 days past due at least once. Panels C and D show the kernel density plots of the observed difference between actual and predicted origination interest rates for different subsamples of the overall dataset. The predicted rates used in Panel C are given by a specification similar to that used in Column (3) of Table 7A, with the *Reported Non-Owner-Occupant* and *Misreported Non-Owner-Occupant* variables removed. The predicted rates used in Panel D are given by a specification similar to that used in Column (3) of Table 7B, with the *Misreported Second* x *CLTV >= 100*, *Misreported Second* x *CLTV < 100*, *Reported Second* x *CLTV >= 100*, and *Reported Second* x *CLTV < 100* variables removed. In both panels, the *x*-axis shows the error term from these regressions (the difference between the actual origination interest rate and the predicted value).

Panel A: Cumulative Default by Occupancy Status



Panel B: Cumulative Default by Second Lien Status

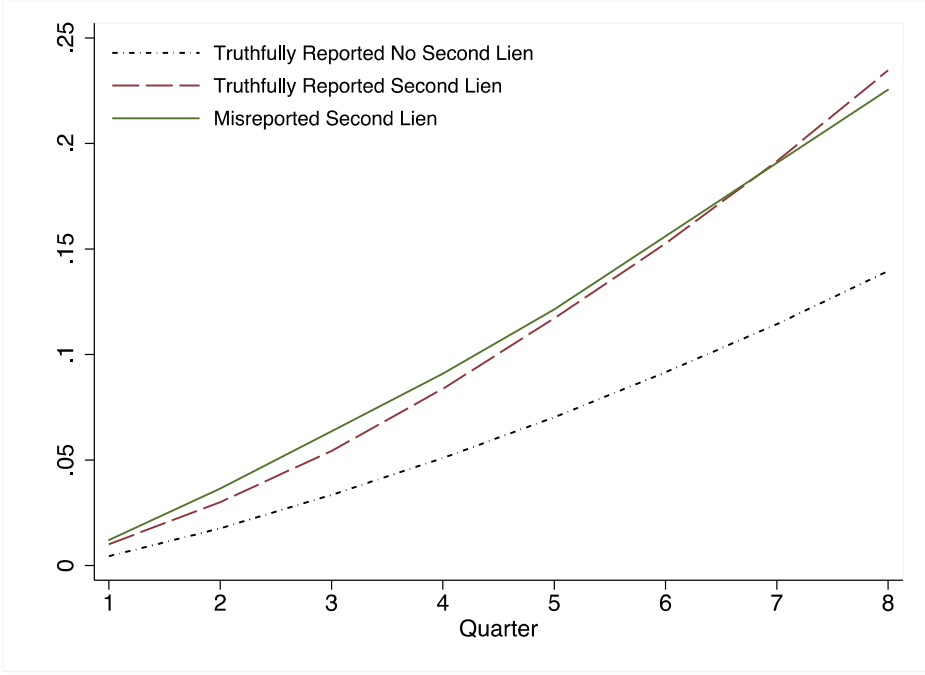
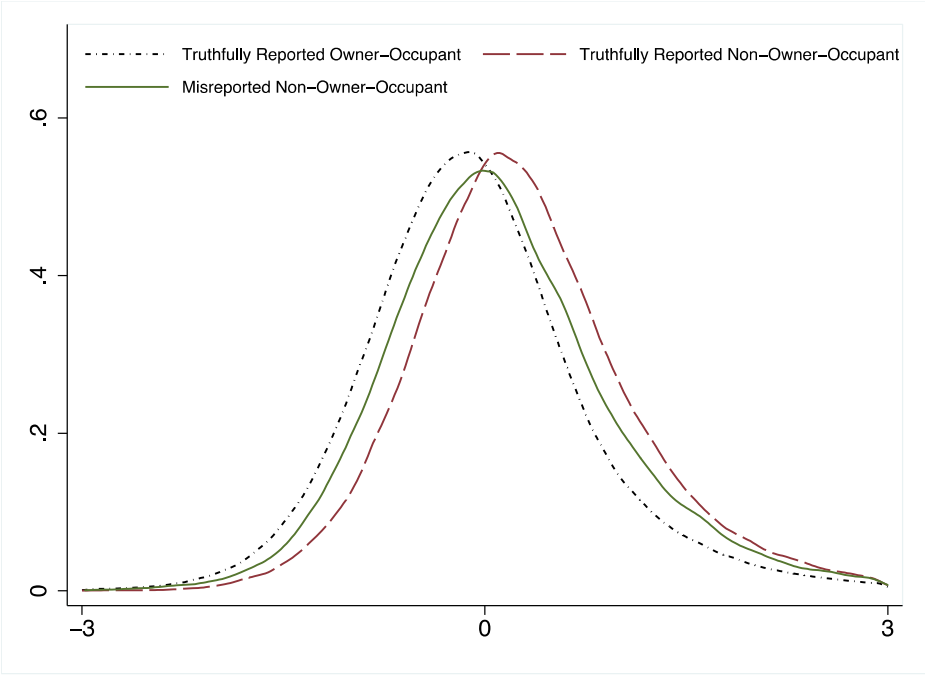


Figure 2 (Continued):

Panel C: Interest Rate Error Terms by Occupancy Status



Panel D: Interest Rate Error Terms by Second Lien Status

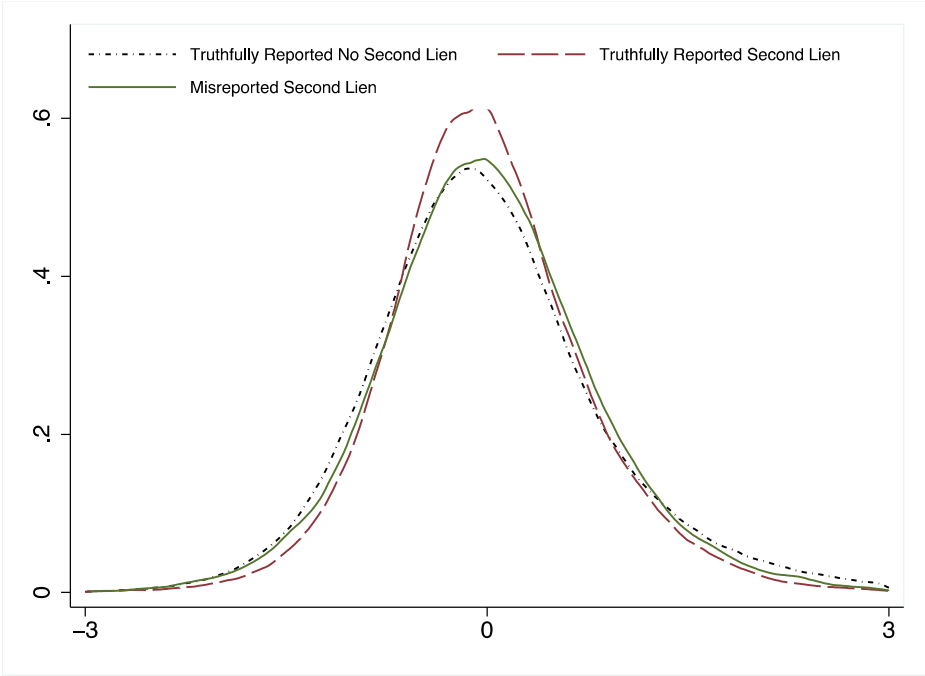


Figure 3: Misreported Non-Owner-Occupants by Underwriter

Figure 3 plots the percentage of misreported non-owner-occupants by underwriter along with a 95% confidence interval. Coefficients result from adding underwriter fixed effects to the specification in Column (3) of Table 2A. These levels are obtained by adding each underwriter fixed effect to the level of misrepresentation for the omitted category (Credit Suisse) with other covariates at their means.

