# A Loan-level Residential Mortgage-backed Security Pricing Model: are CAP CRA loans profit-making for the secondary market?

**Haiou Zhu**

**A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Public Policy in College of Arts and Sciences.**

**Chapel Hill**
**2010**

**Approved by:**
**David K. Guilkey,**
**Mustafa Gültekin,**
**Roberto G. Quercia,**
**Richard Rendleman,**
**Michael A. Stegman.**

# Abstract


HAIOU ZHU: A Loan-level Residential Mortgage Backed Security Pricing Model: are CAP CRA loans profit-making for the secondary market?
(Under the direction of Mustafa Gültekin)

This paper develops an industry comparable loan-level residential mortgage-backed security pricing model. It can be used to design hedging strategies for mortgage portfolio's interest rate risk, and price the cost of guaranteeing RMBS default risk. The loan-level pricing model is designed to address most of the problems with the Government-Sponsored Enterprises' (GSEs') current risk management models that were outlined in the Federal Housing and Finance Administration's 2009 report to Congress. The loan-level pricing model in this paper is able to automatically translate into RMBS prices the slight monthly changes in individual borrowers' prepayment and default risks due to borrower and loan characteristics, macroeconomic conditions, house price changes, and term structure movements. The loan-level model is especially useful for managing low-to-moderate income (LMI) mortgages, which are highly leveraged assets. Applying the loan-level pricing model to the Community Advantage Program (CAP) dataset yields the result that most (i.e. 65% of the purchased CAP loans) of the Community Reinvestment Act (CRA) mortgages in CAP have been profit-making (i.e. positive Option-Adjusted Spreads OAS) for the secondary market, given the market prices Fannie Mae paid. Moreover, the results suggest that the conventional indicators, such as race, income, credit score and loan-to-value at origination, are not reliable in determinants of the mortgage yield. Therefore, avoiding and discriminating

against LMI mortgage pools are not rational. The identification of responsible LMI borrowers or pools and adequate risk based pricing require that the loan-level pricing model be run on each mortgage portfolio. Finally, the loan-level pricing model can help to address one challenge in the overhaul of mortgage finance system pointed out by Geithner, namely pricing the cost of a government guarantee of RMBS default risk. In particular, the expected cost for guaranteeing the default risk of a loan can be calculated as the difference in the OAS between 100% recovery and recovery at the current house price. In short, the loan-level pricing model developed may help the federal government to better meet the financial needs of responsible LMI borrowers, while maintaining the sustainability and soundness of the GSEs.

# Acknowledgements

I am especially grateful for the help of Janneke Ratcliffe, Sarah Riley, and the Center for Community Capital. This paper has also benefited tremendously from the fantastic comments of each of my dissertation committee members. The financial support of the Ford Foundation, and the support from Self Help Credit Union, Department of Public Policy, and Kenan-Flagler Business School have been essential in helping me to complete this dissertation, which I dedicate to my mother.

## Table of Contents

# List of Tables

# List of Figures

## List of Abbreviations

ABS: asset-backed securities

AMI: area median income

ARM: adjustable-rate mortgage

ATS: affine term structure

CAP: Community Advantage Program

CRA: Community Reinvestment Act

CDO: collateralized debt obligations

CMO: collateralized mortgage obligations

CDFIs: Community Development Financial Institutions

CDR: constant default rate

CRPHM: competing risks proportional hazard model

GSEs: government sponsored enterprises

IIA: independence of irrelevant alternatives

LMI: low-to-moderate income

LTV: loan-to-value

MNL: multinomial logit model

OAS: option-adjusted spread

PHM: proportional hazard model

PMMS: Freddie Mac primary mortgage market survey

RMBS: residential mortgage-backed security

SMM: single monthly mortality

UPB: unpaid principal balance

FHFA: Federal Housing Finance Agency

PSM: propensity score matching

ATM: at-the-money

# Chapter 1. Introduction and Policy Motivations.

## Section 1- a. Problems for practitioners after the Crisis.

The subprime crisis has left unsolved problems for the federal government, the Government sponsored enterprises (GSEs) and low-to-moderate income borrowers. The federal government was left with billions of mortgage-backed (both residential and commercial) CMOs (collateralized mortgage obligations), CDOs (collateralized debt obligations) and ABS (asset-backed security) portfolios[1]. These mortgage-backed portfolios are hard to price and hedge risks using the old copula-based pricing models, which were proven problematic during the crisis. Copula-based models were used by some rating agencies to price residential mortgage-backed securities (RMBS) before the subprime crisis, as discussed in the Fitch Ratings report by Hunt (2007). The evidence of the failure of announced ratings as a useful guide in evaluating mortgage risks is provided by Ashcraft, Goldsmith-Pinkham and Vickery (2010), who conduct a study of 3,144 MBS deals. Specifically, they find that "credit risk estimated by simple model is more informative for predicting deal performance than the announced ratings." Furthermore, the paper by Brigo, Pallavicini and Torresetti (2010), a description by industry experts of copula-based CDO pricing models, points out that the Gaussian copula is "a static model that is little more than a static multivariate distribution which is used in credit derivatives (and in particular CDOs) valuation and risk management." Moreover, Brigo et al. mention that some of the

[1] The financial system was pumped with $200 billion of mortgage- linked CDOs in the months before the subprime crisis spread. See Bloomberg news "How Wing Chau Helped Neo Default in Merrill CDOs Under SEC View" available at http://www.bloomberg.com/apps/news?pid=20601109&sid=adplts9scZkg&pos=11.

deficiencies of the copula model have been known for a while as reported by Salmon (2009)[2].

Furthermore, the GSEs have continued to need huge bailout from the government, while they own or guarantee 53 percent of the nation's $10.7 trillion in residential mortgages. According to Bloomberg news reported on July 2010[3], "the cost of fixing Fannie Mae and Freddie Mac, the mortgage companies that last year bought or guaranteed three-quarters of all U.S. home loans, will be at least $160 billion and could grow to as much as $1 trillion after the biggest bailout in American history." Furthermore, according to the same news, "Fannie and Freddie are deeply wired into the U.S. and global financial systems. Figuring out how to stanch the losses and turn them into sustainable businesses is the biggest piece of unfinished business as Congress negotiates a Wall Street overhaul that could reach President Barack Obama's desk by July."

Currently, most information about the possible reasons underlying the continued losses of the GSEs is ex-post and static. For this reason, it cannot be used to facilitate ex-ante decision making and provide methods for reducing losses. For instance, Table 1, which was generated from Fannie Mae's 2010 1st quarter results, provides confusing and inconsistent implications about which product features are likely to be driving the largest losses. Using the third row of Table 1, which presents the percentage of 2009 credit losses relative to the percentage of loans in the guaranty book of business, as a gauge, one can see that the traditionally high-risk categories of "FICO<620," "620<FICO<660," and "OLTV>90%"

---

[2] See Felix Salmon. "Recipe for disaster: the Formula that killed Wall Street." Wired Magazine, 2009. 17.03. Available at http://www.wired.com/techbiz/it/magazine/17-03/wp_quant?currentPage=all

[3] See Bloomberg news "Fannie-Freddie Fix at $160 Billion With $1 Trillion Worst Case" at http://preview.bloomberg.com/news/2010-06-13/fannie-freddie-fix-expands-to-160-billion-with-worst-case-at-1-trillion.html

exhibit very low losses. In contrast, the “Alt-A” and “Subprime” categories with similar FICO scores and OLTV ratios are among the groups with very high losses. Furthermore, the last two rows of Table 1 provide the weighted average FICO scores and percentages of loans with OLTV>90% for each product feature category. If one considers both the weighted average FICO scores and the relative contributions of each product category to 2009 credit losses (i.e. the third row of Table 1), it is clear that the two categories of lower FICO scores, namely "FICO<620" and "620<FICO<660," exhibit low losses. In contrast, the two categories "Alt-A" and "Subprime," which correspond to similar or higher FICO scores, exhibit high losses. Similar conclusions can be obtained by considering both the percentage of loans with OLTV>90% and the relative contributions of each product category to 2009 credit losses. The three groups with the highest percentages of loans with OLTV>90%, namely "FICO<620," "620<FICO<660," and ""OLTV>90%," are the groups with the lowest losses. Moreover, the remaining groups, which have a smaller fraction of loans with OLTV>90%, all exhibit much higher losses. These findings are inconsistent and counter-intuitive; hence, they cannot be used in ex-ante decision making with the goal of reducing losses.

[Insert Table 1.Fannie Mae credit profile by key product features]

The testimonies[4] of assistant secretary Michael Barr and FHFA director Edward DeMarco provide information about the measures that the government has taken to reduce losses. The measures include tightening the GSEs’ underwriting guidelines according to

[4] See testimony by Assistant Secretary for Financial Institutions Michael S. Barr, before Subcommittee on Capital Markets, Insurance, and Government Sponsored Enterprise of House Committee on Financial Services, Written Testimony as Prepared for Delivery - 9/15/2010, at http://www.ustreas.gov/press/releases/tg854.htm And statement of Edward J. DeMarco, Acting Director Federal Housing Finance Agency, at http://www.fhfa.gov/webfiles/16726/DeMarcoTestimony15Sept2010final.pdf

credit scores, loan-to-value (LTV) ratios, and product features; increasing guarantee fees; and adopting loss mitigation measures through loan modifications and foreclosure prevention. However this paper shows that simple indicators, such as product features, and credit score or LTV at origination, are not reliable indicators of mortgage risks and returns, because mortgage risks and yields change constantly over time. Furthermore, since "95% of the mortgages originated in this country are currently financed through either the GSEs or Ginnie Mae," tightening underwriting guidelines using these simple indicators may not completely stem the losses. Nevertheless, this approach could lead to a possible overcorrection via indiscriminate rejection of profitable LMI mortgages, as demonstrated by the profitable CAP CRA loans. The loan-level pricing model developed provides a way to identify profitable mortgages based upon borrower's historical performance, and will help to encourage responsible borrower behavior by means of fair market risk-based pricing.

The practical challenge in adopting the approach in underwriting is that many borrowers have no historical records at origination if they are purchasing houses for the first time. The problem can be solved by predicting the OAS of borrowers without historical records using historical records of similar borrowers in the same portfolio or historical records of similar borrowers constructed using propensity score matching method as in Ding, Quercia, Lei, Ratcliffe (2008). The propensity score matching (PSM) method used by Quercia et al. (2008) can account for observable heterogeneity by pairing borrowers who took out a certain type of loan (who have historical records) with new borrowers (who have no historical records) on the basis of the conditional probability of taking out the specific type of loan, given the observable characteristics of the borrowers. The observable characteristics used by Quercia et al. (2008) include origination variables drawn from the

CAP, McDash and HMDA datasets, including FICO scores, debt-to-income ratios, and neighborhood market dynamics and credit risk.

In addition, this paper shows that loan age is very important in determining both prepayment and default risks. In particular, the default risk of CAP loans continuously increases as a loan seasons. Although the testimonies mentioned above suggests that currently "less than 1% of the losses have come from loans originated in 2009 and 2010," without the adoption of advanced risk management techniques going forward, newly originated and guaranteed loans may incur significant losses when the loans are more seasoned and if the property market continues to decline. Finally, the testimonies also mention that GSE single-family guarantee programs accounted for $166 billion (73%) of the capital lost over that period. Accordingly, the pricing model developed can be used to more accurately price the government guarantee of default risk if good mark-to-market house price indices are provided. Because the expected cost of guaranteeing the default risk of a loan can be calculated as the difference in the option-adjusted spreads (OAS) between 100% recovery and a recovery at the current house price. In short, since the GSEs and the government are practically the "only game in town" in mortgage financing and underwriting, the current measures by the government may not be able to completely solve the problem of GSEs' continued losses. Instead, they may lead to possible indiscriminate rejection of profitable LMI mortgages, as demonstrated by the profitability of CAP loans. Therefore, advanced pricing model and risk management techniques should be used both in underwriting selection based on fair market risk-based pricing, and in portfolio risk management to price default guarantee cost and hedge interest rate risk.

The need for this new methodology is made clear by the Federal Housing Finance Agency (FHFA) 2009 annual report to Congress, which provides information about the poor performance of the pricing models and risk management techniques used by the GSEs. The Report of the Annual Examination of Fannie Mae[5] concludes that "model risk, the risk that model output does not match actual performance remains high". Figure 23 of the same report shows that Fannie Mae had additional $9.1 billion mark-to-market losses in 2008, compared with 2009, due to interest rate volatility. The report also mentions on page 24 that "derivative losses were $9.1 billion lower in 2009 at $6.4 billion as interest rates remained relatively stable in 2009." In particular, "A steep drop in interest rates during the second half of 2008 caused substantial mark-to-market derivative losses in the prior year." Moreover, Figure 31[6] in the report shows that Freddie Mac's derivative losses were $13.1 billion higher in 2008 than in 2009, because "in contrast to the substantial declines in interest rates during the latter half of 2008, rates remained relatively stable in 2009." Among the private banks using advanced pricing models and hedging techniques, such losses due to interest rate volatility can be mostly offset, as shown in the news about the hedging positions on mortgage-servicing rights (MSR) of Wells Fargo and JP Morgan (see later citation). In contrast, the GSEs' poor pricing models and lack of interest rate risk hedging techniques have resulted in continued losses, not only when default risk increases but also when interest rate risk spikes.

[5] See "FHFA 2009 Annual Report to Congress", Section "Report of the Annual Examination of Fannie Mae" Available in pp. 15-38 at http://www.fhfa.gov/webfiles/15784/FHFAReportToCongress52510.pdf

[6] See "Figure 23 Fannie Mae Mark-to-Market Value Gains (Losses)" on pp. 24 and "Figure 31 Freddie Mac Mark-to-Market Value Gains(Losses)" on pp. 48 in "FHFA 2009 Annual Report to Congress".

The FHFA 2009 report points out the following problems with Fannie Mae's risk management[7]. First, according to the report "prepayment models posed significant risk during the year because of an unusually wide primary secondary spread, house price volatility, the lack of credit availability, and uncertainty around the impact of MHA programs." In particular "Prepayment models have continued to predict faster than actual speeds across all major products during the year because the timing and magnitude of the effects of MHA programs are extremely difficult to predict". Second, Fannie Mae has difficulties in predicting interest rate. "Interest rate risk management remained a challenge in 2009 because of high volatility in rates and the mortgage basis, as well as continuing declines in home values." Moreover, "external conditions significantly impeded Fannie Mae's ability to accurately measure and manage interest rate risk exposures." Third, several risk models used by Fannie Mae are built on different assumptions and measures, and they "represent different views along risk dimensions and give conflicting signals".

The above problems pointed out in the FHFA report are solved in the following ways in this paper. First, the poor performance of Fannie Mae's model in predicting prepayment likely stems from the fact that prepayment and default risks are modeled separately. However, numerous scholars using competing risks models have found that the two risks are intertwined, in that an increase in one risk will reduce the other. The Federal Reserve and Treasury purchase of MBS in 2008, which lowered yields and kept rates down, is exactly a case in point of when prepayment can be used to mitigate default risk. Therefore, in this paper, prepayment and default risks are modeled jointly by means of the multinomial logit model (MNL). Second, the many variables and piece wise regression method used in the

[7] See pp. 32-pp.37 of "FHFA 2009 Annual Report to Congress". Section "Report of the Annual Examination of Fannie Mae" at http://www.fhfa.gov/webfiles/15784/FHFAReportToCongress52510.pdf

prepayment and default MNL regression enable the capture and translation of the slight monthly changes in prepayment and default risks that result from different deal structures, different interest rate environment, and different geographical environments. Third, based on the literature concerning non-arbitrage term structure models, interest rate scenarios are predicted by calibrating to daily term structure quotes, in order to back out the market implied interest rate scenarios. Finally, slight monthly changes in prepayment and default risks are consistently translated into prices using fixed-income pricing techniques.

One major reason for the GSEs' continued losses may be poor risk management pointed out in the FHFA report to Congress. However, blames have been readily placed on easy targets, such as low-to-moderate income (LMI) borrowers and Community Reinvestment Act (CRA). LMI borrowers have been blamed for initiating the current scale of mortgage delinquencies and foreclosures. They have also experienced decreased access to credit, due to the more conservative mortgage underwriting recently adopted by the GSEs and the shrinking private labeled market. In this context, Assistant Secretary for Financial Institutions, Michael S. Barr needed to refute claims that the GSEs collapsed because of the government's imposition of affordable housing goals[8] . Moreover, Treasury Secretary Timothy Geithner mentioned that credit is "still quite tight" for some borrowers while expressing "basic confidence" in the U.S. economy. The financial needs of LMI borrowers'

[8]. See Michael S. Barr speech in National Policy Conference 2010 held by the Mortgage Bankers Association in Washington, DC. See also the news"Treasury Refutes Anti-Reform Rhetoric. Outlines Housing Finance Proposals" from MND News Wire, at http://www.mortgagenewsdaily.com/04152010_financial_reform.asp

financial are being partly satisfied, because the Federal Reserve has kept buying MBS[9] , but such Federal Reserve purchases cannot continue indefinitely.

Furthermore, the CRA, which was intended to bring LMI borrowers into the mainstream banking system, is under attack. Laderman and Reid (2009) summarize the attacks that have occurred on the CRA and provide evidences in support of CRA. They use data from the Home Mortgage Disclosure Act (HMDA) and McDash's Lender Processing Services data sets, and they focus on loans originated in California between January 2004 and December 2006 for a total sample of 239,101 observations. They find that loans originated by lenders regulated under the CRA were generally significantly less likely to be in foreclosure than those originated by independent mortgage companies. They also find loans from CRA-regulated institutions certainly performed no worse than loans originated by independent mortgage companies.

Testimony by Michael Stegman before the House Financial Services Committee[10] also details reasons to support the CRA. He argues that it is in the national interest for low and moderate income populations to fully participate in the American economy. Congress and the Federal Housing Finance Agency have imposed a duty to serve the mortgage finance needs of underserved markets. This duty pertains in addition to the GSEs' affordable housing goal purchase requirements. According to the same testimony, a counter argument against CRA is that if the CRA forced covered institutions to offer financial services or credit products that are unprofitable over the long term, then no community reinvestment mandate

[9] According to Bloomberg news "Mortgage-Bond Yields Fall to Low on Fed's Treasury-Buying Plan". Aug 10 2010, at http://www.bloomberg.com/news/2010-08-10/mortgage-bond-yields-that-guide-home-loans-fall-to-lows-on-fed-debt-plans.html

[10] See testimony by Michael A. Stegman. "Remarks before the House Financial Services Committee: 'Proposals to Enhance the Community Reinvestment Act'". September 16, 2009. Available at http://www.house.gov/apps/list/hearing/financialsvcs_dem/stegman.pdf

should impair an institution's safety and soundness. Hence the GSEs may reduce the purchase and underwriting of CRA portfolios.

The fact is mortgage portfolios are one of the most difficult asset categories to manage. The difficulty is due to the competing prepayment and default risks that constantly vary according to different borrowers, different loan terms, different house price expectations, different interest rate expectations and different time periods. Moreover, the pricing of mortgages affects the performance and risks, while the changed performance and risks affect pricing in return. Therefore the solutions used by the private industry in predicting and mitigating losses, in an effect to offset volatility and price declines, are the more advanced and fully automated pricing system and hedging techniques used by Wall Street investment firms, such as Lehman Brothers. The fully automated pricing system facilitates ex-ante decision making in loan origination, in designing hedging strategies, as well as in making modification decisions when loans are in distress.

To calculate the precise price for each loan in any time period, a fully automated pricing framework should be used to analyze the continuous flow of market data including monthly loan-level historical records, daily term structure quotes implying interest rate expectations, and monthly or quarterly systematic macroeconomic and geographical factors. Such a pricing system can translate the ever changing information into prices to facilitate ex-ante decision making, both in normal situations and when loans are in distress. Thus portfolio hedging strategies and loss mitigation measures can be designed accordingly.

An advanced pricing and hedging techniques are essential for underwriters during the market downturns and times of high market volatility, as they are designed to offset volatility and price declines in mortgage portfolios. The hedging results on mortgage- servicing rights

(MSR) achieved by private banks in 2009[11] demonstrate the importance of advanced risk management and hedging techniques, given that private banks normally have higher cost of funds than the GSEs. According to Bloomberg news (2009), “Wells Fargo & Co. earned almost a third of its pretax quarterly profit by hedging mortgage- servicing rights, producing gains similar to those that have helped some of the biggest U.S. banks offset weaker consumer- lending businesses.” In particular, “Wells Fargo’s hedges outperformed write downs it took on the so-called MSRs by $1.5 billion and JPMorgan Chase & Co. came out ahead by $435 million. The two banks, as well as Bank of America Corp. and Citigroup Inc., wrote down MSRs by at least $5 billion in the third quarter as mortgage rates fell by about 0.26 percentage point."

This paper develops a loan-level RMBS pricing model that is industry comparable if not more advanced, and that can be used to better predict and price cost of guarantee of RMBS default risk, and to design hedging strategies for interest rate risk. The loan-level pricing model in this paper addresses most of the problems in the GSEs' current risk management models that were highlighted in the FHFA 2009 report to Congress. Moreover, this model improves upon the copula-based pricing models by automatically translating into RMBS prices the slight monthly changes that occur in individual borrowers' prepayment and default risks, which are due to borrower and loan characteristics, macroeconomic conditions, house price changes, and term structure movements.

Daily term structure quotes used are obtained from Bloomberg financial services. The yield curve data used are swap rates, and the volatility smile data used are at-the-money (ATM) swaption quotes in black volatility. The yield curve and volatility data are sampled

[11] See Bloomberg news "Wells Fargo, JP Morgan Benefit from Servicing Hedging" Oct 22. 2009. Available at http://noir.bloomberg.com/apps/news?pid=newsarchive&sid=azZrwv0uRzpo

for each day in which Fannie Mae purchased mortgage loans from Self Help prior to June 2007. Hence yield and volatility quotes on a total of 687 days are sampled from Bloomberg in total.

The monthly loan-level data come from the loans in the Community Advantage Program (CAP), which is a secondary market program initiated in 1998 by the Ford Foundation, Fannie Mae, and Self-Help. With a Ford Foundation grant of $50 million to Self-Help to fully underwrite each borrower's ability to repay, Self-Help purchases existing portfolios of CRA mortgages from participating lenders that otherwise could not be readily sold in the secondary market. Although the underwriting guidelines are non-traditional, the loans themselves are traditional, as they are prime-priced, 30-year, fixed-rate, lender-originated purchase-money mortgages that are fully underwritten for each borrower's ability to repay. To qualify for the CAP program, borrowers must meet at least of one of the follow requirements: (1)the borrower's income is no more than 80% of the area median income (AMI); (2) the borrower is a minority with an income less than 120% of AMI; (3)the borrower is purchasing a home in a high-minority (>30%) or low-income (<80% of AMI) census tract, in which case income may not exceed 115% of AMI. As of September 2006, Self-Help had purchased 42,694 loans totaling $3.79 billion. With an average loan amount of $88,773, participating lenders appear to be successfully serving the affordable market.

The default risk of CAP loans is completely guaranteed by the Ford Foundation grant, if default occurs within 12 months of loan origination (not 12 months from Self Help's loan purchase date). If a loan goes into serious delinquency or default, Self Help has lender's recourse to return the loan to the originator. After 12 months, any losses due to default are guaranteed by the Ford Foundation grant. The Ford grant allows the CAP loans to be offered at roughly 75 basis points (including all the benefits such as no mortgage insurance needed for loans with LTV above 80%) below the offering rate of a normal Fannie Mae loan with the same characteristics. Hence for any loan that goes into default after the RMBS purchase, investors receive a full refund of their capital and only face the risk of losing advance interest. Hence, from an investor's point of view, defaulted loans are no different than prepaid loans. Nevertheless, modeling the default and prepayment risks of these loans using a competing risks model has significant implications for pricing methodology and for applying the pricing model to other MBS deals that do not bear full default guarantee.

The loan-level pricing model developed in this paper is applied to the whole CAP portfolio of around 46,080 loans that was purchased prior to July 2008 with monthly records ever since origination. The option-adjusted spread (OAS) is calculated using rarely available loan-level Fannie Mae pricing data for the 7,168 loans without missing data by letting the model price equal the market price. Although the pricing model can easily allow a different recovery rate for each loan in the case of default, the full guarantee of default risk in the CAP deal structure makes default the same as prepayment. The unique CAP loan-level data set contain borrower income and race information that is not available in other public data sets, such as McDash and Loan Performance. Hence, regression is used to test whether traditional indicators, such as borrower income and race, reflect mortgage yields in CAP. For more information on demographic characteristics of CAP, see Riley, Ru and Quercia (2009). The default rate in CAP remains low in 2009 after the sample period used, as discussed by Riley and Quercia (2011) about CAP default rate in 2009 comparative to prime ARM and Prime fixed rate loans. In short, there is no significant surge of default risk in CAP after the sample period used in this paper.

The results applying the pricing model suggest that the CRA mortgages in the CAP program are mostly profit-making for the GSEs.  Hence CRA mortgages are not necessarily so unprofitable as to impair an institution's financial safety. This finding is consistent with the earlier findings of Michael Stegman concerning CRA loans that he mentioned in his testimony. Furthermore, the results suggest that conventional indicators of mortgage risks, such as borrower race, borrower income and OLTV, either are not important in determining mortgage option- adjusted spreads or have counter-intuitive signs for predicting mortgage yield. Hence the results demonstrate that mortgage risks and yields are much more complicated than are traditionally recognized. In particular, simple indicators such as race, income, and OLTV are not reliable predictors of yields and risks. Therefore, a blanket avoidance of LMI mortgages is not rational. Accordingly, tightening the GSEs underwriting guidelines using these simple indicators may not completely stem their losses, but may instead lead to the indiscriminate rejection of profitable LMI mortgages, as demonstrated by the CAP CRA loans.  In addition, the results indicate that loan age is a very important factor in determining both prepayment and default risks. Regressions of prepayment and default risks using all CAP loans show that (1) prepayment risk significantly increases until 2 years after origination; and (2) default risk has intermittently significant increases throughout the life of a loan. Therefore, without the adoption of advanced pricing model to accurately price and warn mortgages risks,  newly originated and guaranteed loans by the GSEs may incur significant losses as they become more seasoned and if housing market continues to decline. This concern remains despite the fact that currently “less than 1% of the losses have come from loans originated in 2009 and 2010”[12].

