“Cherry Picking” in Subprime Mortgage Securitizations:

Which Subprime Mortgage Loans Were Sold by Depository Institutions Prior to the Crisis of 2007?

Paul Calem

Division of Banking Supervision and Regulation, Federal Reserve Board

Christopher Henderson

Department of Supervision, Regulation and Credit, Federal Reserve Bank of Philadelphia

and Jonathan Liles

Mission Division, Freddie Mac

July 2010

The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Board, the Federal Reserve Bank of Philadelphia, or Freddie Mac. We thank Robert Avery, Scott Frame, Michael Gordy, Benjamin Keys, Paul Kupiec, Michael LaCour-Little, Doug McManus, Marcelo Rezende, Kate Samolyk, Robert Tsien, George Walker, and Peter Zorn for helpful comments.

“Cherry Picking” in Subprime Mortgage Securitization:

Which Subprime Mortgage Loans Were Sold by Depository Institutions Prior to the Crisis of 2007?

Abstract

Depository institutions may utilize securitization to “cherry pick,” meaning to transfer risk to investors along dimensions that the investors tend to disregard or where their risk assessments are overly optimistic. Using Home Mortgage Disclosure Act data merged with data on subprime loan delinquency by ZIP code, this paper examines the loan sale decision of depository institutions with respect to “high cost” mortgages originated during the subprime lending boom of 2005 and 2006. We find that the likelihood of sale increases with risk along dimensions viewed as indicative of cherry picking. In contrast, along the dimension of mutually observed and priced risk as represented by APR spread, likelihood of sale decreases with risk. In addition, holding constant the APR spread and other factors, likelihood of sale is positively related to the future rate of serious delinquency among subprime loans in the ZIP code where the property is located. Thus, the paper reinforces the view, increasingly prevalent in the literature, that inattention to or misperception of risk by the securitization market was a primary cause of the subprime lending boom and subsequent market collapse. The paper employs a unique approach that offers a relatively robust set of statistical controls.

“Cherry picking” in Subprime Mortgage Securitization:

Which Subprime Mortgage Loans Were Sold by Depository Institutions Prior to the Crisis of 2007?

1. Introduction

Financial innovation has made transferring risk from primary lending markets through securitization a common practice. Banks may securitize a portion of their loan portfolios to minimize regulatory capital requirements or to diversify funding opportunities and risk. Investors seek to purchase asset-backed securities because they are relatively liquid and, in principle, allow the investor ownership claim on a particular bank asset with quantifiable default and prepayment risk. The financial crisis of 2007, however, curtailed most non-agency mortgage securitization activity. We focus on misperception, unawareness, or disregarding of risk in the subprime securitization market as a primary factor leading to the crisis.

Our point of departure is to consider how a bank’s decision to securitize a loan varies in relation to the loan’s credit risk, and what happens if the securitization market is inattentive to, or systematically underestimates, risks considered material by the originating bank. In particular, depository institutions may utilize securitization to “cherry pick;” that is, transfer risk to investors along dimensions that the investors tend to disregard or where their risk assessments are overly optimistic. The decision of depository institutions to sell or retain mortgages provides a unique perspective from which to analyze the performance of the subprime mortgage securitization market overall, although only a small share of subprime mortgages were actually originated by depository institutions. Most were originated by mortgage and finance companies, including bank affiliates and independent organizations, but only depository institutions retained loans for their own portfolios (non-depositories generally relied on an originate-for-sale business model.)

Using Home Mortgage Disclosure Act data merged with data on subprime loan delinquency by ZIP code, this paper examines the loan sale decision of depository institutions with respect to “high cost” loans originated during the subprime lending boom of 2005 and 2006. We find that the likelihood of sale increases with risk along dimensions viewed as indicative of cherry picking. In contrast, along the dimension of mutually observed and priced risk as represented by APR spread, likelihood of sale decreases with risk. In addition, holding constant the APR spread and other factors, likelihood of sale is positively related to the future rate of serious delinquency among subprime loans in the ZIP code where the property is located.

For instance, we find that likelihood of sale increases with measures of neighborhood concentration of credit and collateral risk, such as higher concentration of subprime loans in a neighborhood. Such factors are not incorporated into rating agency models and generally were not considered in pricing of subprime mortgage backed securities, but they are observable to local depository institutions. Thus, the paper reinforces the view, increasingly prevalent in the literature, that misperception and mispricing of risk by the securitization market was a primary cause of the subprime lending boom and subsequent market collapse

The concept of "cherry picking" as described above is similar to but distinct from the traditional concept of a "lemons market" as applied to loan securitization, where investors cannot fully verify the quality of loans being securitized. In the lemons market, buyers are aware of their informational disadvantage and various outcomes are possible, including breakdown of the market, pooling and pricing of risk, or the development of mechanisms to separate the good from the bad.[1] In contrast to the lemons market, where investors are aware of the information asymmetry, cherry picking occurs when investors are inattentive, unaware, or overly optimistic in regard to certain elements of risk.

Distinguishing lemons market effects from cherry-picking is beyond the scope of our study, because we cannot determine whether investors were aware of an information disadvantage or whether loan terms or securitization structure were employed as separating mechanisms. We have no reason to suspect, however, that loan terms or securitization structure varied along the particular risk dimensions we associate with cherry picking. While the results are consistent with either cherry picking or lemons market scenarios, we view cherry picking as more consistent with anecdotal evidence as well as recent research findings concerning lack of awareness of risk within the mortgage securitization market during this period.

Such anecdotal evidence includes facts alleged in recent lawsuits challenging the veracity of lender's claims regarding underwriting quality and corporate governance over credit policy. In a legal action by the U.S. Securities and Exchange Commission (SEC) against Countrywide, for instance, senior

[1] In the literature, these typically arise as pooling or separating equilibriums in models of contracting between parties with asymmetric information.

Countrywide executives are alleged to have engaged in “unsound lending practices” to increase loan volume and market share. In particular, the lawsuit alleges that Countrywide understood that many of the nonprime loans it originated were excessively risky and that this lending would become unsustainable if the secondary market were to become aware of and appropriately price the risk.[2] The recent lawsuits are indicative of the post-crisis awareness that lenders had considerable scope in deciding which nonprime loans to sell or securitize.

Our focus is on the banking side of depository institutions that originated subprime loans and on overall industry patterns, rather than the behavior of any individual institution. In particular, we do not address the intriguing question of why some institutions assumed risks on the trading side (through subprime mortgage-backed security purchases and related activities) that they had previously avoided by securitizing subprime loans originated on the banking side.

The paper is organized as follows. Section 2, which immediately follows, briefly reviews a sampling of related literature on the collapse of the subprime mortgage market and the role of securitization in the subprime lending boom and subsequent market collapse. Section 3 examines the decision to sell or retain a mortgage as it relates to the credit risk of the mortgage, mostly reviewing theories developed in previous studies. Section 4 presents an overview of the data on subprime lending during 2005 and 2006 used for the empirical analysis, and a variety of descriptive statistics constructed from these data. Section 5 develops our empirical analysis of the loan sale decision of depository institutions, involving the estimation of logistic regression equations by year and institution size. Section 6 presents the estimation results and Section 7 concludes.

2. **Literature Review**

The collapse of the subprime market and beginning of the foreclosure crisis in 2007 and subsequent turmoil in mortgage and housing markets have spurred a variety of research on problems in the

---

[2]For example, the lawsuit cites an internal e-mail from Countrywide chairman Angelo Mozilo stating that the option ARM products “are currently mispriced in the secondary market” and the “timing was right” to sell Countrywide Bank’s portfolio of loans. See U.S. District Court for Central California, SEC vs. Angelo Mozilo, David Sambol, and Eric Sieracki, accessed at: http://www.scribd.com/doc/16118724/SEC-Complaint-Against-Former-Countrywide-CEO-Angelo-Mozilo

subprime or broader mortgage market that precipitated the crisis. Much of this research has focused on the deterioration of underwriting standards and house-price depreciation as primary factors (Smith 2007; Demyanyk and van Hemert (2009); Gerardi, Shapiro, and Willen 2008; Hahn and Passell 2008; Sherlund 2008). Haughwout, Peach, and Tracy (2008) focus on early payment default and emphasize that only part of the increase in default during 2007 is attributable to these factors. Demyanyk and van Hemert (2009) argue that the decline in underwriting standards prior to the crisis could have been detected but was masked by rapid house-price appreciation. Coleman, LaCour-Little, and Vandell (2008) present evidence that the expansion of credit resulting from looser underwriting standards contributed to the rise in house prices.

The issue broadly related to the subject of our paper—the role of securitization and associated agency problems—has also garnered attention, with several researchers pointing to securitization as a principal culprit in the crisis. Ashcraft and Schuermann (2008) identify a number of market frictions affecting the subprime mortgage origination and securitization process and argue that the associated misaligned incentives and adverse selection were largely responsible for the market's collapse. A partial list includes agency problems associated with brokers, such as incentives to misrepresent borrower credit quality; cherry picking by portfolio lenders; and rating agency conflicts of interest.[3] Golding, Green, and McManus (2008), Hull (2009) also focus on misaligned incentives of market participants, while Wray (2007) adds lax regulation to the mix. Specific problems discussed include compensation of loan originators and security traders disassociated from subsequent credit performance of the loans, and ratings agencies being paid by the issuers of the securities being rated. These authors put forth recommendations aimed at increasing transparency and reducing moral hazard in both the primary and secondary mortgage markets.

Rajan, Seru, and Vig (2009) draw a distinction between the "hard information" relied on by investors to value securitized loans and "soft information" accessible to originators but not verifiable by a third party. They argue that securitization of subprime mortgages reduced the incentive to collect soft information, resulting in less effective credit screening. In contrast, we focus on the cherry picking that may result from such investor inattention to "soft information". Note that for an individual firm, these are mutually exclusive outcomes, but that is not the case for the market as a whole; some institutions

[3] Ernst, Bocian, and Li (2008) argue that mortgage brokers also often exploit an information advantage relative to the borrower to engage in predatory lending.

may engage in cherry picking, while others may specialize in collecting only hard information and securitize all of the loans they originate.

Ben-David (2007) focuses on the propensity to overstate collateral values by borrowers, intermediaries, and originators when it is advantageous to do so in the presence of asymmetric information. In particular, originators are able to expand their business by securitizing more loans as house prices rise. White (2009) emphasizes the role of overly optimistic evaluations of the credit risk of mortgage-backed securities, in part due to agency problems and in part to inadequate information and "carelessness." Coval, Jurek, and Stafford (2009a, b) point to the amplification of errors in evaluating the risks of the underlying securities of structured finance products. They emphasize the concentration of systemic risk that occurred through these structured products and the mispricing of this risk. While these instruments appeared to be paying a high rate of return, they were, in fact, earning a negative return because of the failure to price "tail" risk. Nakamura (2010a) focuses on better data collection and monitoring of systemic risk concentrations by regulatory agencies as a way to prevent recurrence of this problem.

Mian and Sufi (2009) present evidence that mortgage securitization expanded the supply of credit to riskier segments of borrowers. They show that the massive expansion of credit was disassociated from income growth in areas where it occurred and find an association with subsequent deterioration in credit performance. Their finding of an association between increased securitization and higher default rates at the neighborhood level is similar to that observed in our study, although the empirical approaches differ.[4] Their analysis suggests a direct link between securitization and the recent mortgage crisis.

Elul (2009) compares the repayment performance of securitized residential mortgages to those retained in portfolio, for loans originated during 2003 through 2007. The analysis controls for various loan-level factors including the initial loan-to-value ratio and the borrower's credit score at origination, and for local economic conditions including house price appreciation. The analysis indicates that nonagency prime, adjustable rate securitized loans originated after 2003 (and fixed rate loans originated after 2005) performed significantly worse than similar, non-securitized, loans. The analysis does not

[4] Mian and Sufi (2009) conduct a panel analysis of mortgage lending patterns at the ZIP code level; our analysis is cross-sectional and at the loan level and focuses specifically on the subprime loan sale decision along dimensions likely associated with bank information advantages.

examine the relation between securitization and neighborhood-level risk or other dimensions likely to be associated with cherry picking. According to Elul, the findings “suggest that adverse selection may have been present in the prime nonagency mortgage market and may have contributed to deterioration in underwriting standards.”[5]

Keys, et al. (2010a, 2010b) provide an important contribution to the literature by finding that lenders apply less effort to screen soft information in the low documentation subprime loan market based when originating loans that can more easily be sold to investors. They identify a key point of discontinuity around the FICO score threshold of 620, such that lenders securitized more low-documentation loans with scores above this threshold and screened them less aggressively.[6] This line of research is relevant to our paper in that it provides evidence on "ease of securitization" which may have created opportunities for cherry picking. The finding that the banks could leverage favorable terms of sale to investors based on one risk indicator (FICO) in the absence of documented income suggests that lenders had opportunities to exploit investor inattention to risk.