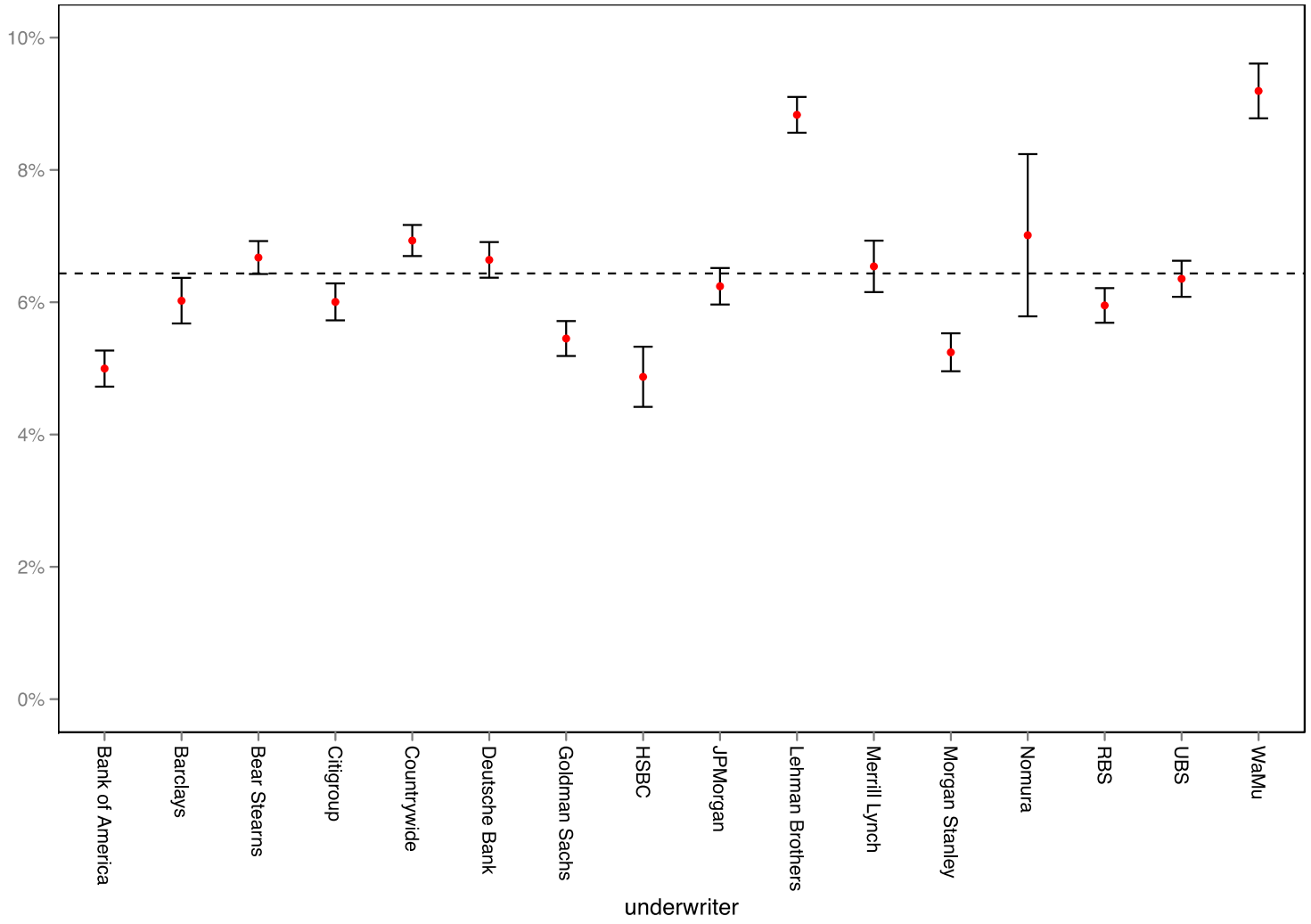


Figure 4: Misreported Second Lien by Underwriter

Figure 4 plots the percentage of misreported second lien by underwriter along with a 95% confidence interval. Coefficients result from adding underwriter fixed effects to the specification in Column (3) of Table 2B. These levels are obtained by adding each underwriter fixed effect to the level of misrepresentation for the omitted category (Credit Suisse) with other covariates at their means.