[12] See pp2. in testimony by Assistant Secretary for Financial Institutions Michael S. Barr, Written Testimony as Prepared for Delivery - 9/15/2010 .  At  http://www.ustreas.gov/press/releases/tg854.htm

In addition, the results suggest that state-specific characteristics dominate the legal environment in determining mortgage yields, because no consistent conclusions can be drawn about whether loans in states with stricter anti-predatory lending laws have higher or lower OAS. Finally, the results indicate that the information contained in private data sets, such as race and income and some neighborhood variables, is important for predicting prepayment and default risks. These effects cannot be explained away regardless of how many additional independent variables are added. Hence, the information collected in private data sets may help to improve pricing and mitigate portfolio losses. Such information may include that collected in longitudinal surveys concerning trigger events, mobility, and neighborhood quality. In short, the identification of responsible LMI borrowers and profitable loans requires that the loan-level model be run on each mortgage portfolio.

The pricing model and hedging techniques are especially important for underwriting LMI mortgages, because they are highly leveraged products that tend to amplify both losses and gains. The loan-level design of the pricing model developed in this paper can identify profit-making LMI mortgages by translating into prices the slight monthly changes of risks associated with each mortgage loan. Moreover, the pricing framework developed here provides a way to incorporate various scholars' research and to translate related results into prices. Such research may concern additional factors that affect prepayment and default risks, advances in term structure theories, and findings that may improve the design of hedge strategies. The loan-level pricing model can be used to estimate the cost of a government guarantee of RMBS default risk if good mark-to-market house price indices (HPI), such as the Case-Shiller Indices, are available. Specifically the expected cost for guaranteeing the default risk of a loan is the difference in the OAS between 100% recovery and a recovery at the current house price. Therefore the cost of a government guarantee should constantly vary with factors that affect mortgage risks and prices, including seasoning, term structure quotes, and house price indices. The accurate pricing of a government guaranteeing cost is important, because "losses in this segment of the Enterprises' (the GSEs' single family credit guarantee business) activities account for $166 Billion of the total $226 billion in losses since year-end 2007, representing 73% of the charges against capital over that period."[13] Therefore the loan-level pricing model developed in this paper, through advanced pricing model and hedging techniques, could help the federal government and the GSEs to better meet the financial needs of responsible LMI borrowers, while maintaining the sustainability and soundness of financial institutions.

### Section 1- b What does the policy literature say?

The arguments that have been advanced in the literature by pioneering policy scholars to promote affordable housing and mortgages are still valid in current economic context. Pioneering policy scholar Michael Stegman's earlier studies provide reasons why it is in the national interest for LMI population to fully participate in the American economy. , In his speech for the World Bank Conference in 2004[14], he pointed out that wealth disparities are greater than income disparities, and that homeownership can help to solve the problem, given that homeownership has historically been a wealth building avenue for low-to-moderate-income households. The paper by Stegman, Freeman and Paik (2007) further

[13] See pp. 3 in statement of Edward J. DeMarco, Acting Director Federal Housing Finance Agency - 9/15/2010. At http://www.fhfa.gov/webfiles/16726/DeMarcoTestimony15Sept2010final.pdf

[14] See "Evolution of Banking & Access to Financial Services in the U.S." by Michael Stegman, for World Bank Conference April 2004, available at http://www1.worldbank.org/finance/assets/images/Michael_Stegman__April_21_.pdf

examines the effect of homeownership on wealth building by exploring wealth differences across a sample of LMI homeowners and renters in the CAP survey panel. They find that homeownership not only affects the likelihood that CAP borrowers' hold assets and debt, but also affects their overall levels of wealth. Other things equal, owning a home increases one's adjusted net worth by almost $37,000. Furthermore, Michael Stegman's speech at 2004 Federal Reserve Bank of Cleveland summit[15] argues that affordable credits can help to prevent LMI borrowers from becoming chronically dependent upon high cost credits, such as payday lending. Making more Affordable credits available can have the effect of bringing millions of LMI Americans into the mainstream banking system and can expand economic literacy. Mortgages and other forms of credits to LMI borrowers have helped banks and other credit institutions (such as CDFIs) to generate tremendous amount of revenues such as fees, besides interest income. In addition, Michael Stegman's speech in the General Accounting Office Planning Conference 2001[16] argues that the nation as a whole benefits as well**:** since more than 90% of anticipated population growth over the next 50 years is expected to be among minority groups, and narrowing income gap between whites and minorities means a dramatic increase in minority spending power. It is increased consumption that has fueled the prior decades of economic expansion.

A series of research papers by Eric Belsky further details the role that homeownership plays in wealth creation and its impact on consumption. Belsky (2008) studies the importance of housing wealth to the balance sheets of more than two thirds of American households, the

[15] See "A Personal Perspective on the Recapitalization of Communities." by Michael Stegman, for 2004 Community Development Policy Summit Federal Reserve Bank of Cleveland. May 2004. Available at http://www.ccc.unc.edu/research.php

[16] See General Accounting Office planning conference: emerging issues in financial markets & community investment. Community assets panel presentation by Michael A. Stegman, February 2001. Available at http://www.ccc.unc.edu/research.php.

connection between housing wealth and consumption, and the substitution of mortgage debt for consumer debt. He also explains the possible impacts that house price declines can have on consumer spending.

Recent research papers by the UNC Center for Community Capital provide new evidences that homeownership continues to be a viable means of creating wealth in some well managed portfolio, the subprime crisis notwithstanding. For instance, Riley, Freeman and Quercia (2009) examine the wealth creation effects of house price appreciation for borrowers whose loans were purchased under CAP. The period of their analysis, which extends from loan origination to April 2009, spans the periods before and after the subprime crisis. Their results indicate that these low-income borrowers have experienced considerable home price appreciation since they purchased their homes, and that they have also accumulated and retained considerable equity, despite the subprime crisis.

Nevertheless, the benefits of the affordable mortgage advocated by policy scholars are only sustainable if the higher risks associated with affordable mortgages are properly priced. That is mispricing, and especially the underestimation of the higher risks of LMI borrowers can result in an excessive supply of credits at prices that are not sufficient to cover the risks associated with these highly leveraged assets. Moreover, it may lead to indiscriminate lending practices, which could impair the safety and soundness of financial institutions. Such concerns form the basis of the opposing arguments against GSEs' continuation in the LMI mortgage market. The challenges and difficulties associate with managing LMI assets were recognized fairly early by Michael Stegman. He pointed out that[17]:

[17] See "Creating Community Wealth." Net Inpact Conference November 3, 2001, available at www.ccc.unc.edu/documents/CC_commWealth_011103.ppt

"we have learned that whether for real estate, development venture capital or micro finance -- placing and recovering capital in low-income communities requires high level of skills, discipline, internal management systems, and development services to find the deals, put them together, and help them succeed".

Therefore, accurate pricing of the LMI mortgages using the model developed in the paper is especially important if the Federal government is to meet the financial needs of responsible LMI borrowers, while also maintaining the sustainability and soundness of GSEs.

The importance of the accurate pricing and risk management of LMI mortgages has been mentioned in recent speeches by Treasury officials in the context of the government policy concerning housing and finance. Treasury Secretary Timothy F. Geithner reiterated the importance of homeownership for LMI borrowers and discussed the challenges facing the secondary market and housing policy in his comments on GSE structure[18],

> "mortgage products should be standardized and support a liquid secondary market with a broad base of investors and 'accurate and transparent pricing'. Government housing policy should aim to promote widely available mortgage credit, financial stability and affordable housing options for lower-income households. "

In addition, Michael S. Barr further stipulated, "the system should distribute the credit and interest rate risk in an efficient and transparent manner that minimizes risk to the broader economic system and does not generate excess volatility or instability[19]".

Nevertheless, the old risk-based pricing models used by rating agencies have failed to measure, predict and correctly price mortgage risks. Ashcraft, Goldsmith-Pinkham and Vickery (2010) study the credit rating of subprime and Alt-A RMBS deals issued between 2001 and 2007. They find strong evidence that ratings become progressively less

---

[18] See Bloomberg news "Geithner Urges Ending Fannie, Freddie 'Ambiguity' (Update3)", available at http://www.bloomberg.com/apps/news?pid=20601087&sid=aOUI4zkc_97c&pos=5

[19] See news "Treasury Refutes Anti-Reform Rhetoric. Outlines Housing Finance Proposals", according to MND News Wire. Available at http://www.mortgagenewsdaily.com/04152010_financial_reform.asp

conservative around the RMBS market peak during 2005-07. They find the credit ratings perform especially poorly among high-risk mortgages (subprime and LMI mortgages), regardless of whether the performance is measured by mortgage default rates, losses or rating downgrades. They also point out that good credit rating should incorporate all relevant information about risk that is available in the information set of credit rating agencies at the time of rating. However their modeling results show that credit rating agencies failed to achieve this goal.

The failure of the rating agencies to predict and price subprime mortgage risks may be partly due to the time varying nature of mortgage prepayment and default risks and the complicated interactions of the two risks on mortgage returns. It is well known that subprime and CRA borrowers have slower prepayment risk but higher default risk. Slower prepayment risk increases the mortgage return in a falling-rate environment, but higher default risk decreases that return. Most importantly, mortgage prepayment and default risks vary constantly. Prepayment risk varies frequently with the interest rate movement expectations, as implied in current market term structure quotes. Similarly, default risk varies with the house price movement expectations and has been shown to be highly correlated in economic downturn and in distressed neighborhoods. Hence, prime loans may not provide as high a return as conventionally believed, due to their high prepayment risk. Moreover, if managed well, subprime or CRA loans may not provide as low a return as usually believed, due to their low prepayment risk. The constant variations of mortgage risks over time and across geographic areas require fully automated pricing model be used to capture monthly changes in risks and translate them into prices.

In short, it is time for a RMBS pricing model to be adopted that can price mortgage risks at the level of the individual borrower and also meet the following principles that have been agreed upon in the literature. First, the pricing model should incorporate all relevant information about risk at the time of purchase. Second, compared to older models, the new pricing model should achieve increasing degrees of granularity, and be able to identify and separate responsible LMI borrowers from irresponsible ones. Finally, the pricing model should provide a way to translate into prices the ever-changing monthly prepayment and default risks of each loan. The loan-level pricing model developed in this paper satisfies all these principles.

The rest of the paper is organized as follows. Chapter 2 reviews the literature on each component of this loan-level pricing model, including general fixed-income pricing frameworks, various term structure models, and modeling prepayment and default risks using competing risks models. Various factors discussed in the literature that affect mortgage prepayment and default risks are also summarized, including geographic factors, borrower characteristics and idiosyncratic loan features. Chapter 3 presents regression results for prepayment and default risks estimations using the whole CAP portfolio. The predictive power of the prepayment and default risks regression is also studied. Chapter 4 details the term structure model of choice, namely one-factor Hull-White model fitted to yield and volatility, and the cash flow discounting method of fixed income pricing. Chapter 5 studies Fannie Mae's pricing practices and the CAP deal structure. Specifically, the OAS results estimated using loan-level Fannie Mae pricing data for 7,168 loans (without missing data) that were purchased on 687 days are presented. Moreover, linear regression is used to study which factors affect the OAS. Chapter 6 summarizes the policy implications of the analysis and provides possible directions for future research.

# Chapter 2. Literature Review.

## Section2- a. Summary of the pricing framework

It is widely known that copula-based models has been used to price CDOs. Hull and White (2004) develop a factor copula model to price CDOs, and the Fitch Ratings report Hunt (2007) shows that copula-based models were used by some rating agencies to price RMBS before the subprime crisis. Moreover, the extensive criticisms of copula models that were voiced after the subprime crisis[20] simply reflect how popular copula models were for pricing CDOs before the crisis. However, the composition of the underlying assets in the CDOs was not very transparent until the after the start of the subprime crisis. According to recent Bloomberg news, Citigroup issued many mortgage-backed CDOs before the subprime crisis, which offered an implicit guarantee of default risk through the "Liquidity Puts" clause [21]. Recent news has also disclosed that some CDO deals issued by Goldman Sachs were also backed by residential MBS[22]. However, using the same pricing models to price underlying assets with very different risk characteristics, such as corporate loans or mortgages, just because they are all in CDO tranches, is a dangerous practice.

Various tranche-level RMBS pricing models were developed some time ago. In his literature review, Sundaresan (2000) mentions that in the MBS market, "complex models of term structure are integrated with fairly intricate models of prepayments to produce valuation

[20] See Sam Jones "The formula that felled Wall Street", Financial Times, April 24, 2009. Available at http://www.ft.com/cms/s/2/912d85e8-2d75-11de-9eba-00144feabdc0.html

[21] See Bloomberg news "Citigroup 'Liquidity Puts' Draws Scrutiny from Crisis Inquiry", available at http://www.bloomberg.com/apps/news?pid=20601087&sid=aZELabu4NReI&pos=1

[22] See Bloomberg news "Goldman Sachs Sued by SEC for Fraud Tied to CDOs" , available at http://www.bloomberg.com/apps/news?pid=20601087&sid=agT1H2ffyJCA

results and risk management inputs for MBS portfolios." Moreover "this is also an area where industry is arguably ahead of the academics in many issues." More recently, Brigo, Pallavicini and Torresetti (2010) summarize the popular CDO tranche-level pricing methods that were used to price corporate loans in the industry before the subprime crisis. According to these industry experts, the various CDO tranche-level models summarized are all based on copula models with small variations in how the default correlations among loans are treated within a tranche and between tranches. However, the copula-based CDO pricing models have been proven to be problematic by the subprime crisis as reported by the somewhat colorful expressions used by the news media to describe these models, such as "the formula that felled Wall Street." Although it is not clear whether the copula-based tranche pricing models have been used by organizations other than Fitch Ratings to price RMBS backed CDO loans, these models should not be used to price RMBS backed CDO loans in the future.

The copula-based tranche-level pricing models have the following problems in comparison with the loan-level pricing model developed in this paper. First, the interest rate is assumed to be deterministic (not stochastic) in CDO tranche-based pricing models. It means there is no model to predict interest rate; hence mortgage interest rate risk is not adequately priced. Second, the default risk and default correlations are mainly captured through a series of systemic state-level factors as in Hunt (2007), which do not take into account the borrower and loan-level factors that have been widely observed in the literature as affecting default risk, for instance, mark-to-market LTVs. Third, the copula-based tranche-level models cannot price the effect of default risk mitigated by prepayment. Such instances were commonly observed in the falling-interest-rate environment that followed the subprime crisis, due to government interventions, such as the Federal Reserve purchase of MBS in

2008 that lowered yields. Finally, mortgage default risk is well known to be highly correlated, that is the most common observations are either concentrated defaults during an economic downturn and distressed neighborhoods, or very few defaults in economic boom. Therefore, if a mortgage pricing model only models default risk like copula-based model, it may produce a wide range of unrealistic prices since tail events are highly likely.

In summary, the tranche-level RMBS pricing models conduct pricing at the aggregate tranche-level and are overly dependent on reduced-form models. These models isolate the performance of underlying assets from prices. As a result, small changes in default risk are not translated into prices, and thus default risk is covered until massive defaults are detected and several subprime mortgage originators went bankrupt, such as Countrywide. In contrast, the loan-level pricing model developed here solves this problem by incorporating all the information that is available at the time of purchase. This information includes the interest rate scenarios implied in current term structure market quotes, prepayment and default behaviors from historical loan-level records that vary due to borrower and loan characteristics, and local macroeconomic conditions and house price movements. Therefore this loan-level RMBS model should be applied not only to CMO, but also to CDO and ABS portfolios whenever the underlying assets are RMBS. To obtain individual tranche price, this loan-level model can be applied to calculate the aggregate price of a portfolio; then tranche prices can be calculated according to tranche waterfall arrangements and deal structures.

According to section 21.1.2. of the book by Brigo and Mercurio (2006), the price of any loan, either mortgage or corporate, is essentially the discounted expected cash flows until the loan terminates. Hence the fundamental pricing equation is:

$$Price_0 = E\left[exp\left(-\int_0^{maturity} r_t dt\right) * Payoff(maturity)\right].$$

In mortgage loan, the payoff depends on both prepayment and default risks. Because the termination risk, due to either prepayment or default, determines when the principal can be returned and how long the interest can be collected. If termination is due to default, then the recovery rate also determines how much principal and interest can be recovered. The interest rate is the most important driver of prepayment risk. The interest rate curve can also predict default risk, since the slope of the yield curve is one indicator of general economic conditions, as well as of the direction of interest rate changes in the future. Moreover, interest rate models generate discount factors that are used to discount predicted cash flows. Hence the interest rate is the first variable that should be modeled in pricing a loan. The interest rate model adopted here is under risk-neutral pricing framework. Hence the whole RMBS pricing accordingly assumes risk-neutrality. That is the physical measure and market prices of risk are not studied here.

The interest rate model is calibrated according to standard term structure theory to the currently observed market term structure level and volatility. This approach is adopted because the current market quotes imply the market expectations of term structure movements. The non-arbitrage feature of the term structure model adopted in this paper means the model is calibrated by taking the observed yield curve and volatility quotes as given. This approach is consistent with the practice of Dunsky and Ho (2007). Hence the interest rate movements do not allow possible arbitrage opportunities in holding a portfolio of bonds at any time. Following the practice of Dunsky and Ho (2007), the one-month rate, 2-year rate and 10-year rate are generated for each node point in the interest rate tree calibration. The 10-year rate is used in calculating the refinance spread variables. The 2-year and 10-year rates are used to generate yield curve spread variables.

Once the interest rate model is calibrated, the prepayment and default risks are calculated for each node point on the tree, which represents the predicted interest rate scenarios backed out from market quotes. The prepayment and default predictions depend not only on simulated interest rate scenarios, but also on many other factors such as borrower and loan characteristics. The regression using multinomial logit (MNL) model is intended to filter out factors that affect the prepayment and default risks, and thus to generate good predictions of termination risks. The predicted prepayment and default risks are calculated from regression predictions based on both historical data (i.e. such as borrower and loan characteristics) and simulated data (i.e. interest rate scenarios). Given the predicted prepayment and default risks, the cash flow can be calculated for each interest rate path based on industry-standard formulas for mortgage cash flow. Finally, the discounted cash flow based on all the interest rate paths is obtained as the model price.

To compare the model price with the market price in a meaningful way, a pricing model that is consistent with the Wall Street firms practice is adopted and emphasized. The ultimate goal of the pricing model presented here is the identification of responsible LMI borrowers whose loans are profit-making for the secondary market via a model that is consistent with Wall Street's prevailing pricing practices. Therefore, the Lehman Brothers option-adjusted spread pricing model for corporate bonds described in Pedersen (2006) is used as an essential reference in developing the loan-level RMBS pricing model. The loan-level pricing model and OAS definition used here are also consistent with the earlier industry reports by Hayre (1999) from Salomon Smith Barney, Beardsell and Liu (2005) from Citigroup, and Breeden (1997) from Smith Breeden with regard to the general pricing framework and RMBS interest rate risk features.

## Section2- b. Prepayment and default estimation by competing risks models.

The literature contains several modeling alternatives for estimating prepayment and default risks by competing risks models. These alternative include the Cox proportional hazard model (PHM), the competing risks proportional hazard model (CRPHM), and the multinomial logit model (MNL).

Clapp, Deng and An (2005) provide a detailed comparison of the econometric efficiency, likelihood functions, and technical details of the PHM, CRPHM, and MNL models. The PHM by Cox (1972) is a continuous-time duration model that allows only one termination event. However mortgage risks involve two events, prepayment and default, and these are competing risks in that a loan in default cannot be prepaid, and vice versa. The CRPHM used by Deng Quigley and Van Order (2000) is designed to allow multiple termination events and competing risk features. However, the multinomial logit model has comparable econometric efficiency to CRPHM according to Clapp, Deng and An (2005). Moreover, it can be readily estimated using most publicly available statistical software packages. For this reason, multinomial logit models are widely used. For instance, Dunsky and Ho (2007), Dunsky and Pennington-Cross (2004), and Pennington-Cross (2010) use the MNL to model competing risks features of mortgage loans.

The multinomial logit (MNL) model is adopted here in estimating the prepayment and default risks for a longitudinal mortgage loan data set. The MNL model is a discrete-time[23] duration model that allows multiple termination events with competing risks feature. The competing risks feature is incorporated in MNL because the total probability of multiple

[23] The difference between discrete-time and continuous-time models depends on whether the dependent variable is a continuous or discrete categorical variable.

events (prepayment, default, remaining current) must sum to one. Therefore, an increase in one risk must be offset by decreases elsewhere. Another feature of the MNL is the assumption of the independence of irrelevant alternatives (IIA). This assumption requires that given the event history of a loan from origination to termination, each monthly observation be treated as though it were independent from the prior observation. In other words, adding or removing one of the available choices should not change the ratios of probabilities for the remaining choices. Furthermore a borrower's prior choices at any point in time are independent of those at any other point in time. Specifically, the monthly conditional prepayment and monthly conditional default rate are defined, respectively, for the $i^{th}$ loan in the $t^{th}$ month. Moreover, the log-likelihood function is given by:

$$\ln L(\beta'_{prep}, \beta'_{def}) = \sum_{t} \sum_{i} \alpha \ln \Pr(Y_{it} = prep) + (1-\alpha) \ln \Pr(Y_{it} = def),$$

$$\Pr(Y_{it} = pre) = \frac{\exp(\beta'_{prep} Z_{prep})}{1 + \exp(\beta'_{prep} Z_{it}) + \exp(\beta'_{def} Z_{it})} \qquad (1)$$

$$\Pr(Y_{it} = def) = \frac{\exp(\beta'_{def} Z_{it})}{1 + \exp(\beta'_{prep} Z_{it}) + \exp(\beta'_{def} Z_{it})}$$

where $Y_{it}$ denotes the $i^{th}$ borrower's decision at time t, and $Z_{it}$ are the observed variables, and ($\beta'_{prep}$ , $\beta'_{def}$) are the vectors of estimation parameters that are presented in the estimation result section, and α is the indicator of whether the event is default or prepayment. The MNL is estimated using maximum-likelihood method by treating restructured discrete-time information for each loan as taken from identical and independent distributions. The log-likelihood function is estimated based on the loan-level longitudinal data set.

## Section2- c. Factors that affect mortgage prepayment and default risks.

Various research papers show that the prepayment and default risks of mortgages are not static but instead vary across different income groups (e.g. prime, subprime, and CRA), across various geographic areas, and across different time periods.

In the pioneering literature review, Quercia and Stegman (1992) summarize the factors affecting default risks from the lender's perspective, the borrower's perspective and the institutional perspective. The importance of many of the variables summarized in the literature review is still being confirmed by later scholars who are using updated data sets and new regression methods. However many factors affecting default risk listed in the paper are not available in any public loan-level dataset. For instance, trigger events, such as borrower employment status, family health problems, or unexpected debts can all trigger default. Moreover, divorce, changes in family size, the presence of school-age children, and environmental problems in the house or neighborhood may trigger borrowers to move and thus prepay a loan. This type of information is generally not available in public data sets but may be found in scattered survey data sets. The private information contained in scattered survey data sets can be used to better predict prepayment and default risks. Furthermore, although the various factors summarized by Quercia and Stegman(1992) have been shown by scholars to affect mortgage default risk, the results of such research cannot be translated into mortgage prices without using a formal loan-level pricing model. The model developed in this paper provides a framework that can solve the problem, and it allows the research of various scholars to be used for mortgage portfolio risk management.

More recent papers identify additional factors that affect default behavior or find different effects for the same factors. For instance, Ding, Quercia, Lei, Ratcliffe (2008) compare the default behavior of borrowers who received Self Help CAP loans to that of similar borrowers who received normal subprime mortgage loans during the subprime crisis. They find that the different loan characteristics imposed on borrowers of similar types appear to be the main driver of the different default behavior. In addition, Ding, Quercia and White (2009) find a lower default rate in neighborhoods in anti-predatory-lending laws states, in states requiring verification of borrowers' repayment ability, in states having broader coverage of subprime loans with high points and fees, and in states having more restrictive regulation on prepayment penalties. Cotterman (2001) studies the effects of neighborhood characteristics on mortgage default, and finds that lower Census-tract median income and higher Census-tract Black composition are associated with higher rates of default, whereas individual borrower race or income are unrelated to default. Dunsky and Pennington-Cross (2004) use multinomial logit model and find that delinquency and default are sensitive to current economic conditions and the state of housing markets. Moreover, credit scores and loan characteristics also play important roles. Danis and Pennington-Cross (2005) study the distressed-pay-in-full phenomenon in a falling rate environment; and find that during their sample period, delinquency predominately leads loan to termination through prepayment while negative equity leads to termination through default.

The loan-level pricing model developed in this paper can also be used to price adjustable-rate mortgages in the future; thus it is interesting to consider the literature findings concerning differences in default behavior between adjustable-rate mortgages and fixed-rate mortgages. Foote, Gerardi, Goette, Willen (2008) use a private data set for the New England area and find that, for ARM loans, most subprime borrowers who defaulted did so well in advance of their reset dates. Their results also show that defaults on subprime ARM loans are more sensitive to declining housing prices than are defaults on fixed-rate loans, and that many borrowers with good credit scores took out subprime loans as the housing boom continued. They find it hard to prove that these borrowers were inappropriately steered into the subprime market, since the loans these borrowers took out were too risky for prime treatment. Finally they also find that 70% of Massachusetts homes recently lost to foreclosure were originally purchased with prime mortgages, and that subprime refinancing has been common for owners with positive equity.

Overall, the literature on prepayment and default risks indicate that many idiosyncratic loan and borrower factors affect mortgages risks, and the impacts of these factors vary constantly depending on different portfolio deal structures, different time periods and different geographical areas. These findings clearly demonstrate the complicated and dynamic nature of mortgage risks. In particular, prepayment depends on a borrower's expectation of refinance opportunities, and default depends on a borrower's expectation of house price trends compared with his unpaid mortgage balance. Moreover, trigger events may force borrowers to prepay or default, even if their expectations are unchanged. In other words, using credit scores or ratings as the only gauge of mortgage risks provides unreliable inference, and prime borrowers may not be as low risk as they are traditionally thought to be. Thus, these literature findings provide additional support for the need for a pricing model on a loan-by-loan and month-by-month basis that can incorporate various scholar findings in the literature to be used in mortgage portfolio risk management.

