Journal of Banking & Finance 33 (2009) 788–799

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Loss given default of high loan-to-value residential mortgages

Min Qi *, Xiaolong Yang

Office of the Comptroller of the Currency, 250 E Street S.W., Washington, DC 20219, USA

ARTICLE INFO

Article history:
Received 25 January 2008
Accepted 8 September 2008
Available online 21 September 2008

JEL classification:
G21
G28

Keywords:
Loss given default
Residential mortgage
Default
Recovery
Downturn
Basel II

ABSTRACT

This paper studies loss given default using a large set of historical loan-level default and recovery data of high loan-to-value residential mortgages from several private mortgage insurance companies. We show that loss given default can largely be explained by various characteristics associated with the loan, the underlying property, and the default, foreclosure, and settlement process. We find that the current loan-to-value ratio is the single most important determinant. More importantly, mortgage loss severity in distressed housing markets is significantly higher than under normal housing market conditions. These findings have important policy implications for several key issues in Basel II implementation.

Published by Elsevier B.V.

1. Introduction

Under the new Basel II capital framework,[1] the calculation of minimum regulatory capital under the advanced internal rating-based (A-IRB) approach requires accurate estimation of parameters that determine the credit risk of banks' financial asset portfolios: probability of default (PD), loss given default (LGD), and exposure at default (EAD).[2] While there has been a growing body of research relevant to the modeling and estimation of PD, there are few studies on LGD (or loss severity, which is equal to 1—the recovery rate) to date, but the number has been increasing rapidly.[3]

The growing literature on LGD has covered several areas, including defining and measuring LGD and the correlation between PD and LGD, both theoretically and empirically. The existing literature has also studied various factors that affect LGD. These include: (1) contract characteristics—seniority and security, credit facility type (loan, bond), term or revolving facility, covenant protection, collateral (type, appraisal date, and results); (2) borrower characteristics—profit margin, debt cushion, leverage; (3) differences across industry and industry conditions; and (4) macroeconomic systematic risk factors. Cyclical effects on LGD are also examined, and LGD during economic downturn periods has been compared to LGD under normal economic conditions. Lastly, research has been carried out to investigate the statistical distribution of LGD.

However, the vast majority of these LGD studies are on wholesale exposures, such as corporate bonds and loans (see, for example, Dermine and Neto de Carvalho, 2006 and the references therein). Partly because of the unavailability of public data, very few studies have been done on retail exposures. In a theoretical credit risk model for large dimensional portfolios such as retail and mortgages, Nyström and Skoglund (2006) moved away from the traditional assumption of fixed recovery rate to models of recovery rate with stochastic collaterals, but their recovery rate models are not calibrated to real data. Clauretie and Herzog (1990) study the effect of state foreclosure laws (judicial procedure, statutory right of redemption, and deficiency judgment) on

* Corresponding author. Tel.: +1 202 874 4061; fax: +1 202 874 5394.
E-mail addresses: min.qi@occ.treas.gov (M. Qi), xiaolong.yang@occ.treas.gov (X. Yang).

[1] International Convergence of Capital Measurement and Capital Standards: A Revised Framework, June 2006, Basel Committee on Banking Supervision. VanHoose (2007) reviews the literature on bank behavior under capital regulation to evaluate the intellectual underpinning for the proposed Basel II system and to assess its effects on bank lending, loan rates, leverage ratio, asset risk, and overall safety and soundness of the banking system.

[2] Effective maturity (M) is also needed for corporate, sovereign, and bank exposures.

[3] Altman et al. (2005a) provide a comprehensive survey of literature on default recovery rates for corporate credit risk. Altman et al. (2005b) contain a collection of papers on recovery risk. Qi (2005) surveys research on LGD in stressed market conditions. In general the estimate of the recovery rates is not easy due to insufficient sample data (Abaffy et al., 2007).

0378-4266/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.jbankfin.2008.09.010

loan losses for mortgages insured privately (i.e., private mortgage insurance (PMI)) and by government (e.g., Federal Housing Administration (FHA)). They find that judicial procedure and statutory right of redemption extend the foreclosure and liquidation processes and thus are associated with larger loan losses. They also show that deficiency judgment reduces loss severity for PMI that has no incentive conflict due to its coinsurance feature, while deficiency judgment has no significant impact on the recovery rate for FHA insurance, with which incentive conflict arises due to the lack of a coinsurance arrangement. Lekkas et al. (1993) empirically test the frictionless form of the options-based mortgage default theory. They find that higher initial loan-to-value (LTV) ratios, regions with higher default rates (Texas), and younger loans are associated with significantly higher loss severities whereas the difference between contract and current interest rates has no impact on loss severities; consequently, they reject the propositions about loss severity implied by the frictionless form of the options-based mortgage default theory. Crawford and Rosenblatt (1995) extend options-based mortgage default theory to include transaction costs and show theoretically and empirically the effect of frictions on the individual strike price that affects loss severity.

The regression analysis in the above three studies can explain only a small portion of the total variations in loan-level mortgage LGD (R^2 ranges from 0.02 to 0.14).[4] More recently, Pennington-Cross (2003) and Calem and LaCour-Little (2004) study determinants of mortgage loss severity based on government-sponsored enterprise (GSE) data, and their regression analysis shows improved explanatory power. The R^2 reported in Calem and LaCour-Little is 0.25, whereas it is 0.95–0.96 in Pennington-Cross (2003). Although the latter study reports very high R^2, it uses a much smaller sample and covers a shorter sample period (1995–1999) that contains no serious housing market depreciation.[5] Coupled with the problems in LGD definition and the timing of the current loan-to-value (CLTV) calculation, the findings of Pennington-Cross (2003) should be interpreted with caution.

Overall the existing studies have found that CLTV or LTV are strongly related to recovery rates (Calem and LaCour-Little, 2004; Pennington-Cross, 2003; Lekkas et al., 1993; Clauretie and Herzog, 1990). The age and size of the loan have also been shown to affect mortgage recovery rates (Calem and LaCour-Little, 2004; Pennington-Cross, 2003; Lekkas et al., 1993). In addition, recovery rates are found to vary with state foreclosure laws (Pennington-Cross, 2003; Clauretie and Herzog, 1990), prime or subprime mortgages (Pennington-Cross, 2003), and the relative median income (Calem and LaCour-Little, 2004). These studies are summarized in Appendix A.

The existing residential mortgage LGD studies, however, have not paid sufficient attention to how LGD would change under housing market downturn conditions, partly because of the lack of reliable mortgage loss data through a complete housing market cycle. The only study we are aware of that quantifies the expected and economic downturn LGD relationship is Calem (2003). However, his results are based on simulated mortgage defaults of a conforming-size residential mortgage portfolio that is hypothetical and geographically diversified. It is not clear whether the same relationship would still hold if actual loan-level loss experiences were used.

In recent years, retail loans have surpassed wholesale loans in dollar amount and have accounted for the largest proportion in total assets among all commercial banks in the US. Furthermore, residential mortgage is now the largest share of aggregate retail loans of all US commercial banks. As of June 2006, the total retail and wholesale loans are around \$2.66 trillion and \$2.42 trillion, respectively, for all commercial banks. Residential mortgages account for 52% of all commercial banks as of June 2006.[6] Given their prominent position in banks' portfolios, retail LGD in general and mortgage LGD in particular have obviously been understudied in the existing literature. The present research intends to fill this gap.