3. Securitization and Credit Risk

Depository institutions have an incentive to sell or securitize loans to diversify funding opportunities and generate liquidity; diversify or manage credit and interest rate risk; and reduce regulatory capital requirements. In examining the relationship between bank loan securitization and credit risk, the literature focuses on incentives deriving from regulation; potential advantages of depository institutions in managing relationships with riskier borrowers (such as loss mitigation advantages); and lemons market type effects associated with seller information advantages relative to the buyer.

Capital regulation. One view is that securitization is favored for lower-risk assets, whereas banks would tend to retain opaque, higher-risk loans, because of incentives arising from regulatory capital

[5] Elul (2009) uses a loan-level data set from LPS Analytics. The study also examines subprime and does not find a difference in performance between securitized and retained loans; however, the data set’s coverage of the subprime market is less representative than for prime and contains relatively few retained subprime loans. Coverage is also lacking for some key risk measures; for instance, the data do not distinguish loans originated with piggyback second liens and do not provide a combined LTV for loans with piggyback seconds.

[6] Keys et al. (2010a) show that low documentation loans just above the 620 threshold defaulted more frequently than their counterparts just below the cutoff, all else equal. These effects were dampened for full documentation loans where additional hard information about the borrowers was collected and screening on soft information may not be as important. Keys, et al. (2010b) find that the amount of time taken to securitize a loan in the low documentation subprime market was systematically different around this threshold, with loans just above 620 securitized more quickly.

requirements. For instance, Calem and Follain (2007) maintain that for lower-risk mortgage loans, existing regulatory capital levels (established under the original Basel Accord) are too high in comparison to economic capital, creating an incentive to securitize these loans. Consistent with this view, Ambrose, LaCour-Little, and Sanders (2005) find evidence in data from a single large lender that risky mortgages are retained on balance sheet, while lower-risk mortgages are sold into secondary markets. They point to incentives arising from capital regulation and reputational risk as likely explanations for their findings.

Specialized risk management. Another view is that banks would tend to securitize lower risk and retain higher-risk loans because of advantages in managing relationships with riskier borrowers. For instance, Hill (1996) develops a model in which the non-securitized assets are higher risk and their credit quality less transparent. According to Hill, securitization "divides the firm into slices which permit more specialized appraisal." With selection of lower-risk assets for securitization, "security investors needn't appraise the particularly costly-to-appraise residual risks," leaving that task to the specialist (bank) that can carry out the more difficult risk assessments at lower cost.

Dewatripont and Tirole (1994, Chapter 10) also suggest that securitized loans are likely to be of higher credit quality because of advantages of depository institutions in managing risk. They argue that banks are uniquely equipped to add value to loans made to higher-risk borrowers by managing the future relationship with the borrower, and that securitizing such loans may weaken this ability or the incentive to apply it.

Macroeconomic uncertainty. Macroeconomic uncertainty might limit the incentive to securitize, further curtailing securitization of riskier loans. Conversely, greater macroeconomic stability, or perceived stability (such as might have occurred during the housing boom) could encourage securitization of more risky loans.

Uncertainty might have such an effect when combined with a cost asymmetry, such that a securitized share that is larger than desired ex-post, given realized economic conditions, tends to be more costly than a securitized share that is smaller than desired ex-post. Such a cost asymmetry the seller or unwinding the securitization and reselling the loans, are likely to be more costly than ex-post securitization of surplus, seasoned, balance sheet loans.

To explain this effect of macroeconomic uncertainty, it is convenient to represent a bank's securitization decision with respect to a pool of recently originated loans by the simple tradeoff depicted in Figure 1. Consistent with the preceding arguments, the marginal benefit of retaining a loan, which reflects regulatory capital and risk management advantages, is represented as increasing with credit risk along the single dimension represented by the risk index R. The marginal cost, of retaining a loan, which reflects the liquidity and diversification advantages of securitization, are represented as more or less independent of an individual loan's credit quality. The bank will securitize only lower risk loans characterized by $R < R^*$. Let the corresponding, optimal share of the pool that is securitized be denoted by S^*.

Macroeconomic uncertainty will have the effect of increasing the marginal return to retention. Macroeconomic uncertainty is introduced into this framework by assuming that the optimal securitized share is uncertain ex-ante because the tradeoff between retention and securitization depends on economic conditions affecting the pool as the loans age. For example, changing macroeconomic conditions might affect the bank's need for liquidity or the value it adds to retained loans through loss-mitigation activities. Let S denote the securitized share (corresponding to a particular credit quality threshold) selected ex-ante by the bank, and let X denote the random variable representing the ex-post optimum, which is dependent on realized economic conditions and has density $f(X)$. A gap between X and S reduces the profitability of securitization from what it would be had X been known with certainty ex-ante so that $S = X$. Denote the profitability conditional on no ex-ante uncertainty by $\Pi(X)$. For convenience, assume that the cost of divergence of S from X is proportional to the absolute size of the gap, . Denote the marginal cost of over-securitizing by c and the marginal cost of under-securitizing by $\alpha \cdot c$, where $\alpha < 1$ reflecting the cost asymmetry. Thus, the bank's optimization problem is to choose S to maximize:

(1)

Simple intuition suggests that the greater the amount of uncertainty associated with the random variable X or the greater the cost asymmetry (smaller α), the smaller the share that will be securitized, to mitigate the higher cost associated with ex-post outcomes involving excess securitization. In the appendix, we establish this result for the case that X has a beta distribution.

Cherry picking. Investor inattention or over-optimism may create opportunities for banks to "cherry pick," meaning, in our context, to retain higher quality loans and sell those with higher expected loss, conditional on information ignored or misperceived by investors. Cherry picking reverses the relationship between loan sales and credit risk implied by the previous arguments. A bank that "cherry picks" transfers risk via loan sale or securitization, along risk dimensions that the buyer or investors tend to disregard or where their risk assessments tend to be imprecise or overly optimistic. Cherry picking may not be in a bank's longer term interest if there are potential reputational consequences.

The risk dimensions associated with cherry picking behavior may correspond to factors not considered by the market, such as elevated probability of default or loss given default associated with particular neighborhoods. Alternatively, it may correspond to factors such as borrower payment capacity which the bank can assess more accurately than investors.

The empirical analysis that follows focuses on particular risk dimensions viewed as likely to be associated with "cherry picking." These include neighborhood-specific risk factors and the borrower's calculated debt payment capacity. Our premise is that depository institutions, because of ties to local markets, superior risk controls, greater diligence, or better aligned incentives, were more mindful of risk along these dimensions than subprime mortgage-backed security investors and rating agencies.

Lemons markets. The "cherry-picking" concept introduced above is similar to but distinct from the traditional concept of a "lemons market" as applied to loan securitization. In the traditional lemons market (Akerlof 1970), sellers (lenders) have more information than buyers (security investors), and each are aware of the information asymmetry.

The information asymmetry in a lemons market may encourage the sale of lower quality goods (securitization of riskier loans), and one possible outcome is market collapse. Other outcomes are possible, however, depending on the structure of the market and the strategies that are feasible. In the context of loan securitization markets, contracts may evolve that give lenders incentives to pool risks, as in Passmore and Sparks (1996), or to separate risks, as in DeMarzo and Duffie (1999). In the separating case , higher-risk loans may be selected for securitization and the lower-risk loans retained, or the opposite outcome could prevail.

For instance, DeMarzo and Duffie (1999) propose a model in which asymmetric information drives securitization of riskier loans. A lender receives a higher price for a lower-risk security (where the credit quality differential is observable only to the lender) by keeping a fraction of the security issue to signal the value of the lender's private information.[7] This signaling through repurchase of the security entails a

[7] DeMarzo (2005) goes further by modeling how informed market participants create derivative securities by tranching pools of loans using specific risk characteristics.

cost and, therefore, provides a disincentive to securitize higher-quality loans. Moreover, the signaling enables the market to distinguish higher- from lower-risk security issues and to price accordingly.

In contrast to the lemons market, where investors are aware of their information disadvantage, cherry picking occurs when investors are inattentive. This might be the case, for instance, in an environment where investors are unaware of certain risks, mistakenly relax their due diligence activity, or are overly optimistic. Although cherry picking might be associated with information advantage of the seller, this is not a prerequisite. Cherry picking may occur when buyers have access to the same information as the seller but do not make use of it.

4. Data

Our empirical analysis examines the loan sale decision of depository institutions regarding subprime home purchase mortgages originated in 2005 and 2006. The analysis relies on loan-level HMDA data, which indicate whether a loan was sold during the year when it was originated; the date of origination; and the type of institution purchasing the loan; and a variety of loan, borrower, and property characteristics that we use to develop our empirical analysis.[8] The latter include the loan amount; loan purpose (home purchase or refinance) and type (conventional or government insured); identity of the institution that originated the loan; income of the borrower, state, county, and census tract location of the property being financed; and ownership status of the property (owner-occupied primary residence or not). For high-cost loans, defined in the HMDA regulation as first-lien loans with APR spread greater than or equal to 300 basis points above the applicable Treasury yield (500 basis points for junior liens), the APR spread is also provided. We supplement the HMDA data with institution-specific information: primary regulator; total assets; whether the institution is a subprime lending specialist as defined by HUD; type of institution; and for banks and thrifts, an indicator for whether the institution has a branch in the county where the loan was originated.[9]

[8] Purchaser categories include Fannie Mae; Freddie Mac; Ginnie Mae; Farmer Mac; commercial or savings bank or savings and loan association; affiliated institution; mortgage, finance, or life insurance company or credit union; other type of purchaser. Note that the disposition of a loan originated late in the year generally is not known, since a sale of the loan would typically not occur until the following year.

[9] Institution types include commercial bank; thrift institution (saving and loan institution or savings bank); bank- or thrift-affiliated mortgage company; and independent mortgage company. We thank Robert Avery of the Federal Reserve Board of Governors for supplying the institution-level data.

We restrict the sample to conventional, first-lien, single-family home purchase loans.[10] We then select for our analysis the entire population of high-cost loans along with all loans (high cost or not) originated by HUD-identified subprime specialists; for convenience we label this population subprime.[11] While most of these loans would typically be considered subprime, some may not be, and some loans that are not included in this population may have subprime characteristics.

We further supplement the data with the following local and neighborhood-level housing and mortgage market variables. Annual house-price appreciation rates by metropolitan statistical area (MSA) and state are calculated using the Federal Housing Finance Agency's (FHFA), formerly the Office of Housing Enterprise Oversight (FHFA)) weighted repeat-sales price index. Percent change in housing starts from the previous year, by MSA and state, are obtained from Economy.com. Number of owner-occupied units in the census tract where the property is located is obtained from the 2000 U.S. census. Several neighborhood variables are constructed by aggregating annual information from HMDA to the tract level. These include, by year, the proportion of mortgages that are high-cost loans; the share of subprime mortgages that are originated by HUD-identified subprime lenders; the proportion of home purchase loans that are for non-owner-occupied properties; and the fraction of high-cost home purchase loans that are second lien.[12]

Finally, we access information on subprime loan and borrower characteristics, along with data on subprime payment performance, at the aggregate ZIP code level using the LoanPerformance TrueStandings Servicing™ online data analytics tool. This platform provides aggregated ZIP-code level information on the current payment status of active mortgages and original loan terms and LTV and FICO score ranges for both paid-off and active mortgages serviced by the top mortgage servicing institutions.[13] In contrast to the more commonly used loan-level subprime securities database of

[10] The restriction to home purchase loans is motivated in part for the sake of brevity and in part by the lack of distinction in HMDA data between two important categories of refinance loans: cash-out refinancing used to extract accumulated home equity, and rate-refinancing used to obtain a lower interest rate or to reduce (at least temporarily) the monthly mortgage payment. Moreover, a prior relationship with the borrower is more likely to be a factor in refinance lending. Inability to distinguish such factors complicates interpretation of empirical results based on refinance loans. Results obtained with refinance loans are similar to those reported below for home purchase loans.

[11] We also exclude loans larger than $1,000,000.

[12] The aggregated tract-level measures from HMDA are defined with respect to conventional, single-family, home purchase, and refinance loans originated by the HMDA-reporting institution. Other than the fraction of second liens, they are calculated with respect to first liens only.

[13] This online business intelligence platform accesses the subprime mortgage database of LoanPerformance, a division of FirstAmerican CoreLogic. Information about TrueStandings Servicing® is available at www.loanperformance.com.

LoanPerformance, the subprime servicing database includes subprime loans retained in bank portfolios as well as those in securities.[14]

Specifically, we obtain estimates of first-lien subprime mortgage delinquency and foreclosure rates as of October 2008 by ZIP code and of the proportion of subprime mortgages originated during 2005 and 2006 in each ZIP code by range of origination LTV ratio, FICO score range, and interest rate or product type (such as fixed or adjustable rate). We merge these ZIP-code-level data into our loan-level HMDA data. Since HMDA data indicate the state, county, and census tract associated with a mortgage, not the ZIP code, we first map each state, county, and census tract into one or more ZIP codes.[15]

Overview of the subprime market in 2005 and 2006. Figures 2 through 8 present descriptive information on the composition of the subprime home purchase loan market in 2005 and 2006, providing background for our analysis. Figures 2 through 5 present summary statistics by type of institution from HMDA data. Figures 6 and 7 provide summary statistics on product mix and loan characteristics from the HMDA population merged with the ZIP code level data from LoanPerformance TrueStandings Servicing™, again with 2004 appended for comparison. Figure 8 presents information from HMDA on proportion of subprime, first-lien home purchase loans originated with "piggyback" second liens in 2004 through 2006.