### Section2- d. Term structure models.

Aside from default risk, interest rate risk is one of the key risks of mortgage for underwriters and investors. According to the FHFA's report (page 24) "derivative losses were $9.1 billion lower in 2009 at $6.4 billion as interest rate remained relatively stable in 2009".

In particular, "a steep drop in interest rate during the second half of 2008 caused substantial mark-to-market derivative losses in the prior year." Moreover, the same report shows that Freddie Mac's derivative losses were $13.1 billion higher in 2008 compared with 2009, because "in contrast to the substantial declines in interest rates during the latter half of 2008, rates remained relatively stable in 2009."

Interest rate risk results from the refinancing (prepayment) behavior of borrowers in response to interest rate volatility. In a falling rate environment, MBS investors collect decreasing interest income due to mortgage prepayment, while the cost of capital is normally fixed. Nevertheless, in a rising rate environment, MBS investors collect fixed interest income since borrowers do not prepay, but they probably face rising borrowing costs. In short, the underwriters or investors may suffer enormous losses as long as the interest rate is volatile. Hence, the term structure model used to predict interest rate scenarios is the key to predicting prepayment risk, because it allows the interest rate risk of mortgages to be priced and hedging strategies to be designed accordingly. Therefore, the interest rate is the first element that needs to be modeled in the pricing framework. The interest rate model not only generates the refinancing scenarios, which are among the key factors for predicting prepayment and default, but it also generates the monthly discount factor. Thus, small changes in the interest rate result in big changes in prices.

In his literature review, Sundaresan (2000) summarizes the theories and methodologies used in default-free term structure models. The major types of models include affine term structure (ATS) models and LIBOR market models. As explained by Sundaresan (2000), the LIBOR market models use discretely compounded forward rates as the numeraire, and this approach has led to theoretically consistent models for valuing caps, options on swaps, and so on. However LIBOR market models requires time-consuming Monte Carlo simulation techniques because of the non-Markovian property of the forward rate process, which may limit the feasibility when applying the loan-level model to mortgage portfolios that may easily have millions of observations in a single month. Furthermore, according to the Lehman Brothers report by Pedersen (2006), the impact of modeling a borrower's prepayment and default sensitivity to the interest rate outweighs the impact of improving the interest rate modeling. Hence, the focus in this paper is on affine term structure models.

According to Sundaresan (2000), in affine term structure models the equilibrium (or arbitrage-free) short rate is an affine function of some underlying state variables of the economy, where the state variables follow an affine diffusion process. The short rate is linearly related to the underlying state variables under both the risk-neutral measure and physical measure. These assumptions allow the derivation of closed-form solutions for a wide variety of fixed-income securities, which greatly simplifies the empirical implementations of ATS models. Egorov, Hong and Li (2006) provide an empirical analysis of the out-of-sample performance of ATS models versus random walk in forecasting the joint conditional probability density of bond yields.  Nevertheless, some scholars, such as Dai and Singleton (2000), argue that at least three factors are required to properly describe the dynamics of the interest rate curve. Egorov, Hong and Li (2006) argue that, first of all, the extensive search for more complicated models using the same data sets may suffer from a so-called "data snooping bias," as pointed out by Lo and MacKinlay (1989) and White (2000). While more complicated models fit a given dataset better than simpler models, they may over fit some idiosyncratic features of the data without capturing the true data-generating-process.

Second, an over-parameterized model contains a large number of estimated parameters and inevitably exhibits excessive sampling variation in parameter estimation. The excessive parameter estimation uncertainty may adversely affect the out-of-sample forecast performance. Third, a model that fits in-sample data well may not forecast the future well because of unforeseen structural changes of regime shifts in the data-generating process. A few studies that consider the out-of-sample performance of ATS models have shown that some of these models fail miserably in forecasting the conditional mean of future bond yields. For example, Duffee (2002) shows that the complete ATS model of Dai and Singleton (2000) have worse forecasts of the conditional mean of bond yields than a simple random walk model in which expected future yields are equal to current yields. Nevertheless, Egorov, Hong and Li (2006) suggest that ATS models may provide good forecasts for the higher order moments, or even for the whole conditional density of bond yields, although they have poor forecasts of the conditional mean dynamics.

Sundaresan (2000) mentions that one solution to the poor in-sample fit of one-factor ATS models is the growth of non-arbitrage pricing models, which can be calibrated to the market data using the shift extension technique. The shift extension technique involves including "time-varying" parameters in these models to allow fitting to the observed initial forward curve and volatility. Shifted Hull-White (HW++[24]), shifted Cox–Ingersoll–Ross model (CIR++), and shifted Black-Karasinski (BK++) models are all popular choices among practitioners. However, the stability of the parameters may be an issue.

Gupta and Subrahmanyam (2005) provide a very comprehensive examination of the pricing and hedging performance of non-arbitrage short rate models, Heath-Jarrow-Morton

[24] "++" means "shifted" in order to differentiate the HW model with shift extension technique from the original HW model.

(HJM) models and the LIBOR market Brace-Gatarek-Musiela (BGM) model. The one-factor models analyzed by Gupta and Subrahmanyam (2005) consist of two non-arbitrage short rate models, namely the Hull-White and the Black-Karasinski models, the HJM general class with five forward rate specifications, and the BGM LIBOR market model. For two-factor models, two alternative forward rate specifications are implemented within the HJM framework. In their paper, the pricing accuracy refers to the ability of a model to price options accurately, conditional on the term structure. It is useful in picking out deviations from arbitrage-free pricing. Hedging performance refers to the ability of the model to capture the underlying movements in the term structure in the future after being initially calibrated to fit current market observables. It is useful for studying whether the interest rate dynamics embedded in the model are similar to those driving the actual economic environment that the model is intended to represent. Gupta and Subrahmanyam (2005) show that one-factor lognormal model (for instance BK) outperforms other competing one-factor models in terms of out-of-sample pricing accuracy. In addition, the estimated parameters of this model are stable. The one-factor BGM model outperforms other models in pricing tests, while two-factor HJM models improve pricing accuracy only marginally. They conclude that, for the accurate pricing of caps and floors, it is more important for the model to fit the skew in the underlying interest rate distribution than to have two stochastic factors in the model. However, they find the hedging performance improves significantly with the introduction of a second stochastic factor in the term structure models, because two-factor models allow a better representation of the dynamic evolution of the yield curve, which is more important for hedging performance than for pricing. However, their results mostly refer to pricing and hedging interest rate caps and floors. For this reason, their results need to be applied with caution when it comes to the RMBS market, because RMBS have much more complicated dynamics, even only with respect to mortgage portfolio interest rate risk.

### Section2- e. Justifications for using the one-factor Hull-White model fitted to yield and volatility.

The term structure model used is the one-factor Hull-White model fitted to yield and volatility. The volatility curve fitted here is based on the market ATM swaption quotes in black volatility, which is not the instantaneous volatility that appears in the continuous-time stochastic differential equation (SDE) of the Hull-White model. Volatility fitting is important because of the growing importance in the literature and in industry practice for modeling the volatility smile. Models not fitted to volatility will be problematic for designing the Vega hedge and thus may not completely hedge the volatility risk.

According to Pedersen (2006) one-factor models fitted to yield and volatility are still widely used in the industry to price corporate loans. One-factor models are especially suitable for pricing mortgage loans, since most public mortgage data sets are huge and easily contain millions of loans for any given month. The tree structure in one-factor models allows fast calibration of the loan-level model, while time-consuming Monte Carlo simulation may limit a model's feasibility in practice. The Hull-White model is chosen in particular, since the refinance rate is simulated in the term structure model and is used as key factors in generating prepayment and default predictions. The refinance rate used is 10 year rate, and this choice is consistent with the practice of Beardsell and Liu (2005) from Citigroup then. Hence the availability of a close form solution is essential for pricing RMBS. Therefore, lognormal models are not suitable for RMBS pricing, and the Hull-While model is the most popular choice among the normal models in practice.

Further discussions of the potential pricing and hedging implications of short rate one-factor models require understanding of the basic features and parametric forms of the one-factor models. The continuous-time presentations of one-factor models under risk-neutral measure are summarized in Table 2. In Table 2, $r(t)$ is the instantaneous short rate at time t, $\theta(t)$ can be considered as time-varying means, $a$ is mean reversion parameter, $\sigma$ is the volatility parameter, and $W(t)$ is one dimensional Brownian motion.

[Insert Table 2. Summary of basic one-factor short rate models]

Although the availability of a close form solution makes the normal model the only feasible choice, the fact remains that compared with the popular log-normal model for interest rate, the normal model will produce better pricing results due to the higher probability assigned to lower rates. This is discussed by Brigo and Mercurio (2006) in their appendix concerning the hedging and pricing performance of various short-rate models. One reason why prepayment risk is so important in RMBS pricing is that the long-term rates have been quite low due to the cheap credits before subprime crisis and the interventions by the Federal Reserve after the subprime crisis. Hence the lognormal model, by giving higher probability to higher rates than does the normal model, will underestimate prepayment risk and thus may lead to persistent bias in pricing.

The comprehensive comparisons by Gupta and Subrahmanyam (2005) provide additional reasons for the popularity of the one-factor models fitted to both yield and volatility in industry practice. Their findings show that for the accurate pricing of caps and floors, it is more important for the model to fit the skew in the underlying interest rate distribution than to have two stochastic factors in the model. Hence one-factor models can be sufficient for pricing purposes. Furthermore, as cited in Gupta and Subrahmanyam (2005), Collin-Dufresne and Goldstein (2002) show the importance of fitting volatility in term structure modeling. Collin-Dufresne and Goldstein (2002) argue that there is a missing stochastic volatility factor that affects the prices of interest rate options, but does not affect the underlying LIBOR or swap rates. They propose models with explicit factors driving volatility, and suggest that cap prices may not be explained well by term structure models that only include yield curve factors.

In terms of hedging performance in real market practice, the discussion by Brigo and Mercurio (2006) in their appendix provides intuitive explanations of how model parametric specifications affect hedging, consistent with the findings of Gupta and Subrahmanyam (2005). According to Brigo and Mercurio (2006), the standard hedging is calculated by shifting the market observable of interest, recalibrating, and computing the difference in prices, divided by the shift amount for the sensitivity of price to the market observable. However if the influence of a local shift in a market observable is distributed globally on the parameters by the calibration, then hedging will be a problem when shifting single points, since the effect is probably lost or confused with other possible causes. For instance, Brigo and Mercurio (2006) mention that a short-rate model with only one time-dependent function, which is to be exactly calibrated to yield curve, has too few parameters to appreciate the influence of local changes in the input volatility structure. Shifting two rather different points may cause the same change in the parameters, due to the flattening of the information implied by the low number of parameters. The problem may be potentially alleviated by introducing additional time-dependent coefficients in the short-rate dynamics (used in fitting to both yield and volatility) or by adding a second stochastic factor (following Gupta and Subrahmanyam 2005). However, the ultimate solution may require LIBOR market models that allow much more sophisticated and flexible forms of forward rates. The intuitive explanations by Brigo and Mercurio (2006) are consistent with the findings of Gupta and Subrahmanyam (2005) that hedging performance improves significantly with the introduction of a second stochastic factor in term structure models, since two-factor models allow a better representation of the dynamic evolution of the yield curve. The intuitive explanations also support the importance of volatility fitting in term structure calibration, when volatility smile modeling becomes more and more emphasized in the literature and in practice. Without a time-dependent parameter just to appreciate the influence of local changes in input volatility structure (the case in two-factor models), the sensitivity with respect to the volatility shift may be miscalculated in designing Vega-hedging, since the effect is probably lost or confused with other possible causes. In other words, if Delta and Gamma hedges are used to completely hedge the yield curve risk, two-factor models will outperform the one-factor shifted HW model used here. However, if Delta and Vega hedges are used to hedge the yield curve risk and volatility risk, then the one-factor model fitted to yield and volatility will outperform two-factor models fitted to yield curve.

Finally, according to Pedersen (2006) the impact of modeling a borrower's prepayment and default sensitivity to the interest rate outweighs the impact of improving the interest rate modeling. Compared with corporate loans or interest rate products, the complication in RMBS pricing is that a different borrower’s prepayment and default sensitivity to the interest rate may vary constantly, due to the information asymmetry that the borrower has, different deal structures and macro-economic and neighborhood conditions. For instance, in the CAP program, mortgage insurance is not needed for loans with an LTV exceeding 80%, and there is no prepayment penalty. Hence, a borrower may not prepay even if he can get slightly better rates by refinancing. Furthermore, it is quite common for a borrower to have a second lien mortgage in addition to the first lien; thus, the actual combined LTV of a borrower may be greater than the observable LTV, which limits the borrower's ability to refinance. Finally, as Foote, Gerardi, Goette, Willen (2008) point out, ARM borrowers with little equity may default when they expect that the rate will be much higher on the next reset date. Hence, modeling the prepayment and default sensitivity of each borrower on each simulated refinance rate path will outweigh the efficiency gain from allowing a non-perfect correlation of interest rates in multifactor short rate models.

# Chapter 3. The Data Set and Prepayment & Default Regressions

## Section 3- a. The loan-level mortgage data and term structure data.

As explained before, the data used come from the Community Advantage Program (CAP), which is a secondary market program initiated in 1998 by the Ford Foundation, Fannie Mae, and Self-Help, a leading Community Development Financial Institution. With a Ford Foundation $50 million grant to underwrite a significant portion of the credit risk, Self-Help purchases existing portfolios of CRA mortgages from participating lenders that otherwise could not be readily sold in the secondary market. These loans feature flexible underwriting and typically include low or no down-payment, higher debt-to-income ratios, approval of borrowers with varied credit records or no established credit, or waiver of the usual requirement that a borrower have at least two months of loan payment available as a cash reserve at the time of closing. As of September 2006, Self-Help had purchased 42,694 loans totaling $3.79 billion. With an average loan of $88,773, participating lenders appear to be successfully serving the affordable market. Ninety-one percent of borrowers earned 80% of AMI or less; 45% are minority; 71% of the loans had an original loan-to-value ratio above 95%, and more than 41% of the borrowers had FICO scores below 660 at the time of origination.

To avoid an arbitrary deletion of loans that could produce bias, all the loans in the CAP program as of the 2nd quarter 2008 are used, along with all of their available monthly records. The total number of loans ever presented in the CAP dataset by June 2008 was 46,080. The earliest monthly record is for November 1983, and the latest monthly record is for June 2008. There are a total of 1,781,650 monthly observations, and a total of 1,483,289 observations are used in the prepayment and default modeling due to missing information.

Term structure data are obtained from Bloomberg financial services, and they are used to calculate the OAS by equating the model price with the market price. Fannie Mae had purchased a total of 8,308 loans by May 2007. However, due to missing data mostly in neighborhood variables, the prices of only 7,168 loans are studied. The 7,168 loans were purchased during the course of 687 days, so the yield and volatility quotes are sampled from each of the 687 days from Bloomberg. The yield curve used is the swap rate, and the volatility smile used are the ATM swaption quote in black volatility, and both choices are consistent with the practice by Dunsky and Ho (2007).

### Section 3- b. Regression specifications and interpretation of results.

The MNL regressions include the following factors that affect prepayment and default risks: seasoning, seasonality, origination cohort, borrower loan and neighborhood characteristics, yield curve slope, refinance ratio and burnout factors. Since the goal of the regression is not to test causality, but to exhaust all the information available so as to have a good prediction of the prepayment and default rates, the explanatory variable selection and formats are a bit different from those used in the traditional regression. The default model does not fit well, since there are too few default observations in the CAP data. As shown in Table 3 there are 1.38% of prepayment observations but only 0.27% of default observations.

[Insert Table 3. Termination events by transaction year]

The modeling results are presented in Table 5. The pseudo R-squares in all the models are relatively low because default and prepayment observations only account for 1.65% of all the observations in the CAP portfolio. Model 3 is the final model includes all the explanatory variables, and Model 1 (referred to as partial model) is similar to Model 3 except that it excludes the additional borrower, loan, and neighborhood characteristics. Hence the differences between these models speak to the effects of these characteristics in risk modeling. Model 2 include all the factors that are included in Model 3, except that the refinance and burnout spreads are entered in simple form rather than in linear spline transformation; hence the differences demonstrate the effect of the linear spline transformation in modeling. The meaning of the linear spline transformation will be explained below. Judging from the consistent significance and signs of the parameter estimates that are common to models, the models are more or less stable. The full model's predictive power is discussed later on. In the following sections, the parameter estimation results for the full model are explained in detail.

[Insert Table 5. MNL regression results]

***Seasoning***

The seasoning effect is captured by describing the prepayment and default rates as a function of the age (in number of months) of the loan. The seasoning variables are age1 through age 12, and they are spline variables, that is a piecewise linear function. Transforming these continuous age variables into spline knots allows a better fit to the categorical dependent variable by allowing a different slope within each piece. A linear function is a function composed of linear segments, i.e. straight lines. One linear segment represents the function for values of x below $x_{0.}$ Another linear segment handles values between $x_0$ and $x_1$, and so on. The linear segments are arranged so that they join at $x_0$, $x_1$,....., which are called the knots. The piece-wise linear function technique is used to improve the model fitting. The coefficients of the spline knots can be interpreted as:

$$\frac{dy}{dage} = \begin{cases} a1 \; if \;\; age < 3 \\ a2 \; if \;\; 3 < age < 6 \\ a3 \; if \;\; 6 < age < 9 \\ a4 \; if \;\; 9 < age < 12 \\ \ldots\ldots\ldots\ldots \ldots\ldots \end{cases}$$

The majority of the age spline parameters are significant in both the prepayment and default modeling. In the prepayment modeling, it is shown that the CAP loan prepayment risk significantly increases until 2 years after origination, and it becomes insignificant between the 2nd and 3rd year. When a loan is seasoned for more than 5 years, it has a significantly lower prepayment risk, which may be because the borrowers holding these loans stayed in CAP long enough that they will not easily refinance. In the default modeling, it is shown that default risk exhibits intermittently but significant increase throughout a loan's life. In short, the results show that loan age is very important in determining both prepayment and default risks. As a loan seasons, prepayment risk increases until some maximum and then decreases or stays constant; however default risk may increase continuously throughout a loan's life. The importance of loan age in determining mortgage risks and thus returns is confirmed by the OAS regressions presented later.

***Seasonality***

Seasonality effect is represented by the transaction month dummy variables, which capture seasonal effects for instance prepayment due to moving. The baseline omitted category is January. The seasonality parameters are all statistically significant in the prepayment modeling. In particular, relative to January, the prepayment rate starts to pick up in February, reaches its peak in July and August, and slides back to its February level in December. In the default model, relative to January, loans are significantly less likely to default throughout the year except in October and December.

***Origination Cohort***

The origination year indicator variables are intended to capture the effects associated with origination year that are not considered in the model, because they are unavailable or unobservable. These effects may result from macroeconomic variables, such as consumer incomes and the unemployment rate, that are unavailable in the dataset. Furthermore, they may result from local economic conditions driving housing prices, as these are rarely directly observed and are time varying. In addition, omitted structural changes in the primary mortgage are also captured by the origination year indicators, which are known to impact prepayment and default behavior. The origination year parameters are mostly statistically significant, the omitted baseline category is loans originated in 2006. Unfortunately, the dummy variables for 2007 and 2008 are automatically dropped due to multicollinearity when additional borrower loan and neighborhood variables are added. In the prepayment model, compared to the 2006 cohort, earlier origination cohorts all prepay significantly faster. In particular, the prepayment rate is highest in the 1995 cohort, mostly because the older origination involves more seasoned loan, hence faster prepayment. In the default case, compared to the 2006 cohort, earlier origination cohorts have significantly lower default risk. The low default rate, which was observed even during the subprime crisis, is an important feature of CAP, as it can be seen in the CDR prediction presented later.

***Refinance Burnout Factors***

The refinance burnout factors are intended to capture a borrower's sensitivity to the refinance spread (i.e. the difference between the current market rate and the market rate at origination) in prepayment and default decisions. When the current market rate is significantly lower than the origination rate, the gains from refinancing at the current market rate will cause a borrower to prepay a loan. The burnout factor, defined as cumulatively mixed refinance opportunities, can help to extract information from a borrower's previous behavior. The burnout factor is designed to back out a borrower's missed refinancing opportunities from the borrower's historical behavior. The burnout variables will help to compensate for the efficiency loss when important information like race and income are missing in most publicly available datasets, since race and income variables are significant in all the MNL regressions. Moreover, according to Sundaresan (2000), MBS prepayment risk is path dependent. For instance, a borrower who persistently missed refinance opportunities in the past is less likely to refinance than the baseline group, given another refinance opportunity. The burn out factors can help to capture the path dependent nature of MBS prepayment risk.

One difference between CAP loans and standard subprime loans is the lack of prepayment penalties, which are present in most subprime loans. The favorable terms of CAP loans to low-to-moderate income borrowers make prepayment penalties unnecessary, which is confirmed by the significantly lower prepayment rate of CAP loans compared with standard subprime loans. Nevertheless, in the prepayment penalty case, the regression can be easily modified by interacting the refinance spread with a dummy variable indicating whether the time period is within the penalty period, following Beardsell and Liu (2005).

The refinance spread and burnout are defined as follows. The refinance and burnout factors are created by simply turning refinance spread and burnout into spline knots.

$$refispd_{i,t} = \left(\frac{PMMS_{i,t=0}}{PMMS_{i,t}}\right) \quad \mathbf{(2)}$$

$$burnout_{i,t} = \sum_{t=0}^{T} Max\left(\frac{PMMS_{i,t=0}}{PMMS_{i,t}} - 1.20, 0\right). \quad \mathbf{(3)}$$

The refinance spread definition is straight-forward. The refinance spread is the ratio of the market rate for the ith loan at origination to the current period market rate. The market rate is the commonly used Freddie Mac Primary Mortgage Market Survey (PMMS) rate. The refinance spread is intended to capture the opportunity cost of refinancing at the current market rate relative to paying the old fixed mortgage note rate on the existing loan. Using the difference between the origination market rate and the current market rate allows isolation of a borrower's response to changes in the market rate, without mixing with the borrower's credit risk and loan features that are correlated with the mortgage note rate.

The burnout factor is designed to capture the missed refinance opportunities, and it is measured by the sum of the significant refinance spread accumulated over the age of the mortgage. The 1.20 threshold is related to the refinance transaction costs, which means that when the refinance spread exceeds 1.20 a significant refinance opportunity occurs. The 1.20 threshold is chosen based on statistical concerns after trial and error. Since the transaction costs of refinancing vary depending on the loan and borrower characteristics, it is hard to come up with a meaningful threshold from reality. The threshold is chosen based on the criterion that the residual of the refinance spread net of 1.20 should follow roughly a normal distribution in the histogram. The 1.20 threshold is a higher than the 1.10 threshold used by Dunsky and Ho (2007) for the LP data set, because LMI borrowers value the luck of getting into the targeted CAP and do not refinance like typical subprime borrowers. Both the refi-spread and the burnout are transformed into linear splines as described in Table 4. The knot points are consistent with those used by Dunsky and Ho (2007).

[Insert Table 4. Refinance and burnout spline knots]

In the regressions analysis presented in Table 5, both the simple refinance spread and the burnout (as in Model 2) and linear spline transformation of the refinance spread and the burnout (as in Model 3) are tested. Both the simple refinance spread and the burnout are significant and have the expected signs, indicating that a higher market rate at origination relative to the current market rate corresponds to faster prepayment, and that a greater cumulative number of missed refinance opportunities corresponds to slower prepayment. In Model 3, when the linear spline transformation of the refinance spread and the burnout are used, the refinance spline knots are still consistent, in that most of them are significant and have positive signs; however, the burnout spline knots look different. When the burnout spread is less than 0.2, it seems counter-intuitive that a higher burnout spread is associated with faster prepayment. However, when burnout spread is above 0.2, the result is again intuitive, in that a higher burnout spread slows prepayment. This effect is marginally significant at the 5% level when the burnout spread is between 0.2 and 0.7. The reason is probably because in the longitudinal regression, most observations have a burnout spread between 0 and 0.2. In other words, as a loan's monthly records grow, the loan will have an increasing burnout spread and prepayment probability. Hence, when the burnout spread is very small (say below 0.2), the regressions just capture the correlation of the burnout spread and prepayment risk. Only when the burnout spread is big enough (borrowers have a significant habit of missing refinance opportunities), does this wood-headed behavior begin to decrease the prepayment rate. Nevertheless, the most important purpose of the regression is to generate a model that fits well, so that the predicted cash flow calculations based on predicted risks calculated later will be more accurate. Model 3 is thus chosen for use in cash flow predictions because of its higher Pseudo R-squared and log-likelihood ratio.

Default behavior is not likely to be affected by the refinance spread and burnout factors. The negative sign and significance of the spline knot when burnout is less than 0.2 are probably again due to the correlation between the growing burnout spread in most observations and decreased default probability. Moreover, they result from both the exceptionally small number of default observation in CAP (0.27%) and the increased prepayment risk, which decreases default risk in the MNL setting.

***FICO Score Effect***

The impact of credit score on prepayment and default is self-evident. Borrowers with low credit scores are more likely to be constrained in their ability to refinance (and thus prepay), and credit score is designed to be an index for a borrower's default risk. Based on existing literature, we expect a positive correlation between credit score and the probability of prepayment and an inverse correlation with probability of default. The credit scores used are updated FICO scores, including both the FICO scores at origination and updated FICO scores that were recorded in January 2005, January 2006, May 2007 and January 2008, the only updated scores available at the time of analysis. Interpolation is used when credit score is missing to reduce the number of missing observations. The potential bias caused by the interpolation is minimized after transforming the continuous credit score variable into spline knots. The credit score spline knots are chosen at 580, 620, 660, and 720, because these categories are widely used in the mortgage industry. A borrower with a credit score above 660 typically would qualify for a prime, conventional loan.