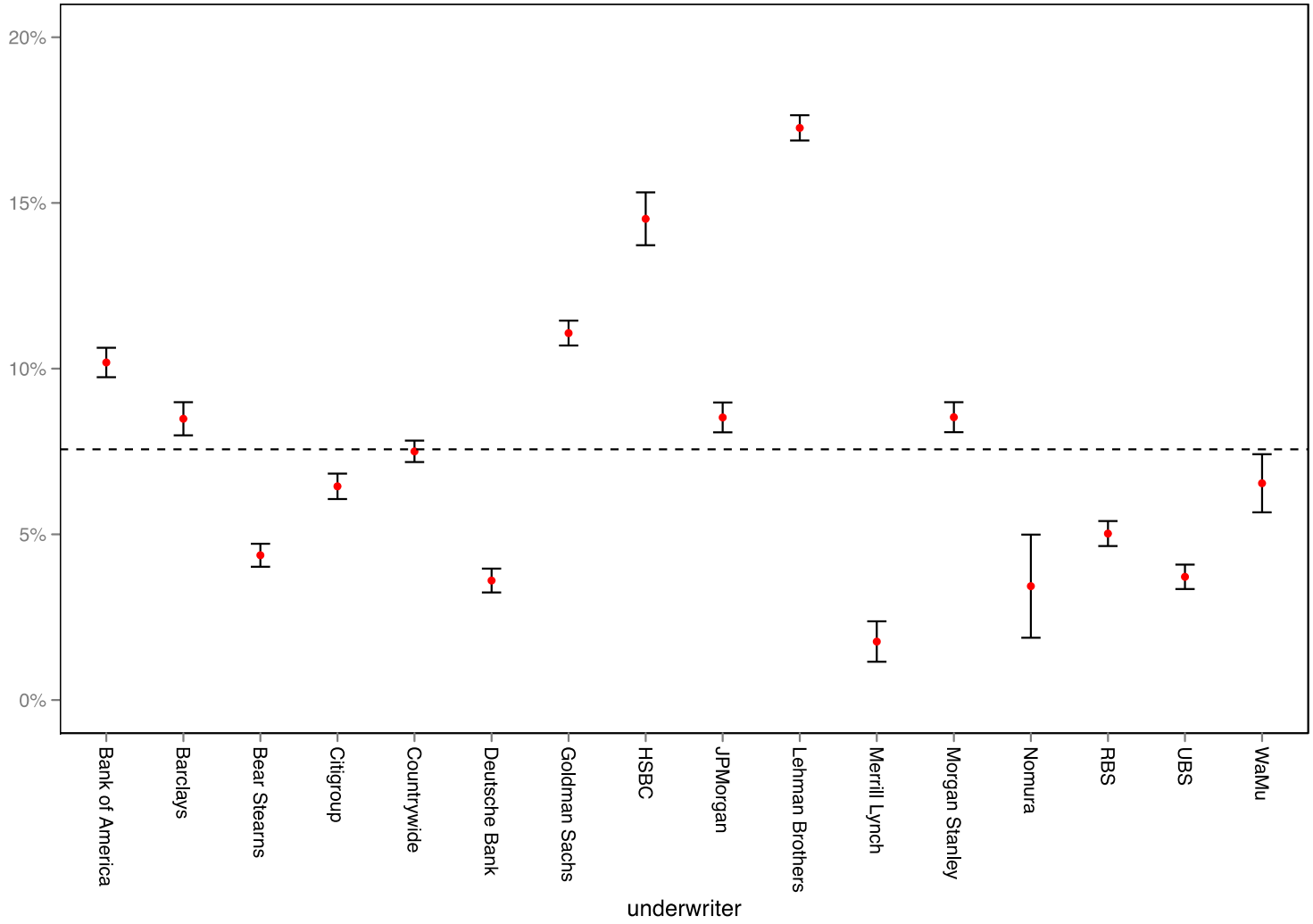
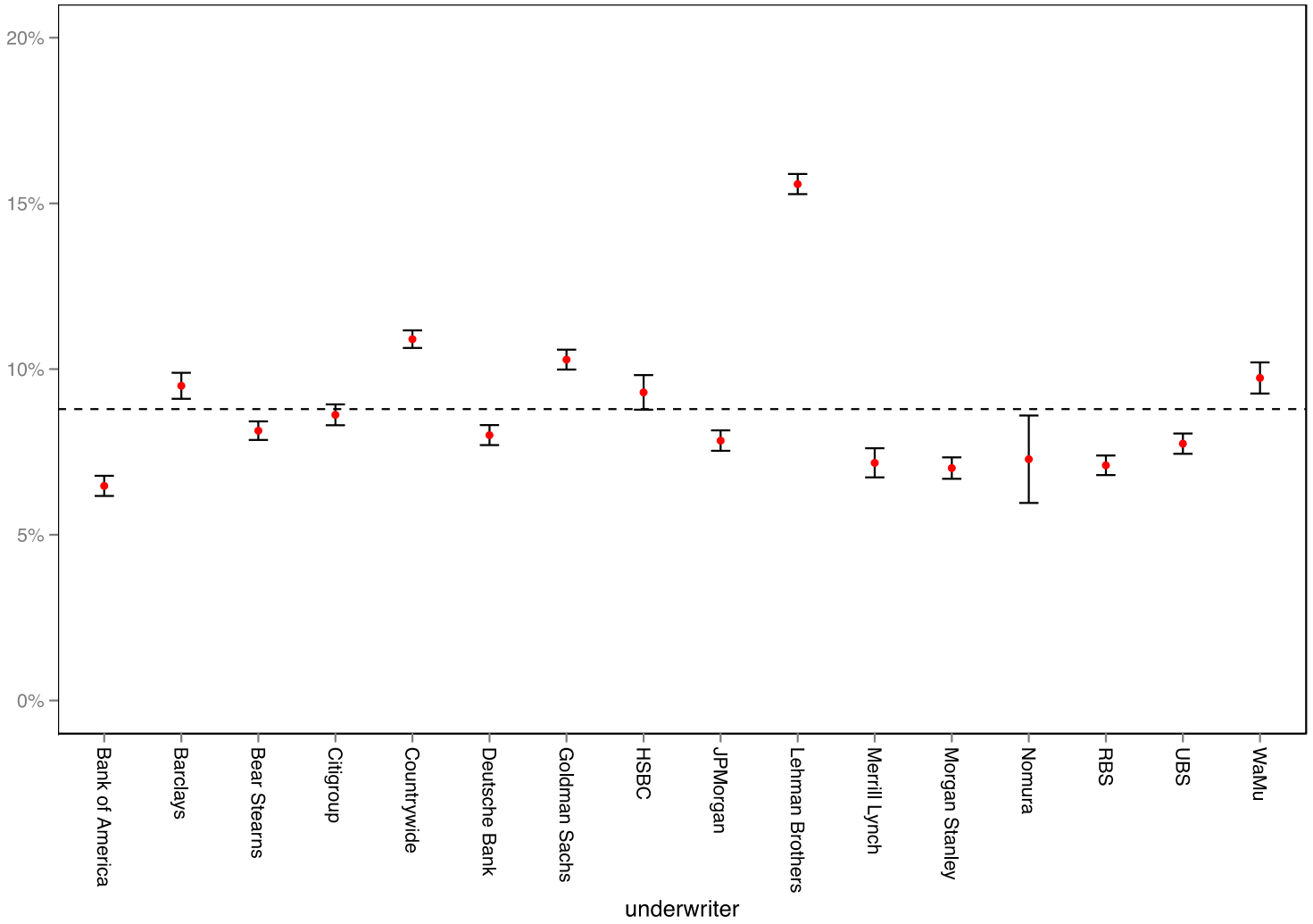


Figure 5: Either Misrepresentation by Underwriter

Figure 5 plots the percentage of loans with either misrepresentation by underwriter along with a 95% confidence interval. These levels are obtained by adding each underwriter fixed effect to the level of misrepresentation for the omitted category (Credit Suisse) with other covariates at their means.



Appendix A: Data

Verifying the Quality of the BlackBox-Equifax Merge

As we mentioned in Section 3, our primary dataset consists of a merge between (i) loan-level mortgage data collected by BlackBox Logic and (ii) borrower-level credit report information collected by Equifax. The merge was performed by Equifax using a proprietary merge algorithm utilizing more than twenty-five variables. Equifax is one of the three largest credit bureaus in the United States (along with Experian and TransUnion) and collects data on more than 400 million consumers across the world. As such, it is one of the very few potential providers for such detailed consumer credit data, and the firm has built a reputation based on data management. It provides a variety of analytical tools, from fraud detection to portfolio analytics, to a diverse group of client companies spanning multiple industrial sectors. Given Equifax's background as a data services company with over $1.5 billion in annual revenue, it certainly has the expertise to perform the merge between the BlackBox and Equifax datasets. Nevertheless, as both of our measures of underwriting misrepresentations rely directly on identifying discrepancies between these two datasets, we also independently verify the quality of the merge.

Equifax reports a merge confidence measure that ranges from 0 to 0.90 (ranging from low to high confidence). The majority of loans fall in the high-confidence buckets, such as between 0.8 and 0.81 or between 0.89 and 0.90 (see Table A1). To further verify the accuracy of the merge, we compare the dynamic loan payment history reported to BlackBox to that reported to Equifax. While BlackBox provides loan-level data, the Equifax data are reported at the consumer level with statistics such as delinquencies either aggregated by debt product (first liens, HELOCs, credit cards, etc.) or separated by account size and product (largest first lien, second largest first lien, etc.). We restrict our analysis to consumers with only one first mortgage in Equifax so that we can compare the correct payment records.[38] Additionally, we limit our analysis to observations where the loan's status in BlackBox is either current or 30, 60, or 90+ days delinquent, as it can be difficult to correctly map other BlackBox statuses such as bankruptcy or prepayment to a "correct" status in Equifax. We report the results in Table A2.