In this paper, we study residential mortgage loss given default using a large set of historical loan-level default and recovery data of high-LTV mortgages from several private mortgage insurance companies. We show that LGD can be largely explained by various characteristics associated with the loan, the underlying property, as well as the default, foreclosure, and settlement process. As expected, CLTV is the single most important determinant. More importantly, mortgage loss severity in distressed housing markets is significantly higher than under normal housing market conditions.

Our study differs from the existing mortgage loss severity studies in several important ways. First, compared to the existing studies on mortgage loss given default, our LGD definition is more comprehensive and closer to the Basel II definition. Besides the unpaid balance and the recovery amount, we also include the accrued interest, foreclosure expenses (legal and courts), property maintenance expenses, sales costs, and repairs. Most importantly, all cash flows (positive or negative) are properly adjusted and discounted to the time of default. Second, we use a unique data set that has the most observations and covers a long period that contains a complete housing market cycle, at least for the New England and the Pacific regions. It allows us to be the first to explicitly and empirically model economic downturn LGD for residential mortgages. Third, our data also contain the most comprehensive information for each defaulted mortgage, making it possible to include more explanatory variables and to explain loss given default better than most of the existing studies. Finally, most of the existing loan-level studies use conforming GSE mortgages of usual LTV ratios, whereas our sample consists largely of high-LTV, PMI-insured mortgages.

This paper has several important policy implications for several key issues in Basel II implementation. First, although LTV at time of loan origination can be used to segment risk, updated LTV (or CLTV) dramatically improves risk segmentation. Second, the LGD mapping function specified in the proposed US. Basel II rules reflects stress effects that are generally greater than what our sample and analysis suggest but is nevertheless appropriate. Finally, after considering mortgage insurance payment, the 10% LGD floor imposed by the US and international Basel II rules for residential mortgage exposures is binding when applied to the average LGD in the MICA sample. However, it becomes non-binding if applied to downturn LGD.

The rest of the paper is organized as the follows. In Section 2, we describe in greater detail the mortgage claim data set that is used in this research. In Section 3, we compare average mortgage loss severity across time, geographic regions, and CLTV ranges. Results of regression analysis are reported in Section 4. Section 5 addresses the implications of our findings on risk-based capital. Conclusions are provided in Section 6.

2. Data and descriptive statistics

We use a large and geographically diverse individual loan-level mortgage default and recovery data set from several major private mortgage insurance companies. The data set was compiled by the Mortgage Insurance Companies of America (MICA), the trade asso-

[4] The adjusted R^2 of 0.56–0.57 reported in Clauretie and Herzog (1990) is from regressions at the state level, not at the loan level.

[5] The sample average LGD in Pennington-Cross (2003) is only 2.1%.

[6] *Source*: "Financial Performance of National Banks", OCC Quarterly Journal 25(3), September 2006.

ciation of the private mortgage insurance industry.[7] Traditionally, lenders have required a down payment of at least 20% of a home's value. PMI expands homeownership opportunities by enabling home buyers to purchase homes with as little as a 3–5% down payment for qualified borrowers. PMI is basically the private sector alternative to FHA and Veterans Affairs (VA) insurance. Unlike FHA, PMI companies do not insure the total loan balance. The mortgage insurance industry shares the risk of default with the financial institution and the secondary market investor. Sharing the risk provides incentive for all parties to keep the loan payments current. In addition, PMI generally costs less than FHA insurance and is available on a wider variety of mortgage loan products, and it is not subject to maximum loan amounts.

Volumes of business for the private and public sectors are cyclical and rise and fall independently of each other. As of 2005, FHA loans made up 19% of the total insured loan dollar volume, VA loans 8%, and MICA member loans 73%. Of the total number of insured loan originations, FHA loans made up 23%, VA loans 7%, and MICA member loans 70%.[8] As of 2005, the total dollar volume of insurance contracts in force of MICA members was $615 billion, which represents roughly 10% of total mortgage outstanding at the time.

The complete data set consists of 241,293 mortgage insurance claims that were settled between 1990 and 2003. It contains information about the loan, such as the original loan amount, and the type of loan (purchase or refinance, conforming or jumbo). It includes the insurance coverage effective date,[9] and it tells where the property is located (state, zip, and census region) as well as what kind of property it is (single family, condo, 2–4 units, etc.). The data set states whether the owner intended to occupy or invest at time of origination, and it includes the original property value and details abut the default (month and year, unpaid principal balance at default, and broker's opinion of property value at default). Further, the data include information about the foreclosure (month and year, whether the property was sold prior to foreclosure, salvage value net of sales costs and repairs[10]) and the settlement date (month and year).

The following descriptive statistics are generated from the entire 241,293 mortgage insurance claims in the data set. The average original loan amount is about $109,000, and the average unpaid balance at default is around $106,000. The average original property value (the lesser of purchase price or appraised value) is $124,000, and the net salvage value accounts for, on average, about 73% of the original property value. The broker's opinion of property values at default averages about $100,000.

About 78% of the loans in the sample are for purchase and 19% for refinance. Most of the loans (91%) are conforming. Most of the properties (81%) are single-family houses and 97.5% are for owner occupancy. Around 27% of the defaulted properties are located in the Pacific region, 19% are in the South Atlantic region, and 13% are in the West South Central region. Among the 50 states plus the District of Columbia, California has the most mortgage insurance claims, representing 22.5% of all claims.

The raw data, compiled by MICA from its member companies, contain many missing values and data errors (e.g., negative loan amount and invalid settlement date). With assistance from MICA experts, the data were cleaned and scrubbed, resulting in

Table 1
Mean and standard error of loss given default (%) by year, region, and CLTV

Default period	Region	CLTV													
		⩽80		⩽90		⩽95		⩽100		⩽110		⩽120		>120	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
1990–1994	USA Remainder	*20.1*	18.4	27.4	14.3	30.6	12.0	32.7	11.9	37.6	11.4	44.0	10.5	56.8	11.9
	New England	*23.2*	19.4	27.5	14.8	30.9	15.0	29.1	10.1	34.7	9.2	38.8	9.5	51.7	9.5
	Pacific	*11.0*	13.7	17.6	8.9	22.5	7.6	25.7	8.2	30.5	8.0	35.3	7.3	44.4	9.1
	National	*18.7*	18.1	25.5	14.0	29.1	11.8	30.6	11.3	35.0	10.7	40.3	10.2	51.7	12.1
1995–1999	USA Remainder	*17.9*	19.0	*20.6*	12.5	23.4	10.6	25.9	9.7	31.1	10.5	39.1	10.9	53.4	12.5
	New England	*13.6*	16.9	*17.9*	11.8	21.5	9.8	23.6	7.0	29.0	8.9	35.6	8.4	48.6	9.3
	Pacific	*5.4*	14.9	*11.7*	8.5	16.4	8.3	20.2	7.8	25.5	7.3	31.8	7.0	42.6	8.6
	National	*16.2*	19.0	*18.3*	12.2	21.5	10.5	24.0	9.4	28.7	9.7	35.1	9.6	47.2	11.5
2000–2003	USA Remainder	*7.8*	16.7	*16.3*	11.6	21.0	10.2	24.5	9.8	30.1	10.4	37.8	10.5	50.6	10.9
	New England	*4.9*	20.6	*11.5*	10.9	16.9	9.7	19.8	8.8	27.8	10.5	34.4	9.6	45.5	10.5
	Pacific	*1.6*	11.8	*9.8*	8.4	15.8	7.8	18.4	8.8	24.5	8.6	31.8	8.5	43.9	9.6
	National	*7.1*	16.5	*15.2*	11.5	20.2	10.0	23.6	9.9	29.4	10.4	37.0	10.4	49.8	10.9
1990–2003	USA Remainder	*15.1*	18.9	*21.2*	13.5	24.8	11.6	26.9	10.8	32.3	11.1	39.9	10.9	53.4	12.0
	New England	*16.0*	20.2	*21.4*	14.4	23.5	12.6	24.6	8.9	31.0	9.6	36.9	9.1	50.0	9.7
	Pacific	*6.9*	14.2	*13.0*	9.1	17.9	8.5	21.4	8.5	26.8	8.0	32.7	7.4	43.2	8.9
	National	*14.0*	18.6	*19.6*	13.2	23.3	11.4	25.5	10.5	30.5	10.5	37.0	10.2	49.2	11.7