The credit score spline knots are most statistically significant at the 10% level in the prepayment models except for the knot between 6.2 and 6.6.  Moreover, it is confirmed that a higher credit score results in faster prepayment and a lower likelihood of default. However it is interesting that when credit score is above 720, borrowers tend to prepay more slowly, which goes against conventional wisdom. This counter-intuitive result may indicate the benefits of the CAP program in improving borrower's credit score. In particular, the group of borrowers who stay in CAP and do not refinance may see improvements in their credit scores as a result of their good payment records. Hence it is observed that when credit score is high prepayment is slow. In the default model, the results are consistent with the intuition that in the first two significant knots, a higher credit score corresponds to lower default risk.

***Unpaid Balance (UPB) Effect***

Homeowners with a larger unpaid balance are more likely to refinance (prepay) and default. Because given a positive option value (default and prepayment options), a greater UPB provides a larger dollar incentive to exercise these options than a smaller one. Furthermore, the fixed costs of refinancing disproportionately reduce the option value for refinancing smaller loans. The continuous unpaid balance variable is transformed into spline knots, and the knots are chosen at 50k, 75k, 100k, 150k. In the prepayment model, the UPB spline knots are mostly statistically significant, except for the knot below 50k, and their signs are consistent with the expectation that a larger UPB should be associated with faster prepayment. In the default model, the knot below 50K and the knot between 75k and 100k are significant at the 5% level. Their signs are consistent with the intuition that a larger UPB corresponds to higher default risk, because of the higher benefits that accrue if the default option is exercised.

### *Mark-to-market Loan-to-Value Ratio Effect*

The MTMLTV ratio, as a measure of the borrower's equity in the property, is constructed as the unpaid balance divided by the current house value[25]. Borrowers with high MTMLTV ratio are expected to be more likely to default but to be constrained in moving and refinancing. In fact, if the MTMLTV ratio exceeds 80%, a higher note rate or mortgage insurance premium will reduce the benefits of prepayment. This effect is expected to be particularly apparent in the CAP portfolio, since most borrowers received loans with LTV ratios above 80% at origination without mortgage insurance but may incur mortgage insurance costs if they refinance. Furthermore, since tapping home equity is a refinancing benefit not captured in the option value, loans with more built-up equity could also see more cash-out refinance activity. In the default case, a higher MTMLTV ratio means less home equity; hence borrowers have less to lose once they default. Therefore, the MTMLTV is expected to be negatively correlated with prepayment risk and positively correlated with default risk. The MTMLTV data used include original LTV and the mark-to-market value obtained from Fannie Mae every quarter from the beginning of 2003 to the second quarter of 2008. The continuous MTMLTV variable is used in modeling, and the significance and sign of MTMLTV in both models confirm the expectations.

### *Yield Curve Slope Effect*

The yield curve slope variable is defined as the 10-year Treasury bill (TB) rate (in percentage term) net of the 2-year TB rate. The yield curve slope is expected to be positively correlated with prepayment risk, in that a steeper yield curve will result in faster prepayment

---

[25] Current market-value estimate are from Fannie Mae's automated valuation model (AVM). Fannie Mae's AVM model consists of three individual models that independently estimate property values based on repeat sales data, property characteristics, and tax assessments, respectively. Fannie Mae then uses a value reconciliation model to compute a best value estimate in the case of multiple model predictions where valuations vary.

of fixed-rate 30-year loans. The intuition is that the relative cost of long-term financing to short-term financing will make borrowers favor short-term financing and thus prepay from 30-year fixed-rate loans. The parameter estimates in the prepayment model confirm the expectation. However, it is interesting that the yield curve slope is also significantly positively correlated with the default rate. This result probably obtains because the two periods during which the default rate spikes, as shown in Figure 3&4, namely the one from Sept 2001 to early 2001 and the other from early 2007 to mid-2008, are both associated with positive yield curve slopes.

***Borrower, Neighborhood, and Loan Characteristics***

One advantage of the CAP dataset is it contains a lot of information that is not available in most publicly available data set, such as Loan Performance (LP) and McDash. In particular the CAP data include borrowers-race, income, and neighborhood information. Hence, the CAP data set can be used to demonstrate how important these factors are for risk modeling and how much predictive power is lost when the information is missing. A comparison between Model 1 (without borrower and neighborhood variables) and Model 2 and Model 3 (with the complete set of variables) illustrates the problem. As shown in Table 5, including the borrower, neighborhood, and loan characteristics greatly improves the goodness-of-fit of the model: the pseudo R2 has greatly improved. Furthermore, borrower race is important for determining both prepayment and default risks, and sex is an important factor in prepayment. It is expected that minority borrowers (African American or Hispanic) will have slower prepayment risk, since they may have fewer opportunities to refinance. However, it goes against conventional perceptions that minority borrowers in CAP are also associated with lower default risk. It may be that the minority borrowers in the sample greatly value the opportunity to be qualified for the targeted CAP and hence do not easily default or refinance. A borrower's debt-to-income ratio (i.e. the back-end ratio) turns out not to be important for either default or prepayment. Interestingly, a borrower's relative status in the census area, measured as the borrower's annual income as percentage of AMI, is an important factor in determining default risk. Moreover, the sign of this effect is consistent with the expectation that a borrower with a higher the income as percentage of AMI will be less likely to default.

Loan characteristics are shown to be important in predicting prepayment and default risks. The credit spread of a loan, specifically the difference between the mortgage note rate and the market PMMS rate at origination, contains information that the mortgage originator knows about borrowers; therefore it should help to predict prepayment and default risks. The results confirm the expectation that a higher mortgage note rate relative to market PMMS rate increases both default and prepayment risks. This result is intuitive because higher credit spread is associated with higher perceived default risk; at the same time, higher credit spread provides greater incentive to refinance into a lower rate loan. The importance of the loan credit spread to both default and prepayment risks also explains the success of the CAP program. The guarantee by the Ford foundation allows Self Help to offer a lower rate to CAP borrowers, which substantially lower prepayment and default risks of CAP loans compared with conventional subprime loans even during the subprime crisis. Hence, lender pricing practices have an important impact on borrower prepayment and default behavior. The effects of the two variables measuring past loan performance are straight-forward: a history of delinquency (30 days) or serious delinquency (60-90 days) greatly limits a borrower's refinance opportunities, and the case of distressed paid-in-full loans, which is observed for other subprime loans, does not seem relevant for CAP loans. Interestingly, a 30 day delinquency does not reveal anything about the default risk of a loan; while the significance and positive sign of serious delinquency in the default model just reflects the correlation between default and serious delinquency.

Neighborhood characteristics greatly affect the prepayment risk as well. Being located in a low-to-moderate income census tract (defined as tract median income less than 80% of the AMI), a minority census tract (defined as a tract with non-Hispanic White less than 50% of the population), or an underserved census tract[26] greatly reduces the prepayment risk, probably because the borrower has fewer opportunities to refinance. However, contrary to conventional wisdom, being in a less favorable neighborhood does not seem to be associated with higher default risk in CAP.

Geography also matters, since loans made in NC and OH and OK have significantly slower prepayment rates, while loans made in CA and FL do not prepay that differently from loans in other states, controlling for all the other factors. With respect to default risk, loans made in NC have significantly lower default risk, while loans made in CA and FL have significantly higher default risk. Overall, loans made in NC have great performance, since both lower prepayment and lower default risks increase a loan's return. The good performance of NC loans is further confirmed and explained in the option-adjusted spread regressions presented later;.

***Default Estimation***

As mentioned earlier according to Table 3, by June 2008, default observations only accounted for 0.27% of the whole sample, while prepayment accounted for 1.38%. Hence,

[26] The "underserved" variable is provided to Self-Help by Fannie Mae, and the definition of underserved follows Fannie Mae's standard definition.

the default modeling does not fit well. Furthermore, defaults are likely to be highly correlated in the sense that either very few defaults are observed in the economic boom or concentrated defaults are observed in the recession. For instance, in two periods defaults are observed to spike, one from September 2001 till early 2001, and the other from early 2007 to mid-2008. Hence, the macroeconomic environment may play an important role in addition to individual loan and borrower characteristics. In other words, default is more likely to be triggered by system-wide risk rather than by individual borrower risk. Hence, the complete pricing of default risk may require the use of counter-party risk pricing framework which assumes that default is not triggered by basic market observables but has an exogenous part that is independent of all the default-free market information. Because it is assumed in counter-party risk pricing that monitoring the default-free market (interest rates, historical loan records, and borrower characteristics) does not give complete information about the default process, and there is no economic rationale behind default concentration. The solution is to back out the market-implied default probability of the default risk guarantor from forward-looking market CDS quotes, and to use it to conduct pricing under a defaultable term structure that is based on the framework developed as by Duffie and Singleton (1999), and more recently by Pan and Singleton (2008).

However, in this paper, the pricing is conducted under the default-free term structure, and the counter-party risk pricing is not studied. Furthermore, the default risk estimation is important from the methodological point of view and when the model is applied to other data sets. However, it has a small impact on the OAS calculation for CAP loans, because the default risk is fully guaranteed by the Ford Foundation grant in the CAP deal structure, and the few defaulted loans represent only 0.27% of observations. Therefore, whether default is modeled under a default-free term structure or a defaultable term structure has a small impact on the OAS of CAP loans. In this paper, the impact of macroeconomic shocks on default risk is captured by a series of macroeconomic variables. Unfortunately, the neighborhood variables seem not to be very important in predicting default in CAP. Nevertheless, even with these concerns, the models used still somewhat track the default pattern, as shown in Figure 3&4. The significance of the individual loan-level and borrower and neighborhood characteristics in the default model indicates that a simple reduced-form model like copula is not sufficient to determine the termination of a mortgage, because such a model assumes that the event of the termination of a mortgage is isolated from other individual specific effects.

### Section 3- c. In-sample prediction results.

Hereinafter, the in-sample predictive power of the models for both default and prepayment risks is analyzed. Figure 1&2 show one-quarter-ahead and one-month-ahead single monthly mortality (SMM) predictions based on scheduled and actual balances. The SMM is defined as[27]:

$$SMM = \frac{Scheduled\ balance - Actual\ Balance}{Scheduled\ Balance}. \quad \mathbf{(4)}$$

The scheduled balance is the expected balance given the amortization schedule, last month's balance, and no prepayment or default. Actual balance is the remaining balance after the scheduled balance is adjusted for prepayments.

[27] See pp.199, The Handbook of Mortgage Backed Securities.

In Figure 1&2, the actual SMM is calculated using the actual balance from the dataset. The variable "qrschSMM" is the one-quarter-ahead prediction based on the scheduled balance, meaning the last quarter’s scheduled balance is combined with the predicted prepayment probability over the quarter to calculate the monthly SMM. The "qractSMM" is created using last quarter’s actual balance combined with the predicted prepayment probability over the quarter to calculate the monthly SMM. The "monschSMM" is created using the last month’s scheduled balance combined with the predicted prepayment probability over the month to calculate the monthly SMM. The "monactSMM" is created using the last month’s scheduled balance combined with the predicted prepayment probability over the month to calculate the monthly SMM. The results show that prediction utilizing information about the last period's (month’s or quarter’s) actual balance can capture all spikes in the prepayment rate, while prediction using the last period’s scheduled balance does not fit the data perfectly but still captures the trend in prepayment quite well. The finding is important because information about the last period's actual balance is often unavailable at the time the pricing are conducted. Therefore prediction based on scheduled balance may be more feasible.

In the default risk prediction presented in Figure 3&4, the variable "qrCDRpred" is the one-quarter-ahead prediction of the constant default rate (CDR) based on the number of loans existing at the end of last quarter. The variable "mCDRpred" is the one-month-ahead prediction based on the number of loans existing at the end of last month. In the CDR prediction, the last-period information about the number of loans existing is always utilized, and the prediction results capture the actual CDR trend pretty well. The definition of CDR is as follows:

$$CDR = \frac{\#of\ loans\ defaulted\ in\ period\ t}{\#\ of\ loans\ outstanding\ at\ the\ beginning\ of\ period\ t}, \textbf{(5)}$$

Overall, the prediction results capture the actual prepayment and default trends pretty well, especially in the one-month-ahead prediction. The result is important because it demonstrates the predictive power of the model in generating monthly cash flow projections in the pricing model with a monthly step size.

[Insert Figure 1. One-quarter-ahead SMM prediction based on scheduled and actual balance]

[Insert Figure 2. One-month-ahead SMM prediction based on scheduled and actual balance]

[Insert Figure 3. One-quarter-ahead CDR prediction based on actual size]

[Insert Figure 4. One-month-ahead CDR prediction based on actual size]

# Chapter 4. Term Structure Calibration and Cash Flow Discounting.

## Section 4- a. Continuous-time specifications of the shifted Hull-White model.

One important goal of the loan-level pricing model is to incorporate all the information available at the time of purchase when evaluating whether LMI mortgages provide a positive return to the secondary market. It is widely agreed that current market term structure quotes contain information about future yields and volatility and the state of the economy. Therefore, it is important to use non-arbitrage term structure models so that the RMBS are consistently priced and marked-to-market, so as to prevent arbitrage opportunities. Hence, the non-arbitrage term structure model used is the one-factor Hull & White (HW++) model fitted to the yield curve and volatility, because close form solution is required for generating long term rate as explained before. The basic strategy used to fit the initial yield and volatility curves is the inclusion of "time-varying" parameters in the model.

As summarized by Gupta and Subrahmanyam (2005), the generalized one-factor spot rate assumes that the instantaneous short-rate process evolves under the risk-neutral measure according to:

$$df(r) = [\theta(t) - \alpha f(r)]dt + \sigma dW(t), \qquad \mathbf{(6)}$$

where f(r) is some function of the short rate r, θ(t) is a function of time chosen so that the model provides an exact fit to the initial term structure, usually interpreted as a time-varying mean, α is mean reversion parameter, and σ is volatility parameter. When f(r)= r(t), the resulting model is the basic HW model fitted to the yield curve which is also the extended Vasicek Model:

$$dr(t) = [\theta(t) - \alpha r(t)]dt + \sigma dW(t). \quad \mathbf{(7)}$$

When f(r)= ln r(t) , the resulting model is BK fitted to the yield curve:

$$dlnr(t) = [\theta(t) - \alpha lnr(t)]dt + \sigma dW(t). \quad \mathbf{(8)}$$

The HW one-factor model fitted to both yield and volatility has the form:

$$dr(t) = [\theta(t) - \alpha(t)r(t)]dt + \sigma dW(t), \quad \mathbf{(9)}$$

where θ(t) and $\alpha(t)$ are deterministic functions of time that can be chosen so as to exactly fit both the observed yield curve and the volatility structure.

Hull and White (1990-1994) solved the stochastic differential equation (SDE) in (9) by the explicit finite difference method. This method solves for the parameters by equating the moment conditions of the trinomial tree with the continuous-time process and requiring that the transition probabilities sum to one. According to Hull and White (1994a), the instantaneous short rate r(t) conditional on $\mathcal{F}_t$ ( i.e. the information available up to time t ) is normally distributed with mean and variance given by:

$$E\{r(t)|\mathcal{F}_t\} = r(s)e^{-a(t-s)} + \alpha(t) - \alpha(s)e^{-a(t-s)}, \qquad \mathbf{(10)}$$

$$Var\{r(t)|\mathcal{F}_t\} = \frac{\sigma^2}{2a}\left[1 - e^{-2a(t-s)}\right], \qquad \mathbf{(11)}$$

where

$$\alpha(t) = f^M(0,t) + \frac{\sigma^2}{2a^2}[1 - e^{-at}]^2. \quad \mathbf{(12)}$$

The advantage of the normally distributed interest rate model is that there exists a close-form solution for the pure discount bond (zero-coupon bond), which follows a lognormal distribution. Future bond prices, at time T, dependent on the current term structure, the level of the short rate at time T, and the constant parameters of the short-rate process are given by:

$$P(t,T) = A(t,T)e^{-B(t,T)r(t)}$$

where

$$B(t,T) = \frac{1}{a}\left(1 - e^{-a(T-t)}\right)$$

$$lnA(t,T) = ln\frac{P(0,T)}{P(0,t)} - B(t,T)\frac{\partial lnP(0,T)}{\partial T} - \frac{\sigma^2}{4a^3}(1 - e^{-a(T-t)})^2(1 - e^{-2at}). \quad \mathbf{(13)}$$

Fitted to volatility (as well as yield) is important because modeling volatility smile has been increasingly emphasized in practice, according to both Brigo and Mercurio (2006; section 3.6) and Pedersen (2006). Hence using a term structure model not fitted to volatility will not be able to produce satisfactory hedges in the future, when the volatility term structure is a key input in the industry practice. Moreover, according to Pedersen (2006), one-factor models fitted to yield and volatility are still widely used in the industry to price corporate loans. Therefore, it is important to study the volatility skew features of one-factor models. According to Brigo and Mercurio (2006; section 3.6), one important criterion of a satisfactory interest rate model is that it should allow for a humped shape in the term structure of volatility,  the shape of the volatility skew typically observed in the market. The "term structure of volatility" mentioned above refers to the model-implied volatility. The model-implied T volatility $v_T^{model}$means the deterministic solution of volatility that makes the model price equal to the observed market price, where T is the maturity date. The term structure of the volatility implied by the short rate model is the graph of the model-implied T volatility against the time T, which is observed to be humped shape most of the time in the market. However there is a relationship between the model-implied volatility and the related absolute instantaneous volatility. When the zero coupon curve is increasing or slightly inverted, the term structure can feature large humps if the related absolute instantaneous volatilities of instantaneous forward rates that expressed as follows:

$$T \rightarrow \sqrt{\frac{Var\big(df(t,T)\big)}{dt}} = \sigma_f(t,T), \qquad \mathbf{(14)}$$

allows for a hump themselves. In short, the relationships between humps of term structure of volatilities and the humps in the instantaneous forward rates are as follows:

- no humps in $T \rightarrow \sigma_f(t,T)$, imply that only small humps for $T \rightarrow v_T^{model}$ are possible;
- humps in $T \rightarrow \sigma_f(t,T)$, imply that large humps for $T \rightarrow v_T^{model}$ are possible.

Brigo and Mercurio (2006) examine the ability of various popular one-factor models to produce a humped shape in the term structure of volatility. They find that the HW model gives rise to a more pronounced volatility skew than is usually observed. They also examine the CIR++ model and calculate the absolute volatility of instantaneous forward rates. They find that $T \rightarrow \sigma_f(t,T)$ is monotonically decreasing, thus the model-implied cap volatility calibrated to cap data displays a slightly humped shape. The cap volatility implied by the BK++ model is monotonically decreasing most of the time except when the forward yield curve is decreasing. Finally, they find that models with extra parameters in a suitable time-dependent function help to better recover the humped shape of the market cap-volatility curve. However, including additional time-dependent parameters will cause a parameter stability problem and affect the hedging results, as studied in Gupta and Subrahmanyam (2005). Therefore, there essentially exists a trade-off between better fitting the initial yield and volatility on the one hand and parameter stability on the other.

### Section 4- b. Cash flow projection using the MNL model.

The prepayment and defaults risks are estimated using the MNL model. In the discrete-time duration setting on MNL, the probability is calculated as follows. Consistent with the notations in equation (1), let $Y_{i,t}$ denote outcome observed for individual i at time t, where:

$$Y_{i,t} = \begin{cases} 0 \; if \; current \\ 1 \; if \; prepay \\ 2 \; if \; default \end{cases}$$

Hence for individual i

$$prob\left(Y_{i,t} = 0\right) = \frac{1}{1 + \sum_{j=1}^{2} \exp(Z_{i,t}\beta_j)}$$

$$prob\left(Y_{i,t} = m | Y_{i,t-1} = 0\right) = \frac{\exp\left(Z_{i,t}\beta_m\right)}{1 + \sum_{j=1}^{2} \exp(Z_{i,t}\beta_j)}, m = 1,2. \quad (\mathbf{15})$$

Combining all the observations across time and across individuals, the likelihood function in equation (1) is obtained.

For every node on the trinomial tree, the prepayment and default risks are calculated according to the MNL model results. The predicted values are the monthly conditional prepayment and default rates. Using the predicted prepayment and default risks, the cash flows are generated according to industry standard formulas. The scheduled balance remaining at the end of month n is:

$$SB_n = \frac{B_0[(1 + C)^N - (1 + C)^n]}{[(1 + C)^N - 1]}, \qquad (\mathbf{16})$$

The monthly payment at month n is:

$$M_n = \frac{B_{n-1} * C}{1 - \frac{1}{(1+C)^\wedge(T-n)}} \quad (\mathbf{17})$$

$where \; B_{n-1} is \; actual \; balance \; at \; the \; end \; of \; month \; n - 1; C \; is \; monthly \; coupon \; rate;$
$T \; is \; original \; term \; in \; months, i.e. 360.$

The interest payment at month n is:

$$I_n = C * B_{n-1} \textbf{ (18)}$$

The scheduled principal pay down at month n is:

$$SP_n = M_n - I_n \ (\mathbf{19})$$

The unscheduled principal pay down at month n is

$$UP_n = [SMM(n) + CDR(n) * \gamma]B_{n-1} \ \ (\mathbf{20})$$

where SMM(n) is the monthly conditional prepayment rate, CDR(n) is the monthly conditional default rate, and γ is the recovery rate in the event of default. The cash flow at month n is:

$$CF_n = I_n + SP_n + UP_n \ \ (\mathbf{21})$$

The actual balance at the end of month n is:

$$B_n = B_{n-1} - SP_n - UP_n \ \ (\mathbf{22})$$

The cash flows are discounted using the one-month discount rate according to the standard option pricing method.

### Section 4- c. Z-spread and Option-Adjusted Spread calculations.

The OAS and Z-spread calculations in what follows are consistent with those in conducted in Lehman Brothers' report by Pedersen (2006). According to Pedersen (2006), it is standard in the industry that a positive Z-spread or OAS indicates that the security is cheap for the buyers, and a negative Z-spread or OAS indicates that the security is expensive. Furthermore, according to Dunsky and Ho (2007), "the OAS can be interpreted as the gross profit of funding a mortgage loan". To be more specific, "it is the interest income net of the combined prepayment and default options sold to the mortgagors". Hence a positive OAS indicates that the buyer is making a profit on the security by paying the purchase price, and a negative OAS means that the buyer encounters a loss by paying the purchase price. This

OAS interpretation is also consistent with earlier industry reports by Hayre (1999) from Salomon Smith Barney and Beardsell and Liu (2005) from Citigroup concerning the general pricing framework and RMBS interest rate risk features.

The Z-spread (i.e. zero-volatility-spread) is the constant spread added to the initial yield curve such that the model price equals the market price paid. The Z-spread is the constant spread Z that satisfies the following equation:

$$P = \sum_{i=1}^{N} \frac{CF_i}{1 + R_i^Z + Z} \quad , \mathbf{(23)}$$

where P is the price, $CF_i$ is the predicted cash flow at time i based on the prepayment and default predictions, and $R_i^Z$ is the initial yield for maturity i. Moreover, N is the mortgage term in months.

The Z-spread is a relative measure, such that a positive Z-spread indicates that the security is cheap while a negative Z-spread indicates that the security is relatively expensive. For bonds with credit risks, the Z-spread to the initial yield curve should be positive to reflect the credit premium required. The higher the credit risk, the higher the Z-spread to the risk-free bond. The Z-spread is designed to solve the problem of the yield spread whereby all cash flows of different periods are discounted at the same rate. Hence, the yield curve is not detailed enough to allow a proper comparison of two bonds with different coupons even if the maturities are similar. The more the yield curve deviates from a flat curve, the more important it becomes to use the Z-spread instead of yield spread. The Z-spread is the excess return that can be earned from buying the bond and holding it to maturity, assuming that the issuer does not default and that coupons can be reinvested at the risk-free rate plus the Z-spread.

However, for bonds with embedded options the Z-spread is often not meaningful, since stochastic term structure model is used, and the OAS is the measure of spread used under the stochastic term structure. The OAS is simply the constant spread added to all the spot rates on all interest rate paths, and it makes the average present value of the paths equal the market price. The OAS is the constant spread that satisfies the following equation:

$$P = \sum_{j=1}^{J} W_j \sum_{i=1}^{N} \frac{CF_{i,j}}{1 + R_{ij}^{OAS} + OAS} \ , \mathbf{(24)}$$

where $W_j$ is the probability of rate path j, $CF_{i,j}$ is the predicted cash flow in period i along rate path j based on the prepayment and default predictions, and $R_{ij}^{OAS}$ is the zero coupon rate in period i along rate path j.

The OAS can be thought of as a Z-spread that has been adjusted for any option embedded in the bond, and for a bond without an embedded option the OAS is equal to the continuously compounded Z-spread. It is useful to compare a callable bond to a portfolio with positions in two hypothetical securities. One such security is the identical bond stripped of its embedded call option, called a stripped bond; the other security is the option on the stripped bond with the same call schedule as the option-embedded bond. Under the above assumption the value of the bond becomes:

$$Market\ price\ of\ bond\ with\ options = value\ of\ striped\ bond + value\ of\ options.$$

If the value of the option is known, it can be subtracted from the market price of the callable bond to arrive at a market-implied value of the stripped bond. Base on this, the Z-spread of the stripped bond can be calculated and reported as the OAS of the bond with embedded options. This is essentially the approach used to calculate the OAS when a stochastic term structure model is used. Overall, the OAS is the shift of all interest rates in all scenarios generated in the stochastic term structure model to correctly price the underlying stripped bond. The OAS is positive when the model price is greater than the market price, and vice versa.

Finally, the pricing model can be used to calculate the cost of guaranteeing default risk for RMBS loan underwriters, if good mark-to-market HPIs are provided. For an investor, the difference between purchasing a loan with default risk guarantee and purchasing one without is whether the recovery rate is 100% or the current house price in the case of default. Hence the expected cost of guaranteeing the default risk of a loan is the difference in the OAS between 100% recovery and a recovery at the current house price. According to Dunsky and Ho (2007), the guarantee cost for the default risk guarantor can be calculated as:

$$GC = OAS(100\%\ recovery) - OAS(recovery\ at\ current\ house\ price), (\mathbf{25})$$

# Chapter 5. Fannie Mae RMBS Pricing Practices and OAS Results Interpretations.

## Section 5- a. Fannie Mae RMBS pricing practices.

This section describes the secondary market pricing practices of Fannie Mae, and the typical deal structure of the GSEs RMBS in securitization. Understanding these practices is helpful for future modeling recovery risk in a counter-party risk framework.