Although the empirical analysis below focuses on the subprime loan sale decision of depository institutions, most subprime loans, however, were originated by nondepository institutions, as shown in figure 2, which provides the distribution of high-cost home purchase loans by type of institution. In 2005, a slim majority of the loans were originated by subprime specialists (among which are only a few, relatively small depository institutions); nonspecialist nondepository institutions had the next highest share, nearly 30 percent, followed by large (more than $10 billion in total assets) and small depository institutions, respectively. In 2006, the share of subprime specialists declined to about 36 percent, while the share of nonspecialist nondepository institutions increased to match that of the specialists. The shares of large and small depositories increased as well, from around 20 to about 30 percent combined share.

---

[14] Loans assigned to the subprime database are serviced by institutions that specialize in servicing subprime loans or are identified as subprime by the servicing institution. Despite the recent demise of most subprime specializing institutions, the subprime database continues to track the performance of active subprime loans, because the servicing of these loans has largely been transferred to other institutions that contribute to the database.

[15] Where a census tract traverses more than one ZIP code, we allocate the tract across the ZIP codes in proportion to loan counts observed in Freddie Mac internal data.

Figure 3 indicates the rationale for restricting attention to depositories: nondepository institutions generally did not retain loans but relied on an originate-to-sell business model. Figure 3 shows the disposition of subprime home purchase loans originated in 2005 and 2006 (whether sold or retained by the end of the year, and type of purchaser), by type of institution originating the loan. We excludes loan originated in November and December because of the relatively short period for observing the loan sale. Similar patterns are observed each year. More than 90 percent of the originations of subprime specialist institutions are reported sold to nonaffiliated non-banking institutions; it is likely that most or all of these loans were packaged in securities and sold to investors. Less than 10 percent are reported as not sold, some of which may be in the securitization "pipeline" or are "warehouse" loans being held for sale, while some may be held for investment. At nonspecialist non-depository institutions (including independent mortgage companies as well as mortgage subsidiaries of depository institutions or of bank or thrift holding companies), the share sold to nonaffiliated non-banking institutions was somewhat smaller, closer to 70 percent, reflecting a somewhat larger share sold to affiliated institutions and banking organizations. The share reported as not sold was again less than 10 percent. In contrast, about half of the loans of large depository institutions and less than 20 percent of the originations of small depository institutions are reported sold to nonaffiliated non-banking institutions. More than 80 percent of the originations of small depository institutions are reported as not sold.

Figures 4 and 5 suggest differences in lending practices across types of institution categories, particularly between small depository institutions and others. Figure 4 indicates the average ratio of loan amount to income by type of institution in 2005 and 2006. This ratio is highest at subprime specialists, close to 2.5 in each year, suggesting relatively high risk exposure with respect to the repayment capacity of borrowers. It is marginally smaller at large depository institutions and nonspecialist nondepository institutions, in the range of 2.2 to 2.4, and much lower, about 1.5, at small depository institutions.[16] Average loan sizes show a similar pattern by institution category, as indicated in Figure 5. When viewed alongside the disposition patterns in Figure 3, these differences are broadly consistent with the previously cited studies suggesting that securitization encouraged lax screening of borrowers.

Figure 6 indicates that about half of the dollar volume of subprime home purchase mortgages originated in 2005 was in hybrid ARMs (mostly 2-28 and 3-27); about 30 percent was standard ARMs;

[16] Note that the decline in the average loan-amount-to-income ratio between 2005 and 2006 does not necessarily imply declines in payment-to-income ratios, since subprime product mix changed, and shorter-term interest rates rose.

about 10 percent were fixed-rate (FRMs); and the remainder (less than 10 percent) was a mix of nontraditional products (balloon and others, including interest-only mortgages). From 2005 to 2006, substantial shifts occurred from hybrid ARMs to the nontraditional products and from FRMs to standard ARMs, a continuation of a shift occurring between 2004 and 2005.

As shown in Figure 7, around a third of subprime home purchase mortgages in 2005 and 2006 were categorized as low documentation. A slim majority had FICO scores greater than 620, indicating that borrower credit rating was only one of a number of factors distinguishing subprime from prime loans. Fewer than 20 percent have an LTV of 90 percent or greater. These proportions were similar in 2004.

Note, however, that these data indicate only the LTV associated with the first lien. Figure 8 provides estimates of the proportion of first-lien, subprime home purchase loans associated with "piggyback" second liens in 2004, 2005, and 2006.[17] About half of first-lien subprime loans originated in 2005 and 2006 had a "piggyback" second, a considerable increase compared with 2004 and even more so compared with previous years.[18]

5. Empirical Approach

We analyze the disposition of subprime loans originated by depository institutions in 2005 and 2006 in relation to factors associated with credit risk, factors associated with transactions costs, and a variety of control variables, by estimating logit regression models. In particular, we examine the disposition of loans in relation to APR spread, which reflects the markets assessment of credit risk based on generally observed risk characteristics, and also in relation to particular risk dimensions viewed as likely to be associated with "cherry picking." A finding that likelihood of sale is inversely related to credit quality along the latter dimensions would be consistent with cherry picking behavior.

We estimate separate equations for large and small institutions (greater or less than $10 billion in assets). We include institution-specific fixed effects in the equation for large institutions and a vector of institution-specific characteristics in the equation for small institutions.

[17] Our estimate of proportion of first liens with a piggyback second is based on HMDA data with matched first and second liens (generously provided by Robert Avery of the Federal Reserve Board) and is calculated as follows. Let N = total number of matched seconds (subprime or prime) in these data; M = total number of subprime firsts; n1= number matched to a subprime first; n2 = number matched to a prime first; and Y= total no. of unmatched seconds (subprime or prime). We impute y1=Y*(n1/N) to be the number of unmatched seconds that are piggybacks to a subprime first. Then the total matched to subprime first-lien (actual + imputed) = n1 + y1, and the proportion of firsts with a piggyback second is (n1 + y1)/M.

[18] Lien status is not provided in HMDA data prior to 2004. However, monthly 2004 HMDA data indicate rapid growth in second-lien loan originations during the year.

Sample restrictions. Most of the original HMDA sample (75 percent of the 2005 sample and 70 percent for 2006) is excluded by the restriction to loans originated by depository institutions, as shown in Table 1, lines 1 and 2a. We further restrict the sample by excluding loans sold to affiliates and or to non-affiliate depository institutions, since the meaning of the sale is ambiguous and the ultimate disposition of the loan is not known (they may or may not be resold in securities.) In addition, we exclude the depository subsidiary of AIG and large institutions that originate fewer than 100 loans.[19] The effect of these exclusions on sample size is shown in Table 1, lines 2b and 2c.

As shown in line 3, about 15 percent of the remaining HMDA sample for each year cannot be merged with the LoanPerformance data, because information on loans originated in the ZIP code is lacking in the LoanPerformance database. The post-merge sample sizes are shown in line 4.

Most of the loans excluded on the basis of sale to affiliated or nonaffiliated depository institutions were sales to affiliates by large depository institutions. Although in the aggregate, about 20 percent of the subprime loans originated by large depository institutions in both 2005 and 2006 were sold to affiliates, most of these sales were by institutions that sell all or nearly all of their loans. In estimating the fixed effects model, where institution-specific effects are accounted for, such institutions are necessarily excluded from the sample. Thus, we need not be concerned about censoring bias arising from these exclusions.[20]

Because we allow for systematic differences across institutions through inclusion of fixed effects, we exclude from the large bank sample institutions that sell more than 90 percent or fewer than 5 percent of the loans they originate. Row 5 of Table 1 reports the impact of these exclusions on sample size.[21]

---

[19] These institutions are viewed as nonrepresentative of large depository institutions that are active in the mortgage market. The AIG subsidiary is excluded because its parent is an insurance company. The large institutions that originate fewer than 100 loans are not active mortgage lenders and their loans represent a negligible share of the large institution sample. Moreover, it is convenient to exclude them for the purpose of estimating a fixed effects model. Note that depository subsidiaries of investment banks also end up excluded from the final sample, on the basis of criteria introduced below, because they sell more than 90 percent of the loans they originate.

[20] With exclusion of the institutions that sell all or nearly all of their loans, the share of the large bank sample excluded on the basis of sale-to-affiliate is 7 percent in 2005 and 8 percent in 2006.

[21] We apply a weaker criterion at the upper end to account for the fact that sales of loans originated near the end of the year are underreported. Three institutions in 2005 and four in 2006 were excluded because they had fewer than 5 percent sold loans, and ten institutions in 2005 and eight in 2006 were excluded because they had more than 90 percent sold loans. In addition, for consistency, for both years we exclude Fremont Mortgage, which straddled the 90 percent threshold, meeting it in 2005 and falling just short of it in 2006.

The final large bank samples are composed of 24 and 31 institutions, respectively, for 2005 and 2006. A list of these institutions is available from the authors on request.

Risk variables. Our empirical analysis distinguishes between dimensions of mutually observed and competitively priced risk and dimensions likely to be disregarded by the securitization market and therefore be associated with cream-skimming. Risk factors observable to the market as a whole we expect would be reflected in the loan's APR spread, which we capture with a set of dummy variables:

- APR spread—a set of dummy variables indicating the ranges [3.0, 3.25), [3.25, 5.25), [5.25, 7.0), or > 7.0

Large APR spreads are associated with borrowers who are perceived (by both the originator and the market) to have greater risk of default. They may also be associated with elevated prepayment risk; that is, greater sensitivity of prepayment speeds to changing interest rates.[22] Very high APR loans may also involve legal risks arising from predatory lending laws and regulations.

An important limitation of the APR spreads reported in HMDA data is that they are based on nominal maturities, and therefore are not necessarily comparable across fixed-rate and variable-rate mortgage products having the same nominal maturity. Fixed-rate mortgages are tied to longer term market interest rates, whereas adjustable-rate mortgages are tied to shorter term market rates.[23] Moreover, because interest rate type is not reported in HMDA data, we cannot control for these differences. This problem is mitigated in our study by the fact the flat yield curve environment during the period examined. During the first half of 2005, long-term rates exceeded short term rates by 75-150 basis points. During the second half of 2005, the gap closed as short-term rates rose faster than long-term rates, and in 2006 the yield curve was very flat, with short-term rates close to and at times marginally higher than long-term rates. Moreover, we conduct our analysis separately for 2005 and 2006. As described below, results are very similar for both years.

The empirical analysis incorporates several, neighborhood-specific risk factors viewed as likely to be associated with "cherry picking." Our premise is that depository institutions, primarily because of ties to local markets but also possibly because of greater diligence or better aligned incentives, had been more mindful of risk along these dimensions than subprime mortgage-backed security investors and rating

[22] Credit risk is generally regarded as the primary factor determining the pricing of subprime mortgages. Credit and prepayment risk are likely to have similar impacts on a depository institution's incentive to sell or securitize a mortgage; for instance, the incentive to sell higher-risk loans that derives from regulatory capital requirements applies to both credit and prepayment risk.

[23] See Avery, Brevoort, and Canner 2006 for discussion of this issue.

agencies. We employ two alternative approaches to test for cherry picking related to neighborhood-specific information. The first approach, termed Model 1, includes each of the following neighborhood risk factors as explanatory variables in the logit regression equations:

- Percent of home purchase loans originated in census tract that are subprime
- Percent of census tract's subprime loans that are originated by subprime specialists
- Housing market depth (log of the number of owner-occupied units in the census tract)

The second approach, Model 2, includes the single neighborhood variable:

- Future (January 2008) subprime mortgage default rate in the census tract[24]

Neighborhood concentrations of subprime loans can generate elevated neighborhood default rates with adverse spillover effects on collateral values.[25] Concentrations of loans originated by subprime specialists can have similar effects.[26] Neighborhoods with fewer home sales will tend to have fewer informed appraisals and, therefore, a higher default risk associated with a particular measured LTV ratio.[27] If investors tended to disregard risk along these dimensions, enabling depository institutions to cherry pick, then the likelihood of loan sale would increase in relation to the first two measures and decrease in relation to the third.[28] Similarly, cherry picking implies that the likelihood of sale would be positively related to the ex-post neighborhood default rate.

In both Models 1 and 2, we include two additional variables that we associate with cherry picking:

- An indicator for whether the institution has a branch in the county where the property is located
- The ratio of loan amount to borrower income

---

[24] The default rate is measured as number of loans 90-days or more delinquent or in foreclosure, divided by number of active loans.

[25] See Lee (2008) for a literature review of price-related spillover effects.

[26] About 10 percent of the loans originated by subprime specialists each year in 2005 and 2006 were not high cost as defined in HMDA data but would have been higher risk than loans originated by prime lenders. Subprime specialists may also have tended to serve higher-risk segments among borrowers obtaining high-cost loans.