Italic numbers are not significant at 5% level.

[7] MICA has six members: GE Mortgage Insurance, Mortgage Guaranty Insurance Corporation, PMI Mortgage Insurance Co., Republic Mortgage Insurance Company, Triad Guaranty Insurance Corporation, and United Guaranty Corporation, which represent the majority of the PMI companies in the United States.

[8] http://www.privatemi.com/news/factsheets/2006-2007.pdf.

[9] The insurance coverage effective month and year are generally the same as the loan origination month and year.

[10] Salvage value is actual sale price if known; otherwise it is the regression-adjusted broker's opinion of the property value.

106,857 clean observations that are used for the analysis contained in the rest of this paper. All data exclusion criteria are listed in Appendix C, and descriptive statistics from the cleaned data are provided in Tables 2 and 3. Although more than half of the original 241,293 observations were lost, most of the data losses are due to missing values.[11] As to the data losses from cleaning and scrubbing, we feel the exclusion criteria suggested by MICA experts make sense. Comparing some of the descriptive statistics in Table 3 to those in the raw data, we are reasonably confident that the cleaned sample is still representative to the raw data. For example, in the cleaned data around 76% of loans are for purchase and 85% of the underlying properties are single-family house, these are close to the 78% for purchase and 81% single-family house in the raw data.

3. Mortgage loss severity by period, region, and CLTV

The mortgage risk factors of LTV and CLTV are calculated as original loan amount divided by property value at origination and unpaid balance at default divided by property value at default, respectively. We define loss given default (LGD) as[12]

$$\mathrm{LGD} = 100 \times \frac{\mathrm{CUPB} + \mathrm{ACRINT} + \mathrm{FCLEXP} + \mathrm{PROEXP} - \mathrm{NETREC}}{\mathrm{CUPB}}, \tag{1}$$

where CUPB is unpaid balance at default; ACRINT is the interest accrued on CUPB for 3 months at a monthly average of the 30-year fixed conventional commitment rates based on the Freddie Mac weekly survey; FCLEXP is foreclosure expense (servicing and legal costs incurred from default to foreclosure) and is assumed to be \$6000; property maintenance expenses (PROEXP) is assumed to be 3% of net recovery (NETREC), where NETREC = min(NETSALVAGE, 1.5ORIGVAL), where NETSALVAGE and ORIGVAL are the salvage value net of sales costs and repairs and original property value, respectively.[13] All cash flows are discounted at the 1-year LIBOR plus 3% from the foreclosure or settlement date to the time of default.[14]

We use the repeat-sales house price index (HPI) reported by the Office of Federal Housing Enterprise and Oversight (OFHEO) as a proxy for the housing market conditions. Fig. 1 plots HPI in Panel A, and the corresponding house price ratio (HPR, defined as the current HPI as a percent of HPI 18 months previous[15]) in Panel B. We consider a housing market is in a downturn if HPR is less than 100. Based on this economic downturn definition and for the sample period from 1990 to 2003, the New England region was in a downturn during the entire period of 1990–1994, and the Pacific region was in a housing market downturn during the latter half of the 1990–1994 period. The Middle Atlantic region was briefly in a downturn from 1990 to 1991 and again from 1994 to 1995. According to the analysis of the claim rate in each calendar year by MICA, the weighted average claim rates are 0.62%, 0.58%, and 0.51% for the periods from 1990–1994, 1995–2000, and 2000–2003, respectively.[16] Given the periods of housing market downturns shown in Fig. 1 and the higher average claim rate, the period of 1990–1994 is considered as our economic downturn period.[17]

Table 1 provides the mean and standard error of loss given default for different CLTV segments, three geographic regions and the United States, during three sub-periods and the entire sample period. Table 1 shows the loss severity varies with CLTV. For the same region and during the same period, higher mean loss severity is often observed with higher CLTV. For loans with CLTV $\leqslant$ 80, LGD is not significant at 5% level across all regions and time periods and for loans with 80 < CLTV $\leqslant$ 90, LGD is not significant at 5% level in nearly all but the stress period, whereas for loans with CLTV > 90, LGD is significant across all regions and time periods. This is consistent with the theoretical and empirical results documented in the existing literature. It can also be observed from Table 1 that loss severity is generally higher during the economic downturn period of 1990–1994 as compared to other periods. For example, for loans with 95 < CLTV $\leqslant$ 100, the nation-wide mean LGD is 30.6% in 1990–1994, which is 5.1 percentage points higher than the nation-wide mean LGD of 25.5% during the entire sample period of 1990–2003. This is consistent with the notion that loss severity should vary with the housing market condition.

Table 1 does not control for other factors that may affect loss severity, such as the loan age and amount, property type, loan purpose, and whether the property was sold prior to foreclosure. In Table 2, we report the mean and standard error of loss given default by these factors.

We can observe the following from Table 2. First, there is a positive correlation between CLTV and LGD. The mean LGD of the highest CLTV bucket (49.2%) is more than three times as large as that of the lowest CLTV bucket (14%). Second, the initial LTV seems to be positively related to LGD, although to a much less degree compared to CLTV. The mean LGD of the highest LTV bucket (31.7%) is only 2.5 percentage points higher than the mean LGD of the lowest LTV bucket (29.2%). Third, LGD is negatively related to normalized loan size.[18] Fourth, LGD increases as the loan ages. Fifth, LGD also seems to vary with property type, loan purpose, whether the owner intends to live in the property at origination, and whether the property was sold prior to foreclosure. Finally, there are some variations in LGD with the state foreclosure laws.

Tables 1 and 2 should be interpreted with caution because the mean loss severity of some of the cells in the table is calculated from a limited number of observations. In the next section, we study the impact of these and other determinants of mortgage loss severity in a multiple regression framework.

4. Determinants of loss given default

Descriptive statistics of the key variables from the cleaned data are provided in Table 3 (variable definitions are given in Appendix B). In Table 3, LGD shows considerable amount of variation ranging

[11] For example, more than half of the raw data (128,343 observations) do not have the month in which foreclosure started, and 90,475 observations do not have broker's opinion of the property value.