The market prices considered are those that Fannie Mae paid to purchase mortgages in the CAP program from Self Help for securitization into RMBS. Fannie Mae has purchased a total of 8,308 loans in total as of May 2007, and hence all the 8,308 loan-level price data are studied in this section. The full guarantee of default risk in the CAP deal structure translates into a 100% recovery rate in pricing, which means the purchase prices should not be too far from par. The market price data confirm the expectation.

Although the full guarantee of default in CAP eliminates the need to allow a different recovery rate for each loan, it is worth mentioning how the recovery rate could be modeled. If good mark-to-market HPIs are available, a different recovery rate for each loan can be easily allowed in the loan-level model to translate the recovery rate into prices. Accordingly, the loan-level pricing model can be used to calculate the costs of guaranteeing the default risk of RMBS loans. In particular, in the recent heated discussion of how to overhaul the U.S. mortgage finance system, Treasury Secretary Geithner commented that[28]: “The challenge is

---

[28] See Bloomberg news "U.S. Treasury, Mortgage-Lenders Seek to Keep Government Role in Housing Fix". Aug 18, 2010, available at http://www.bloomberg.com/news/2010-08-18/u-s-treasury-morgage-lenders-seek-to-keep-government-role-in-housing-fix.html

to make sure that any government guarantee is priced to cover the risk of losses and structured to minimize taxpayer exposure". As explained previously, given good mark-to-market HPIs, such as Case-Shiller indices, the cost of a government guarantee of RMBS default risk can be easily calculated using the loan-level pricing model as:

$$GC = OAS(100\%\, recovery) - OAS(recovery\ at\ current\ house\ price).$$

With time-varying mortgage risks, the cost of a government guarantee should change constantly with factors that affect mortgage risks and prices, such as loan seasoning, term structure quotes, and house price indices.

Another way to price the recovery risk is to model it in a counter-party risk framework. In a typical Fannie Mae deal structure (not CAP loans), the recovery risk is guaranteed by Fannie Mae and mortgage insurers. In practice, mortgage insurance is required for loans with an LTV above 80%, for which losses are most likely in the case of foreclosure. For these loans, the mortgage insurer guarantees 75% of the house value that is reflected in the principal balance; hence the difference between the unpaid principal balance and 75% of the house value is guaranteed by Fannie Mae. Therefore, as became clear during the subprime crisis, recovery risk can be viewed as a form of counter-party risk, which is reflected in the probability that the guarantor of the recovery risk (i.e. mortgage insurers or broker dealers) will default from their responsibilities. Hence, one way to model recovery risk is to treat it as a counter-party risk and to model the likelihood that a guarantor of mortgage recovery risk will default using the CDS quote.

Table 6 provides a summary of the prices Fannie Mae paid. Fannie Mae's pricing practice is to pay roughly the same price for loans that it purchases on a given purchase date. As shown in the Table 6, the standard deviation of prices does not increase with number of loans purchased. For instance, the standard deviation of prices is not especially high in quarter 2 of 2001 when 1,921 loans are purchased, in quarter 2 of 2003 when 715 loans are purchased, and in quarter 4 of 2005 when 1,670 loans were purchased.

[Insert Table 6. Total loan purchase by purchase quarter]

A possible reason behind Fannie Mae's pricing practice of offering roughly the same price for loans purchased on the same date is that these loans are packaged as a pool, and for this reason are sold at the same price as a pool in the secondary market. Furthermore, the term structure of interest rate is roughly the same within a given day or month, except during the subprime crisis in 2008-09. It is true that small variations in term structure will result in big variations in prices, not only due to variations in discount rate but also because of changes in market expectation of prepayment risk as a result of shifts in the term structure. Hereinafter, an analysis is provided of the OAS and Z-spread of purchased CAP loans obtained using the loan-level model.

### Section 5- b. OAS and Z-spread results interpretation.

As discussed in definitions of the OAS and the Z-spread in Section 4.c, a positive OAS (or Z-spread) indicates that the model price is higher than the market price, and vice versa. Intuitively, a positive OAS means that the yield of the mortgage is still undervalued by the market price according to the pricing model. In particular, the higher the OAS the more the loan is undervalued. Due to missing information in the loan-level dataset especially in the neighborhood variables, the OAS is calculated for 7,168 loans out of the 8,308 loans with loan-level purchase prices. Table 7 summarizes the distribution of the OAS and the Z-spread for the 7,168 loans. According to Table 7, 35% of the 7,168 loans have a negative OAS, meaning that the model prices are less than the market prices and thus, the market prices overestimate the yields according to the pricing model. In addition, 65% of the 7,168 loans have a positive OAS and yields that are underestimated by market prices, while18% have an especially high OAS above 100 bps. In short, considering all the information available at the time of purchase, by the purchase prices the 7,168 loans have quite good profit-making performance for the underwriter and investors. The Z-spread is shown to be much higher than the OAS for a given loan, and this result is consistent with Z-spread and OAS quotes that are commonly observed in the market.

[Insert Table 7. Summary of OAS and Z-spread]

Table 8 and Table 9 present the tests using simple linear regression of whether CRA borrower characteristics, such as race, income, low credit score and high LTV at origination, are significantly correlated with a lower OAS and Z-spread, at least for issued RMBS composed of CAP loans. The adjusted R-squareds are quite low in all the OAS regressions, because the linear regressions mix a high-frequency dependent variable (the OAS) with independent categorical dummy variables. The low goodness-of-fit should not be a problem because the linear regressions are not intended to identify factors that capture variations in the OAS. All the variables that can possibly affect the OAS and the Z-spread are already included in the MNL regression and term structure calibration. Hence, the linear regressions are mostly intended to test whether CRA features and conventional wisdoms are reliable for predicting mortgage yields. In addition, the coding developed by Ding, Quercia and White (2009) and Ding and Quercia (2009) is used to test whether the state legal environment and the share of subprime origination are important for determining mortgage yields. Since several different coding methods are used by Ding, Quercia and White (2009), the regressions are run separately.  The dummy variables for state market coding are generated from Table 1 of the paper by Ding and Quercia(2009). The market coding scales take on values from 1 to 4, and states with a market code of 4 are those having the smallest share of subprime loans among all originations. Dummy variables for states with prepayment penalty, repayment ability and in effect coding are generated from Table 1 of the paper by Ding, Quercia and White (2009). The prepayment penalty scales ranges from 0 to 4, and the states with a prepayment penalty code of 4 are those having strongest laws against prepayment penalty. The repayment ability variable is binary, and a value of 0 indicates states with laws that impose repayment ability standards but only on loans above HOEPA triggers, or states that do not regulate mortgage repayment ability. Dummy variables for ineffect, ineffecttb and Pennington's ineffect[29] are coded slightly different but are all intended to identify states with a mortgage status that could plausibly have an impact on high-cost or subprime mortgage lending.

The OAS regressions in Table 8 show that the factors affecting OAS are quite different from those predicted by conventional wisdom, and indicators at origination are not entirely reliable in predicting mortgage yield. The parameter estimation results are explained in detail below.

[Insert Table 8. Linear regression of Option-Adjusted Spread]

***Panel A: Borrower and loan characteristics***

Loan age (in months) at the time of purchase turns out to be very important in predicting OAS, since more seasoned loans have lower prepayment and default risks due to their smaller UPB and LTV. A loan with a higher UPB (in thousand $) at time of purchase has a lower OAS, since a higher UPB will make a borrower more likely to prepay due to the larger absolute savings, net of transaction costs of refinancing.

[29] Pennington's ineffect variable is defined by Pennington_Cross, Bostic, Chomsisengphet, Engel, McCoy, and Wachter (2008).

However, many of the race, income, and variables at origination indicating CRA borrower characteristics turn out to have signs that are contrary to conventional expectations. For instance a loan with a higher LTV at origination has a significantly higher OAS. A loan to a borrower with FICO score at origination missing or less than or equal to 620 has a significantly higher OAS, while a loan to a borrower with FICO score at origination greater than or equal to 720 has a significantly lower OAS. A loan for which the borrower's income exceeds 50% of AMI has a significantly lower OAS. In particular, a loan made to an African American has a significantly higher OAS. In short, the results consistently show that in CAP, the traditionally perceived high risk loans have higher OAS instead. The counter-intuitive results come mainly from the fact that CAP provides a full guarantee of recovery risk, as this guarantee makes default the same as prepayment. Hence prepayment risk is somewhat the major risk here. Therefore, the CAP deal structure makes traditionally considered high risk loans no longer high risk. On the contrary the lower prepayment risk of CAP borrowers causes their loans to provide investors with higher returns. The loan-level model, by design, correctly captures this phenomenon. In summary, the results show that CRA features are not necessarily significantly correlated with lower mortgage yields, but can be associated with higher yields in CAP due to this type of loan’s lower prepayment risk. Furthermore, indicators at origination, such as LTV and credit score at origination, are not entirely reliable in predicting OAS, because the risks and yields of mortgage loan change constantly over time. Identifying profitable LMI mortgage requires the loan-level model be run on each mortgage portfolio.

***Panel B: Purchase year cohort***

The cohort effect of purchase year is proven to be important. Compared with a baseline of loans purchased in the year 2005, the cohorts of loans purchased in 2002, 2003, 2004 and 2006 have a significantly higher OAS; and loans purchased in 2001 have a significantly lower OAS. This result obtains because term structure quotes do not vary much within a cohort of loans purchased in the same year, and the term structure quotes not only generate the discount factor but also generate simulated refinance scenarios that drive prepayment risk. This cohort effect simply confirms the importance of term structure modeling in RMBS pricing.

***Panel C: State legal environment***

In addition, the signs and significance of the state legal environment variables in Table 8 show that state idiosyncratic characteristics dominate the legal environment in determining the OAS, since there is no universal answer to the question of whether stricter anti-predatory lending laws lead to a higher OAS. Variables for the share of subprime origination are mostly significant except in Model 2 where the "St ineffect" variable is used. Compared with the baseline of states having an above-average share of subprime originations, the states with the most subprime originations have the highest OAS, and states with the least subprime originations have the lowest OAS. This result suggests that the OAS may be largely driven by property market performance in CAP, since states with the most subprime originations are mostly those states having booming property markets. The variables for prepayment penalty restrictions are insignificant. Laws governing repayment ability turn out not to be important in determining the OAS. States with a mortgage status that could plausibly have an impact on high-cost or subprime mortgage lending turn out to have significantly lower OAS except in Model 2, which means that "ineffect" is probably a less consistent measure than the other two. After controlling for the state legal environment and the borrower and loan characteristics, loans in NC turn out to have a significantly higher OAS. Probably because Self Help is headquartered in NC and, therefore, acquire better information about local borrowers and can better service local loans. The findings on the effects of state laws are somewhat different from those of Ding, Quercia and White (2009). Mostly because state anti-predatory lending laws govern default and foreclosure risks but the full guarantee of default risk in the CAP deal structure and the very small number of defaults in the CAP make default and foreclosure risks less important in determining the OAS. Hence the research design of Ding, Quercia and White (2009) is more suitable for studying the impact of the state anti-predatory lending laws. Nevertheless, the above results show that idiosyncratic state characteristics seem to dominate in the determination of the OAS. Therefore, states should be given more autonomy in enacting and enforcing consumer protection laws based on their idiosyncratic situations.

For the sake of completeness, Table 9 provides comparison of regression results of the Z-spread using identical independent variables with OAS regression of Model 1 in Table 8. Since Z-spread is highly correlated with the OAS, the signs and significance of mostly of the variables in the Z-spread regression is consistent with the OAS regression. The most important difference in the Z-spread regression compared with the OAS regression is the high goodness of fit indicated by the high adjusted R-squared. The high goodness of fit is probably because the Z-spread is much less volatile than the OAS, as the Z-spread is calculated assuming the interest rate is non-stochastic.

[Insert Table 9. Comparison of OAS and Z-spread regressions]

### Section 5-c. OAS regressions with bundling effects.

Table 11 and Table 12 provide linear regression results for the OAS with tests of bundling effects. These regressions are intended to test whether the bundling effect in the market price data, i.e. the prices being exactly the same for loans bundled in the same tranche, have biased the previous arguments about the profitability of CAP CRA loans. Table 11 compares the results of the original regression, with repeating the linear regression of Model 1 (in Table 8) using only unbundled loans, and with repeating the regression using only bundled loans. Table 12 tests the bundling effect for the whole sample in another way by adding dummy variables for bundling using Model 1, and comparing the results with those for the original Model 1 in Table 8. A bundled loan is defined as a loan that has exactly the same price and the same purchase date as at least one other loan. Due to the importance of time variation in determining the mortgage yield, a dummy variable is created to identify whether the loan is bundled in each year with a bulk purchase of more than 100 loans. Hence in Table 12 in the regression using dummy variables for bundling on each bulk purchase year, the dummy variables for purchase year are removed due to redundancy. Table 10 provides a summary of the percentage of loans that were bundled in each purchase year, and a total of 5,917 (83%) of the 7,168 loans are bundled loans. Most variables become insignificant in the regression based on only unbundled loans, because the sample of 1,252 unbundled loans is very small. Despite that most variables are insignificant in Table 11, a few variables, i.e. age at purchase, the dummies for credit score at origination missing and less than or equal to 620, the dummy for African-American, and the dummies for purchase year cohort of 2006, 2003 and 2002, are still significant and have the same signs for the small unbundled sample as in the whole sample. Hence the effects of loan age, credit score at origination, African American and purchase year cohort on OAS that were found in the model on whole sample are still quite strong even for the small unbundled sample. For the regression based on only bundled loans, most variables have the same signs and significance as in the original regression for the whole sample. Therefore, the tests in Table 11 are not inconsistent with the previous arguments, but are not conclusive, because the very small size of the sample of unbundled loans makes most variables insignificant.

[Insert Table 10. Percentage of loans bundled by purchase year]

[Insert Table 11. Comparison of OAS regression on unbundled loans]

[Insert Table 12. Comparison of OAS regression with bundling effect using dummies]

The alternative test of the bundling effect presented in Table 12 shows that previous arguments about CAP loan profitability are still valid based on the effects of dummy variables controlling for the bundling effect. Table 12 compares the OAS regression of Model 1 with dummy variables for bundling with the original Model 1 in Table 8. The signs and significance of most of the variables remain unchanged in the new regression, except the dummies for bundling year. The signs of bundling year dummies are different mostly because different baselines are used. Moreover, if a dummy variable on bundling has a significantly positive sign, it means that being bundled in that particular year correlates with a significantly higher OAS compared with the baseline of the purchased CAP loans that are not bundled in that year, even after controlling for borrower and loan characteristics, and state legal environment. The dummy variables for bundling in 2002 and 2006 have significantly positive signs, and those for bundling in 2001 and 2005 have significantly negative signs. These results may be explained by noting that the good performance of CAP loans made the whole CRA tranches bundled in 2002 and 2006 profit-making, and bundling these CRA CAP loans together facilitated the sales of these tranches to investors. Because one of the important principles of bundling adopted by underwriters is to bundle loans in a way that make them the easiest to sell to investors.

Hence, whole tranches composed of CRA CAP loans do not necessarily have significantly lower yields, and the yields of whole tranches of CAP loans vary from year to year. Therefore, using the loan-level model is essential for identifying profitable loans, and for appropriately managing and hedging risks accordingly.

Overall, the above tests do not provide conclusive evidence concerning the profitability of the bundled CAP CRA loans relative to that of then unbundled loans. Testing and interpreting the relative profitability of bundled loans versus unbundled loans is difficult for the following reasons. First, the sample of CAP-issued MBS is limited to 8,308 loans and does not contain sufficient information about the unbundled loans. Most importantly, a loan that is not observed to be bundled with other loans in CAP is not necessarily priced and sold individually, because the loans may be bundled by Fannie Mae with other loans that do not come from CAP. Table 10 shows that unbundled loans were concentrated in years when a small number of CAP loans were purchased. In contrast, in years 2001, 2003 and 2005 when bulk purchases of CAP loans occurred, most purchased loans were bundled. It is possible that, in years when a few CAP loans were purchased, some CAP loans were bundled with loans in other portfolios because there was insufficient volume to bundle the CAP loans by themselves. Hence the fact that a loan is not bundled with other CAP loans does not necessarily mean that it did not end up bundled with loans from any other portfolio. Second, the limited CAP data set does not allow the construction of a comparable sample of unbundled loans that would allow the OAS for these loans to be compared with the OAS of the bundled loans in CAP. If a public loan-level data set, such as Loan Performance, is made available to us, then the propensity score matching method (PSM) as used by Quercia et al. (2008), can be easily applied to construct a comparable sample of similar borrowers receiving the same type of loans. Then, the OAS for this sample could be compared with the OAS of the bundled loans in CAP. Obtaining a good comparable sample using the PSM method used by Quercia et al. (2008) requires a public loan-level dataset, such as Loan Performance or McDash, which contains a wide choice of loans and records since origination. Therefore, drawing a definitive conclusion requires more data and more information about GSEs bundling guidelines.

If a public dataset, such as Loan Performance, is made available to us, the PSM method can also be used in secondary market underwriting to predict the OAS of borrowers without historical records by constructing a sample of similar borrowers receiving the same type of loans. For underwriting purpose, the OAS prediction can be obtained quickly and directly using linear regression as presented in Table 8. However this approach requires a linear model with high goodness of fit. The linear regression with high goodness-of-fit can be achieved by using continuous independent variables rather than categorical variables in Table 8, and by running the linear regression on a public data set with sufficiently many observations. Therefore, in order to facilitate underwriting selection, the availability of a public loan-level data set is essential for generating good predictions of the OAS for borrowers without historical records.

# Chapter 6. Policy Implications and Future Extensions

The policy implications of this paper are summarized as follows.

- The results provide strong evidence that CAP CRA mortgages can be quite profitable for the secondary market and for investors. Specifically, 65% of the 7,168 issued MBS from CAP that are studied in this paper have a positive OAS and 18% have an especially high OAS in excess of 100 basis points. The good result mostly can be attributed to good servicing and to the joint efforts of the Ford foundation, Self Help and Fannie Mae. In particular, the significantly higher OAS of NC loans is probably attributed to Self Help, being headquartered in NC, acquiring better information about local borrowers and better servicing local loans.
- The OAS and Z-spread regressions show that conventional perceptions of LMI borrower risks and returns, which are based largely on simple indicators like income, race, credit score and loan-to-value at origination, are not reliable for identifying profitable LMI mortgages. Furthermore, CRA tranches composed mostly of CAP loans do not necessarily have significantly lower yields, as shown by the results that tranches bundled in 2002 and 2006 exhibit a higher OAS after borrower characteristics, loan characteristics and the state legal environment are controlled for. Therefore, avoiding or discriminating against LMI mortgage pools is not rational.
- Since the risks of mortgage loans changes constantly, accurate pricing and effective risk management require adoption of the loan-level model developed here that can

automatically analyze the continuous flow of market data, including daily term structure quotes, monthly loan-level data, and monthly state macroeconomic environments. On the basis of accurate pricing model developed here, the cost of government guarantee of RMBS default risk can be precisely estimated, and strategies for hedging interest rate risk can be designed accordingly. Therefore, the model developed can be used by the federal government to better meet the financial needs of LMI borrowers while also maintaining the sustainability and soundness of the GSEs.

- The loan-level pricing model developed here provides a way to identify profitable mortgages for underwriting based on historical borrower performance. This approach can help to avoid the indiscriminate rejection of profitable LMI loans. For borrowers with no historical records in underwriting, the historical records of similar borrowers can instead be used and the "similar borrowers" sample can be constructed using propensity score matching (PSM) method by Ding, Quercia, Lei, Ratcliffe (2008). The PSM method is able to pair borrowers having historical records with new borrowers having no historical records on the basis of the conditional probability of getting a certain type of loan, given the observable characteristics.
- The pricing model developed in this paper can be used to estimate the cost of government guarantee of RMBS default risk if good mark-to-market HPIs, such as the Case-Shiller indices, are available. Moreover, this model can help to address one challenge outlined by Treasury Secretary Geithner in the discussion of how to overhaul the U.S. mortgage finance system—pricing of government guarantee of RMBS default risk, because the GSEs single-family MBS guarantee programs

accounted for $166 billion (73%) of the capital lost over the period, according to FHFA report.

- Idiosyncratic state-level factors seem to be the primary drivers of mortgage OAS. Therefore, states should be given more autonomy in enacting and enforcing consumer protection laws based on their idiosyncratic situations.

The proposed future extensions of this pricing framework are as follows.

- Effective strategies for hedging interest rate risk are currently under development using the loan-level model. Both the multifactor LIBOR market model and various one-factor short-rate models will be used in designing hedging strategies. Specifically, the multifactor LIBOR market model may have significant value for improving hedging efficiency as discussed in the term structure literature review.
- More data sets are essential to fully demonstrate the benefits of the loan-level pricing model developed. Moreover, PSM method adopted by Ding, Quercia, Lei, Ratcliffe (2008) can be used to construct samples of similar borrowers for the purpose of OAS prediction or comparison. For precise estimation of the cost of government guarantee of RMBS default risk, good mark-to-market house price indices are necessary. Furthermore, the precise and fast prediction of OAS for borrowers without historical records, for secondary-market underwriting purpose, requires construction of a sample of similar borrowers matched by PSM method and a linear OAS regression with high goodness of fit. A public loan-level dataset, like Loan Performance, is essential for both PSM method and linear regression with high goodness of fit.

- The private information contained in scattered survey data sets may facilitate better prediction of prices and the hedging of risks associated with LMI mortgage portfolios. As discussed previously, some factors summarized by Quercia and Stegman (1992) as affecting default risk are not available in any public loan-level dataset. For example, such factors include trigger events, such as borrower employment status, family health problems, or unexpected debts. Furthermore, divorce, changes in family size, or the addition of school-age children, as well as residential or neighborhood-level environmental problems may trigger borrowers to move and thus prepay their loans. The private information contained in scattered survey data sets can be used to improve pricing and hedging.

# Appendix A

## Figure 1. One-quarter-ahead SMM prediction based on scheduled and actual balance

This figure shows one-quarter-ahead single monthly mortality (SMM) prediction based on scheduled and actual balance using all loans in the CAP portfolio till July 2008. The variable "qrschSMM" is the one-quarter-ahead prediction based on the scheduled balance, meaning the last quarter's scheduled balance is combined with the predicted prepayment probability over the quarter to calculate the monthly SMM. The "qractSMM" is created using last quarter's actual balance combined with the predicted prepayment probability over the quarter to calculate the monthly SMM.