[27] It is commonly argued that housing market depth is associated with the accuracy of appraised values and the resulting LTV ratios (Nakamura 2010b). Because of potential endogeneity of number of home sales, we employ the log of the number of owner-occupied units as an instrument.

[28] Alternatively, subprime lenders may have relied disproportionately on brokers, in which case subprime share or share originated by subprime specialists may be correlated with share originated by brokers. As discussed below, broker channel may be a source of risk concerning which investors may have incomplete information, creating an opportunity for cherry picking.

Loans originated "out-of-market" (where the institution has no branch presence) more likely had been originated through mortgage brokers compared to loans in areas covered by the bank's branch network. Broker originations have been associated with agency problems and elevated credit risk (Jiang, Nelson, and Vytlacil 2009, Keys et al. 2009). The ratio of loan amount to income is a proxy for debt-repayment capacity, where a larger loan amount in relation to income indicates a greater payment burden and increased credit risk.

The rationale for associating these risk factors with cherry picking merits some discussion. If origination channel is not consistently reported to investors, or investors disregard this information, then the possibility of cherry picking based on origination channel arises.[29] Such cherry picking would be reflected in the likelihood of sale being positively related to branch presence. A large percentage of the subprime and other subprime mortgages originated in 2005 and 2006 were adjustable rate mortgages (ARMs) that had artificially low initial (teaser) interest rates or were subject to an initial, interest-only period.[30] Often, the recorded debt payment-to-income ratios used in credit risk evaluations were based on the initial interest rate; for instance, this is why the banking regulatory agencies in June 2007 released a "Statement on Subprime Mortgage Lending" requiring lenders to qualify borrowers on the basis of the fully indexed rate. Thus, loan-to-income ratio is a dimension along which credit risk is likely to have been misinterpreted by rating agencies and security investors relying on reported debt payment-to-income ratios.

Depository institutions may have gained an opportunity to cherry pick by applying more robust measures of repayment capacity of subprime borrowers than those used by ratings agencies or security investors. Note that the advantage here relates to how information is used rather than its availability, and as such could reflect superior controls over information processes. Cherry picking in this case would imply that the likelihood of sale increases with the loan-to-income ratio.

---

[29] Origination channel is a requested but not a required field for rating agency RMBS models. Even when loans originated through the wholesale channel are identified, the investor or rating agency likely would not be able to distinguish between third-party originators who have long-term relationships with the bank from those that do not. The latter would be more strongly associated with agency problems and increased credit risk (Jiang, Nelson, and Vytlacil 2009).

[30] Reeves and Weaver (2007) find that the holding the index rate constant, the bulk of securitized subprime ARM mortgages originated in 2005 and 2006 would migrate to higher debt payment-to-income ratios following the reset date, reflecting the presence of an initial teaser rate. For example, more than 90 percent of mortgages with initial debt payment-to-income ratios in the 40-45 percent range would migrate to a higher debt payment-to-income bucket.

A caveat is that we cannot control for loan-level risk factors that are likely correlated with the loan-to-income ratio but less readily associated with cherry picking because they are standard variables for assessing mortgage credit risk. For instance, loan-to-income ratio likely has a positive correlation with the loan-to-value ratio and inverse correlation with borrower credit score. Regardless of such correlation, loan-to-income ratio was a dimension likely associated with flawed risk assessments in the securitization market during this period, for the reasons just discussed. Thus, despite this caveat, a finding that likelihood of sale increases with the loan-to-income ratio would seem to be evidence of cherry picking.

Other local market variables. Models 1 and 2 also include two measures of local area housing market conditions:

- MSA or (for non-MSA areas) state annual house-price appreciation
- MSA or state annual percent change in housing starts

Possibly, the securitization market was less mindful of these local area risk dimensions than banks, although not likely to the same degree as at the neighborhood level. The relationship of these variables to perceived credit risk during 2005 and 2006 is somewhat ambiguous. While rising house prices and growth in housing starts usually imply reduced risk of default, they can also be associated with overheated or volatile markets.

Model 1 incorporates some additional neighborhood (census tract) variables such that, although they may capture neighborhood risk factors, their expected relationship to cherry picking is somewhat ambiguous. (We include them in Model 2 only when checking robustness.) They are:

- Percent of subprime home purchase loans that are junior lien
- Percent of subprime, first- lien home purchase loans tract that are high LTV ($LTV \geq 90$)
- Percent of subprime, first- lien home purchase loans that are low FICO ($FICO < 620$)

On the one hand, neighborhood concentrations of junior-lien, high-LTV, or other higher-risk subprime loans can generate elevated neighborhood default rates, providing an incentive to cherry pick.[31] On the

[31] Further, there are some indications that rating agency models underestimated the credit risk of loans originated with a piggyback second lien; if so, an opportunity to "cherry pick" based on piggyback status could arise. See http://online.wsj.com/article/SB118714461352698015.html

other hand, these variables might proxy for the underlying risk composition of the subprime loans originated in the neighborhood by depository institutions.[32] Since piggyback status, LTV, and FICO score are standard variables for assessing mortgage credit risk , they are not as readily associated with cherry picking. Moreover, an observed, inverse relationship between the borrower's equity stake in the home and the likelihood of sale, particularly at smaller banks, might reflect a risk diversification rather than cherry picking motive. Banks seeking to eliminate geographic concentrations of loans in their portfolios in order to mitigate collateral risk tied to house-price volatility would tend to focus on loans where the borrower's equity stake is small, for which such risks are magnified.

Bank characteristics. As noted, the model estimated for large banks incorporates institution-specific fixed effects. For small banks, the following institution-specific characteristics are controlled for:

- Dummy variable identifying subprime specialist institutions[33]
- Set of dummy variables indicating primary regulator (OCC, Federal Reserve, FDIC, or OTS)
- Dummy variable for thrift institutions[34]
- Institution size (log of total assets)

In particular, we expect that smaller institutions will face higher transactions costs associated with loan sale or securitization, so that the likelihood of sale will increase with size of the institution.

Other control variables. Additional control variables included in both Models 1 and 2 are:

- Loan size—a set of dummy variables indicating the ranges (in $1,000) < 55, [55,155), [155,255), and >255
- A dummy variable indicating whether the subject property is located in a metropolitan area
- A dummy variable identifying loans originated by subprime specialists with APR spread < 3.0
- An indicator for nonprimary residence (non-owner-occupied or second home)

---

[32] Piggyback status is potentially endogenous, because the decision to slice the loan into senior and junior pieces may be associated with a decision to sell one piece and retain the other. The results are robust to dropping this variable.

[33] There are seven such institutions in the small bank sample in both 2005 and 2006.

[34] Thrift institutions are self-identified in the HMDA data; they may have savings and loan or savings bank charters and be OTS or FDIC insured.

In addition, when estimating Model 1, for census tracts with missing values of the two variables from the LoanPerformance data (percent high LTV and percent low FICO), we set these value equal to zero. We then include two dummy variables identifying these cases. Similarly, when estimating Model 2, if the future subprime mortgage default rate in the census tract is missing, we set it equal to zero, and we include a dummy variable for these cases.

Row 6 of Table 1 indicates the number of observations excluded from the estimation samples due to missing data on borrower income, and Rows 7 and 8 provide the final sample sizes for Models 1 and 2, respectively. Table 2 provides sample mean and median values of the continuous variables included in the estimated logit equations, by loan disposition (retained or sold). Summary statistics for the large bank sample are shown in panel A, and for the small bank sample in panel B. Table 3, panels A and B, provides frequencies for the categorical variables (other than institution fixed effects) by value of the variable and loan disposition, for large and small banks, respectively.

Limitations of the analysis. We have already noted potential limitations of the analysis arising from our inability to control at the loan level for loan and borrower characteristics related to credit risk or loan pricing but not reported in HMDA data. These include interest rate type or loan product category, borrower FICO score, and LTV or (for loans with piggyback seconds) combined LTV ratio.

Another limitation is that we cannot observe the extent to which banks may have provided credit enhancements or retained first loss or residual risk positions when selling or securitizing loans, although, anecdotally, sale and securitization of subprime mortgages by depository institutions generally involved significant risk transfer.[35] Moreover, we address only the loan sale or securitization decision of the banking side of the organization—many of the same subprime loans that were packaged into ABS may have cycled back to the trading or investment portfolios of these organizations.

6. Results

Estimation results for Model 1 are presented in Table 4. Results for the small bank sample are shown in panel A, and for the large bank sample in panel B. Estimation results for Model 2 are presented in Table 5 panels A and B for small and large banks, respectively.

---

[35] Typically, the investment bank arranging the sale of the security or a third party would retain first loss or residual risk positions. Covenants and warrants were generally uncommon for most mortgage ABS tranches.

For both models in both years, for both large and small banks, the estimation results indicate that the likelihood of sale declines with APR spread. This relationship conforms to the view that banks tend to retain loans that market participants in general regard to be higher risk. These results are consistent with those of Ambrose, LaCour-Little, and Sanders (2005), who find that banks hold higher-risk loans on book.

In contrast, along the risk dimensions posited to be associated with cherry picking, the findings indicate that depository institutions retain higher-quality loans and sell the lower-quality loans. Thus, the results are consistent with cherry picking in subprime mortgage securitization.

Looking first at the loan-to-income ratio, in all cases the estimated coefficient is statistically significant, with similar magnitude for large and small banks and between Models 1 and 2. A full unit increase in loan-to-income ratio (say, from 2 to 3) is associated with roughly a 20 percent increase in the likelihood of sale in 2005 and around 10 percent in 2006.

All three neighborhood risk factors in Model 1are statistically significant and show the likelihood of sale increasing with risk, both in 2005 and 2006 for both large and small banks. The relationship of log number of owner-occupied units to the likelihood of sale is notably stronger for small compared to large institutions. Possibly, large banks are achieving geographic diversification benefits from retaining loans originated in smaller metropolitan areas, and these benefits partly offset the cherry picking incentive.

In Model 2, for both the large and small bank samples in both years, the likelihood of sale exhibits a statistically significant, positive relation to the future subprime mortgage default rate in the neighborhood. In 2005, for instance, large banks were 18 percent more likely to securitize and small banks 32 percent more likely to securitize per 10-percentage-point increment in future delinquency rate of the census tract. The relationship between loan sale and the expected future default rate suggests that banks anticipated credit deterioration at the neighborhood level and used this information to select which loans to retain and which to securitize.

Results for out-of-market origination are mixed. For small banks, the estimated coefficient on the out-of-market dummy variable is statistically significant in both years and both models and indicates roughly a 30-percentage-point lower likelihood of sale for loans originated out-of-market. For large banks, no relationship is evident. One possible explanation of the different findings for large and small banks is that large banks may more consistently record origination channel. In that case, large banks would more consistently identify loans originated by brokers, which would preclude cherry picking.

Another possible explanation is that large banks achieve greater geographic diversification benefit from retaining loans originated outside of the counties where they have branches, offsetting the incentive to cherry pick these loans.[36]

Other local market variables. Estimation results for local area housing market conditions and for the additional neighborhood risk variables included in Model 1 are mixed. Most notably, for the small bank sample, a larger share of high-LTV and piggyback subprime home purchase loans in a neighborhood is strongly associated with increased likelihood of sale. This result is consistent with the risk diversification motive posited above. Also for small banks, likelihood of sale is strongly, positively related to local area house price appreciation; for large banks, the inverse relationship holds. A possible explanation is that small banks are most dependent on loan sales for funding mortgage originations and in markets where the pace of home sales is high, whereas the liquidity needs of large banks are less tied to local market conditions.

Other control variables. The estimated relationship of loan size to the likelihood of sale differs between small banks, which are most likely to retain the smallest loans, and large banks, which are most likely to retain the largest loans. One possible explanation is that small and large banks face differing types of transactions costs because they access the securities market in different ways. Small banks typically access the market indirectly by selling loans to aggregators who combine the loans of a number of banks into a security. This process may involve a relatively high transactions cost per loan, generating a preference toward selling larger loans. Moreover, selling larger loans may provide greater risk diversification benefit, which may be more important for small than for large banks. Larger banks typically access the securities market directly by packaging their loans into pools, with pool-level transactions costs likely to predominate. This process may generate a preference toward retaining larger loans in order to reduce the per-loan cost of issuing a security.

The estimated relationship to occupancy status also differs between small banks, which are more likely to retain loans for non-primary residence, and large banks, which are more likely to sell these loans. A possible explanation is the role of relationship lending and private information, whereby smaller institutions may have ongoing banking relationships with borrowers financing investment

[36] Out-of-market originations by large banks are likely to be more wide-ranging and distant than those of small banks.

properties or second homes, who tend to be higher quality based on characteristics observable only to their bank. At larger banks, other factors such as risk diversification motives may dominate.