[12] The loss severity defined in Eq. (1) is before Mortgage Insurance (MI) claim. The mortgage insurance companies have the option of either paying the maximum percentage of the claim amount or paying the claim in full and taking title to the property. If the MI company exercises the option to pay the claim in full, the loss to the investor after MI is the small difference between total loss and MI claim amount.

[13] Foreclosure and property expenses are not reported in the MICA data, nor are the accrued interest, the mortgage rate, or the discount rate. These numbers are chosen based on conversations with experts.

[14] LIBOR (London Interbank Offered Rate) is widely used as a reference rate for funding cost. In Basel II context, an appropriate discount rate will reflect the uncertainty of recovery cash flows and the presence of undiversifiable risk, which implies that the appropriate discount rate for IRB purposes likely will differ from the interest rate required under FAS 114 for accounting purposes. As such, for discounting purpose, we add a 3% risk premium to the 1-year LIBOR.

[15] An 18-month window was chosen because a short-term drop in housing price might not cause a surge of housing defaults; on the other hand, if the window is too long, one might not observe any drop in HPI.

[16] The claim rate is defined as the number of claims in each calendar year divided by the number of contracts in force at the beginning of each calendar year. The claim rate of each calendar year is then weighted by the number of contracts in force at the beginning of each calendar year to obtain the weighted average claim rate for a multi-year period.

[17] Downturn period is defined under Basel II as a period of high probability of default (not high loss given default). In mortgage markets, however, defaults are often driven by drops in house prices, and thus periods of high default rates are often associated with declines in house prices.

[18] Loan amount is divided by the median home price at loan origination in each metropolitan statistic area or state median price at origination if the former is not available.

Table 2
Mean and standard error of loss given default (%) by key factors

	Mean	Std		Mean	Std		Mean	Std
CLTV			*LTV*			*LNPURP*		
CLTV080	*14.0*	18.6	LTV080	29.2	17.4	Purchase	30.7	16.6
CLTV090	*19.6*	13.2	LTV090	29.7	15.8	Refinance	29.7	15.5
CLTV095	23.3	11.4	LTV090+	31.7	16.7	Other	34.4	16.4
CLTV100	25.5	10.5						
CLTV110	30.5	10.5	*PROPTYPE*			*PRESALE*		
CLTV120	37.0	10.2	SFD	29.8	16.2	Yes	25.2	12.2
CLTV120+	49.2	11.7	CONDO	34.3	16.5	No	29.4	16.0
			DUP	38.0	17.8	Unknown	34.4	17.5
LNSZN			*AGE*			*SRR*		
LNSZN060	37.9	19.2	24	28.7	15.5	No	30.4	16.2
LNSZN080	31.2	15.8	48	28.4	14.5	Yes	31.6	18.0
LNSZN110	27.9	14.3	84	30.0	15.6			
LNSZN110+	25.6	13.2	>84	35.9	19.3			
OCCUP			*JUDICIAL*			*NODJ*		
Owner Occup	30.4	16.3	No	29.4	15.7	No	31.1	17.1
Investment	34.6	18.7	Yes	32.5	17.3	Yes	29.1	14.4

Variable definitions are in Appendix B. Italic numbers are not significant at 5% level.

Table 3
Descriptive statistics (variable definitions are in Appendix B)

Variable	Mean	Std	Min	Max	Variable	Mean	Std	Min	Max
LGD	30.53	16.40	−31.96	99.97	LTV	90.11	5.42	70.86	108.15
CLTV	104.42	20.30	38.73	170.89	LTV080	0.11	0.31	0	1
CLTV080	0.07	0.26	0	1	LTV090	0.45	0.50	0	1
CLTV090	0.16	0.37	0	1	LTV090+	0.44	0.50	0	1
CLTV095	0.12	0.33	0	1	PROPTYPE1SFD	0.85	0.36	0	1
CLTV100	0.12	0.33	0	1	PROPTYPE2CON	0.12	0.32	0	1
CLTV110	0.19	0.39	0	1	PROPTYPE3DUP	0.02	0.13	0	1
CLTV120	0.13	0.33	0	1	LNPURP1P	0.76	0.43	0	1
CLTV120+	0.20	0.40	0	1	LNPURP2R	0.24	0.42	0	1
HPR	104.34	5.43	78.32	135.56	OCCUP1O	0.98	0.16	0	1
HPR100	0.21	0.40	0	1	PRESALE1Y	0.11	0.31	0	1
HPR105	0.30	0.46	0	1	PRESALE2N	0.57	0.49	0	1
HPR110	0.39	0.49	0	1	AGE	58.90	39.30	0	290
STRESS	0.21	0.40	0	1	AGE24	0.18	0.38	0	1
LNSZN	0.92	0.48	0.10	9.94	AGE48	0.32	0.47	0	1
LNSZN060	0.24	0.42	0	1	AGE84	0.29	0.45	0	1
LNSZN080	0.25	0.44	0	1	JUDICIAL	0.37	0.48	0	1
LNSZN110	0.26	0.44	0	1	SRR	0.10	0.30	0	1
LNSZN110+	0.25	0.43	0	1	NODJ	0.28	0.45	0	1

from around −32.0% to 100.0%, with a mean of 30.5% and standard deviation of 16.4%. CLTV also varies wildly from 38.8% to 170.9% with an average of 104.4%. The average initial LTV is 90.1%, and only 11% of the loans have an LTV below 80%. These percentages are expected from private mortgage insurance data.

More than half (51.7%) of the defaulted mortgages in our sample have a CLTV greater than 100%. This is consistent with the "ruthless" default explanation from the options-based mortgage default theory, which considers default as an optimal decision of rational consumers. The rational borrower will default only when the value of the collateral falls below the mortgage value by an amount equal to the net transaction costs, such as the costs of moving, future deficiency payments, and the stigma associated with default (Crawford and Rosenblatt, 1995).

The CLTV of slightly less than half of the defaulted mortgages in our sample is $\leqslant$100%, and 23.6% and 7.4% have CLTV $\leqslant$90% and $\leqslant$80%, respectively.[19] Some of the defaults are triggered by unexpected non-financial reasons, such as job loss, a significant change in health status, and change in family structure, and especially divorce. In these cases the default option is exercised even while it is not "in-the-money" (Ambrose et al., 1997; Pennington-Cross, 2006). Other defaults might be optioned by rational borrowers who consider selling expenses (brokerage fees and taxes) and additional benefit of default[20], and hold an unbiased estimation of the property value[21], when determining whether to sell the property, to maintain the mortgage, or to default.

Also from Table 3, HPR ranges from 78.3% to 135.6% with a mean of 104.3% and a small standard deviation of 5.4%. This indicates that on average the house price indices in states where the defaulted properties are located experienced an 18-month appreciation of 4.3% with little variation at the state level. The economic downturn indicator has a mean of 0.206, suggesting that 20.6% of the defaults happened in states where house prices depreciated in the past 18 months. Only 10.9% of the properties were sold prior to foreclosure for sure and for 57.2% of the properties were not pre-

[19] In comparison, Pennington-Cross (2003) shows that 99.7%, 86.2% and 38.1% of a random sample of Fannie Mae and Freddie Mac foreclosed properties from 1995 to 1999 have CLTV $\leqslant$ 100%, $\leqslant$90%, and $\leqslant$80%, respectively.