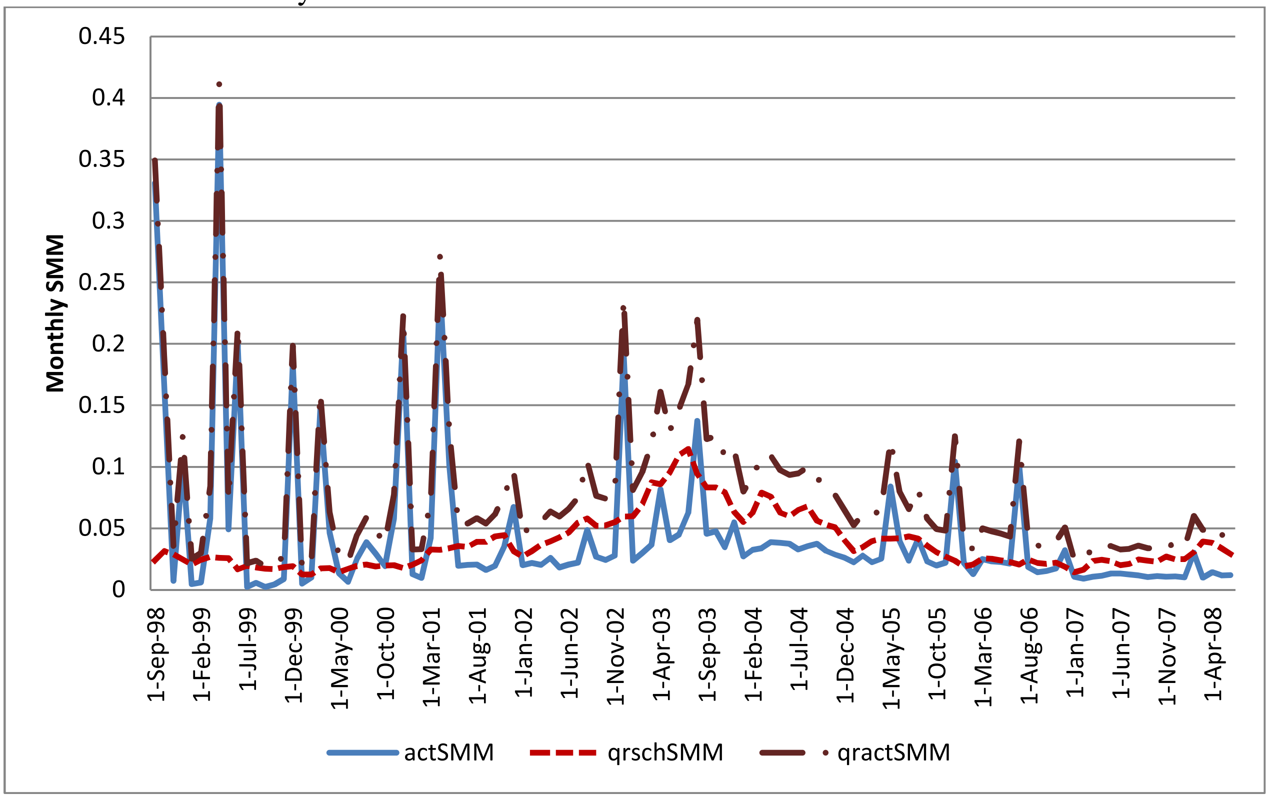
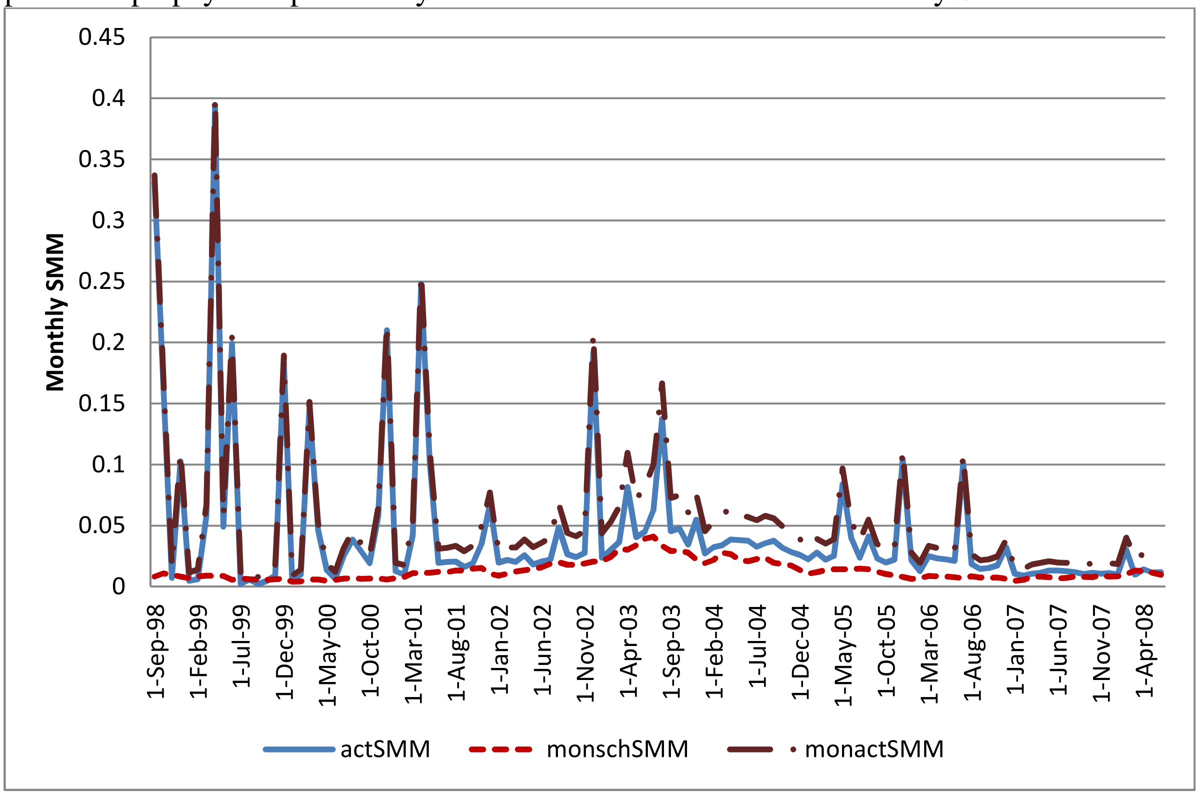


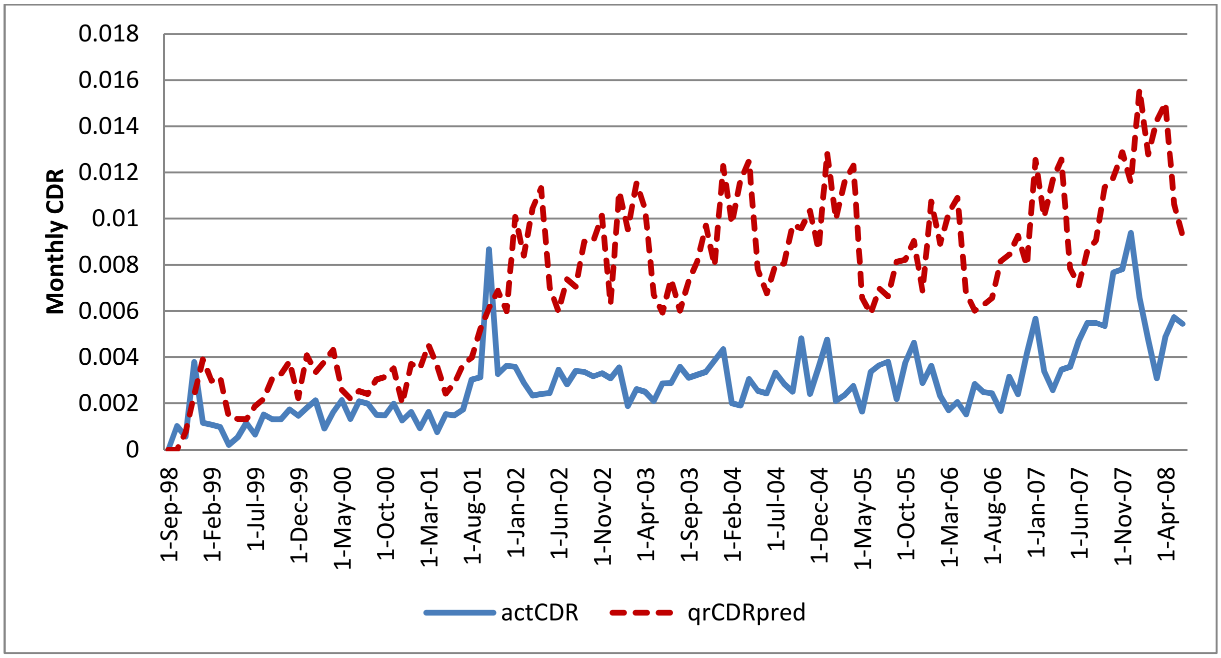
**Figure 2. One-month-ahead SMM prediction based on actual and scheduled balance**

This figure shows one- month-ahead single monthly mortality (SMM) prediction based on scheduled and actual balance using all loans in the CAP portfolio till July 2008. The "monschSMM" is created using the last month's scheduled balance combined with the predicted prepayment probability over the month to calculate the monthly SMM. The "monactSMM" is created using the last month's scheduled balance combined with the predicted prepayment probability over the month to calculate the monthly SMM.



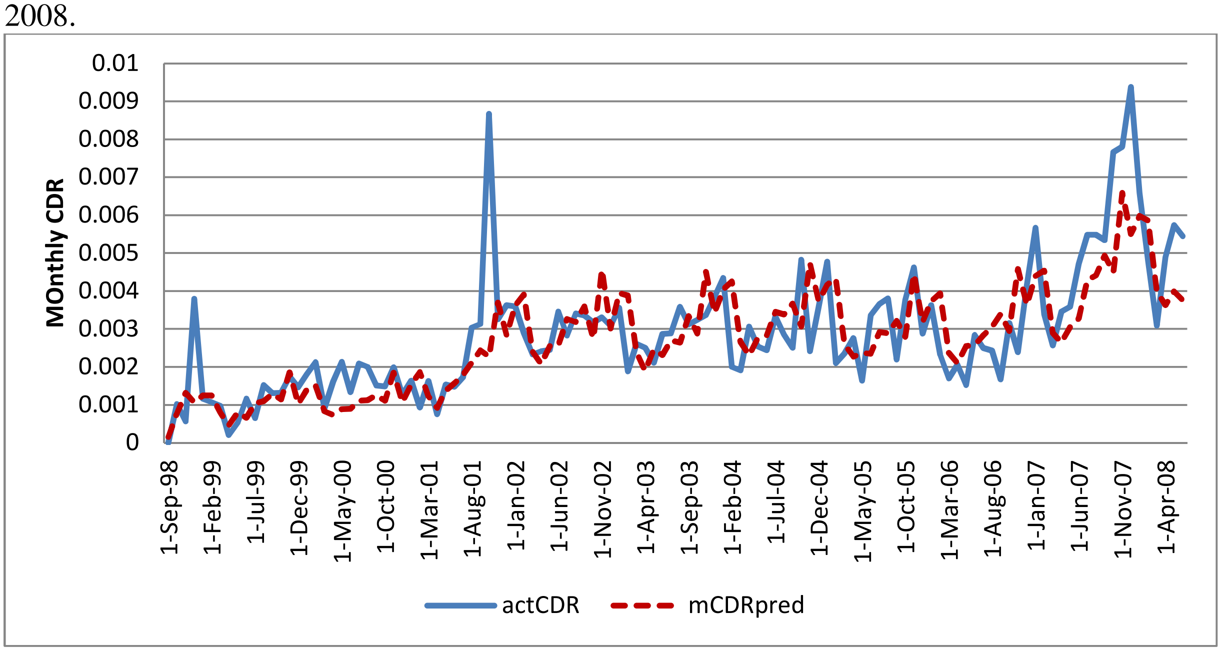
### Figure 3. One-quarter-ahead CDR prediction based on actual size

Figure 3 shows one-quarter-ahead prediction of constant default rate (CDR) based on number of loans existing at the end of last quarter. It is generated using all loans in the CAP till July 2008.




### Figure 4. One-month-ahead CDR prediction based on actual size

Figure 3 shows one-month-ahead prediction of constant default rate (CDR) based on number of loans existing at the end of last month. It is generated using all loans in the CAP till July 2008.



## Table 1. Fannie Mae credit profile by key product features[30]

The table shows each product feature's contribution to the normalized credit losses during 2007-2009. The rows titled "Single Family conventional guaranty book", "2009 credit loss", "2008 credit loss", "2007 credit loss" exhibit each product feature's share in respective categories. The rows titled "weighted average FICO", "Original LTV>90%" exhibit average FICO and percentage with Original LTV>90% for each product feature. Above data are obtained directly from Fannie Mae's 2010 1st quarter results. The rows titled "%2009creditloss/%guarantybook", "%2008creditloss/%guarantybook", "%2007creditloss/guarantybook" are intended to normalize the credit loss by the product's share of single family conventional guaranty book. They represent the percentage of credit losses in that year relative to the percentage of loans in the guaranty book of business.

| As of March 31 2010 | Neg Amrt Loans | Interest only | FICO <620 | 620< FICO <660 | OLTV >90% | FICO<620 &OLTV >90% | Alt_A | Sub-prime |
|---|---|---|---|---|---|---|---|---|
| Single Family conventional guaranty book | 0.5% | 6.3% | 3.8% | 8.0% | 9.4% | 0.8% | 8.5% | 0.3% |
| 2009 credit loss | 2.0% | 32.6% | 8.8% | 15.5% | 19.2% | 3.4% | 39.6% | 1.5% |
| 2009creditloss /guarantybook | 4.0x | 5.2x | 2.3x | 1.9x | 2.0x | 4.3x | 4.7x | 5.0x |
| 2008 credit loss | 2.9% | 34.2% | 11.8% | 17.4% | 21.3% | 5.4% | 45.6% | 2.0% |
| 2008creditloss /guarantybook | 5.8x | 5.4x | 3.1x | 2.2x | 2.3x | 6.8x | 5.4x | 6.7x |
| 2007 credit loss | 0.9% | 15.0% | 18.8% | 21.9% | 17.4% | 6.4% | 27.8% | 1.0% |
| 2007creditloss /guarantybook | 1.8x | 2.4x | 4.9x | 2.7x | 1.9x | 8.0x | 3.3x | 3.3x |
| weighted average FICO | 706 | 725 | 588 | 641 | 700 | 592 | 717 | 622 |
| Original LTV>90% | 0.3% | 9.1% | 21.9% | 20.7% | 100% | 100% | 5.4% | 6.8% |

[30] See pp.6 of the Fannie Mae's 2010 1st quarter result is available at http://www.fanniemae.com/ir/pdf/sec/2010/q1credit_summary.pdf

## Table 2. Summary of basic one-factor short rate models.

This table summarizes the continuous-time presentations of one-factor models under risk-neutral measure. In the table, r(t) is the instantaneous short rate at time t, θ(t) can be considered as time-varying means, α is mean reversion parameter, σ is the volatility parameter, and W(t) is one dimensional Brownian motion.

| Model | Continuous time | Distribution | Analytical solution |
|---|---|---|---|
| HW | $dr(t) = [\theta(t) - \alpha r(t)]dt + \sigma dW(t)$ | Normal | Yes |
| BK | $dlnr(t) = [\theta(t) - \alpha lnr(t)]dt + \sigma dW(t)$ | Lognormal | No |
| CIR | $dr(t) = [\theta(t) - \alpha r(t)]dt + \sigma\sqrt{r(t)}dW(t)$ | Normal | yes |

## Table 3. Termination events by transaction year

This table shows the percentage of current, prepayment, and default observations of the whole CAP portfolio (including those with some missing observations) by transaction year from 1998 till July 2008. In total, there are 1.38% (of all observations) prepayment observations and 0.27% default observations.

| Transaction Year | Current(%) | Prepaid(%) | Default(%) | Total(#) |
|---|---|---|---|---|
| 1998 | 97.84 | 2.01 | 0.15 | 16211 |
| 1999 | 99.20 | 0.65 | 0.15 | 58590 |
| 2000 | 99.27 | 0.56 | 0.17 | 94282 |
| 2001 | 98.37 | 1.36 | 0.28 | 185405 |
| 2002 | 98.13 | 1.57 | 0.30 | 204422 |
| 2003 | 96.85 | 2.86 | 0.30 | 217689 |
| 2004 | 97.88 | 1.83 | 0.29 | 190893 |
| 2005 | 98.37 | 1.35 | 0.29 | 189933 |
| 2006 | 98.86 | 0.91 | 0.22 | 222785 |
| 2007 | 98.98 | 0.74 | 0.29 | 226790 |
| 2008 | 98.98 | 0.73 | 0.29 | 108179 |
| Total | 98.35 | 1.38 | 0.27 | 1715179 |

## Table 4. Refinance and burnout spline knots

This table summarizes the transformation of the refi-spread and burnout into linear spline as described in equations below. Transforming the continuous refinance spread and burnout variables into spline knots allows a better fit to the categorical dependent variable in MNL regression by allowing a different slope within each piece. The spline knot explanations and knot point choices are discussed in Section 3- b.

$$Burn_k = I\{L(k-1) < burnout_{i,t} < L(k)\} * \left(burnout_{i,t} - L(k-1)\right)$$

$$+ I\{burnout_{i,t} \geq L(k)\} * \left(L(k) - L(k-1)\right), (for\ k = 1, \ldots, 4).$$

$$Burn_k = I\{L(4) < burnout_{i,t}\} * \left(burnout_{i,t} - L(4)\right), (for\ k = 5).$$

| k | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| L(k) | 0 | 0.2 | 0.7 | 1.2 | 1.7 | |

$$Refi_h = I\{L(h-1) < refispd_{i,t} < L(h)\} * \left(refispd_{i,t} - L(h-1)\right)$$

$$+ I\{refispd_{i,t} \geq L(h)\} * \left(L(h) - L(h-1)\right), (for\ h = 1, \ldots, 5).$$

$$Refi_h = I\{L(5) < refispd_{i,t}\} * \left(refispd_{i,t} - L(5)\right), (for\ h = 6).$$

| h | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| L(h) | 0 | 1 | 1.1 | 1.2 | 1.3 | 1.4 |

## Table 5. MNL regression results

This table provides multinomial logit regressions of prepayment and default risks modeling of all loans in the CAP portfolio from 1998 until July 2008 using explanatory variables grouped into panels. The panels of explanatory variables include seasoning, seasonality, origination cohort, FICO score effect, UPB effect, yield curve slope, MTMLTV, refinance burnout factor, and borrower, neighborhood and loan characteristics. Interpretations of these panels are provided in Section 3- b.

| | Model 1 (partial) | | Model 2 (simple refi-spread) | | Model 3 (Final ) | |
|---|---|---|---|---|---|---|
| | obs | 1589836 | obs | 1483289 | obs | 1483289 |
| | PseudoR$^2$ | 0.1305 | PseudoR$^2$ | 0.1368 | PseudoR$^2$ | 0.1382 |
| | LR $\chi^2$ | 38766.09 | LR $\chi^2$ | 38317.06 | LR $\chi^2$ | 38711.20 |
| | DF | 126 | DF | 126 | DF | 150 |
| | Log likelihood | -129116 | Log likelihood | -120867 | Log likelihood | -120669 |
| | Coef. | P>\|z\| | Coef. | P>\|z\| | Coef. | P>\|z\| |
| | **Prepay** | | **Prepay** | | **Prepay** | |
| ***Seasoning*** | | | | | | |
| Age1 in (0,3] | 1.75210 | 0.010 | 1.70407 | 0.012 | 1.72800 | 0.010 |
| Age2 in (3,6] | 0.23771 | 0.014 | 0.25106 | 0.011 | 0.25861 | 0.009 |
| Age3 in (6,9] | 0.18407 | 0.001 | 0.18133 | 0.002 | 0.18677 | 0.001 |
| Age4 in (9,12] | 0.19330 | 0.000 | 0.19956 | 0.000 | 0.20512 | 0.000 |
| Age5 in (12,18] | 0.03895 | 0.000 | 0.05560 | 0.000 | 0.04819 | 0.000 |
| Age6 in (18,24] | 0.01706 | 0.040 | 0.02410 | 0.004 | 0.01780 | 0.036 |
| Age7 in (24,30] | 0.00658 | 0.355 | 0.01406 | 0.050 | 0.00887 | 0.223 |
| Age8 in (30,40] | -0.00046 | 0.907 | 0.00285 | 0.475 | 0.00043 | 0.915 |
| Age9 in (40,50] | -0.00184 | 0.639 | -0.00222 | 0.572 | -0.00233 | 0.561 |
| Age10 in (50,60] | -0.00809 | 0.027 | 0.00208 | 0.572 | -0.00804 | 0.032 |
| Age11 in (60,90] | -0.01725 | 0.000 | -0.00837 | 0.000 | -0.01367 | 0.000 |
| Age12 in (90, 290] | -0.00695 | 0.000 | -0.00391 | 0.001 | -0.00737 | 0.000 |

## Table 5. MNL regression results (cont'd)

| | Coef. | P>\|z\| | Coef. | P>\|z\| | Coef. | P>\|z\| |
|---|---|---|---|---|---|---|
| ***Seasonality*** | | | | | | |
| Feb | 0.14362 | 0.000 | 0.12486 | 0.001 | 0.13655 | 0.000 |
| March | 0.36397 | 0.000 | 0.34212 | 0.000 | 0.36058 | 0.000 |
| April | 0.38953 | 0.000 | 0.41358 | 0.000 | 0.40354 | 0.000 |
| May | 0.39554 | 0.000 | 0.40029 | 0.000 | 0.40087 | 0.000 |
| June | 0.43093 | 0.000 | 0.42174 | 0.000 | 0.43094 | 0.000 |
| July | 0.47876 | 0.000 | 0.52848 | 0.000 | 0.49563 | 0.000 |
| Aug | 0.51260 | 0.000 | 0.58967 | 0.000 | 0.53182 | 0.000 |
| Sept | 0.36892 | 0.000 | 0.42241 | 0.000 | 0.37784 | 0.000 |
| Oct | 0.32688 | 0.000 | 0.38303 | 0.000 | 0.34522 | 0.000 |
| Nov | 0.34676 | 0.000 | 0.39771 | 0.000 | 0.36072 | 0.000 |
| Dec | 0.18862 | 0.000 | 0.21491 | 0.000 | 0.19220 | 0.000 |
| ***Origination Cohort*** | | | | | | |
| 1995 | 1.37192 | 0.000 | 1.26076 | 0.000 | 1.55193 | 0.000 |
| 1996 | 1.24395 | 0.000 | 1.18512 | 0.000 | 1.38505 | 0.000 |
| 1997 | 1.05069 | 0.000 | 1.12620 | 0.000 | 1.26459 | 0.000 |
| 1998 | 1.05188 | 0.000 | 1.13484 | 0.000 | 1.22222 | 0.000 |
| 1999 | 0.74153 | 0.000 | 0.90178 | 0.000 | 0.92664 | 0.000 |
| 2000 | 0.85989 | 0.000 | 0.91181 | 0.000 | 0.92083 | 0.000 |
| 2001 | 0.70591 | 0.000 | 0.93157 | 0.000 | 0.84374 | 0.000 |
| 2002 | 0.59483 | 0.000 | 0.84160 | 0.000 | 0.78062 | 0.000 |
| 2003 | 0.58863 | 0.000 | 0.63399 | 0.000 | 0.78490 | 0.000 |
| 2004 | 0.61818 | 0.000 | 0.53063 | 0.000 | 0.68155 | 0.000 |
| 2005 | 0.34171 | 0.001 | 0.23723 | 0.032 | 0.37488 | 0.001 |
| 2007 | -0.61702 | 0.029 | | | | |
| 2008 | | | | | | |

## Table 5. MNL regression results (cont'd)

| | Coef. | P>\|z\| | Coef. | P>\|z\| | Coef. | P>\|z\| |
|---|---|---|---|---|---|---|
| ***FICO Score Effect*** | | | | | | |
| ***Cscore/100 in*** | | | | | | |
| ( min, 5.8] | 0.10617 | 0.004 | 0.07714 | 0.037 | 0.07895 | 0.033 |
| (5.8, 6.2] | 1.37884 | 0.000 | 1.28244 | 0.000 | 1.29309 | 0.000 |
| (6.2, 6.6] | -0.06829 | 0.414 | -0.11430 | 0.183 | -0.11205 | 0.192 |
| (6.6, 7.2] | 0.08183 | 0.071 | 0.08029 | 0.086 | 0.08359 | 0.074 |
| (7.2, max] | -0.64885 | 0.000 | -0.55380 | 0.000 | -0.53624 | 0.000 |
| ***UPB Effect*** | | | | | | |
| ***upb/1000*** | | | | | | |
| (0, 50] | -0.00061 | 0.640 | -0.00088 | 0.524 | -0.00128 | 0.350 |
| (50, 75] | 0.01584 | 0.000 | 0.01603 | 0.000 | 0.01594 | 0.000 |
| (75, 100] | 0.00994 | 0.000 | 0.00892 | 0.000 | 0.00944 | 0.000 |
| (100, 150] | 0.00686 | 0.000 | 0.00449 | 0.000 | 0.00466 | 0.000 |
| (150, 407.666] | 0.00236 | 0.001 | 0.00251 | 0.001 | 0.00261 | 0.001 |
| ***Yield Curve Slope*** | | | | | | |
| Tbill10yr-- 2 yr | 0.26380 | 0.000 | 0.24186 | 0.000 | 0.22965 | 0.000 |
| ***Mark-to-market LTV*** | | | | | | |
| MTMLTV | -0.01038 | 0.000 | -0.01009 | 0.000 | -0.00941 | 0.000 |

## Table 5. MNL regression results (cont'd)

| | Coef. | P>\|z\| | Coef. | P>\|z\| | Coef. | P>\|z\| |
|---|---|---|---|---|---|---|
| ***Refi Burnout factor*** | | | | | | |
| ***Refispread spline*** | | | Refi_spread | | | |
| refi1 in (0,1] | 3.51127 | 0.000 | 2.44699 | 0.000 | 4.57086 | 0.000 |
| refi2 in (1,1.1] | 1.51818 | 0.000 | | | 2.03636 | 0.000 |
| refi3 in (1.1, 1.2] | 1.77834 | 0.000 | | | 1.92868 | 0.000 |
| refi4 in (1.2, 1.3] | 0.16991 | 0.632 | | | 0.31423 | 0.392 |
| refi5 in (1.3, 1.4] | 0.25322 | 0.555 | | | 0.58825 | 0.194 |
| refi6 in (1.4, max] | 1.37810 | 0.000 | | | 1.99551 | 0.000 |
| ***Burnout spread spline*** | | | Burn_out spread | | | |
| burn1 in (0, 0.2] | 2.31735 | 0.000 | -0.02571 | 0.000 | 2.28482 | 0.000 |
| burn2 in (0.2, 0.7] | -0.16586 | 0.039 | | | -0.16093 | 0.052 |
| burn3 in (0.7, 1.2] | 0.17084 | 0.110 | | | 0.16279 | 0.139 |
| burn4 in (1.2, 1.7] | -0.09488 | 0.325 | | | -0.01965 | 0.845 |
| burn5 in (1.7, max] | -0.01916 | 0.005 | | | -0.01126 | 0.128 |
| ***Borrower, Neighborhood and Loan characteristics*** | | | | | | |
| is low_to_mod inc track | -0.03087 | 0.116 | -0.05178 | 0.025 | -0.04779 | 0.039 |
| is minority track | -0.14689 | 0.000 | -0.05478 | 0.015 | -0.04389 | 0.052 |
| is underserved area | -0.09347 | 0.000 | -0.05633 | 0.005 | -0.05666 | 0.005 |
| is worst ever delin 30 days | -0.44781 | 0.000 | -0.44451 | 0.000 | -0.45406 | 0.000 |
| is worst ever delin above 30 days | -1.67211 | 0.000 | -1.71760 | 0.000 | -1.72191 | 0.000 |
| is african american | | | -0.54375 | 0.000 | -0.55345 | 0.000 |
| is hispanic | | | -0.14920 | 0.000 | -0.17715 | 0.000 |
| is female | | | -0.08273 | 0.000 | -0.08330 | 0.000 |
| is rural | | | -0.07161 | 0.001 | -0.07212 | 0.001 |
| back end ratio | | | 0.01524 | 0.153 | 0.01339 | 0.232 |
| annual inc as %AMI | | | 0.00010 | 0.217 | 0.00010 | 0.235 |
| median track inc as %AMI | | | 0.08565 | 0.037 | 0.08021 | 0.051 |
| is NC | | | -0.11399 | 0.000 | -0.12991 | 0.000 |
| is CA | | | 0.00706 | 0.822 | 0.01735 | 0.583 |
| is FL | | | -0.07926 | 0.065 | -0.02867 | 0.506 |
| is OH | | | -0.05951 | 0.156 | -0.07567 | 0.071 |
| is OK | | | -0.33488 | 0.000 | -0.40697 | 0.000 |
| orig coupon- market PMMS | | | 0.44919 | 0.000 | 0.43775 | 0.000 |
| constant | -17.08240 | 0.000 | -16.04867 | 0.000 | -18.25059 | 0.000 |