We find that for entire sample, the BlackBox and Equifax payment statuses exactly match 93.8% of the time and are within 30 days of each other in 98.0% of cases. We include the second measure to account for potential lags in reporting across the two databases. We observe that status-matching rates increase monotonically with Equifax's confidence measure, with only 0.3% of the highest confidence loans having significantly different BlackBox and Equifax delinquency statuses. It is worth noting that among loans with a confidence measure of 0.89 or

[38] Consider a borrower that has two first mortgages: we would observe the delinquency status of both mortgages, but we would have to make assumptions regarding which status corresponded to the matched BlackBox loan.

greater, about 16% of loans do not have matching BlackBox and Equifax zip codes.[39] This reinforces the notion that Equifax uses unique identifiers in performing its merge, given the difficulty of performing a merge across millions of loans without requiring a match on zip code.

Based on these findings, we conclude that the most reliable subsample of the merged data consists of loans with a confidence of 0.89 or higher (44.4% of the sample). For the owner-occupancy analysis, this results in a sample of 1,563,223 mortgages, down from 3,549,858 loans originated from 2005 to 2007 that are reported to BlackBox as owner-occupied (a 55.96% reduction). For the analysis of misreported second liens, this results in a sample of 854,959 mortgages, down from 1,759,519 loans originated from 2005 to 2007 that are reported as having no second liens in BlackBox (a 51.41% reduction).

Next, we examine how restricting to loans with a high Equifax merge confidence impacts our sample. In Panel A of Table A3 we display summary statistics for loans reported as owner-occupied in the BlackBox dataset. The first two columns report these statistics for all loans in the BlackBox dataset that are reported as owner-occupied. The last two columns report statistics for the subsample of these loans limited to mortgages with highest Equifax merge confidence (0.89 or higher). On average, loans in the high-quality merged sample have somewhat higher quality observable characteristics. In particular, these loans have lower interest rates, lower CLTV ratios, and higher FICO scores and origination balances. Figure A1 displays kernel density plots comparing this high-quality merged sample to the sample of BlackBox loans with reported owner-occupied status that do not meet the merge confidence restriction. We see that despite the differences in mean values, the two samples have similar underlying distributions of important observable characteristics.

We repeat this sample comparison analysis for sample of loans that report no presence of second liens. Panel B of Table A3 reports statistics for loans for which BlackBox reports that the CLTV and LTV are equal. The first two columns report these statistics without any restrictions, while the last two columns report the statistics for high-quality merged sample. Again, imposing these restrictions leads to a sample with slightly higher quality observables. FICO scores and origination balances increase, while interest rates and CLTV decrease. Figure A2 displays kernel density plots comparing the merged sample with a sample of loans with no reported second liens in the BlackBox that do not meet the Equifax merge confidence restrictions. Again, the shapes of the distributions are quite similar, demonstrating that these two samples are also not meaningfully different.

Finally, we conduct a placebo test to assess whether there is a relationship between incorrectly merged records and the subsequent performance of the loans. To identify incorrectly merged

[39] This group of loans includes mortgages reported to investors as for non-owner-occupied properties.

records, we utilize the origination loan balances that Equifax reports each month for the borrower's two largest active first mortgages. We compare these two amounts over the loan's first six observations to the origination loan balance reported by BlackBox to construct the *Balance Mismatch* variable, which takes the value of 1 if neither Equifax balance is ever within two percent of the BlackBox origination balance and is zero otherwise.[40] We note that the balance of the securitized first mortgage is unlikely to be misreported to investors because servicers verify and report on a monthly basis outstanding loan amount and payments to the securitization trust; hence, such records may indicate incorrectly merged loans across the two databases. Because only two largest mortgage balances are reported in Equifax database, we remove borrowers that have three or more first mortgages reported in their credit file. The reason is that we are not sure if the loan in BlackBox would be part of the two largest mortgage balances that are reported in the Equifax database.

We find that 2.01% of the loans have mismatched origination loan balances. To examine the relationship between default likelihood and incorrectly merged records, we run an OLS regression where the dependent variable is the same 90-day delinquency dummy variable used in Table 5. The right hand variables include *Balance Mismatch*, the two previous misrepresentation measures, origination cohort fixed effects, and the same vector of controls used in prior loan-level regressions. Column (2) clusters standard errors by state level.

Table A4 reports the results. While we observe a small positive coefficient for *Balance Mismatch*, the effect is economically small relative to what we find for our primary variables. Moreover, the effect is not statistically significant when standard errors are clustered at the state level. This suggests that there is no economically meaningful relationship between the potential merge accuracy of the two databases and subsequent loan performance. Importantly, the inclusion of *Balance Mismatch* does not reduce the economic or statistical significance of the misrepresentation variables. This evidence provides further support that our methodology allows us to identify actual misrepresentations and does not reflect incorrectly merged records.

Constructing Underwriter-Level Measures

This subsection discusses the construction of the underwriter-level variables used in Table 10. The measure of the relative importance of the RMBS underwriting business is constructed by dividing the aggregate dollar value of non-agency RMBS underwritten by an underwriter in 2005 by the underwriter's total assets as of the end of 2005, as reported by Compustat. This variable has a mean of 0.243 and a standard deviation of 0.334. The years of experience in underwriting subprime deals is a deal-level variable capturing the underwriter's experience in the subprime

[40] Results are quantitatively similar regardless of how we define *Balance Mismatch*, with respect to both the 6-month time window and the strictness of the balance match (within 2%, 5%, $100, etc.).

MBS market. To construct this measure, we take each pool's securitization year and subtract the year that the deal's underwriter entered the subprime MBS market, using ABSNet's universe of subprime MBS deals from 1999 onward. This variable has a mean of 5.98 and a standard deviation of 1.45.

The commercial bank underwriter variable, which accounts for differences between commercial and investment banks, is a dummy that takes a value of 1 if the underwriter's charter identifies it as a commercial bank and is 0 otherwise. The underwriter bonus-to-salary ratio is computed by dividing the cash bonus of each underwriter's five highest paid executives by their base salary in 2005 (in some cases only information for top six highest paid executives is available and used). This data comes from Computstat's Execucomp database where available and is otherwise constructed using firms' public financial reporting. This variable has a mean of 14.98 and a standard deviation of 15.53. The Risk Management Index (RMI), based on Ellul and Yerramili, is computed as the first principal component of the following six risk management variables: Credit Risk Officer (CRO) is present, CRO is an executive, CRO is a top five executive, CRO compensation centrality, Risk committee experience, and Active risk committee.

Table A1: Distribution of Equifax Merge Confidence

The table shows the distribution of loans in the BlackBox dataset across the Equifax's merge confidence level. No loans have a reported merge confidence greater than 0.90.