[20] Additional benefit of default includes occupying the property "rent free" from default to foreclosure.

[21] It is often the case that the true value of defaulted property is considerably below the fair market value of similar properties in the same neighborhood. Indeed, the net salvage value averages 73% of the original property value in the complete MICA sample, although the worst 18-month house price index depreciation was less than 10% in nine census regions from 1990 to 2003.

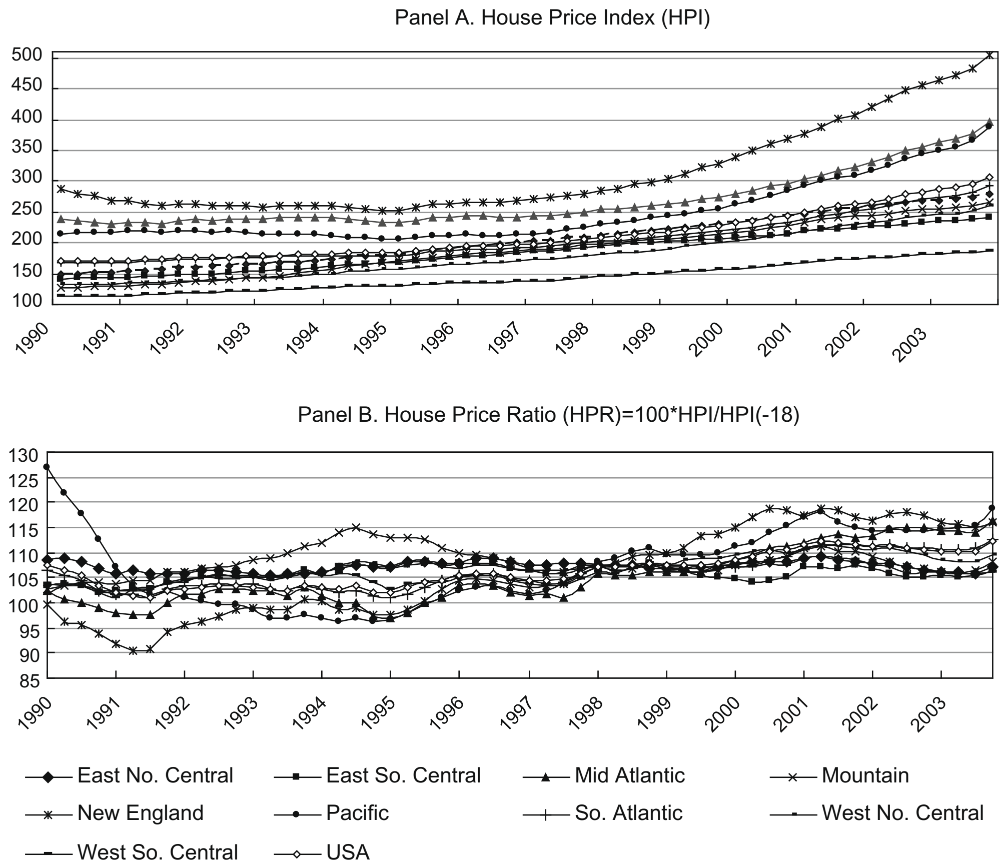

Fig. 1. Repeat-sales house price index reported by OFHEO in nine census regions and the United States (1990–2003) and 18-month house price ratio.

sold. The average time from loan origination to foreclosure (or settlement if the foreclosure date is missing) is around five years. About 18% of defaulted properties in the sample were foreclosed (settled) within two years, and 50% within four years.[22]

To get more insight into how the key variables are related, we report in Table 4 the correlation matrix for the following variables: LGD, CLTV, LTV, HPR, loan size, and the number of months from origination to foreclosure (AGE) and from default to foreclosure (FCTIME). There is a very high positive correlation between LGD and CLTV (0.699), a modest negative correlation between LGD and HPR (−0.172), and a modest negative correlation between CLTV and HPR (−0.184) as expected. Consistent with intuition, Table 4 also shows that initial normalized loan amount is negatively correlated with LGD (−0.220) and with LTV (−0.082), LGD is positively related to the length of the foreclosure process (.173), and larger initial loan amount is associated with earlier default (−.099). There is little correlation (0.002) between CLTV and LTV, and it is statistically insignificant. Our sample consists of only *defaulted* mortgages, a subset of all mortgages originated. For this subpopulation CLTV is largely driven by house price movements after origination, thus it may have little correlation with the original LTV.

[22] Loan age is the number of months between the origination date and the foreclosure date (or the settlement date if the foreclosure date is missing). The foreclosure date is missing for around 15% of our sample. The average might have been driven up by some outliers, for example, the maximum span from origination to default (or foreclosure) is 290 months in our sample, and the standard deviation is close to 40 months.

4.1. Regression with CLTV

Loss severity can be statistically characterized by conditional means and variances, which are in turn contingent on housing market conditions and on loan and property characteristics. We specify a general regression equation relating loss given default to loan and property characteristics and housing market conditions as

$$\mathrm{LGD}_{it} = \alpha + \sum_{j=1}^{J} \beta_j X_{ijt} + \sum_{k=1}^{K} \gamma_k Z_{ik} + \varepsilon_{it}, \tag{2}$$

where LGD_{it} is the loss given default of the ith defaulted mortgage measured at time of default t,calculated as in Eq (1); X_{ijt} is the value of the jth time-varying explanatory variable for the ith defaulted mortgage at time t, such as CLTV dummies, economic downturn indicator, whether the property was sold prior to foreclosure, and loan age indicators; Z_{ik} is the value of the kth non-time-varying explanatory variable for the ith defaulted mortgage, such as LTV dummies, loan size dummies, loan purpose, property type, owner occupancy, and state foreclosure law dummies, that are observed at mortgage origination. The detailed variable definitions can be found in Appendix B. Since CLTV, LTV, and loan size may have nonlinear effect on LGD, we use CLTV, LTV, and loan size dummies instead of the continuous variables themselves in the regression. These dummy indicators are also more relevant for retail segmentation analysis.

Our sample is very large and there are no obvious violations of the classic regression assumptions. Thus the model was estimated using ordinary least squares as in Clauretie and Herzog (1990), Calem and LaCour-Little (2004), and all other studies on mortgage

Table 4
Pearson correlation coefficient

	LGD	CLTV	LTV	HPR	LNSZN	AGE	FCTIME
LGD	1.000	**0.699**	**0.065**	**−0.172**	**−0.220**	**0.178**	**0.173**
CLTV		1.000	0.002	**−0.184**	0.003	**−0.059**	**−0.013**
LTV			1.000	**0.061**	**−0.082**	**0.166**	**−0.021**
HPR				1.000	**−0.065**	**−0.118**	**−0.104**
LNSZN					1.000	**−0.099**	0.008
AGE						1.000	**0.240**
FCTIME							1.000

Bold numbers are significant at 0.01% level.

loss severity. The regression parameter estimates, corresponding *p*-values, and goodness of fit measures are shown in Table 5. The model shows relatively high explanatory power ($\bar{R}^2 = 0.610$).