For both models in both years, for both large and small banks, estimation results show that the likelihood of sale is lower for loans originated outside of metropolitan areas. Among small banks, this relationship may reflect higher transactions costs of securitization for banks located in rural areas. Among larger banks, it may reflect geographic diversification benefits associated with retaining loans originated in nonmetropolitan areas. Alternatively, it may reflect cherry picking, since loans originated outside of metropolitan areas have been better performing.[37]

A number of control variables for type of institution included in the estimated equations for small banks are each statistically significant. Institution size exhibits a positive relation to the likelihood of sale, suggesting lower transactions costs of securitization for larger institutions. FDIC-regulated institutions (which are state-chartered and not members of the Federal Reserve System) have the lowest likelihood of loan sale, possibly tied to higher transactions costs. Thrift institutions have the highest likelihood of loan sale, which probably reflects a stronger risk-diversification motive due to the typically narrower product and geographic scope of their lending activities.

Robustness. We explored robustness of the estimation results in a number of ways. To begin with, we dropped from the sample all loans originated in November or December of each year, for information on loan disposition is less complete. In addition, we looked at excluding from the sample loans sold to another banking organization. In each case the results are substantially unchanged.[38]

Dropping from Model 1 the potentially endogenous neighborhood variable percent of census tract subprime home purchase loans that are junior lien has little impact on the estimated coefficients of other variables. Estimating Model 1 using the full HMDA sample, after omitting the two neighborhood variables derived from LoanPerformance data (percent of first-lien subprime home purchase loans that are high LTV and percent low FICO) also has little impact on the estimated coefficients of other variables (with or without inclusion of percent junior lien). Model 1 estimation results also are robust to

---

[37] Mean values for the January 2008 subprime default rate are 19.4 and 14.7 in the 2006 large bank sample and 16.8 and 13.9 in the 2006 small bank sample, for metropolitan and nonmetropolitan areas, respectively, and are nearly the same in the 2005 samples.

[38] Including observations with missing income (by setting the loan-to-income ratio to zero for these observations, and identifying them with a dummy variable in the estimated equations) also has no appreciable impact on any of the Model 1 or Model 2 estimation results.

inclusion of neighborhood measures of subprime product mix (shares of standard ARMs, hybrid ARMs, and nontraditional products) derived from LoanPerformance data.

7. Conclusion

Depository institutions may utilize securitization to “cherry pick,” meaning to transfer risk to investors along dimensions that the investors tend to disregard or where their risk assessments are naïve or overly optimistic. This paper provides evidence of cherry picking by depository institutions in the subprime mortgage securitization market. Using Home Mortgage Disclosure Act (HMDA) data merged with data on subprime loan delinquency by ZIP code, we examine the bank decision to sell (securitize) subprime mortgages originated in 2005 and 2006. We find that the likelihood of sale increases with risk along dimensions observable to banks but not likely observed or considered by investors.

In particular, the likelihood of sale increases with the ratio of loan amount to borrower income, consistent with the premise that depository institutions applied more robust measures of borrower repayment capacity than those used by ratings agencies or investors. The likelihood of sale also increases along neighborhood dimensions of credit risk, including the proportion of loans that are high cost and the proportion originated by subprime specialists in the ZIP code where the property is located. Likewise, it is inversely related to the depth of the neighborhood housing market. In addition, among small depository institutions, the likelihood of sale is greater for loans originated out-of-market. For both large and small institutions, the likelihood of sale is positively related to the ex-post (January 2008) rate of serious delinquency among subprime loans in the ZIP code where the property is located, when this is substituted for individual neighborhood risk factors in an alternative model specification. These findings are consistent with the premise that depository institutions had been more mindful of risk along these dimensions than subprime mortgage-backed security investors and rating agencies.

While these findings are consistent with either cherry picking, they would also be consistent with lemons market-type explanations. Distinguishing lemons market effects from cherry-picking is beyond the scope of our study, because we cannot determine whether investors were aware of an information disadvantage or whether loan terms or securitization structure were employed as separating mechanisms. We favor the cherry picking view for several reasons. First, we have no reason to suspect, that loan terms or securitization structure varied along the particular risk dimensions we associate with cherry picking. Second, the cherry picking explanation is consistent with anecdotal evidence concerning lack of attentiveness to risk within the mortgage securitization market during this period. Third, it seems more consistent with the ex-post outcome in this market, where credit losses have greatly exceeded expectations. Finally, the cherry picking explanation also aligns with recent research including Keys *et al.* (2010a,b) highlighting gaps in risk detection among ratings agencies and investors.

Since depository institutions' share of the subprime loan origination market was relatively small, cherry picking by these institutions cannot be considered a major factor precipitating the subsequent breakdown of the nonagency mortgage securitization market. Nonetheless, the findings suggest that investors had been inattentive to risk or ill-informed along various dimensions, including the repayment capacity of the borrower and neighborhood concentrations of credit risk. Alternatively, they had less effective controls around information processes than depository institution originators, because of a breakdown of due diligence or misaligned incentives. From this perspective, the findings contribute to an understanding of the dynamics leading to the market collapse, and they are consistent with and complementary to other recent research highlighting the role of agency problems in the subprime mortgage securitization market.

A caveat is that we cannot observe the extent to which banks may have provided credit enhancements or retained first loss or residual risk positions when selling or securitizing loans. Also, we address only the loan sale or securitization decision of the banking side of the organization. Many of the same subprime loans that were packaged into ABS may have cycled back to the trading or investment portfolios of these organizations.

Table 1: Sample Sizes (Loan Counts) at Various Stages of Sample Development

Stages of Sample Development	2005	2006
1. Initial HMDA sample[1]	1,404,749	1,289,036
2. HMDA Exclusions		
(a) Non-Depository institution[2]	(1,063,743)	(891,255)
(b) Sold to affiliate or depository	(43,453)	(48,114)
(c) Other exclusions[3]	(17,160)	(2,984)
(c) Remaining	280,393	346,683
3. Merge to LoanPerformance TrueStandings Data[4]		
Missing data		
(a) Large institutions	(22,853)	(31,426)
(b) Small institutions	(18,588)	(21,590)
4. Remaining	238,952	293,667
(a) Large institutions	181,101	230,090
(b) Small institutions	57,851	63,577
5. Percent sold exclusions		
Large institutions only		
(a) Sold < 5% or Sold >90%	(121,888)	(120,311)
(b) Remaining	59,213	109,779
6. Missing income		
(a) Large institutions	(4,366)	(7,365)
(b) Small institutions	(4,792)	(5,763)
7. Final Sample: Model 1 with Fixed Effects[5]		
Large institutions[6]		
(a) Observations read	54,847	102,414
(b) Observations used	54,798	102,239
(c) Total sold	25,007	58,474
Small institutions		
(d) Observations read	53,059	57,814
(e) Observations used	53,026	57,631
(f) Total sold	10,128	11,495
8. Final Sample: Model 2 without Fixed Effects[5]		
Large institutions[6]		
(a) Observations read	54,847	102,414
(b) Observations used	54,847	102,414
(c) Total sold	25,024	58,587
Small institutions		
(a) Observations read	53,059	57,814
(b) Observations used	53,051	57,687
(c) Total sold	10,141	11,513

[1] The HMDA base population consists of subprime, home purchase, first-lien originations, excluding loans in U.S. territories, missing geography, and outliers for loan amount, rate spread, and loan-to-income ratio.

[2] Depository institutions include commercial banks, savings banks, and credit unions. Nondepository institutions refer to commercial bank subsidiaries, subsidiaries of commercial bank holding companies, thrift institution subsidiaries, subsidiaries of a thrift holding company, liquidated commercial bank or thrift institutions, subsidiaries of a credit union, or independent mortgage banks.

[3] Other exclusions were large institutions with less than 100 subprime loans and subprime loans originated by the depository subsidiary of AIG.

[4]The number reported here is the number dropped from the estimation sample for Model 1, which is slightly different from that for Model 2.

[5] An additional number of observations were excluded due to missing information on number of owner-occupied units.

Table 2: Mean and Median Values of Continuous Variables

Panel A: Banks with assets < 10 billion

Variable	Retained/ Sold	2005			2006	
		mean	median		mean	median
Loan amount	Retained	90.99	56.00		109.72	70.00
	Sold	150.66	122.00		172.16	138.00
APR spread	Retained	4.42	4.01		4.33	3.91
	Sold	4.22	3.86		4.04	3.57
Future delinquency rate in census tract	Retained	15.4%	14.9%		15.8%	15.0%
	Sold	17.4%	16.6%		17.2%	16.4%
Tract pct second lien	Retained	20.7%	20.0%		20.7%	20.0%
	Sold	30.0%	32.0%		30.1%	31.3%
Tract pct FICO< 620	Retained	53.0%	53.1%		54.6%	55.3%
	Sold	47.8%	48.1%		49.5%	50.3%
Tract pct LTV > 90	Retained	19.2%	18.2%		22.7%	22.2%
	Sold	17.1%	16.1%		20.4%	19.1%
Tract pct subprime specialist	Retained	43.7%	44.9%		32.7%	33.3%
	Sold	53.5%	54.8%		38.5%	39.0%
Tract pct high cost	Retained	32.7%	31.0%		34.6%	33.2%
	Sold	31.7%	29.2%		33.8%	31.8%
Loan-to-income ratio	Retained	1.42	1.09		1.52	1.22
	Sold	2.19	2.17		2.13	2.12
Log of census tract owner-occupied units	Retained	7.24	7.29		7.23	7.29
	Sold	7.16	7.21		7.16	7.22
Local area house-price appreciation rate	Retained	9.9%	7.4%		6.5%	6.0%
	Sold	12.8%	8.7%		6.6%	6.4%
Local area pct change in housing starts	Retained	7.8%	5.7%		-8.9%	-9.3%
	Sold	5.5%	4.4%		-12.3%	-13.9%

Table 2, cont'd.

Panel B: Banks with assets > 10 billion

Variable	Retained/ Sold	2005			2006	
		mean	median		mean	median
Loan amount	Retained	167.61	129.00		204.94	159.00
	Sold	203.74	159.00		203.43	165.00
APR spread	Retained	4.44	4.17		4.49	4.04
	Sold	4.81	4.84		5.06	5.11
Future delinquency rate in census tract	Retained	17.9%	17.3%		18.2%	17.5%
	Sold	19.4%	19.0%		19.2%	18.6%
Tract pct second lien	Retained	28.5%	30.8%		29.6%	31.3%
	Sold	33.2%	36.0%		32.9%	35.2%
Tract pct FICO< 620	Retained	46.8%	47.1%		47.6%	48.4%
	Sold	43.4%	43.1%		46.0%	46.7%
Tract pct LTV > 90	Retained	16.6%	15.5%		19.0%	17.4%
	Sold	14.8%	13.7%		18.5%	16.8%
Tract pct subprime specialist	Retained	53.9%	55.4%		39.3%	39.8%
	Sold	59.3%	60.6%		43.4%	43.8%
Tract pct high cost	Retained	31.9%	29.3%		34.5%	32.3%
	Sold	32.8%	30.6%		37.5%	35.7%
Loan-to-income ratio	Retained	2.18	2.15		2.16	2.13
	Sold	2.72	2.72		2.43	2.47
Log of census tract owner-occupied units	Retained	7.14	7.21		7.15	7.23
	Sold	7.10	7.17		7.07	7.15
Local area house-price appreciation rate	Retained	13.6%	10.4%		6.7%	6.4%
	Sold	14.0%	12.1%		6.2%	6.1%
Local area pct change in housing starts	Retained	6.5%	4.8%		-12.2%	-11.7%
	Sold	5.5%	4.4%		-13.7%	-13.5%

Table 3: Mean Values of Categorical Variables

Panel A: Banks with assets < 10 billion

Variable	value	Frequencies		% Sold	
		y2005	y2006	y2005	y2006
Indicator for loan sale	0	76.7%	76.2%	0.0%	0.0%
	1	23.3%	23.8%	100.0%	100.0%
Insufficient data for measuring delinquency rate	0	97.8%	98.5%	23.6%	24.0%
	1	2.2%	1.5%	10.8%	13.1%
Bank has branch in county of origination	0	12.7%	10.8%	33.1%	35.6%
	1	87.3%	89.2%	21.9%	22.4%
Non-owner-occupied property	0	73.8%	72.5%	25.7%	26.0%
	1	26.2%	27.5%	16.7%	18.0%
Insufficient data for measuring LTV	0	99.8%	99.7%	23.4%	23.8%
	1	0.2%	0.3%	11.7%	15.4%
Metropolitan area indicator	0	23.2%	23.2%	13.4%	14.0%
	1	76.8%	76.8%	26.4%	26.8%
HUD subprime specialist	0	96.9%	97.7%	22.4%	23.3%
	1	3.1%	2.3%	54.4%	46.4%
OCC-regulated banks	1	22.9%	24.5%	25.7%	27.8%
FRB	2	11.8%	9.7%	34.2%	29.3%
FDIC	3	41.8%	43.1%	9.1%	13.6%
OTS	4	18.4%	18.4%	46.9%	39.9%
Thrift institution	0	77.8%	77.2%	18.1%	19.9%
	1	22.2%	22.8%	41.7%	37.0%

Table 3, cont'd.