Merge Confidence	Number of Loans
[0.00,0.80)	3.6%
[0.80,0.81)	41.8%
[0.81,0.89)	10.2%
Highest Confidence Available: [0.89,0.90)	44.4%
Total	100%

Table A2: Delinquency Status Matching by Equifax Merge Confidence

The table shows the fraction of observations with delinquency statuses that match or are within 30 days of each other in BlackBox and Equifax. The sample is restricted to borrowers with only one first mortgage record in Equifax and to observations where the loan's status in BlackBox is either current or 30, 60, or 90 days delinquent. A "Status Match" is defined as an exact match between the reported BlackBox and Equifax delinquency statuses. "Status within 30 Days" is defined as two statuses that are less than 60 days apart, e.g. 90 days delinquent in BlackBox and 60 days delinquent in Equifax.

Merge Confidence	Fraction Payment Status Match	Fraction Payment Status Within 30 Days
[0.00,0.80)	0.841	0.924
[0.80,0.81)	0.904	0.964
[0.81,0.89)	0.942	0.991
[0.89,0.90)	0.976	0.997

Table A3: Comparison of Descriptive Statistics

The table presents a comparison of summary statistics of key variables. The first two columns of each panel present statistics for all loans in a given sample, and the second two columns restrict to loans that have the highest merge confidence. Panel A presents mortgages reported as for owner-occupied properties, while Panel B shows these statistics for loans reported as having no second liens to the RMBS trustee.

Panel A: Sample of Loans Reported as for Owner-Occupied Properties

	All Sample		Merged Sample	
	Mean	SD	Mean	SD
Interest Rate	6.990	1.947	6.605	2.024
FICO	662.6	74.95	680.2	73.20
Balance	291.0	254.2	324.6	248.3
CLTV	84.33	11.44	83.72	11.71
Purchase	0.483	0.500	0.496	0.500
No Cash Out Refi	0.149	0.356	0.136	0.343
Cash Out Refi	0.360	0.480	0.358	0.480
Low or No Doc.	0.419	0.493	0.463	0.499
ARM	0.606	0.489	0.527	0.499
Option ARM	0.0991	0.299	0.103	0.304
Number of Loans	3,549,858		1,563,223	

Panel B: Sample of Loans Reported as Having No Second Liens

	All Sample		Merged Sample	
	Mean	SD	Mean	SD
Interest Rate	6.925	2.236	6.458	2.424
FICO	652.8	76.48	663.5	76.90
Balance	277.6	225.3	293.1	238.9
CLTV	79.79	10.03	78.93	9.955
Purchase	0.366	0.482	0.361	0.480
No Cash Out Refi	0.117	0.321	0.121	0.326
Cash Out Refi	0.506	0.500	0.507	0.500
Low or No Doc.	0.472	0.499	0.514	0.500
ARM	0.597	0.491	0.475	0.499
Option ARM	0.113	0.317	0.169	0.375
Number of Loans	1,741,606		854,959	

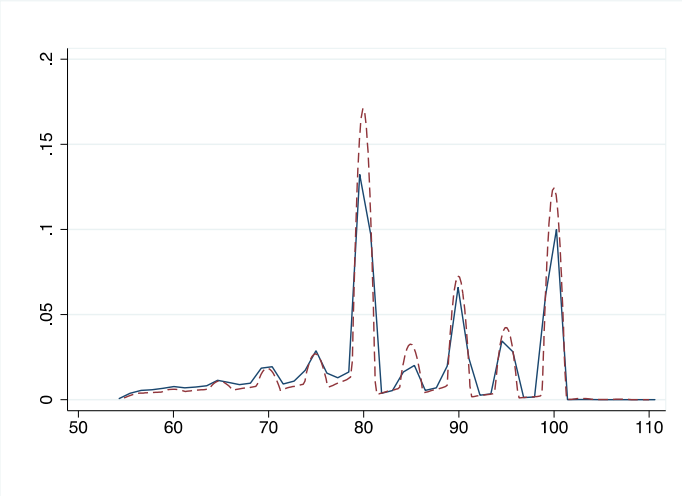
Table A4: Balance Mismatch Placebo Table

The table presents the OLS estimates from regressions where the dependent variable takes the value of 1 if the loan defaults (90+ days delinquent) in the first 2 years since origination and is zero otherwise. *Balance Mismatch* takes a value of 1 if the origination loan balance reported by BlackBox is not within two percent of any first mortgage origination loan balance reported to Equifax during the loan's first six months, and is zero otherwise. Other Controls include origination variables used in Table 3 such as FICO, interest rates, and LTV ratios. Squared and cubed terms for FICO and CLTV ratio are also included to account for potential non-linear effects. The estimates are in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

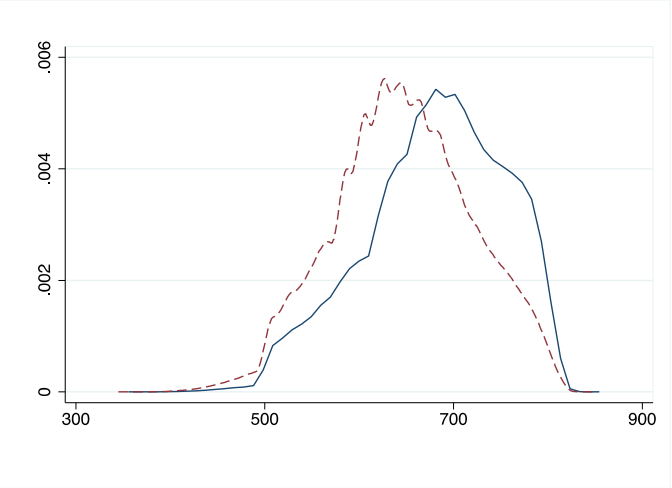
	(1)	(2)
Balance Mismatch	0.709***	0.709
	(3.70)	(1.60)
Misreported Non-Owner-Occupant	7.485***	7.485***
	(60.55)	(13.85)
Misreported Second	8.214***	8.214***
	(51.31)	(9.56)
Other Controls	Yes	Yes
Half-Year Origination Cohorts	Yes	Yes
SEs Clustered by State	No	Yes
Number of Loans	1487121	1487121
Percent 90 DPD	14.44	14.44
Percent Balance Mismatch	2.010	2.010
R-Squared	0.155	0.155

Figure A1: Comparability of Owner-Occupied Loans by Merge Restrictions – Kernel Density of Observables

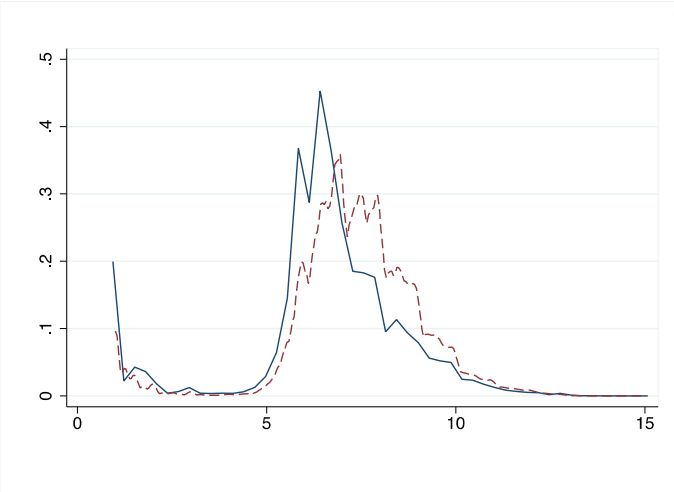
The figure shows the kernel density plots for loan origination CLTV ratios (a), FICO credit scores (b), interest rates (c), and loan balances (d) among loans that are reported as owner-occupied in the BlackBox dataset. The group of loans meeting the Equifax merge restrictions is represented by the solid line, and the group that fails these restrictions is represented by the dashed line.