### Table 5. MNL regression results (cont'd)

| | Coef. **Default** | P>\|z\| | Coef. **Default** | P>\|z\| | Coef. **Default** | P>\|z\| |
|---|---|---|---|---|---|---|
| ***Seasoning*** | | | | | | |
| Age1 in (0,3] | 1.14186 | 0.087 | 0.72134 | 0.267 | 0.73184 | 0.260 |
| Age2 in (3,6] | 0.76892 | 0.000 | 0.78834 | 0.000 | 0.78985 | 0.000 |
| Age3 in (6,9] | 0.01600 | 0.780 | 0.05393 | 0.434 | 0.05010 | 0.467 |
| Age4 in (9,12] | 0.15702 | 0.000 | 0.16839 | 0.001 | 0.16665 | 0.001 |
| Age5 in (12,18] | 0.05216 | 0.004 | 0.04724 | 0.012 | 0.05190 | 0.006 |
| Age6 in (18,24] | 0.02012 | 0.222 | 0.01861 | 0.264 | 0.02156 | 0.200 |
| Age7 in (24,30] | 0.05757 | 0.000 | 0.05035 | 0.001 | 0.05977 | 0.000 |
| Age8 in (30,40] | 0.01780 | 0.045 | 0.01335 | 0.133 | 0.01533 | 0.090 |
| Age9 in (40,50] | 0.02220 | 0.014 | 0.02493 | 0.006 | 0.02722 | 0.003 |
| Age10 in (50,60] | 0.01765 | 0.040 | 0.01133 | 0.184 | 0.01587 | 0.068 |
| Age11 in (60,90] | 0.01700 | 0.000 | 0.01300 | 0.000 | 0.01666 | 0.000 |
| Age12 in (90, 290] | 0.02076 | 0.000 | 0.01782 | 0.000 | 0.02079 | 0.000 |

## Table 5. MNL regression results (cont'd)

| | Coef. | P>\|z\| | Coef. | P>\|z\| | Coef. | P>\|z\| |
|---|---|---|---|---|---|---|
| ***Seasonality*** | | | | | | |
| Feb | -0.50960 | 0.000 | -0.51129 | 0.000 | -0.51279 | 0.000 |
| March | -0.69172 | 0.000 | -0.67882 | 0.000 | -0.68035 | 0.000 |
| April | -0.50464 | 0.000 | -0.52960 | 0.000 | -0.51495 | 0.000 |
| May | -0.54486 | 0.000 | -0.55852 | 0.000 | -0.54146 | 0.000 |
| June | -0.32664 | 0.000 | -0.34746 | 0.000 | -0.32099 | 0.000 |
| July | -0.32387 | 0.000 | -0.36504 | 0.000 | -0.32157 | 0.000 |
| Aug | -0.20900 | 0.002 | -0.26819 | 0.000 | -0.22358 | 0.001 |
| Sept | -0.41208 | 0.000 | -0.42630 | 0.000 | -0.40185 | 0.000 |
| Oct | 0.04519 | 0.468 | 0.03275 | 0.607 | 0.05430 | 0.396 |
| Nov | -0.17720 | 0.007 | -0.20426 | 0.002 | -0.18428 | 0.006 |
| Dec | -0.03173 | 0.612 | -0.06030 | 0.349 | -0.04151 | 0.520 |
| ***Origination Cohort*** | | | | | | |
| 1995 | -3.61356 | 0.000 | -3.01054 | 0.000 | -3.21735 | 0.000 |
| 1996 | -3.24941 | 0.000 | -2.90022 | 0.000 | -3.02903 | 0.000 |
| 1997 | -3.00127 | 0.000 | -2.66583 | 0.000 | -2.74983 | 0.000 |
| 1998 | -2.81250 | 0.000 | -2.56679 | 0.000 | -2.59835 | 0.000 |
| 1999 | -2.44327 | 0.000 | -2.20256 | 0.000 | -2.20583 | 0.000 |
| 2000 | -1.73909 | 0.000 | -1.71315 | 0.000 | -1.67313 | 0.000 |
| 2001 | -1.89041 | 0.000 | -1.95081 | 0.000 | -1.85669 | 0.000 |
| 2002 | -1.78867 | 0.000 | -1.76957 | 0.000 | -1.71722 | 0.000 |
| 2003 | -1.88786 | 0.000 | -1.68203 | 0.000 | -1.72730 | 0.000 |
| 2004 | -1.24456 | 0.000 | -1.20094 | 0.000 | -1.22604 | 0.000 |
| 2005 | -0.75965 | 0.000 | -0.81164 | 0.000 | -0.78950 | 0.000 |
| 2007 | 0.81862 | 0.000 | | | | |
| 2008 | | | | | | |

## Table 5. MNL regression results (cont'd)

| | Coef. | P>\|z\| | Coef. | P>\|z\| | Coef. | P>\|z\| |
|---|---|---|---|---|---|---|
| ***FICO Score Effect*** | | | | | | |
| ***Cscore/100 in*** | | | | | | |
| ( min, 5.8] | -0.25104 | 0.000 | -0.24568 | 0.000 | -0.24882 | 0.000 |
| (5.8, 6.2] | -0.41645 | 0.005 | -0.34891 | 0.022 | -0.36542 | 0.017 |
| (6.2, 6.6] | -0.29557 | 0.194 | -0.26926 | 0.245 | -0.25784 | 0.265 |
| (6.6, 7.2] | 0.27347 | 0.181 | 0.21651 | 0.295 | 0.21010 | 0.310 |
| (7.2, max] | 0.20906 | 0.456 | 0.43457 | 0.125 | 0.42497 | 0.134 |
| ***UPB Effect*** | | | | | | |
| ***upb/1000*** | | | | | | |
| (0, 50] | 0.00401 | 0.135 | 0.00624 | 0.028 | 0.00610 | 0.031 |
| (50, 75] | 0.00001 | 0.997 | 0.00217 | 0.345 | 0.00218 | 0.341 |
| (75, 100] | 0.00392 | 0.141 | 0.00688 | 0.014 | 0.00671 | 0.016 |
| (100, 150] | 0.00058 | 0.769 | -0.00083 | 0.712 | -0.00069 | 0.758 |
| (150, 407.666] | 0.00027 | 0.856 | -0.00144 | 0.504 | -0.00147 | 0.495 |
| ***Yield Curve Slope*** | | | | | | |
| Tbill10yr- 2 yr | 0.17118 | 0.000 | 0.20112 | 0.000 | 0.17578 | 0.000 |
| ***Mark-to-market LTV*** | | | | | | |
| MTMLTV | 0.00591 | 0.000 | 0.00861 | 0.000 | 0.00772 | 0.000 |

## Table 5. MNL regression results (cont'd)

| | Coef. | P>\|z\| | Coef. | P>\|z\| | Coef. | P>\|z\| |
|---|---|---|---|---|---|---|
| ***Refi Burnout factor*** | | | | | | |
| ***Refispread spline*** | | | | | | |
| refi1 in (0,1] | -0.29391 | 0.641 | | | 0.16213 | 0.803 |
| refi2 in (1,1.1] | -2.27692 | 0.002 | Refi_spread | | -2.17929 | 0.005 |
| refi3 in (1.1, 1.2] | 0.73038 | 0.311 | -1.18409 | 0.000 | 0.76448 | 0.299 |
| refi4 in (1.2, 1.3] | 1.68989 | 0.050 | | | 1.42261 | 0.106 |
| refi5 in (1.3, 1.4] | -2.49698 | 0.009 | | | -2.71518 | 0.006 |
| refi6 in (1.4, max] | -1.27261 | 0.210 | | | 0.35341 | 0.752 |
| ***Burnout spread spline*** | | | | | | |
| burn1 in (0, 0.2] | -1.93257 | 0.000 | Burn_out spread | | -1.74333 | 0.000 |
| burn2 in (0.2, 0.7] | -0.29053 | 0.157 | 0.00419 | 0.668 | -0.20709 | 0.318 |
| burn3 in (0.7, 1.2] | 0.14230 | 0.586 | | | 0.14937 | 0.568 |
| burn4 in (1.2, 1.7] | -0.21259 | 0.345 | | | -0.30368 | 0.19 |
| burn5 in (1.7, max] | -0.01548 | 0.143 | | | 0.00229 | 0.852 |
| ***Borrower, Neighborhood and Loan characteristics*** | | | | | | |
| is low_to_mod inc track | 0.04796 | 0.212 | 0.00632 | 0.897 | 0.00544 | 0.911 |
| is minority track | -0.05664 | 0.138 | 0.01958 | 0.660 | 0.01818 | 0.683 |
| is underserved area | 0.05576 | 0.146 | 0.05333 | 0.232 | 0.05136 | 0.249 |
| is worst ever delin 30 days | -25.80673 | 1.000 | -8.20772 | 1.000 | -15.72557 | 1.000 |
| is worst ever delin above 30 days (60-90days) | 23.36588 | 0.000 | 26.96790 | 0.000 | 29.45130 | 0.000 |
| is african american | | | -0.14863 | 0.000 | -0.14621 | 0.000 |
| is hispanic | | | -0.19696 | 0.001 | -0.19670 | 0.001 |
| is female | | | -0.03379 | 0.283 | -0.03334 | 0.290 |
| is rural | | | 0.03792 | 0.374 | 0.03678 | 0.388 |
| back end ratio | | | 0.15029 | 0.324 | 0.17296 | 0.255 |
| annual inc as %AMI | | | -0.00310 | 0.001 | -0.00298 | 0.002 |
| median track inc as %AMI | | | -0.11655 | 0.292 | -0.11647 | 0.291 |
| is NC | | | -0.15265 | 0.000 | -0.14666 | 0.000 |
| is CA | | | 0.40187 | 0.002 | 0.37614 | 0.003 |
| is FL | | | 0.30765 | 0.003 | 0.25816 | 0.014 |
| is OH | | | -0.06120 | 0.302 | -0.05979 | 0.314 |
| is OK | | | 0.00327 | 0.962 | 0.02846 | 0.684 |
| orig coupon- market PMMS | | | 0.23111 | 0.000 | 0.23765 | 0.000 |
| constant | -30.88015 | . | -32.90117 | . | -36.63604 | . |

### Table 6. Total loan purchase by purchase quarter

This table provides a summary of the frequency, mean and standard deviation of the prices Fannie Mae paid for securitized CAP loans by purchase quarters. The price is per $100 unpaid principal balance.

| Purchase quarter | Frequency | Mean(price) | SD(price) |
|---|---|---|---|
| 1996Q4 | 1 | 98.6966 | |
| 1998Q3 | 2 | 100 | 0 |
| 1999Q2 | 5 | 100 | 0 |
| 2000Q3 | 8 | 100.7600 | 0.54994 |
| 2000Q4 | 17 | 100.5959 | 0.58931 |
| 2001Q1 | 4 | 101.4084 | 0.65203 |
| 2001Q2 | 1921 | 100.0992 | 0.94910 |
| 2001Q3 | 62 | 101.1622 | 0.62246 |
| 2001Q4 | 447 | 100.7286 | 0.60018 |
| 2002Q1 | 26 | 101.5192 | 0.90931 |
| 2002Q2 | 42 | 101.3217 | 0.77097 |
| 2002Q3 | 30 | 101.4407 | 0.57816 |
| 2002Q4 | 203 | 101.2111 | 1.35878 |
| 2003Q1 | 57 | 101.3399 | 0.61367 |
| 2003Q2 | 715 | 101.9637 | 0.76258 |
| 2003Q3 | 106 | 99.4671 | 2.79799 |
| 2003Q4 | 500 | 100.0898 | 1.20027 |
| 2004Q1 | 131 | 101.6689 | 0.87679 |
| 2004Q2 | 177 | 100.4401 | 1.74237 |
| 2004Q3 | 129 | 101.5938 | 1.26628 |
| 2004Q4 | 163 | 101.3971 | 1.28593 |
| 2005Q1 | 122 | 100.9179 | 0.87333 |
| 2005Q2 | 398 | 100.5026 | 0.71966 |
| 2005Q3 | 142 | 100.8089 | 0.99461 |
| 2005Q4 | 1670 | 101.7239 | 0.49844 |
| 2006Q1 | 128 | 100.7226 | 0.81870 |
| 2006Q2 | 206 | 100.5360 | 1.13967 |
| 2006Q3 | 245 | 100.8273 | 0.93031 |
| 2006Q4 | 371 | 100.7255 | 0.83117 |
| 2007Q1 | 196 | 100.8469 | 0.88929 |
| 2007Q2 | 84 | 100.7776 | 0.68229 |
| total | 8308 | 100.8911 | 1.17313 |

## Table 7. Summary of OAS and Z-spread

This table summarizes the distribution of the OAS and the Z-spread in basis points for the 7,168 loans Fannie Mae purchased and without missing information from 1999 quarter 2 until 2007 quarter 2. The Z-spread is shown to be much higher than the OAS for a given loan, which is consistent with the Z-spread and OAS quotes that are commonly observed in the market.

| Value in bps | OAS | | Z-spread | |
|---|---|---|---|---|
| | Freq | Percentage | Freq. | Percentage |
| <0 | 2499 | 34.86 | 352 | 4.91 |
| 0-50 | 1995 | 27.83 | 632 | 8.82 |
| 50-100 | 1412 | 19.70 | 1355 | 18.90 |
| 100-max | 1262 | 17.61 | 4829 | 67.37 |
| Total | 7168 | 100 | 7168 | 100 |

## Table 8. Linear regression of Option-Adjusted Spread

This table tests if CRA features are significantly correlated with lower OAS in cross sectional sample of issued RMBS of CAP loans by simple linear regressions, using cross sectional purchased loan data. Interpretations of independent variables in panels A, B, and C are provided in Section 5-b.

| | | Model 1 | | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|---|
| | | **Final model** | | **Model use "Ineffect"** | | **"Pennington's Ineffect"** | |
| | | # obs | 7168 | # obs | 7168 | # obs | 7168 |
| | | Adj-$R^2$ | 0.1102 | Adj-$R^2$ | 0.1048 | Adj-$R^2$ | 0.1073 |
| | | RMSE | 89.419 | RMSE | 89.687 | RMSE | 89.561 |
| | | **Coeff** | **P>\|t\|** | **Coeff** | **P>\|t\|** | **Coeff** | **P>\|t\|** |
| | age_at_purchase | 0.16928 | 0.001 | 0.15683 | 0.003 | 0.17063 | 0.001 |
| **P** | OrigLTV | 80.20307 | 0.000 | 74.12736 | 0.000 | 74.66801 | 0.000 |
| **a** | OrigFICO_missing | 9.46382 | 0.053 | 11.75022 | 0.016 | 10.10587 | 0.039 |
| **n** | OrigFICO<=620 | 19.97848 | 0.000 | 20.34418 | 0.000 | 20.07853 | 0.000 |
| **e** | OrigFICO>=720 | -11.32277 | 0.000 | -11.05331 | 0.000 | -11.10977 | 0.000 |
| **l** | income>50%AMI | -4.69713 | 0.050 | -5.19671 | 0.031 | -4.94811 | 0.039 |
| **A** | Borrower AfriAmer | 9.12455 | 0.001 | 10.61787 | 0.000 | 10.15939 | 0.000 |
| | Borrower Hispanic | 4.93078 | 0.225 | 4.43456 | 0.283 | 3.90922 | 0.337 |
| | Borrower OthMinor | -4.61957 | 0.219 | -4.69156 | 0.213 | -4.51188 | 0.230 |
| | UPB in thousand | -0.17520 | 0.000 | -0.15939 | 0.000 | -0.16291 | 0.000 |
| | purchase in2007 | 10.77801 | 0.634 | 15.97662 | 0.481 | 13.02351 | 0.565 |
| **P** | purchase in2006 | 49.67040 | 0.000 | 51.18404 | 0.000 | 49.83857 | 0.000 |
| **a** | purchase in2004 | 15.34754 | 0.004 | 16.81652 | 0.002 | 16.36130 | 0.002 |
| **n** | purchase in2003 | 25.76140 | 0.000 | 24.77559 | 0.000 | 24.35734 | 0.000 |
| **e** | purchase in2002 | 63.82826 | 0.000 | 67.35736 | 0.000 | 65.71283 | 0.000 |
| **l** | purchase in2001 | -28.33827 | 0.000 | -23.07306 | 0.000 | -26.61027 | 0.000 |
| **B** | purchase in2000 | 7.57017 | 0.703 | 8.45095 | 0.671 | 8.27975 | 0.677 |
| | purchase in1999 | -8.94456 | 0.842 | -5.89270 | 0.896 | -7.40986 | 0.869 |
| | St wt market code1 | 31.27844 | 0.003 | 24.85475 | 0.018 | 31.21196 | 0.003 |
| **P** | St wt market code2 | 9.23873 | 0.124 | 2.25524 | 0.711 | 10.30582 | 0.094 |
| **a** | St wt market code4 | -13.73483 | 0.051 | -14.72863 | 0.038 | -20.58189 | 0.004 |
| **n** | St wt preppenal0 | -12.40584 | 0.172 | -2.73806 | 0.763 | -12.59870 | 0.172 |
| **e** | St wt preppenal1 | 6.27979 | 0.325 | 10.98734 | 0.119 | 2.82948 | 0.665 |
| **l** | St wt preppanal3 | 2.76517 | 0.658 | 8.42948 | 0.179 | -1.92440 | 0.772 |
| **C** | St wt preppenal4 | -3.18298 | 0.436 | -3.53873 | 0.409 | -3.64465 | 0.373 |
| | St repayabil | 2.48394 | 0.521 | -5.00654 | 0.379 | -1.67429 | 0.660 |
| | St ineffecttb | -38.68450 | 0.000 | | | | |
| | St ineffect | | | 3.06549 | 0.623 | | |
| | PenningtonIneffect | | | | | -24.49457 | 0.000 |
| | NC | 26.06751 | 0.000 | 23.92746 | 0.001 | 32.51978 | 0.000 |
| | CA | -8.66175 | 0.460 | -15.62976 | 0.212 | -8.42900 | 0.474 |
| | constant | -6.97040 | 0.554 | -38.27624 | 0.001 | -14.16332 | 0.238 |

## Table 9. Comparison of OAS and Z-spread regressions

This table provides comparison of regression results of the Z-spread using identical independent variables with OAS regression of Model 1 in Table 8.

| | | Model 1 | | Model 1 | |
|---|---|---|---|---|---|
| | | **OAS dependent Var** | | **Z-spread Dependent Var** | |
| | | # obs | 7168 | # obs | 7168 |
| | | Adj- $R^2$ | 0.1102 | Adj- $R^2$ | 0.3777 |
| | | RMSE | 89.419 | RMSE | 76.34 |
| | | **Coeff** | **P>\|t\|** | **Coeff** | **P>\|t\|** |
| | age_at_purchase | 0.16928 | 0.001 | 0.17245 | 0.000 |
| | OrigLTV | 80.20307 | 0.000 | 97.92960 | 0.000 |
| **P** | OrigFICO_missing | 9.46382 | 0.053 | 6.57893 | 0.115 |
| **a** | OrigFICO<=620 | 19.97848 | 0.000 | 14.58129 | 0.000 |
| **n** | OrigFICO>=720 | -11.32277 | 0.000 | -6.12147 | 0.005 |
| **e** | income>50%AMI | -4.69713 | 0.050 | -5.41552 | 0.008 |
| **l** | borrower AfriAmer | 9.12455 | 0.001 | 5.53149 | 0.024 |
| **A** | Borrower Hispanic | 4.93078 | 0.225 | 2.42556 | 0.484 |
| | Borrower OthMinor | -4.61957 | 0.219 | -3.89916 | 0.224 |
| | UPB in thousand | -0.17520 | 0.000 | -0.13618 | 0.000 |
| | purchase in2007 | 10.77801 | 0.634 | -19.26858 | 0.318 |
| **P** | purchase in2006 | 49.67040 | 0.000 | 0.06969 | 0.986 |
| **a** | purchase in2004 | 15.34754 | 0.004 | 108.40670 | 0.000 |
| **n** | purchase in2003 | 25.76140 | 0.000 | 118.55620 | 0.000 |
| **e** | purchase in2002 | 63.82826 | 0.000 | 148.57910 | 0.000 |
| **l** | purchase in2001 | -28.33827 | 0.000 | 66.49767 | 0.000 |
| **B** | purchase in2000 | 7.57017 | 0.703 | -10.23635 | 0.546 |
| | purchase in1999 | -8.94456 | 0.842 | 3.37294 | 0.930 |
| | St wt market code1 | 31.27844 | 0.003 | 13.21634 | 0.138 |
| | St wt market code2 | 9.23873 | 0.124 | 10.67207 | 0.037 |
| **P** | St wt market code4 | -13.73483 | 0.051 | 14.78508 | 0.014 |
| **a** | St wt preppenal0 | -12.40584 | 0.172 | -11.09236 | 0.153 |
| **n** | St wt preppenal1 | 6.27979 | 0.325 | -0.91417 | 0.867 |
| **e** | St wt preppanal3 | 2.76517 | 0.658 | 3.39930 | 0.524 |
| **l** | St wt preppenal4 | -3.18298 | 0.436 | 4.86943 | 0.162 |
| **C** | St repayabil | 2.48394 | 0.521 | -23.16562 | 0.000 |
| | St ineffecttb | -38.68450 | 0.000 | -7.66110 | 0.128 |
| | NC | 26.06751 | 0.000 | -7.71423 | 0.214 |
| | CA | -8.66175 | 0.460 | -10.25965 | 0.305 |
| | constant | -6.97040 | 0.554 | 8.81482 | 0.380 |

### Table 10. Percentage of loans bundled by purchase year

This table provides a summary of the percentage of loans that were bundled in each purchase year. A total of 5,917 (82.54%) of the 7,168 loans are bundled loans.

| Purchase year | # bundled | # total | % bundled |
|---|---|---|---|
| 1999 | 4 | 4 | 100.00% |
| 2000 | 3 | 21 | 14.29% |
| 2001 | 2334 | 2373 | 98.36% |
| 2002 | 164 | 257 | 63.81% |
| 2003 | 1086 | 1283 | 84.65% |
| 2004 | 159 | 437 | 36.38% |
| 2005 | 1794 | 2069 | 86.71% |
| 2006 | 366 | 709 | 51.62% |
| 2007 | 7 | 16 | 43.75% |
| total | 5917 | 7169 | 82.54% |

## Table 11. Comparison of OAS regression on unbundled loans

This table compares the result of the original OAS regression with repeating the linear regression of model 1 (in Table 8) on only unbundled loans, and only bundled loans. Interpretations of the results are provided in Section 5-c.

| | | Model 1 | | Model 1_Unbundled | | Model 1_bundled | |
|---|---|---|---|---|---|---|---|
| | | **Final model** | | **Unbundled 1,252** | | **Bundled 5,916** | |
| | | # obs | 7168 | # obs | 1252 | # obs | 5916 |
| | | Adj-$R^2$ | 0.1102 | Adj-$R^2$ | 0.0576 | Adj-$R^2$ | 0.1080 |
| | | RMSE | 89.419 | RMSE | 74.131 | RMSE | 92.106 |
| | | **Coeff** | **P>\|t\|** | **Coeff** | **P>\|t\|** | **Coeff** | **P>\|t\|** |
| | age_at_purchase | 0.16928 | 0.001 | 2.22991 | 0.000 | 0.18719 | 0.001 |
| | OrigLTV | 80.20307 | 0.000 | 32.56610 | 0.328 | 84.83821 | 0.000 |
| **P** | OrigFICO_missing | 9.46382 | 0.053 | 19.25725 | 0.071 | 6.33346 | 0.251 |
| **a** | OrigFICO<=620 | 19.97848 | 0.000 | 14.78869 | 0.021 | 21.44905 | 0.000 |
| **n** | OrigFICO>=720 | -11.32277 | 0.000 | -6.73292 | 0.189 | -11.98447 | 0.000 |
| **e** | income>50%AMI | -4.69713 | 0.050 | 2.35613 | 0.633 | -6.03545 | 0.026 |
| **l** | Borrower AfriAmer | 9.12455 | 0.001 | 11.96354 | 0.056 | 8.12851 | 0.012 |
| **A** | Borrower Hispanic | 4.93078 | 0.225 | 7.51358 | 0.456 | 3.64484 | 0.420 |
| | Borrower OthMinor | -4.61957 | 0.219 | -9.15828 | 0.345 | -2.49044 | 0.548 |
| | UPB in thousand | -0.17520 | 0.000 | -0.02609 | 0.720 | -0.21172 | 0.000 |
| | purchase in2007 | 10.77801 | 0.634 | 18.02840 | 0.475 | -11.30019 | 0.748 |
| **P** | purchase in2006 | 49.67040 | 0.000 | 33.39985 | 0.000 | 58.37655 | 0.000 |
| **a** | purchase in2004 | 15.34754 | 0.004 | 5.66478 | 0.382 | 22.43933 | 0.007 |
| **n** | purchase in2003 | 25.76140 | 0.000 | 19.57205 | 0.006 | 28.11463 | 0.000 |
| **e** | purchase in2002 | 63.82826 | 0.000 | 40.90469 | 0.000 | 76.54004 | 0.000 |
| **l** | purchase in 2001 | -28.33827 | 0.000 | 2.71130 | 0.834 | -27.49074 | 0.000 |
| **B** | purchase in 2000 | 7.57017 | 0.703 | 7.93048 | 0.663 | -18.88057 | 0.724 |
| | purchase in 1999 | -8.94456 | 0.842 | | | -6.88647 | 0.882 |
| | St wt market code1 | 31.27844 | 0.003 | -28.38463 | 0.545 | 32.01796 | 0.007 |
| | St wt market code2 | 9.23873 | 0.124 | 6.27297 | 0.587 | 8.58437 | 0.239 |
| **P** | St wt market code4 | -13.73483 | 0.051 | 9.21166 | 0.595 | -15.40021 | 0.052 |
| **a** | St wt preppenal0 | -12.40584 | 0.172 | -2.51130 | 0.886 | -15.10474 | 0.158 |
| **n** | St wt preppenal1 | 6.27979 | 0.325 | 2.60303 | 0.837 | 7.06430 | 0.395 |
| **e** | St wt preppanal3 | 2.76517 | 0.658 | 3.99883 | 0.771 | 1.43189 | 0.843 |
| **l** | St wt preppenal4 | -3.18298 | 0.436 | 11.31081 | 0.307 | -5.25317 | 0.266 |
| **C** | St repayabil | 2.48394 | 0.521 | -15.25242 | 0.186 | 2.50638 | 0.569 |
| | St ineffecttb | -38.68450 | 0.000 | -11.98559 | 0.402 | -40.20777 | 0.000 |
| | NC | 26.06751 | 0.000 | -18.75757 | 0.423 | 26.00976 | 0.002 |
| | CA | -8.66175 | 0.460 | | | -6.02563 | 0.656 |
| | constant | -6.97040 | 0.554 | 7.92250 | 0.819 | -7.04345 | 0.585 |