Panel B: Banks with assets > 10 billion

Variable	value	Frequencies		% Sold	
		y2005	y2006	y2005	y2006
Indicator for loan sale	0	19.8%	30.6%	0.0%	0.0%
	1	80.2%	69.4%	100.0%	100.0%
Insufficient data for measuring delinquency rate	0	271.7%	214.5%	80.2%	69.4%
	1	2.0%	1.1%	70.5%	56.7%
Bank has branch in county of origination	0	77.0%	49.9%	79.7%	68.2%
	1	196.7%	165.7%	80.3%	69.7%
Non-owner-occupied property	0	239.0%	171.1%	82.3%	71.2%
	1	34.7%	44.5%	65.4%	62.4%
Insufficient data for measuring LTV	0	273.6%	215.4%	80.2%	69.4%
	1	0.2%	0.3%	69.5%	56.5%
Metropolitan area indicator	0	24.2%	21.2%	67.2%	58.0%
	1	249.6%	194.4%	81.4%	70.6%
HUD subprime specialist	0	271.8%	214.2%	80.1%	69.3%
	1	1.9%	1.4%	97.0%	82.4%
OCC-regulated banks	1	142.4%	77.8%	77.3%	64.0%
FRB	2	5.1%	7.8%	31.9%	39.5%
FDIC	3	99.4%	66.7%	90.5%	84.3%
OTS	4	26.8%	63.2%	66.3%	63.9%
Thrift institution	0	149.4%	88.8%	74.9%	60.0%
	1	124.3%	126.8%	86.5%	75.9%

Table 4: Model 1 Estimation Results

Panel A: Depository institutions with assets less than $10 billion Dependent Variable: Subprime loan was sold (1,0)						
Variable	Description	2005			2006	
		Odds Ratio	Chi Square		Odds Ratio	Chi Square
Specialist	Lender is subprime specialist	0.30	59.9*		0.05	70.6*
OCC1	Lender is a national bank	0.89	0.0		0.99	67.9*
FRB1	Lender is state-chartered bank and member of the Federal Reserve system	0.93	1.8		0.82	0.3
FDIC1	Lender is a state-chartered nonmember bank	0.43	901.5*		0.47	615.1*
OTS1	Lender is an S&L	1.57	104.4*		0.87	2.8
Thrift	Thrift institution indicator	1.41	27.0*		1.87	140.9*
Tract % high cost	Fraction of tract loans that are high cost	2.70	96.3*		3.60	183.0*
Specialist rate0	Lender is subprime specialist; loan not high cost	8.35	35.1*		43.58	47.6*
Rate1	APR spread in [3.0, 3.25)	1.74	49.8*		2.40	187.3*
Rate2	APR spread in (3.25,5.25]	1.34	21.0*		1.70	73.8*
Rate3	APR spread in (5.25,7.0]	1.25	8.1*		1.81	73.7*
Investor	Loan for nonprimary residence	0.79	48.6*		0.71	144.2*
Loan-to-income ratio	Ratio of loan amount to borrower income	1.17	147.9*		1.08	40.1*
Log owner units	Log of census tract owner-occupied units	0.71	219.3*		0.78	150.3*
Lnsize1	Loan amount < $55,000	0.44	192.7*		0.41	307.4*
Lnsize2	Loan amount in [$55,000, $155,000)	1.10	3.5		1.06	2.1
Lnsize3	Loan amount in [$155,000, 255,000)	1.21	13.2*		1.12	7.6*
Metro	Property in an MSA	1.09	5.5**		1.12	12.3*
Branch	Indicator for loan originated in county where bank has a branch	0.72	90.5*		0.63	178.8*
Log assets	Log of institution total assets	1.25	576.4*		1.12	191.2*
Tract % second	Fraction of tract's high cost home purchase loans that are 2nd lien	18.32	485.3*		32.43	866.5*
Tract % Specialist	Subprime specialist share of subprime loans in tract	4.87	194.2*		2.57	73.8*
% LTV GT 90	Fraction of census tract's subprime home purchase loans with LTV ³ 90	4.09	81.6*		4.07	122.5*
LTV Dummy	Insufficient data for measuring LTV distribution	1.34	0.6		1.81	6.5**
% Low FICO	Fraction of census tract's subprime home purchase loans with FICO < 620	0.93	0.3		0.80	4.8**
HPI change	Annual rate of change in local area HPI in year	1.51	9.5*		7.55	43.5*
MSASTARTS	Annual rate of change in local area housing starts in year	0.33	171.6*		1.09	1.1
Sample Size			53,026			57,631
C statistics			0.820			0.775
* Statistically significant at the 1 percent level						
** Statistically significant at the 5 percent level						

Table 4, cont'd.

Panel B: Depository institutions with assets greater than $10 billion Dependent Variable: Subprime loan was sold (1,0)					
Variable	Description	2005		2006	
		Odds Ratio	Chi Square	Odds Ratio	Chi Square
Tract % high cost	Fraction of loans in census tract that are high cost	2.42	132.6*	1.84	101.1*
Rate1	APR spread in [3.0, 3.25)	7.42	480.9*	2.11	267.1*
Rate2	APR spread in [3.25, 5.25)	5.58	373.9*	1.59	121.2*
Rate3	APR spread in [5.25, 7.0)	3.22	166.3*	1.54	102.3*
Investor	Loan indicator for nonprimary residence	1.44	150.9*	1.32	157.0*
Loan-to-income ratio	Ratio (loan-level) created by dividing loan amount by income	1.18	209.2*	1.09	95.2*
Log owner units	Natural log of census tract owner-occupied units	0.94	13.4*	0.97	5.6**
Loan size1	Loan size (0, 55K]	2.48	331.3*	1.96	324.6*
Loan size2	Loan size (55, 155K]	2.78	745.6*	2.31	1098.8*
Loan size3	Loan size (155, 255K]	1.92	327.1*	1.69	533.8*
Metro	Metro area loan indicator	1.11	13.2*	1.19	53.4*
Branch	Indicator for county where bank has a branch	0.98	0.6	1.02	1.5
Tract % Second	Percent of 2nd liens in census tract	1.02	0.0	1.78	51.5*
Tract % Specialist	Percent of HUD subprime specialists in the census tract	1.47	17.2*	1.57	35.5*
% LTV GT 90	Fraction of subprime home purchase loans in the census tract with LTV greater than 90%	2.48	47.1*	2.19	69.9*
LTV dummy	Insufficient data for estimating LTV distribution	1.39	1.3	1.02	0.0
% Low Fico	Fraction of subprime home purchase loans in the census tract with FICO less than 620	1.27	5.9**	0.63	45.8*
HPI change	Annual house-price rate of change for given year	0.39	53.6*	0.38	24.6*
MSA Housing starts	Annual housing starts rate of change for the metropolitan area in given year	0.87	4.5**	1.24	15.2*
Sample Size			57,498		102,239
C statistics			0.777		0.782
* Statistically significant at the 1 percent level					
** Statistically significant at the 5 percent level					

Table 5: Model 2 Estimation Results

Panel A: Depository institutions with assets less than $10 billion Dependent Variable: Subprime loan was sold (1,0)					
Variable	Description	2005		2006	
		Odds Ratio	Chi Square	Odds Ratio	Chi Square
Specialist	Lender is subprime specialist	0.43	30.1*	0.06	60.3*
OCC1	Lender is a national bank	0.83	0.00	1.01	152.7*
FRB1	Lender is state-chartered bank and member of the Federal Reserve system	0.94	13.6*	0.71	0.9
FDIC1	Lender is state-chartered nonmember bank	0.38	1082.5*	0.41	804.6*
OTS1	Lender is an S&L	1.34	77.5*	0.75	0.2
Thrift	Thrift institution indicator	1.58	50.2*	2.05	199.2*
Branch	Indicator for loan originated in county where bank has a branch	0.72	90.0*	0.65	166.1*
Specialist rate0	Lender is subprime specialist; loan not high cost	8.75	38.8*	30.9	40.5*
Rate1	APR spread in [3.0, 3.25)	1.72	49.3*	1.85	100.8*
Rate2	APR spread in (3.25,5.25]	1.42	23.0*	1.32	21.8*
Rate3	APR spread in (5.25,7.0]	1.33	13.1*	1.51	37.6*
Log assets	Log of institution total assets	1.28	735.6*	1.13	248.9*
Investor	Loan for nonprimary residence	0.87	18.4*	0.78	77.8*
Loan-to-income ratio	Ratio of loan amount to borrower income	1.20	217.2*	1.10	63.8*
Lnsize1	Loan amount < $55,000	0.37	334.4*	0.34	555.4*
Lnsize2	Loan amount in [$55,000, $155,000)	1.02	0.2	0.94	3.2
Lnsize3	Loan amount in [$155,000, 255,000)	1.18	11.4*	1.06	0.6
Metro	Property in an MSA	1.49	146.3*	1.50	180.0*
HPI change	Annual rate of change in local area HPI in year	2.53	36.5*	13.67	79.5*
MSASTARTS	Annual rate of change in local area housing starts in year	0.34	181.7*	0.78	10.5*
Tract % Bad	Subprime 90+ delinquency rate in census tract as of Jan. 2008	3.21	35.4*	1.99	13.7*
Bad rate dummy	Insufficient data for measuring tract delinquency rate	0.73	7.8*	0.82	2.8
Sample Size			53,081		57,687
C statistics			0.802		0.752
* Statistically significant at the 1 percent level					
** Statistically significant at the 5 percent level					

Table 5, cont'd.

Variable	Description	2005 Odds Ratio	2005 Chi Square	2006 Odds Ratio	2006 Chi Square
Panel B: Depository institutions with assets greater than $10 billion Dependent Variable: Subprime loan was sold (1,0)					
Branch	Bank has a branch in county were loan was originated	0.99	0.19	1.02	1.4
Rate1	APR spread in [3.0, 3.25)	7.03	458.5*	2.08	264.2*
Rate2	APR spread in [3.25, 5.25)	5.37	359.5*	1.58	120.6*
Rate3	APR spread in [5.25, 7.0)	3.15	161.1*	1.55	106.3*
Investor	Loan indicator for nonprimary residence	1.48	175.4*	1.34	172.2*
Loan-to-income ratio	Ratio (loan-level) created by dividing loan amount by income	1.19	227.7*	1.11	121.6*
Loan size1	Loan size (0, 55K]	3.31	699.3*	2.05	471.3*
Loan size2	Loan size (55, 155K]	3.36	1201.4*	2.37	1493.2*
Loan size3	Loan size (155, 255K]	2.09	430.9*	1.70	573.3*
Metro	Metro area loan indicator	1.15	23.1*	1.28	107.7*
HPI change	Annual house-price rate of change for given year	0.27	128.6*	0.49	13.1*
MSA Housing starts	Annual housing starts rate of change for the metropolitan area in given year	0.90	2.6	1.19	10.4*
Tract % bad	Subprime 90+ days delinquent in census tract as of Jan. 2008	1.73	12.2*	1.70	18.0*
Bad rate dummy	Insufficient data for measuring delinquency rate	0.88	1.6	0.85	2.9
Sample Size			54,847		102,414
C statistics			0.773		0.778
* Statistically significant at the 1 percent level					
** Statistically significant at the 5 percent level					