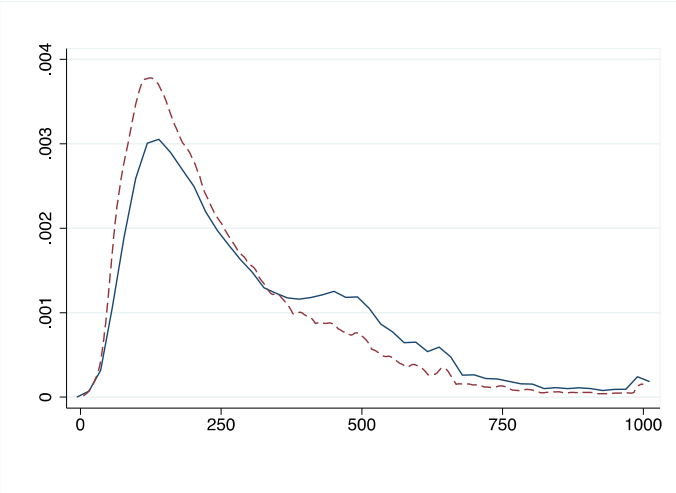
(a): Origination CLTV Ratios



(b): Origination FICO Scores



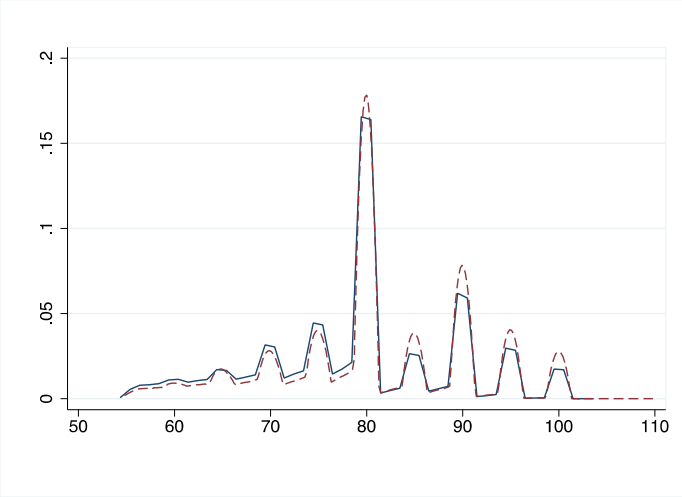
(c): Origination Interest Rates



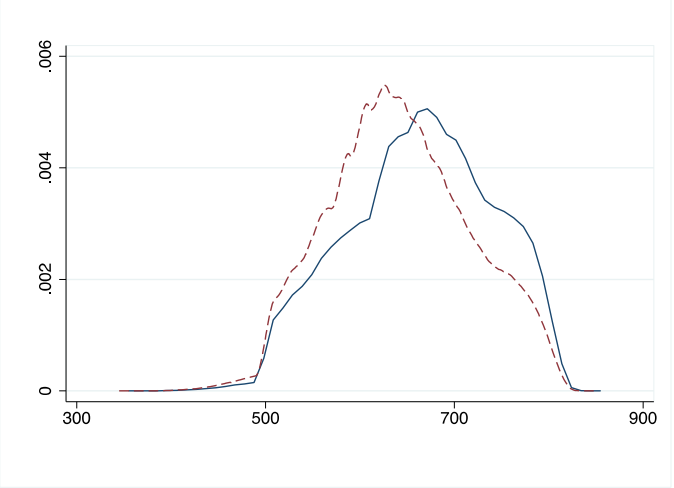
(d): Origination Loan Balance

Figure A2: Comparability of Loans without a Second Lien by Merge Restrictions – Kernel density of observables

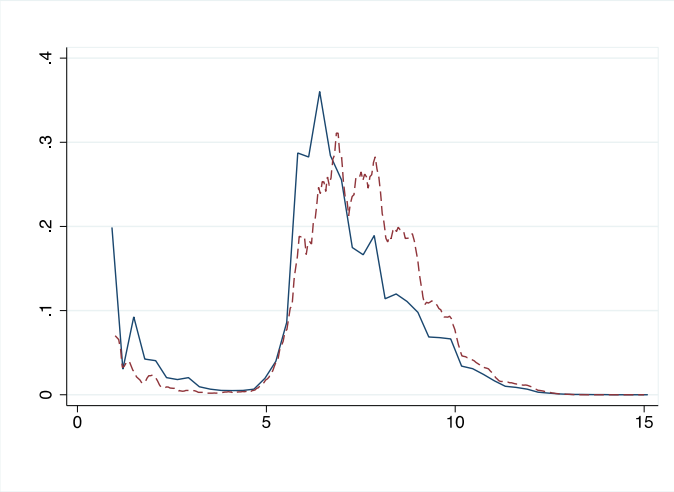
The figure shows the kernel density plots for loan origination CLTV ratios (a), FICO credit scores (b), interest rates (c), and loan balances (d) among loans that are reported as not having any second liens in the BlackBox dataset. The group of loans meeting the Equifax merge restrictions is represented by the solid line, and the group that fails these restrictions is represented by the dashed line.



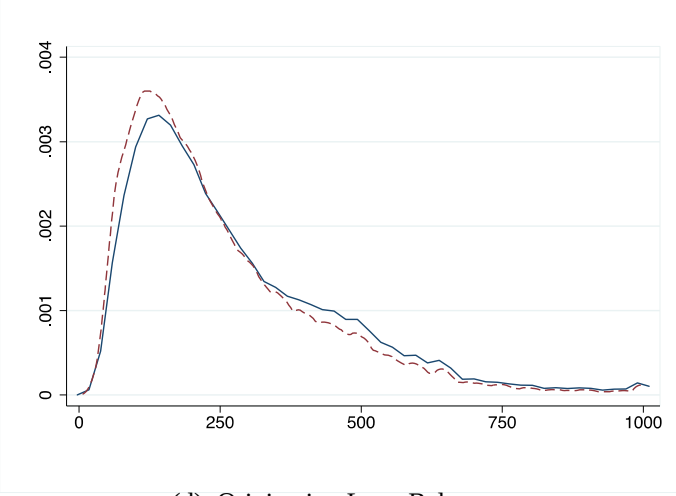
(a): Origination CLTV Ratios



(b): Origination FICO Scores



(c): Origination Interest Rates



(d): Origination Loan Balance

Appendix B: Non-Linear Specifications

Table B1: Misrepresentations and Loan Characteristics

The table presents the probit estimates from regressions where dependent variable takes a value of 1 if the loan has the given misrepresentation, and is 0 otherwise. Columns (1) and (2) show the results with *Misreported Non-Owner-Occupant* as the dependent variable, while Columns (3) and (4) show results with *Misreported Second Lien* as the dependent variable. The estimates show the marginal effects in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. These estimates can be compared to the OLS estimates presented in Table 3.