## Table 12. Comparison of OAS regression with bundling effect using dummies

This table tests the bundling effect on the whole sample by adding dummy variables on bundling using model 1, and compares the results with the original model 1 in Table 8. Variables in Penal B (bundling year) are interpreted in Section 5-c.

| Model 1_original | | | Model 1_dummy on bundling | | |
|---|---|---|---|---|---|
| # obs | 7168 | | # obs | 7168 | |
| Adj-$R^2$ | 0.1102 | | Adj-$R^2$ | 0.1089 | |
| RMSE | 89.419 | | RMSE | 89.481 | |
| | **Coeff** | **P>\|t\|** | | **Coeff** | **P>\|t\|** |
| **Panel A** | | | **Panel A** | | |
| age_at_purchase | 0.16928 | 0.001 | age_at_purchase | 0.20539 | 0.000 |
| OrigLTV | 80.20307 | 0.000 | OrigLTV | 78.60672 | 0.000 |
| OrigFICO_missing | 9.46382 | 0.053 | OrigFICO_missing | 9.04366 | 0.065 |
| OrigFICO<=620 | 19.97848 | 0.000 | OrigFICO<=620 | 21.01714 | 0.000 |
| OrigFICO>=720 | -11.32277 | 0.000 | OrigFICO>=720 | -11.17579 | 0.000 |
| income>50%AMI | -4.69713 | 0.050 | income>50%AMI | -4.29025 | 0.074 |
| Borrower AfriAmer | 9.12455 | 0.001 | Borrower AfriAmer | 9.22269 | 0.001 |
| Borrower Hispanic | 4.93078 | 0.225 | Borrower Hispanic | 3.84146 | 0.347 |
| Borrower OthMinor | -4.61957 | 0.219 | Borrower OthMino | -2.83178 | 0.454 |
| UPB in thousand | -0.17520 | 0.000 | UPB in thousand | -0.17593 | 0.000 |
| **Panel B** | | | **Panel B (Bundling year)** | | |
| purchase in2007 | 10.77801 | 0.634 | | | |
| purchase in2006 | 49.67040 | 0.000 | is bundled in 2006 | 29.89418 | 0.000 |
| purchase in2004 | 15.34754 | 0.004 | is bundled in 2005 | -29.83227 | 0.000 |
| purchase in2003 | 25.76140 | 0.000 | is bundled in 2004 | -5.13506 | 0.500 |
| purchase in2002 | 63.82826 | 0.000 | is bundled in 2003 | -0.94741 | 0.841 |
| purchase in2001 | -28.33827 | 0.000 | is bundled in 2002 | 48.29923 | 0.000 |
| purchase in2000 | 7.57017 | 0.703 | is bundled in 2001 | -55.39627 | 0.000 |
| purchase in1999 | -8.94456 | 0.842 | | | |
| **Panel C** | | | **Panel C** | | |
| St wt market code1 | 31.27844 | 0.003 | St wt market code1 | 30.76564 | 0.003 |
| St wt market code2 | 9.23873 | 0.124 | St wt market code2 | 8.58737 | 0.156 |
| St wt market code4 | -13.73483 | 0.051 | St wt market code4 | -12.00328 | 0.091 |
| St wt preppenal0 | -12.40584 | 0.172 | St wt preppenal0 | -12.77882 | 0.160 |
| St wt preppenal1 | 6.27979 | 0.325 | St wt preppenal1 | 6.38422 | 0.329 |
| St wt preppanal3 | 2.76517 | 0.658 | St wt preppanal3 | 1.44142 | 0.818 |
| St wt preppenal4 | -3.18298 | 0.436 | St wt preppenal4 | -2.95820 | 0.474 |
| St repayabil | 2.48394 | 0.521 | St repayabil | 2.53612 | 0.516 |
| St ineffecttb | -38.68450 | 0.000 | St ineffecttb | -38.12976 | 0.000 |
| NC | 26.06751 | 0.000 | NC | 22.99158 | 0.002 |
| CA | -8.66175 | 0.460 | CA | -6.10920 | 0.603 |
| constant | -6.97040 | 0.554 | constant | 19.47696 | 0.118 |

# Appendix B

## Table 13. Summary descriptive statistics of variables in MNL

This table provides summary descriptive statistics[31] of variables used in MNL regressions in Table 5. These MNL regressions use all loans in the CAP portfolio from 1998 until July 2008.

| Variable | Freq | Mean | Std. Dev | Min | Max |
|---|---|---|---|---|---|
| ***Seasoning*** | | | | | |
| Age1 in [0,3] | 62501 | 0.91847 | 1.20408 | 0 | 3 |
| Age2 in (3,6] | 41164 | 5.04778 | 0.81565 | 4 | 6 |
| Age3 in (6,9] | 51388 | 8.05036 | 0.81587 | 7 | 9 |
| Age4 in (9,12] | 61412 | 11.02959 | 0.81562 | 10 | 12 |
| Age5 in (12,18] | 135515 | 15.53056 | 1.70814 | 13 | 18 |
| Age6 in (18,24] | 144244 | 21.53112 | 1.70709 | 19 | 24 |
| Age7 in (24,30] | 145147 | 27.47694 | 1.70678 | 25 | 30 |
| Age8 in (30,40] | 220333 | 35.40597 | 2.87244 | 31 | 40 |
| Age9 in (40,50] | 192796 | 45.35701 | 2.86418 | 41 | 50 |
| Age10 in (50,60] | 153695 | 55.27799 | 2.85974 | 51 | 60 |
| Age11 in (60,90] | 279923 | 73.48826 | 8.46875 | 61 | 90 |
| Age12 in (90, 290] | 193924 | 116.28840 | 22.05668 | 91 | 290 |
| ***Fico Score Effect*** | | | | | |
| creditscore/100 | | | | | |
| ( , 5.8] | 253594 | 5.24219 | 0.62659 | 0.04 | 5.80 |
| (5.8, 6.2] | 190321 | 6.01422 | 0.11653 | 5.80 | 6.20 |
| (6.2, 6.6] | 271401 | 6.40781 | 0.11473 | 6.20 | 6.60 |
| (6.6, 7.2] | 422138 | 6.89406 | 0.17159 | 6.60 | 7.20 |
| (7.2, 8.5] | 500724 | 7.62121 | 0.25817 | 7.20 | 8.50 |
| ***UPB Effect*** | | | | | |
| upb/1000 | | | | | |
| (0, 50] | 419933 | 37.11486 | 9.91361 | 0.01 | 50.00 |
| (50, 75] | 547771 | 62.41088 | 7.15953 | 50.00 | 75.00 |
| (75, 100] | 362405 | 85.85227 | 6.97727 | 75.00 | 100.00 |
| (100, 150] | 253950 | 118.85700 | 13.15955 | 100.00 | 150.00 |
| (150, 407.666] | 64201 | 183.58540 | 33.85787 | 150.00 | 407.33 |
| ***Yield Curve Slope*** | | | | | |
| Tbill10yr-- 2 yr | 1715864 | 1.05913 | 0.94791 | -0.41 | 2.59 |
| ***Mark-to-market LTV*** | | | | | |
| mtmltv | 1714374 | 82.95535 | 17.98869 | 1.90 | 187.44 |

[31] The some of the variables are transformed into spline knot format in the MNL regression.

**Table 13. Summary descriptive statistics of variables in MNL (Cont'd)**

| Variable | Freq | Mean | Std. Dev | Min | Max |
|---|---|---|---|---|---|
| ***Refi Burnout factor*** | | | | | |
| refi_spread | 1715864 | 1.11091 | 0.15125 | 0.77 | 2.58 |
| burn_out spread | 1715864 | 0.84232 | 2.19670 | 0.00 | 46.11 |
| refi1 in (0,1] | 471690 | 0.93696 | 0.05018 | 0.77 | 1.00 |
| refi2 in (1,1.1] | 381122 | 1.04986 | 0.02897 | 1.00 | 1.10 |
| refi3 in (1.1, 1.2] | 399667 | 1.14883 | 0.02828 | 1.10 | 1.20 |
| refi4 in (1.2, 1.3] | 260426 | 1.24315 | 0.02826 | 1.20 | 1.30 |
| refi5 in (1.3, 1.4] | 134796 | 1.34441 | 0.02785 | 1.30 | 1.40 |
| refi6 in (1.4, max] | 68163 | 1.46659 | 0.06837 | 1.40 | 2.58 |
| burn1 in (0, 0.2] | 1150654 | 0.01506 | 0.03943 | 0.00 | 0.20 |
| burn2 in (0.2, 0.7] | 205311 | 0.40815 | 0.14169 | 0.20 | 0.70 |
| burn3 in (0.7, 1.2] | 89597 | 0.90472 | 0.13762 | 0.70 | 1.20 |
| burn4 in (1.2, 1.7] | 49212 | 1.41497 | 0.14322 | 1.20 | 1.70 |
| burn5 in (1.7, max] | 221090 | 5.39821 | 3.58384 | 1.70 | 46.11 |
| ***Borrower, Neighborhood and Loan characteristics*** | | | | | |
| back end ratio | 1600117 | 1.18016 | 32.27178 | 0 | 2317 |
| annual inc as %AMI | 1627255 | 62.93549 | 52.12175 | 0 | 5504 |
| median track inc as %AMI | 1711390 | 0.91887 | 0.26562 | 0 | 3.65 |
| orig coupon- market PMMS | 1715788 | 0.27563 | 0.66435 | -7.13 | 3.41 |

## Table 13. Summary descriptive statistics of variables in MNL (Cont'd)

| Categorical Variables | Freq | % of sample |
|---|---|---|
| ***Seasonality*** | | |
| Jan | 145138 | 8.45860 |
| Feb | 144421 | 8.41681 |
| March | 144397 | 8.41541 |
| April | 148762 | 8.66980 |
| May | 150023 | 8.74329 |
| June | 152367 | 8.87990 |
| July | 135132 | 7.87545 |
| Aug | 136624 | 7.96240 |
| Sept | 138247 | 8.05699 |
| Oct | 137126 | 7.99166 |
| Nov | 135842 | 7.91683 |
| Dec | 147785 | 8.61286 |
| ***Origination Cohort*** | | |
| 1995 | 143760 | 8.37829 |
| 1996 | 72901 | 4.24865 |
| 1997 | 163414 | 9.52372 |
| 1998 | 211468 | 12.32429 |
| 1999 | 132070 | 7.69700 |
| 2000 | 194759 | 11.35049 |
| 2001 | 233222 | 13.59210 |
| 2002 | 174778 | 10.18601 |
| 2003 | 121278 | 7.06804 |
| 2004 | 107738 | 6.27894 |
| 2005 | 71054 | 4.14100 |
| 2006 | 38516 | 2.24470 |
| 2007 | 15679 | 0.91377 |

## Table 13. Summary descriptive statistics of variables in MNL (Cont'd)

| Categorical Variables | Freq | % of sample |
|---|---|---|
| ***Borrower, Neighborhood and Loan characteristics*** | | |
| is low_to_mod inc track | 549741 | 32.03873 |
| is minority track | 518096 | 30.19447 |
| is underserved area | 1047828 | 61.06708 |
| is worst ever delin 30 days | 328104 | 19.12180 |
| is worst ever delin above 30 days | 256846 | 14.96890 |
| is african american | 419533 | 24.45025 |
| is hispanic | 196389 | 11.44549 |
| is female | 759842 | 44.28335 |
| is rural | 293481 | 17.10398 |
| is NC | 733859 | 42.76907 |
| is CA | 98226 | 5.72458 |
| is FL | 71674 | 4.17714 |
| is OH | 113885 | 6.63718 |
| is OK | 106797 | 6.22409 |

### Table 14. Summary descriptive statistics of variables in linear regression

This table provides summary descriptive statistics of variables used in linear regressions in Table 8-12. These linear regressions use cross sectional sample of loans with prices that Fannie Mae paid for securitized CAP loans from 1999 quarter 2 until 2007 quarter 2.

| Variable | Freq | Mean | Std.Dev. | Min | Max |
|---|---|---|---|---|---|
| **Panel A** | | | | | |
| age_at_purchase | 7169 | 22.01479 | 28.44360 | 0 | 190 |
| OrigLTV | 7169 | 0.93547 | 0.12280 | 0.12 | 1.24 |
| UPB in thousand | 7169 | 80.70805 | 42.68700 | 3.21 | 403.01 |
| Categorical variables | Freq | % of sample | | | |
| **Panel A** | | | | | |
| OrigFICO_missing | 477 | 6.65365 | | | |
| OrigFICO<=620 | 1192 | 16.62714 | | | |
| OrigFICO>=720 | 2048 | 28.56744 | | | |
| income>50%AMI | 4720 | 65.83903 | | | |
| Borrower AfriAmer | 1700 | 23.71321 | | | |
| Borrower Hispanic | 651 | 9.08076 | | | |
| Borrower OthMinori | 928 | 12.94462 | | | |
| **Panel B** | | | | | |
| purchase in2007 | 16 | 0.22318 | | | |
| purchase in2006 | 709 | 9.88980 | | | |
| purchase in2004 | 437 | 6.09569 | | | |
| purchase in2003 | 1283 | 17.89650 | | | |
| purchase in2002 | 257 | 3.58488 | | | |
| purchase in2001 | 2373 | 33.10085 | | | |
| purchase in2000 | 21 | 0.29293 | | | |
| purchase in1999 | 4 | 0.05580 | | | |

### Table 14. Summary descriptive statistics of variables in linear regression (Cont'd)

| Categorical variables | Freq | % of sample |
|---|---|---|
| **Panel C** | | |
| St wt market code1 | 723 | 10.08509 |
| St wt market code2 | 1655 | 23.08551 |
| St wt market code4 | 1877 | 26.18217 |
| St wt preppenal0 | 282 | 3.93360 |
| St wt preppenal1 | 1249 | 17.42223 |
| St wt preppanal3 | 989 | 13.79551 |
| St wt preppenal4 | 1970 | 27.47943 |
| St repayabil | 3326 | 46.39420 |
| Pennington's ineffect | 6675 | 93.10922 |
| St ineffecttb | 6868 | 95.80137 |
| St ineffect | 4182 | 58.33450 |
| is NC | 1497 | 20.88157 |
| is CA | 526 | 7.33715 |
| **Bundling year** | | |
| is bundled in 2006 | 366 | 5.10531 |
| is bundled in 2005 | 1794 | 25.02441 |
| is bundled in 2004 | 159 | 2.21788 |
| is bundled in 2003 | 1086 | 15.14856 |
| is bundled in 2002 | 164 | 2.28763 |
| is bundled in 2001 | 2334 | 32.55684 |

# References

Ashcraft, Adam. Goldsmith-Pinkham, Paul. Vickery, James (2010). "MBS ratings and the mortgage credit boom." February 12, 2009. Available at http://www.econ.rutgers.edu/dmdocuments/AdamAshcraft.pdf

Beardsell, Mark A. and Liu, Wei. (2005) "Measuring Loan-Level Option Cost for Sub-Prime and Alternative-A Fixed Rate Mortgages" American Real Estate Society 2005 Annual Meeting. Both authors are from Citigroup Consumer Analytics & Modeling Unit at that time.

Ben-Shahar Danny. (2006) "Screening Mortgage Default Risk: A Unified Theoretical Framework." *Journal of Real Estate Research* by American Real Estate Society. 28 (2006) Issue 3. pp.215-240. Available at http://ideas.repec.org/a/jre/issued/v28n32006p215-240.html

Belsky, Eric S. (2008). "Housing wealth effects and the course of the US economy: theory, evidence and policy implications." Nov. 2008. W08-7. Joint Center for Housing Studies, Harvard University.

Bernanke, Ben S.(2009), " US federal reserve chairman Ben at the federal serve system's sixth biennial community affairs research conference." Washington, D.C., April 17, 2009. Financial innovation and consumer protection.

Breeden, Douglas T. (1997), "Convexity and empirical option costs of mortgage securities". March 1997. The Journal of Fixed Income.

Brigo, Damiano and Mercurio, Fabio. (2006). Interest Rate Models - Theory and Practice. With Smile, Inflation and Credit. Springer Finance. 2nd ed. 2006, LVI, 981 p. 124 illus., SBN: 978-3-540-22149-4.

Brigo, Damiano. Pallavicini, Andrea. and Torresetti, Roberto (2010). "Credit Models and the Crisis, or How I learned to stop worrying and love the CDOs." Working paper February 2010, available at http://ideas.repec.org/p/arx/papers/0912.5427.html. Also see: Credit Models and the Crisis: A journey into CDOs, Copulas, Correlations, and Dynamic Models. Wiley. Chichester, 2010.

Clapp, John M., Deng, Yongheng and An, Xudong, (2005). "Unobserved Heterogeneity in Models of Competing Mortgage Termination" (August 22, 2005). Available at SSRN: http://ssrn.com/abstract=512624

Collin-Dufresne. P. Goldstein, R.S., 2002. "Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. Journal of Finance 57, 1685-1730.

Cotterman, Robert F. "Neighborhood Effects in Mortgage Default Risk." (March 2001). Prepared for HUD office of Policy Development and Research. Unicon research corporation. Available at http://www.huduser.org/Publications/PDF/default_full.pdf

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985): "A Theory of the Term Structure of Interest Rates,".Econometrica, 53, 385-467.

Cox, D. (1972) "Regression Models and Life Tables." Journal of the Royal Statistical Society, Series B, 34,1972, pp.187-220.

Danis, Michelle A. and Pennington-Cross, Anthony N.,(2005) "The Delinquency of Subprime Mortgages" (March 2005). FRB of St. Louis Working Paper No. 2005-022A Available at SSRN: http://ssrn.com/abstract=761804

Danis, Michelle A. and Pennington-Cross, Anthony N., (2005)"A Dynamic Look at Subprime Loan Performance" (May 2005). FRB St. Louis Working Paper No. 2005-029A Available at SSRN: http://ssrn.com/abstract=761152

Dai, Qiang and Singleton, Kenneth J. (2000) "Specification Analysis of Affine Term Structure Models". The Journal of Finance, • Vol. LV, No. 5 • Oct. 2000.

Deng, Y., J. M. Quigley and R. Van Order. (2000). "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options". Econometrica 68(2): 275-307.

Ding, Lei. Quercia, Roberto G. Lei, Wei, Ratcliffe, Janneke. (2008) "Risky Borrowers or Risky Mortgages: Disaggregating Effects Using Propensity Score Models." Working paper, Center for Community Capital, UNC-Chapel Hill.

Ding, Lei. Quercia, Roberto and White, Alan (2009). "State anti-predatory lending laws: impact and federal preemption phase I descriptive analysis." Research report. Oct 5, 2009. Center for Community Capital, UNC-Chapel Hill.

Ding, Lei, and Quercia, Roberto (2009). "Tailoring loan modifications: when is principal reduction desirable?" Working paper, Center for Community Capital, UNC-Chapel Hill.

Duffee, Gregory R.(2002) "Term premia and interest rate forecasts in Affine Models", Journal of Finance 57 (2002), pp. 405–443.

Duffie, Darrell, and Kenneth Singleton, (1999). "Modeling term structure of defaultable bonds", Review of Financial Studies 12, 1999, 687–720.

Dunsky. R. and A. Pennington-Cross. 2004. "Estimation, development and back testing of default and prepayment equations." Working paper.

Dunsky, Rovert M. and Ho, Thomas S. (2007) "Valuing fixed rate mortgage loans with default and prepayment options". Spring 2007. The Journal of Fixed Income.

Egorov, Alexei V. Hong, Yongmiao, and Li, Haitao (2006). "Validating forecasts of the joint probability density of bond yields: Can affine models beat random walk?" Journal of Econometrics, Vol 135. Issue 1-2, Nov-Dec 2006. pp. 255-284.

Fabozzi, Frank J. The Handbook of Mortgage Backed Securities.

Foote, Christopher L. Gerardi, Kristopher. Goette, Lorenz. Willen, Paul S. (2008). "Just the facts: An initial analysis of subprime's role in the housing crisis." Journal of Housing Economics. 17 2008, pp. 291-305.

Gupta, Anurag and Subrahmanyam, Marti G., (2005), Pricing and hedging interest rate options: Evidence from cap-floor markets, Journal of Banking & Finance, 29, issue 3, p. 701-733,

Jarrow, Robert. Li, Haitao, and Zhao, Feng. (2007). "Interest Rate Caps Smile Too! But Can the LIBOR Market Models Capture It?" Journal of Finance. Vol 62, Issue 1. pp. 345-382.

Hayre, Lakhbir. (1999), “Guide to Mortgage-backed Securities”, Fixed income research. Salomon Smith Barney. March 1999.

Hull, J., White, A.(1990a). "Valuing Derivative Securities Using the Explicit Finite Difference Method." Journal of Financial and Quantitative Analysis 25, 87-100.

Hull, J., White, A.(1990b) "Pricing Interest Rate Derivative Securities." The Review of Financial Studies 3, 573-592

Hull, J., White, A.(1993d) "One-factor Interest Rate Models and the Valuation of Interest Rate Derivative Securities." Journal of Finance and Quantitative Analysis 28, 235-254.

Hull, J., White, A.(1994a) Branching out. Risk 7, 34-37.

Hull, J., White, A.(1995b) "A Note on the Models of Hull and White for Pricing Options on the Term Structure: Response." The Journal of Fixed Income 5. (September, 1995), 97-102.

Hull, J., White, A.(2004). "Valuation of a CDO and an nth to Default CDS Without Monte Carlo

Simulation." Journal of Derivatives, 12,2, Winter 2004.

Hunt, Bill. (2007) "ResiLogic: U.S. Residential Mortgage Loss Model Technical Document." Quantitative Finance Research Special Report. Fitch Ratings. January 18, 2007. Available at http://www.securitization.net/pdf/Fitch/ResiLogic_18Jan07.pdf

Laderman, E., and C. Reid. (2009). "CRA Lending During the Subprime Meltdown." In Revisiting the CRA: Perspectives on the Future of the Community Reinvestment Act, pp. 115-133. A Joint Publication of the Federal Reserve Banks of Boston and San Francisco. http://www.frbsf.org/publications/community/cra/cra_lending_during_subprime_meltdown.pdf [accessed March 1, 2009]

Lo, A. and MacKinlay, A.C. 1989. "Data-snooping biases in tests of financial asset pricing models" , Review of Financial Studies 3 (1989), pp. 175–208.

Pan, Jun and Singleton, Kenneth J. (2008). "Default and Recovery Implicit in the Term Structure of Sovereign CDS Spreads", The Journal of Finance. Vol LXIII, No. 5, Oct. 2008, 2345-2384.

Pedersen, Claus M. (2006). "Explaining the Lehman Brothers Option-Adjusted Spread of a Corporate Bond. " February 27, 2006. Fixed Income Quantitative Credit Research. Lehman Brothers.

Pennington-Cross, Anthony. (2010). "The duration of foreclosures in the subprime mortgage market: a competing risks model with mixing" Journal of Real Estate Finance and Economics, Springer, vol 40 (2), pages 109-129, February.

Pennington-Cross, Anthony. (2003). "Credit History and the Performance of Prime and Non-prime Mortgages." Journal of Real Estate Finance and Economics, 27:3, 279-301.

Pennington-Cross, A., S. Chomsisengphet, R. Bostic, K. C. Engel, P. A. McCoy, and S. Wachter. (2008). "Mortgage Product Substitution and State Anti-Predatory Lending Laws: Better Loans and Better Borrowers?" U of Penn, Inst for Law & Econ Research Paper No. 09-27.

Quercia, Roberto and Ding, Lei (2009). "Tailoring loan modifications: when is principal reduction desirable?" working paper on Aug 14, 2009. Center for Community Capital, UNC-Chapel Hill.

Quercia Roberto, and Stegman, Michael (1992). "Residential mortgage default: a review of the literature". Journal of Housing Research. Vol 3. Issue 2. pp.341-379.

Riley, Sarah. Freeman, Allison. and Quercia, Roberto (2009). "Navigating the housing downturn and financial crisis: Home appreciation and equity accumulation among community reinvestment homeowners." working paper Nov. 2009. Center for Community Capital, UNC-Chapel Hill.

Riley, Sarah. Ru, HongYu, and Quercia, Roberto (2009). "The Community Advantage Program Database: Overview and Comparison with the Current Population Survey." Cityscape 11(3):247-255.

Riley, Sarah and Quercia, Roberto (2011). "Navigating the Housing Downturn and Financial Crisis: Home Appreciation and Equity Accumulation among Community Reinvestment Homeowners." In: Susan Wachter and Marty Smith (eds), Reinventing the American Mortgage System: Rethink, Recover, Rebuild. University of Pennsylvania Press.

Stegman, Michael. "Remarks before the House Financial Services Committee: 'Proposals to Enhance the Community Reinvestment Act'". Available at http://www.house.gov/apps/list/hearing/financialsvcs_dem/stegman.pdf

Stegman, Michael. "Evolution of Banking & Access to Financial Services in the U.S", April 2004; "A personal perspective on the recapitalization of communities", May 2004; "Creating Community Wealth." Net Inpact Conference Nov 3, 2001; "GAO planning conference: emerging issues in financial markets & community investment. Community assets panel." Feb 2001. All available at http://www.ccc.unc.edu/research.php.

Stegman, Michael A. Freeman, Allison. Paik, Jong-Gyu, (2007). "The portfolios and wealth of low-income homeowners and renters: findings from an evaluation of Self-Help Ventures Fundâ€™s Community Advantage Program," Community Development Investment Center Working Paper 2007-02, Federal Reserve Bank of San Francisco.

Sundaresan, Suresh. M. (2000), "Continuous-Time Methods in Finance: A Review and an Assessment." The Journal of Finance, 2000, 55: 1569–1622.

Thomas. Ho and S. Lee. (1986). "Term structure movements and pricing interest rate contingent claims." Journal of Finance, 42:11291142, 1986

Thomas. Ho and S. Lee. (2004). The Oxford Guide to Financial Modeling, Applications for Capital Market, Corporate Finance, Risk Management, and Financial Institutions. Oxford University Press. 2004.

White, H. 2000. "A reality check for data snooping", Econometrica 69 (2000), pp. 1097–1127.

Van Order, Robert, and Peter Zorn. (2004). "The Performance of Low Income and Minority Mortgages: A Tale of Two Options," Freddie Mac Working Paper.