Figure 1: The Sell/Retain Decision in Relation to Credit Quality

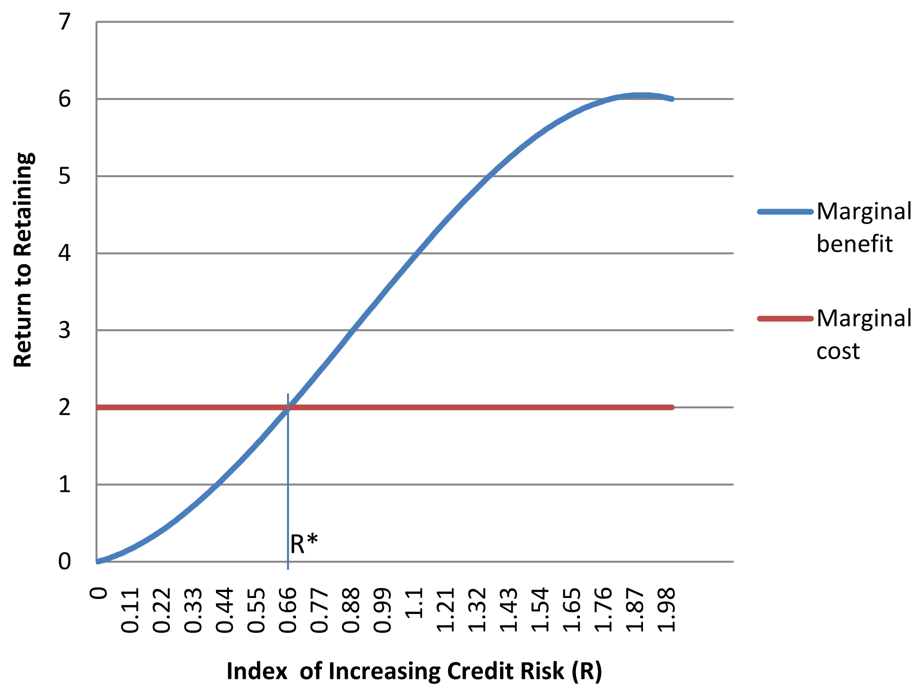


Figure 2: Distribution of high-cost loans by type of institution

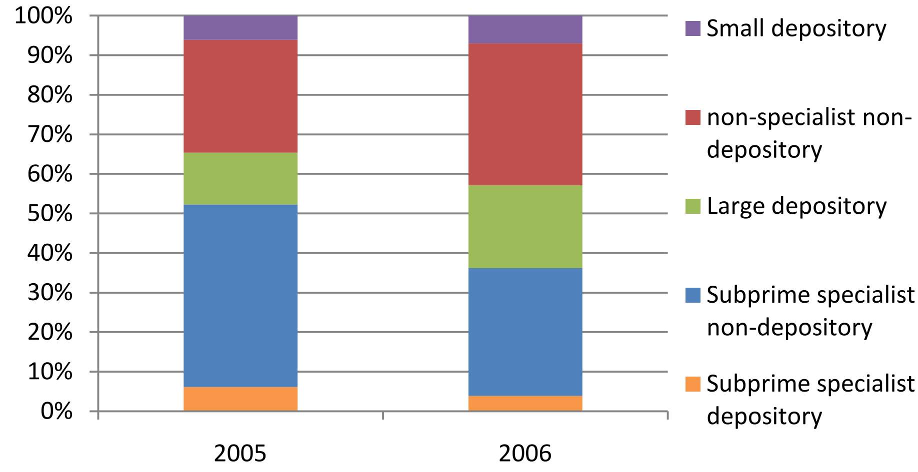
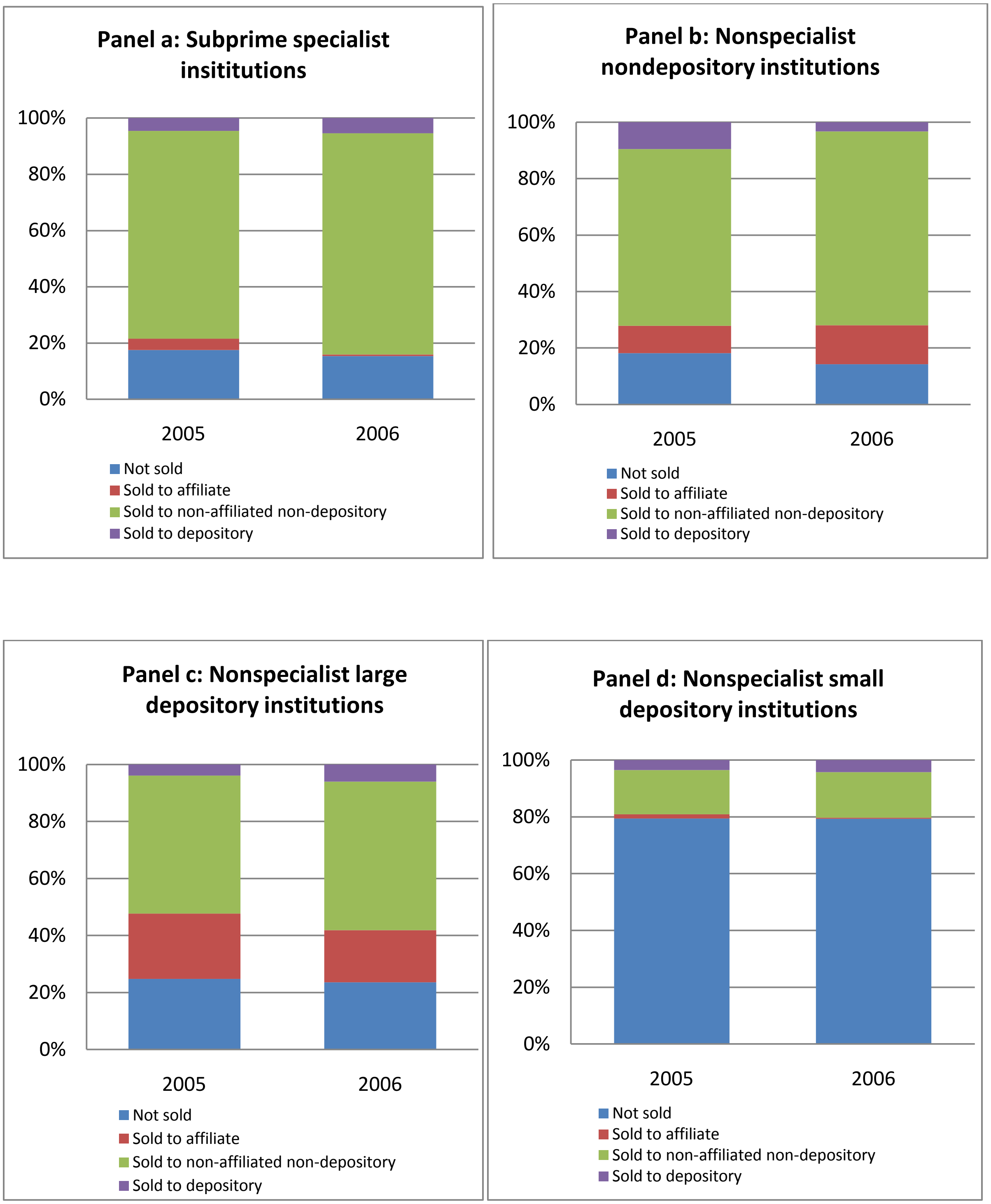


Figure 3: Disposition of Subprime Loans (purchaser type)



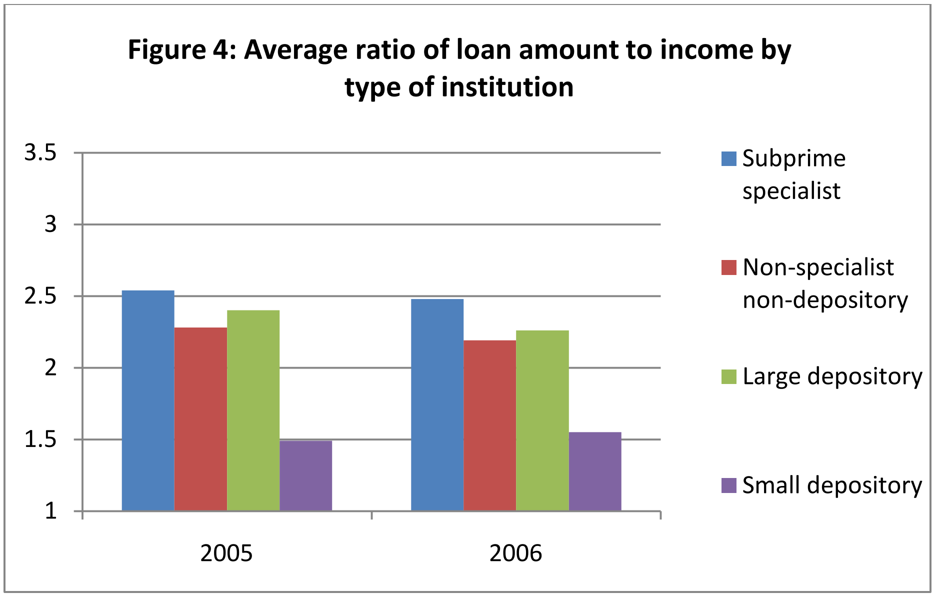


Figure 4: Average ratio of loan amount to income by type of institution
3.5
3
2.5
2
1.5
1
2005
2006
Subprime specialist
Non-specialist non-depository
Large depository
Small depository

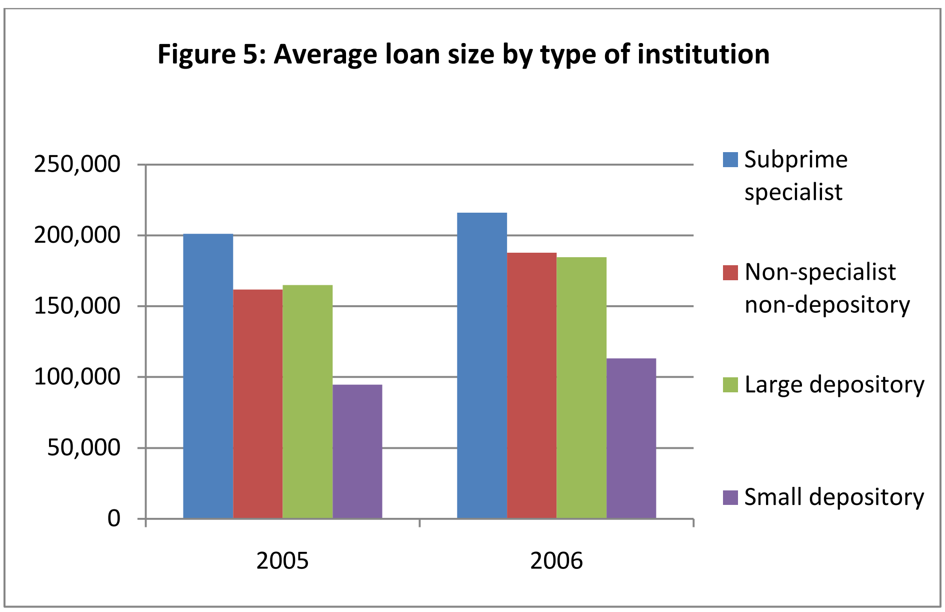


Figure 5: Average loan size by type of institution
250,000
200,000
150,000
100,000
50,000
0
2005
2006
Subprime specialist
Non-specialist non-depository
Large depository
Small depository

Figure 6: Product mix

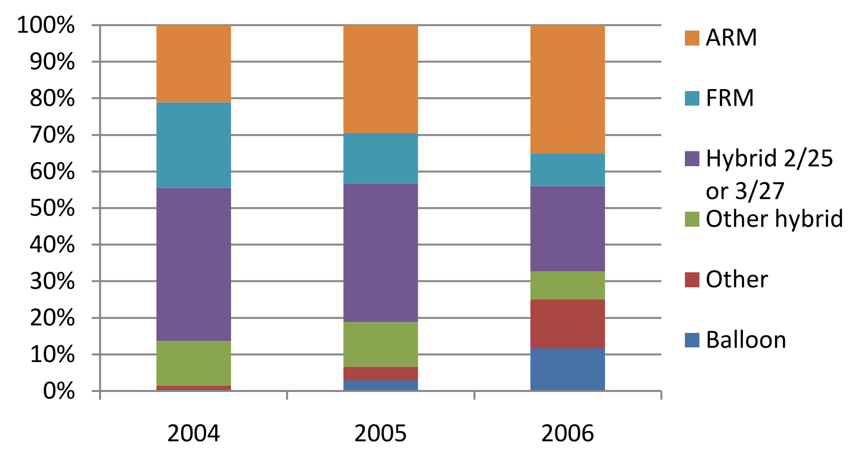
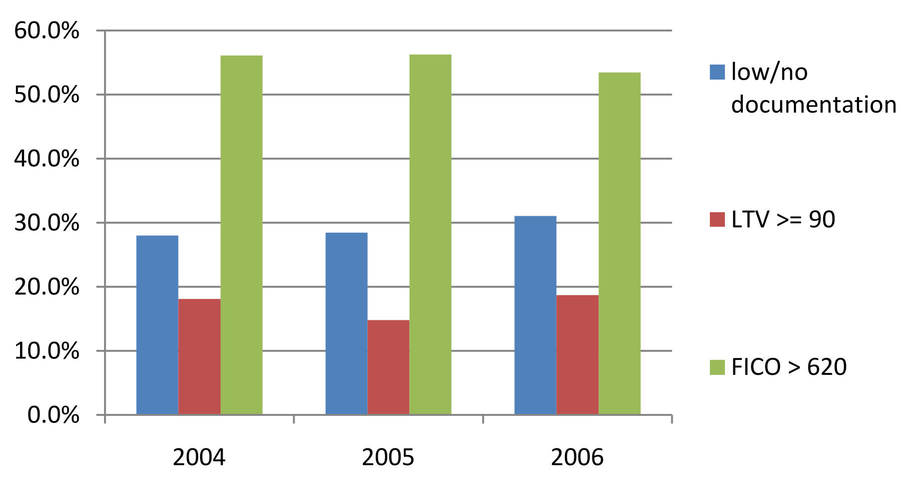


Figure 7: Documentation level, LTV, and FICO



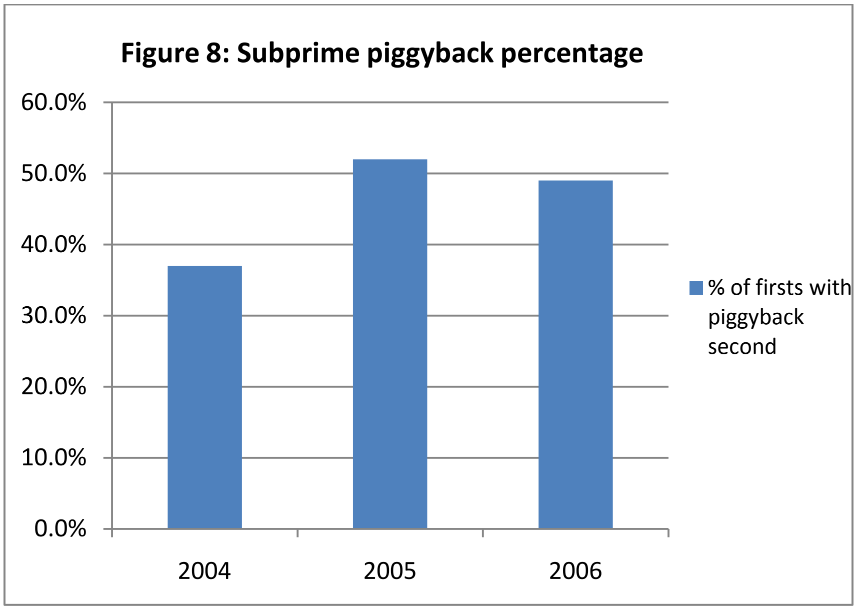


Figure 8: Subprime piggyback percentage
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
2004
2005
2006
% of firsts with piggyback second

References

Akerlof, George A. (1970), “The Market for ’Lemons’: Quality Uncertainty and the Market Mechanism,” *Quarterly Journal of Economics* 84:3, pp. 488-500.