The impact of the housing market condition is captured in the regression analysis by the explanatory variables of CLTV and economic downturn scenarios. Consistent with the existing studies, loss severity rates are significantly positively related to CLTV (Pennington-Cross, 2003; Calem and LaCour-Little, 2004) and are significantly higher in distressed housing markets (Clauretie and Herzog, 1990). This is also consistent with economic intuition: loans with lower CLTV will have a higher equity value which leads to a higher recovery rate and hence lower loss severity, and vice versa. In principle, if the "true" CLTV could be observed at time of default without noise, and the actual timing and amount of foreclosure and property expenses were available, one might be able to explain close to 100% of the variations in LGD. The coefficient of the downturn indicator is positive (2.87) and significant, indicating that, other things being equal, loss severity will be 2.87 percentage points higher in distressed housing markets.

Normalized loan size has a negative impact on loss severity rates. If every other variable is held constant, loans of size less than or equal to 80% and 60% of median home price will have higher loss severity rates, by 4.36 percentage points and 10.97 percentage points, respectively, than those that are greater than 110% of median home price. These findings are consistent with observations from the foreclosure process, which contains relatively fixed cost components (for example, attorney and title fees) at foreclosure and sale of the property regardless of the loan size. These fixed cost components are likely to result in larger LGDs for smaller loans. Alternative explanations are also possible if normalized loan size is considered to be a proxy for some desirable characteristics of the underlying property (e.g., location and amenities) or the borrower (e.g., income, education and wealth).

Property types of single family and condo have lower loss severity rates, by 0.40 and 2.13 percentage points, respectively, compared to other property types such as multiple-unit properties. The LGD of owner-occupied properties is 1.70 percentage point lower than the LGD of investment properties. Presale of the property in the process of default to foreclosure will result in a lower LGD by 6.07 percentage points. This finding is consistent with the industry observation in the process of default to foreclosure. Presale will, in general, incur smaller sales and repair costs. The age of the loan has a positive effect on loss severity, consistent with the finding of Calem and LaCour-Little (2004) but inconsistent with those of Lekkas et al. (1993) and Pennington-Cross (2003).

Finally, we find that the LGD is higher in states with judicial foreclosure process and statutory rights of redemption. These observations are largely in line with those found in the existing literature. Our results show that the LGD is lower in states where deficiency judgments are prohibited, contrary to what has been found in the existing literature. Based on the framework of Ambrose et al. (1997), in states where deficiency judgment is prohibited, lenders are likely to try hard to shorten the delay between default and foreclosure to reduce the period of the "free rent" and thus the probability of default.[23] This could actually result in lower LGD. Thus our results are consistent with the theoretical framework of Ambrose et al. (1997). Another reason may be that deficiency judgments are rare even when they are permitted because defaulting homeowners are unlikely to have many assets aside from the home and they often protect themselves against deficiency judgments by filing for bankruptcy (Pence, 2006). One more possible explanation is that since California is one of the very few states that prohibit deficiency judgment and it accounts for the largest percent (22.5%) of our sample, deficiency judgment may act as a proxy for California, or some other variables that are not already included in our model.

4.2. Regression with LTV

The US interagency Basel IA Notice of Proposed Rulemaking (NPR) (2006) and Advance Notice of Proposed Rulemaking (ANPR) (2005) on risk-based capital guidelines suggest basing risk weights for residential mortgages on LTV ratios. This will make capital requirements sensitive to risk and will unlikely increase regulatory burden for banks since LTV data are readily available and are often used in loan approval and pricing as well as in managing mortgage portfolios. More importantly, core banks adopting A-IRB often include LTV as a risk driver in their mortgage segmentation system or in their loan level credit risk models. To further assess the relevance of LTV in determining loss severity, we drop the CLTV dummies from the previous regression. The resulting model is shown in Table 6.

In the event of default and foreclosure, the homeowner equity is a function of the initial LTV and the subsequent course of house prices, which vary by geographic region and time period. In particular, the loss severity will increase as the defaulted loan experiences a subsequent house price decline for the 12–18 months starting from delinquency. The theory is clearly demonstrated in the empirical estimates from Table 6 as loss severity increases by about 7.43 percentage points in the distressed housing markets where HPI shows decline from its level 18 months ago. The regression results also show that higher LTV at origination leads to higher loss severity at default. The impact of other explanatory variables

Table 5
LGD regression with CLTV

Variable	Coefficient	*p*-value	Variable	Coefficient	*p*-value
Intercept	12.744	<.0001			
CLTV090	8.925	<.0001	PROPTYPE1SFD	−0.403	0.0279
CLTV095	13.925	<.0001	PROPTYPE2CON	−2.134	<.0001
CLTV100	17.037	<.0001	LNPURP1P	0.347	<.0001
CLTV110	22.179	<.0001	OCCUP1O	−1.703	<.0001
CLTV120	28.333	<.0001	PRESALE1Y	−6.069	<.0001
CLTV120+	38.871	<.0001	PRESALE2N	−2.009	<.0001
STRESS	2.867	<.0001	AGE24	−7.561	<.0001
LTV090	0.695	<.0001	AGE48	−6.383	<.0001
LTV095	1.705	<.0001	AGE84	−5.426	<.0001
LNSZN060	10.969	<.0001	JUDICIAL	1.924	<.0001
LNSZN080	4.363	<.0001	SRR	3.029	<.0001
LNSZN110	1.958	<.0001	NODJ	−4.081	<.0001
Adj. R^2	0.610		Observation	106,857	
F	6956.75	<.0001			

[23] According to "Foreclosure Prevention" (Fannie Mae, 1997), the allowable time between referral to the foreclosure attorney and the foreclosure sale date ranges from three to seven months in eight states prohibiting deficiency judgments, whereas it ranges from three to 12 months in other states. Indeed, out of the eight states prohibiting deficiency judgments, six do not follow a judicial procedure, and five do not have statutory rights of redemption. It has been documented in the literature that judicial procedure and statutory rights of redemption extend the foreclosure and liquidation processes and are associated with higher LGD.

Table 6
Alternative LGD regression with LTV

Variable	Coefficient	*p*-value	Variable	Coefficient	*p*-value
Intercept	35.812	<.0001			
STRESS	7.433	<.0001	OCCUP1O	−3.339	<.0001
LTV090	2.489	<.0001	PRESALE1Y	−6.898	<.0001
LTV090+	5.022	<.0001	PRESALE2N	−3.094	<.0001
LNSZN06	10.542	<.0001	AGE24	−4.511	<.0001
LNSZN08	4.287	<.0001	AGE48	−5.393	<.0001
LNSZN11	1.469	<.0001	AGE84	−4.211	<.0001
PROPTYPE1SFD	−3.206	<.0001	JUDICIAL	1.754	<.0001
PROPTYPE2CON	−1.867	<.0001	SRR	1.330	<.0001
LNPURP1P	−2.389	<.0001	NODJ	−2.172	<.0001
Adj.R^2	0.145		Observation	106,857	
F	1008.29	<.0001			

largely follows the same pattern as observed in Table 5. However, the estimates reported in Table 6 may be biased since the model suffers from omitted variable problem as CLTV, an important explanatory variable, was not included. As a result, the coefficient to STRESS may have a positive bias since the omitted variable CLTV is positively correlated with the dependant variable LGD and the included variable STRESS.[24]

Overall the following observations can be made from Tables 5 and 6. First, the following factors—current loan-to-value ratio, stress factor, loan size, property type (single family, condo, etc.), loan purpose (purchase or refinance), whether the owner intended to occupy at origination, whether the property was sold prior to foreclosure, the age of the loan, and the state foreclosure laws—jointly can explain about 61% of the variation in the loss severity in the MICA data, which consist largely of high-LTV mortgages (average LTV around 90%). Second, CLTV is the single most important determinant of LGD—the higher the CLTV, the higher the LGD. After dropping CLTV dummies, the adjusted R^2 decreases dramatically from 61% to 15%. Third, LGD during housing market downturns is statistically significantly higher. The stress factor is especially important in the absence of CLTV; LGD is about 7.43 percentage points higher during economic downturn periods, partly because of the positive bias induced by omitting CLTV in the model. Fourth, CLTV is a much better predictor of LGD than LTV.