	Misreported Non-Owner-Occupant		Misreported Second Lien	
	(1)	(2)	(3)	(3)
Interest Rate	0.600***	0.564***	-0.588***	-0.603***
	(48.36)	(41.46)	(-31.24)	(-29.72)
FICO	0.00794***	0.00777***	0.00523***	0.00531***
	(23.43)	(22.81)	(12.99)	(13.20)
Balance	0.00209***	0.00200***	-0.00783***	-0.00741***
	(28.71)	(26.91)	(-61.79)	(-58.31)
CLTV	0.0222***	0.0213***	-0.118***	-0.112***
	(11.80)	(11.30)	(-43.64)	(-41.47)
No Cash Out Refi	-3.85***	-3.87***	-3.74***	-3.59***
	(-108.69)	(-109.48)	(-88.60)	(-84.45)
Cash Out Refi	-5.43***	-5.45***	-9.78***	-9.57***
	(-138.92)	(-139.26)	(-157.89)	(-155.48)
Low or No Doc.	2.61***	2.61***	-1.05***	-1.04***
	(65.71)	(65.57)	(-20.92)	(-20.95)
ARM	2.13***	2.15***	5.14***	5.10***
	(52.09)	(52.44)	(86.02)	(85.98)
Option ARM	8.27***	8.02***	-2.76***	-2.87***
	(59.46)	(56.18)	(-32.76)	(-33.66)
Confidence Match Controls	Yes	Yes	Yes	Yes
Half-Year Origination Cohorts	No	Yes	No	Yes
Number of Loans	1,563,223	1,563,223	854,959	854,959
Percent Misrepresented	6.426	6.426	7.131	7.131

Table B2: Misrepresentations and Loan Default

The table presents the probit estimates from regressions where the dependent variable takes a value of 1 if the mortgage ever defaults (ever goes 90 days past due on payments) in the first two years since origination, and is 0 otherwise. Panel A shows the results with *Misreported Non-Owner-Occupant* and *Reported Non-Owner-Occupant* as control variables. Panel B shows results with *Misreported Second* x *CLTV >= 100*, *Misreported Second* x *CLTV < 100*, *Reported Second* x *CLTV >= 100*, and *Reported Second* x *CLTV < 100* as control variables. "Other Controls" include origination variables used in Table 3 such as FICO, interest rates, and LTV ratios. Squared and cubed terms for FICO and CLTV ratios are also included to account for potential nonlinear effects. Column (1) also includes half-year origination fixed effects. Column (2) incorporates property state fixed effects. Column (3) clusters standard errors by state. The estimates show the marginal effects in percentage terms; *t*-statistics are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. These estimates can be compared to the OLS estimates in Tables 5C and 5D.

Panel A: Mortgage Default and Misreported versus Reported Non-Owner Occupants

	(1)	(2)	(3)
Misreported Non-Owner-Occupant	9.03***	8.34***	8.34***
	(69.91)	(66.10)	(10.64)
Reported Non-Owner-Occupant	3.89***	3.72***	3.72***
	(47.86)	(45.95)	(4.87)
Other Controls	Yes	Yes	Yes
Half-Year Origination Cohorts	Yes	Yes	Yes
State Fixed Effects	No	Yes	Yes
SEs Clustered by State	No	No	Yes
Number of Loans	1,827,497	1,827,497	1,827,497
Percent Default	14.89	14.89	14.89
Percent Misrepresented	6.424	6.424	6.424

Panel B: Mortgage Default and Misreported versus Reported Second Lien

	(1)	(2)	(3)
Misreported Second x CLTV < 100	8.20***	8.95***	8.95***
	(21.86)	(23.39)	(19.18)
Reported Second x CLTV < 100	8.81***	9.19***	9.19***
	(55.57)	(57.36)	(10.36)
Misreported Second x CLTV >= 100	13.4***	14.3***	14.3***
	(50.93)	(53.17)	(8.72)
Reported Second x CLTV >= 100	15.5***	16.2***	16.2***
	(104.68)	(107.85)	(7.66)
Other Controls	Yes	Yes	Yes
Half-Year Origination Cohorts	Yes	Yes	Yes
State Fixed Effects	No	Yes	Yes
SEs Clustered by State	No	No	Yes
Number of Loans	1,109,250	1,109,250	1,109,250
Percent Default	17.10	17.10	17.10
Percent Misrepresented	5.212	5.212	5.212

Appendix C: List of RMBS Lawsuits Fillings

This Appendix presents a list of recent lawsuits fillings regarding the material breach of representations and warranties by the underwriters and originators of residential mortgage-backed securities. Lawsuits are organized by plaintiff and filing year. The 58 lawsuits presented here span the years 2008-2012 (2008 =4, 2009 = 5, 2010 = 9, 2011 = 35, 2012 = 5).

Filed by Insurers

Filed in 2008. *MBIA Insurance Corp. v. Countrywide Home Loans, Inc., et al.*, New York State Supreme Court, County of New York. No. 602825/08.

Filed in 2008. *MBIA Insurance Corp. v. Residential Funding Company, LLC,* New York State Supreme Court, County of New York. No. 603552/08.

Filed in 2009. *Financial Guaranty Insurance Co. v. Countrywide Home Loans Inc.*, New York State Supreme Court, County of New York, No. 650736/09.

Filed in 2009. *MBIA Insurance Corp. v. Bank of America Corp., et. al.*, California Superior Court, County of Los Angeles., No. BC-4157572.

Filed in 2009. *MBIA Insurance Corp. v. IndyMac, F.S.B., et al.*, District of Columbia District Court, No. 1:09-cv-01011.

Filed in 2009. *Syncora Guarantee Inc. v. Countrywide Home Loans Inc., et al.*, New York State Supreme Court, County of New York, No. 650042/09.

Filed in 2010. *Ambac Assurance Corp., et al. v. Countrywide Home Loans, Inc., et al.*, New York State Supreme Court, County of New York, No. 651612/10.

Filed in 2010. *Ambac Assurance Corp., et al. v. DLJ Mortgage Capital, Inc., et al.*, New York State Supreme Court, County of New York, No. 600070/10.

Filed in 2010. *Assured Guaranty Municipal Corp. v. DLJ Mortgage Capital, Inc., et al.*, New York State Supreme Court, County of New York, No. 600070/10.

Filed in 2010. *MBIA Insurance Corp. v. GMAC Mortgage, LLC,* New York State Supreme Court, County of New York. No. 600837/10.

Filed in 2011. *United Financial Casualty Co., et al. v. Countrywide Financial Corp., et al.*, California Central District Court, No. 2:11-cv-04766.

Filed in 2012. *MBIA Insurance Corp. v. Ally Financial Inc., et al.*, Minnesota District Court, County of Hennepin, No. 0:12-cv-02563.

Filed by Government

Filed in 2010. *Federal Home Loan Bank of San Francisco v. Deutsche Bank Securities, Inc., et al.*, California Northern District Court, No. 3:10-cv-03039.