Ambrose, Brent, Michael LaCour-Little, and Anthony Sanders (2005), “Does Regulatory Capital Arbitrage, Reputation, or Asymmetric Information Drive Securitization?” *Journal of Financial Services Research*, 28:1, pp. 113-133.

Ashcraft, Adam, and Til Schuermann (2008), “Understanding the Securitization of Subprime Mortgage Credit,” Federal Reserve Bank of New York Staff Report no. 318.

Avery, Robert B., Kenneth P. Brevoort, and Glenn B. Canner (September 2006), “Higher-Priced Home Lending and the 2005 HMDA Data,” *Federal Reserve Bulletin* 92 pp. A123-A166.

Ben-David, Itzhak (2007), “Financial Constraints, Inflated Home Prices, and Borrower Default during the Real-Estate Boom,” Working Paper no. 2009-03-001, Fisher College of Business, Ohio State University.

Cagan, Christopher L. (March, 2007), “Mortgage Payment Reset the Issue and the Impact,” *First American* CoreLogic.

Calem, Paul S., and James R. Follain (2007), “Regulatory Capital Arbitrage and the Potential Competitive Impact of Basel II in the Market for Residential Mortgages,” *Journal of Real Estate Finance and Economics* 35 (2), pp. 197-219.

Calem, Paul S., and Michael LaCour-Little (2004), “Risk-Based Capital Requirements for Mortgage Loans,” *Journal of Banking & Finance* 28, pp. 647-672.

Coleman, Major, Michael LaCour-Little, and Kerry D. Vandell (2008), “Subprime Lending and the Housing Bubble: Tail Wags Dog?” *Journal of Housing Economics* 17, pp. 272-290.

Coval, Joshua, Jakub Jurek, and Erik Stafford (2009a), “The Economics of Structured Finance,” *Journal of Economic Perspectives* 23(1), pp. 3-25.

Coval, Joshua, Jakub Jurek, and Erik Stafford (2009b), “Economic Catastrophe Bonds,” *American Economic Review* 99 June, 628-666.

DeMarzo, Peter (2005), “The Pooling and Tranching of Securities: A Model of Informed Intermediation,” *Review of Financial Studies Mini Issues* 18(1), pp. 1-35.

DeMarzo, Peter, and Darrell Duffie (January, 1999), “A Liquidity-Based Model of Security Design,”*Econometrica* pp. 65-99.

Demyanyk, Yuliya, and Otto Van Hemert (2009), “Understanding the Subprime Mortgage Crisis,” *Review of Financial Studies Mini Issues,* published online May 4, 2009.

Dewatripont, Mathias, and Jean Tirole (1994), *The Prudential Regulation of Banks.* Cambridge, Massachusetts: The MIT Press

Elul, Ronel (2009), "Securitization and Mortgage Default: Reputation vs. Adverse Selection," Working Paper no. 09-21, Federal Reserve Bank of Philadelphia.

Ernst Keith, Debbie Bocian, and Wei Li (2008), "Steered Wrong: Brokers, Borrowers, and Subprime Loans," Center for Responsible Lending.

Gerardi, Kristopher, Adam Hale Shapiro, and Paul S. Willen (2008), "Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures," *Federal Reserve Bank of Boston Working Papers No. 07-15.*

Golding, Edward, Richard K. Green and Douglas A. McManus (February 2008), "Imperfect Information and the Housing Finance Crisis," *Joint Center for Housing Studies at Harvard University* UCC08-6.

Hahn, Robert and Peter Passell (February, 2008), "Better that the Fed Regulates Subprime Mortgages" *Economist Voice.*

Haughwout, Andrew, Richard Peach, and Joseph Tracy (August 2008), "Juvenile Delinquent Mortgages: Bad Credit or Bad Economy," Federal Reserve Bank of New York Staff Report No. 341.

Hill, Claire A. (Winter, 1996), "Securitization: A Low-Cost Sweetener for Lemons," Washington University *Law Quarterly* pp. 1061-1120.

Hull (2009), "The Credit Crunch of 2007: What Went Wrong? Why? What Lessons Can Be Learned?" Joesph L. Rotman School of Management, University of Toronto.

Jiang, Wei, Ashlyn Nelson, and Edward Vytlacil, (2009), "Liar's Loan? Effects of Loan Origination Channel and Loan Sale on Delinquency," manuscript, Columbia University.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig (2009), "Financial Regulation and Securitization: Evidence from Subprime Loans," *Journal of Monetary Economics*, July, 700-720.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig (2010a), "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *Quarterly Journal of Economics*, 125, February 2010.

Lee, Kai-yan (2008), **"**Foreclosure's Price-Depressing Spillover Effects on Local Properties: A Literature Review," Federal Reserve Bank of Boston Community Affairs Discussion Paper no. 2008-01.

Mian, Atif, and Amir Sufi (2009), "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *Quarterly Journal of Economics,* 124:4, pp. 1449-1996.

Nakamura, Leonard (2010a), "Durable Financial Regulation: Monitoring Financial Instruments as a Counterpart to Regulating Financial Institutions," draft, Federal Reserve Bank of Philadelphia.

Nakamura, Leonard (2010b), "How Much Is That Home Really Worth? Appraisal Bias and House-Price Uncertainty," *Federal Reserve Bank of Philadelphia Business Review* First Quarter.

Passmore, S.W. and R. Sparks (1996), "Putting the Squeeze on a Market for Lemons: Government-Sponsored Mortgage Securitization," *Journal of Real Estate Finance and Economics*, 13 (1), pp. 27-43.

Rajan, Uday, Amit Seru, and Vikrant Vig (2009), "The Failure of Models That Predict Failure: Distance, Incentives and Defaults," University of Chicago Graduate School of Business Research Paper no. 08-19.

Sherlund, Shane M. (2008), "The Past, Present, and the Future of Subprime Mortgages," Federal Reserve Board Finance and Economics Discussion Series 2008-63.

Smith, Brent C. (2007), "The Subprime Mortgage Market: A Review and Compilation of Research and Commentary," *The Homer Hoyt Institute*.

White, Lawrence J. (2009), "The Credit Rating Agencies and the Subprime Debacle," *Critical Review* 21 (2-3), pp. 389-399.

Wray, L. Randall (December 2007), "Lessons from the Subprime Meltdown," Working Paper no. 522, Levy Economics Institute of Bard College.

Appendix 1: Macroeconomic Uncertainty and Securitized Share

In this appendix we develop a simple representation of the choice of share of loans securitized as a decision under uncertainty, where excess securitization (relative to the ex-post realized optimal share) is more costly than retaining too large a share. This cost asymmetry reflects the simple intuition that it is easier to sell surplus balance-sheet assets at a later date than it is to unwind a securitization or replace assets on balance sheet. We show that under plausible conditions, the proportion securitized will decrease as the cost asymmetry and degree of uncertainty increases.

Let X denote the ex-post target share, that is, the share that would have been securitized conditional on a realization of macroeconomic or sector-specific economic conditions. Ex-ante, X is a random variable with density f(x). Let S denote the rate of securitization chosen by the bank ex-ante. Recall from the discussion in the main text that the bank's optimization problem is to choose S to maximize:

(1) , where $\alpha < 1$

We shall assume that X has a beta distribution, denoted F(X, *a, b*), with mean μ and standard deviation σ given by:

(2) $\mu = a/(a+b)$; $\sigma^2 = ab/(a+b)^2(a+b+1) = \mu(1-\mu)/(a+b+1)$; $a > 0$; $b > 0$

The beta distribution has advantages of being bounded between 0 and 1 and encompassing a variety of possibilities, such as uniform, unimodal and bimodal symmetric distributions, and unimodal, asymmetric long-tailed distributions.

We characterize the uncertainty inherent in the beta distribution in terms of its median and interquartile range. Thus, between two beta distributions with the same median, the one with a larger interquartile range we characterize as incorporating more uncertainty. Also, between two distributions both having infinite density at zero and zero density at one (or vice versa), the one with the larger interquartile range (or, equivalently, the one closer to uniform) incorporates more uncertainty. We shall rely on the following results, derived later in this Appendix:

Proposition 1: Let $F(x, a_1, b_1)$ and $F(x, a_2, b_2)$ be two beta distributions with identical medians x^m. If $F(x, a_1, b_1)$ has a wider interquartile range, then $F(x, a_1, b_1) > F(x, a_2, b_2)$ for all $x < x^m$.

Proposition 2: Let $F(x, a_1, b_1)$ and $F(x, a_2, b_2)$ be two beta distributions such that, either $a_i < 1$ and $b_i \geq 1$ for i=1,2, or $a_i \geq 1$ and $b_i < 1$ for i=1,2. If $F(x, a_1, b_1)$ has a wider interquartile range, then $F(x, a_1, b_1) > F(x, a_2, b_2)$ for all $0 < x < 1$.

It is straightforward to verify that the solution to the bank's optimization problem, denoted S*, satisfies:

(3) $F(S^*, a, b) = \alpha/(1 + \alpha)$

Let F^{-1} denote the inverse of the beta distribution $F(X, a, b)$; then:

(4) $S^* = F^{-1}(\alpha/(1 + \alpha), a, b)$

Since $\alpha < 1$, we have $S^* < x^m$, where x^m denotes the median. Therefore, it follows from Proposition 1 that across beta distributions having the same median, S* declines as the interquartile range widens. Likewise, across beta distributions as characterized in Proposition 2, S* declines as the interquartile range widens. Thus, in each case the rate of securitization declines as the degree of ex-ante uncertainty, represented by the interquartile range of the distribution, increases.[39]

It remains to verify Propositions 1 and 2. To verify Proposition 1, let $F(x, a_1, b_1)$ and $F(x, a_2, b_2)$ be two beta distributions with identical medians x^m. Then $F(x, a_1, b_1)$ and $F(x, a_2, b_2)$ intersect at x=0, where $F(x, a_1, b_1) = F(x, a_2, b_2) = 0$; at x=1, where $F(x, a_1, b_1) = F(x, a_2, b_2) = 1$; and at $x = x^m$, where $F(x, a_1, b_1) = F(x, a_2, b_2) = 1/2$. Next, we show that these are the only points at which the two distributions will intersect.

The beta distribution is defined as F(x, a, b)=B(x, a, b)/B(0, a, b), where B(x, a, b) is the beta function. Calculating the second derivative of the beta distribution, we obtain:

(5) $\partial^2F/\partial x^2 = [(a-1)x^{a-2}(1-x)^{b-1} - (b-1)x^{a-1}(1-x)^{b-2}]/B(x, a, b)$

It follows that:

(6) $\partial^2F/\partial x^2 > (=) (<) 0$ if and only if $[(a-1)(1-x) - (b-1)x] > (=) (<) 0$.

[39] Letting δ denote the interquartile range, we have $dS^*/d\delta = -(\partial F/\partial\delta)/(\partial F/\partial x) < 0$.

Hence, there can be at most one x_0 such that $\partial^2F/\partial x^2 = 0$; i.e., at most one inflection point of the distribution, implying that any two beta distributions can intersect at most once in (0, 1). It follows that if $F(x, a_1, b_1)$ and $F(x, a_2, b_2)$ intersect at a common median value x^m, then either $F(x, a_1, b_1) > F(x, a_2, b_2)$ for all $x < x^m$, or vice versa. It then follows immediately that $F(x, a_1, b_1) > F(x, a_2, b_2)$ for all $x < x^m$ if and only if $F(x, a_1, b_1)$ has the wider interquartile range.

To verify Proposition 2, let $F(x, a_1, b_1)$ and $F(x, a_2, b_2)$ be two beta distributions. From (6), if $a_i < 1$ and $b_i \geq 1$ for i=1 and 2, then the distributions are strictly concave; if $a_i \geq 1$ and $b_i < 1$ for i=1 and 2 they are strictly convex. In either case, the distributions will intersect only at 0 and 1, from which it follows that $F(x, a_1, b_1) > F(x, a_2, b_2)$ for all x in (0, 1) if and only if $F(x, a_1, b_1)$ has the wider interquartile range.