5. Implications for several issues in Basel II implementation

Our findings have important implications for risk-based regulatory capital requirements and Basel II implementation. In particular, our results and analysis should shed light on the following issues: whether LTV or CLTV should be used in determining Basel II capital; whether the proposed supervisory LGD mapping function for downturn LGD is appropriate; and whether the 10% supervisory LGD floor for residential mortgage exposures is binding. In this section, we discuss these one by one.

5.1. Use LTV to segment risk

In the US Basel IA ANPR (2005), the US financial regulatory agencies seek comments on the use of LTV to determine risk weights for residential mortgages and on whether LTV should be updated periodically. Basel IA NPR (2006) proposed more granular LTV buckets for first mortgage risk-weight categories and combined LTV for junior lien mortgages. In July 2007 the agencies decided to drop the Basel IA proposals and agreed to issue a proposed rule that would provide non-core banks that would not be adopting A-IRB with the option to adopt the standardized approach included in the New Accord. Nevertheless, core banks adopting A-IRB often use LTV as one of the risk drivers in their mortgage segmentation system or in their loan level credit risk models for risk parameter quantification.

Our empirical results in Table 6 show that LTV is statistically and economically significantly related to LGD, and higher LTV is associated with higher LGD. Since regulatory capital is linearly related to LGD, our statistical results thus support the use of LTV to segment risk and the notion that the higher the LTV, the higher the risk weights.

5.2. Whether LTV should be updated periodically

CLTV has a much higher correlation with LGD than LTV (Table 4), and not surprisingly, regression with CLTV has much better explanatory power than that without CLTV (Tables 5 and 6). Furthermore, the average LTV at origination is around 90% whereas at time of default, the average CLTV jumps up to 104% (Table 3), reflecting a significant decrease in homeowner equity. Therefore, in our opinion, LTV should be updated periodically to better segment risk. This is also consistent with Basel II rules which require banks to periodically review and update risk parameters and minimum regulatory capital on an on-going basis based on the current portfolio, no less frequently than annually.

However, to calculate CLTV, lenders need to update the property value periodically, which could be quite costly. Alternative approaches could be explored to get a timely update on property values. For example, the lender could use automated valuation models (AVMs) through vendors or tax assessment, build internal models, or at least adjust the property value using a local house price index. All these alternatives bear some model risk, i.e., the updated property value might differ from the true market value of the property. When the difference is too big, CLTV might become inferior to LTV. Therefore, it is important for banks to follow validation standards on property valuation models (including vendor models) to ensure the model risk is in check.

5.3. Supervisory LGD mapping function

Paragraph 468 of the Basel II Framework requires that LGD be measured "to reflect economic downturn conditions where necessary to capture the relevant risks". The Basel Committee released in July 2005 new guidance regarding a "principles-based" approach to satisfying the requirements of paragraph 468. However, due to lack of historical loan-level LGD data, many core banks will have difficulty estimate downturn LGD at this time and in the near future. A supervisory LGD mapping function, LGD = 0.08 + 0.92 ×

[24] In the OLS framework, bias in the coefficient of one variable resulting from omitting another relevant variable is equal to the coefficient of the omitted variable (had it been included in the model) multiplied by the regression coefficient in a regression of the omitted variable on the variable under consideration. See Maddala (2001) for explanation.

ELGD, that transforms the long-run default-weighted average LGD (or ELGD) into economic downturn LGD had been proposed in the US Basel II NPR (2006). A-IRB banks that are unable to develop acceptable internal downturn LGD estimates would be required to use the supervisory mapping function to calculate their downturn LGD for Basel II regulatory capital. Although the US financial regulatory agencies did not include the supervisory LGD mapping function in the US Basel II Final Rule (2007), the agencies continue to believe that the function (and the associated estimation of the long-run default-weighted average LGD) is one way a bank could address difficulties in estimating downturn LGD. In this section, we examine the accuracy of this mapping function based on the MICA data.

Considering the mean loss severity of the entire sample period (1990–2003) as observed ELGD, and the mean loss severity of the period 1990–1994 as the observed downturn LGD. In Fig. 2, we plot the observed downturn LGDs (blue squares) against observed ELGDs (green triangles) by region (New England, Pacific, and National) and by CLTV buckets taken from Table 1. We also plot the estimated downturn LGDs from the supervisory mapping function (red round dots) against observed ELGDs.

In Fig. 2, almost all except for two blue squares lie between the green triangles and the red round dots, suggesting that in general the supervisory mapping function is somewhat too strict, i.e., it produces higher estimated downturn LGDs than the observed downturn LGDs from the MICA sample. The differences range from −1.28% for New England with $90 < \text{CLTV} \leqslant 95$ to 3.35% for Pacific with $\text{CLTV} \leqslant 80$, with a mean of 1.59% and a median of 1.82%.

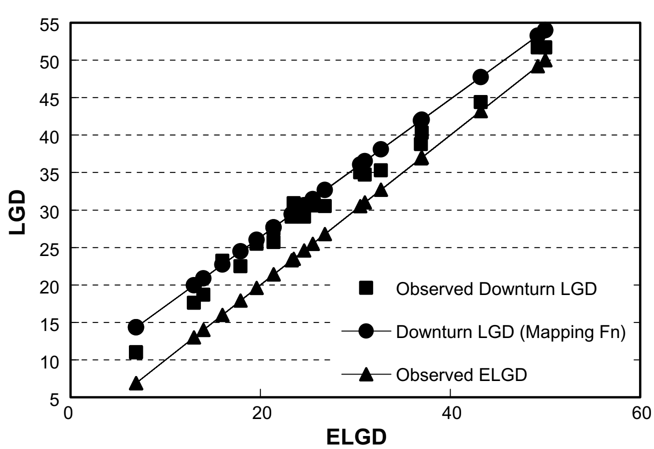

Fig. 2. Relevance of the proposed LGD mapping function based on observed LGD.

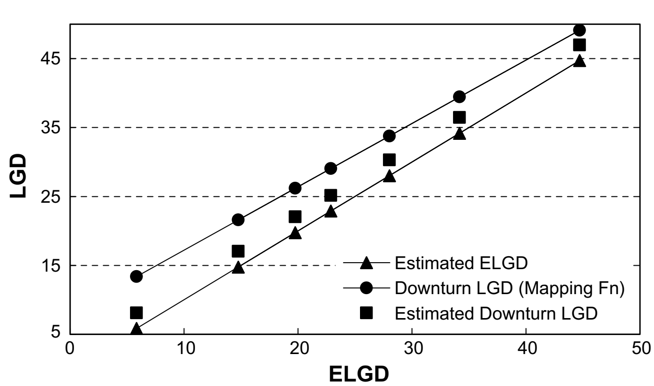

Fig. 3. Relevance of the proposed LGD mapping function based on the regression model.