Filed in 2011. *Federal Housing Finance Agency v. Ally Financial Inc., et al.*, New York State Supreme Court, County of New York, No. 1:11-cv-07010.

Filed in 2011. *Federal Housing Finance Agency v. Bank of America Corp., et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06195.

Filed in 2011. *Federal Housing Finance Agency v. Barclays Bank PLC, et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06190.

Filed in 2011. *Federal Housing Finance Agency v. Citigroup, Inc., et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06196.

Filed in 2011. *Federal Housing Finance Agency v. Countrywide Financial Corp., et al.*, New York State Supreme Court, County of New York, No. 11-cv-06916.

Filed in 2011. *Federal Housing Finance Agency v. Credit Suisse Holdings (USA), Inc., et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06200.

Filed in 2011. *Federal Housing Finance Agency v. Deutsche Bank AG, et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06192.

Filed in 2011. *Federal Housing Finance Agency v. First Horizon National Corp., et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06193.

Filed in 2011. *Federal Housing Finance Agency v. General Electric Co., et al.*, New York State Supreme Court, County of New York, No. 1:11-cv-07048.

Filed in 2011. *Federal Housing Finance Agency v. Goldman, Sachs & Co., et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06198.

Filed in 2011. *Federal Housing Finance Agency v. HSBC North America Inc., et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06189.

Filed in 2011. *Federal Housing Finance Agency v. JPMorgan Chase & Co., et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06188.

Filed in 2011. *Federal Housing Finance Agency v. Merrill Lynch & Co., et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06202.

Filed in 2011. *Federal Housing Finance Agency v. Morgan Stanley, et al.*, New York State Supreme Court, County of New York, No. 1:11-cv-06739.

Filed in 2011. *Federal Housing Finance Agency v. Nomura Holding America Inc., et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06201.

Filed in 2011. *Federal Housing Finance Agency v. Royal Bank of Scotland Group PLC, et al.*, United States District Court, District of Connecticut, No. 3:11-cv-01383.

Filed in 2011. *Federal Housing Finance Agency v. SG Americas, Inc., et al.*, United States District Court, Southern District of New York, No. 1:11-cv-06203.

Filed in 2011. *Federal Housing Finance Agency v. UBS Americas, Inc., et al.*, United States District Court, Southern District of New York, No. 1:11-cv-05201.

Filed in 2011. *National Credit Union Administration Board v. Goldman Sachs and Co. et al.,* California Central District Court, County of Los Angeles, No. 2:11-cv-06521.

Filed in 2011. *National Credit Union Administration Board v. RBS Securities, Inc., et al.,* California Central District Court, County of Los Angeles, No. 2:11-cv-05587.

Filed in 2011. *National Credit Union Administration Board v. J.P. Morgan Securities LLC, et al.,* United States District Court, District of Kansas, No. 2:11-cv-02341.

Filed in 2011. *National Credit Union Administration Board v. RBS Securities, Inc., et al*, United States District Court, District of Kansas, No. 2:11-cv-02340.

Filed in 2011. *National Credit Union Administration Board v. Wachovia Capital Markets, LLC,* United States District Court, District of Kansas, No. 2:11-cv-02649.

Filed in 2011. *U.S. Securities and Exchange Commission v. Citigroup Global Markets, LLC,* United States District Court, Southern District of new York, No. 11-cv-07387.

Filed in 2012. *National Credit Union Administration Board v. Barclays Capital, et al.,* United States District Court, District of Kansas, No. 2:12-cv-02631.

Filed in 2012. *National Credit Union Administration Board v. Credit Suisse Securities (USA), et al.,* United States District Court, District of Kansas, No. 2:12-cv-02648.

Filed in 2012. *National Credit Union Administration Board v. UBS, et al.,* United States District Court, District of Kansas, No. 2:12-cv-02591.

Individual and Class Action Lawsuits

Filed in 2007. *David H Luther, et al. v. Countrywide Home Loans Servicing LP, et al.*, California Superior Court, County of Los Angeles, No. BC-380698.

Filed in 2008. *New Jersey Carpenters Health Fund, et. al v. DLJ Mortgage Capital, et al.*, United States District Court, Southern District of New York, No. 08-cv-5653.

Filed in 2008. *Public Employees' Retirement System of Mississippi, et al. v. Merrill Lynch & Co., Inc., et al.,* United States District Court, Southern District of New York, No. 08-cv-10841.

Filed in 2009. *Public Employees' Retirement System of Mississippi, et al. v. Goldman Sachs Group Inc., et al.,* United States District Court, Southern District of New York, No. 09-cv-01110.

Filed in 2010. *Allstate Insurance Co., et al. v. Countrywide Financial Corp., et al.,* United States District Court, Southern District of New York, No. 10-cv-09591.

Filed in 2010. *Footbridge Limited Trust, et al. v. Countrywide Home Loans, et al.,* United States District Court, Southern District of New York, No. 09-cv-04050.

Filed in 2010. *Maine State Retirement System v. Countrywide Financial Corp., et al.*, United States District Court, Central District of California, No. 10-cv-00302.

Filed in 2010. *Stichting Pensioenfonds ABP v. Countrywide Financial Corp., et al.*, United States District Court, Central District of California, No. 2:10-cv-07275.

Filed in 2011. *American International Group, Inc., et al. v. Bank of America Corp., et al.,* United States District Court, Southern District of New York, No. 1:11-cv-06212.

Filed in 2011. *Capital Ventures International v. UBS Securities LLC, et al.*, United States District Court, District of Massachusetts, No. 1:11-cv-11937.

Filed in 2011. *Massachusetts Mutual Life Insurance Co. v. Credit Suisse First Boston Mortgage Securities Corp., et al.*, United States District Court, District of Massachusetts, No. 3:11-cv-30048.

Filed in 2011. *Massachusetts Mutual Life Insurance Co. v. DLJ Mortgage Capital, Inc., et al.*, United States District Court, District of Massachusetts, No. 3:11-cv-30047.

Filed in 2011. *Massachusetts Mutual Life Insurance Co. v. RBS Financial Products, et al.,* United States District Court, District of Massachusetts, No. 3:11-cv-30044.

Filed in 2011. *Massachusetts Mutual Life Insurance Co. v. Residential Funding Co., et al.,* United States District Court, District of Massachusetts, No. 3:11-cv-30035.

Filed in 2011. *National Integrity Life Insurance Co. v. Countrywide Financial Corp. et al.,* United States District Court, Southern District of New York, No. 1:11-cv-08077.

Filed in 2011. *The Union Central Life Insurance Co. et al. v. Credit Suisse Securities et al.,* United States District Court, Southern District of New York, No. 1:11-cv-02327.

Filed in 2011. *Walnut Place LLC et al. v. Countrywide Home Loans, Inc. et al.,* New York Supreme Court, County of New York, No. 650497/11.

Filed in 2011. *Western and Southern Life Insurance Co. et al. v. Countrywide Financial Corp. et al.*, United States District Court, Southern District of Ohio. No. 1:11-00267.

Filed in 2012. *Western Conference of Teamsters Pension Trust Fund v. Countrywide Financial Corp. et al.,* United States District Court. No. 2:12-cv-05122.