We further examine the mapping function based on the regression analysis in Section 4. Fig. 3 plots the downturn LGD against ELGD based on the regression results in Table 5 for seven hypothetical mortgage defaults from seven different CLTV buckets, but with the same and the most typical values for the rest of explanatory variables—the loan size is between 0.8 and 1.1 times the median house price in the area, the loan is for a single-family house, the loan purpose is purchase, the owner intends to live in the property, there is no presale before foreclosure, and the loan age is between two and four years. As discussed in Section 4.1, since the CLTV dummies have captured most of the variations in mortgage loss severity, the STRESS indicator has a coefficient of only 2.87. This implies that, other things being equal, the loss severity of a mortgage defaulted during the downturn period will be only 2.87 percentage points higher than a mortgage defaulted during other periods. Consequently, the estimated downturn LGDs (blue squares) are 2.28 ($=2.87(1-0.206)$) percentage points above the estimated ELGD (green triangles) and all are below the estimated downturn LGDs based on the LGD supervisory mapping function (red round dots). On average the LGD mapping function gives a downturn LGD that is 3.78 percentage points higher than that from the MICA sample; the median difference is 3.89 percentage points. Thus the supervisory LGD mapping function is somewhat conservative based on the regression model with CLTV dummies.

5.4. Ten percent supervisory LGD floor

In the proposed as well as the final Basel II rules, there is a 10% floor for LGD of residential mortgage exposures, except for those guaranteed by a sovereign entity such as FHA or VA. Before the private mortgage insurance claim benefit is factored in, our sample average LGD is 30.5%. Loss severity is less than 10% for 8.44% of the 106,857 defaults.

The MICA data does not contain information on mortgage insurance claim benefit, but it can be reasonably well inferred based on common PMI coverage levels and claim handling practices. We assume the coverage level is 12% of the claimed losses on mortgages with LTV less than or equal to 85%, 25% on mortgages with LTV less than or equal to 90%, 30% on 95%, and 35% on 100%. The mortgage insurance companies have the option of either paying the maximum coverage percentage of the claim amount or paying the claim in full and taking the title to the property. If the mortgage insurance company exercises the option to pay the claim in full, the loss to the lender after factoring in the mortgage insurance benefit is close to zero.

After considering the private mortgage insurance payment calculated as above, the sample average LGD becomes 6.13%. Around 70% of the 106,857 mortgage defaults in our sample have an LGD that is less than 10%. Therefore, the 10% LGD floor is binding for most of the mortgage defaults in the MICA sample. Note, however, these statistics are based on LGDs of the entire sample of mortgage claims. When applied to the downturn LGD of 13.64% ($=8\% + 0.92 \times 6.13\%$, based on the supervisory LGD mapping function), the 10% floor becomes non-binding.

6. Conclusions

Using a large set of historical loan-level default and recovery data of high-LTV mortgages from several private mortgage insurance companies, we find that the following factors—the current loan-to-value ratio, loan-to-value at origination, downturn factor (measured by a decline in house price index from a year and a half previously in the state where the property is located), loan size, property type (single family, condo, etc.), loan purpose (purchase or refinance), whether the owner intended to occupy at origination, whether the property was sold prior to foreclosure, the age of the loan, and the state foreclosure laws—jointly can explain most of the variations in the loss severity in our sample (adjusted R^2 of 0.610). We also find that CLTV is the single most important determinant of LGD—the higher the CLTV, the higher the LGD. Dropping the CLTV dummies causes the adjusted R^2 to decrease dramatically from 0.610 to 0.145. Loss severity in distressed housing markets is found to be statistically significantly higher. In the absence of CLTV, the omitted variable problem causes the stress factor appears to be especially large—LGD is about 7.43 percentage points higher under economic downturn conditions. Finally, LTV is positively related to LGD, but CLTV is a much better predictor of LGD than LTV.

The implications of our study on risk-based capital are the following: LTV at the time of loan origination can be used to segment risk; updated LTV (or CLTV) is the single most important predictor for residential mortgage LGD and thus should be used to segment risk if it is available. Furthermore, the proposed supervisory LGD mapping function appears to be somewhat conservative across geographic regions and current loan-to-value ratios of the exposures, based on both a default-weighted average approach and a regression analysis approach. The conservatism of the supervisory LGD mapping function should give banks incentive to develop their internal downturn LGD estimates and get them approved by their primary supervisors. Finally, the 10% supervisory LGD floor is binding for a large percent of the mortgage defaults covered in the MICA sample but may become less binding if only defaults in downturn housing markets were considered.

It is important to note that our empirical results and conclusions are based on defaulted mortgages that are privately insured by MICA members. These mortgages generally have high LTV and represent nearly 10% of the US residential mortgage market in terms of both the total outstanding (a little more than 10 trillion as of 2005 year-end) and the total originations (about 3 trillion in 2005).[25] Should PMI mortgage default pattern and loss severity differ from those of other mortgages, such as FHA- or VA-insured, second lien, or mortgages requiring no private mortgage insurance, our findings and conclusions might not apply to other mortgages.

Acknowledgements

The authors are especially grateful to Ted Durant for sharing his time and expertise, and to Basil Petrou and Mitch Stengel for help in making this study possible. We also thank Mike Carhill, Souphala Chomsisengphet, Dennis Glennon, Mark Levonian, Mitch Stengel, Gary Whalen of the OCC, Tsuyoshi Oyama and Masao Yoneyama of Bank of Japan, the participants in the 2006 Quantitative Risk Forum at the Federal Reserve Bank of Philadelphia and the Basel II Accord Implementation Group Validation Subgroup meeting in May 2007, and Charles Calhoun for comments that have improved this work. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Office of the Comptroller of the Currency (OCC), or the US Treasury Department.

Appendix A

Summary of existing studies on residential mortgage loss severity

Study	Obs.	Sample Period	Source	LGD Definition	$\bar{R}^2$	Major Findings
Clauretie and Herzog (1990)	408	1980–1987	PMI via Moody's	Direct loss paid/previous year end risk	0.56–0.57	$\Delta r(-)$, HPA(−), $\Delta U(+)$, PS(−), SRR(+), NODJ(+)
Clauretie and Herzog (1990)	85,000	1972–1988 (claims paid)	FHA	(UPB-house value)/original loan amount	0.04–0.05	$\Delta r(-)$, HPA(−), LTV(+), PS(−), SRR(+)
Lekkas et al. (1993)	9457	1975–1990 (originated)	Freddie Mac	(UPB-Appraised value)/UPB	0.06–0.07*	LTV(+), Age(−), Texas(+), Odds Ratio(+)
				(UPB-Sale price)/UPB		
Crawford and Rosenblatt (1995)	1191	1988–1992 (foreclosed)	A large northeastern thrift	(UPB-REO sale price)/UPB	0.02–0.03*	$\Delta r(-)$, SRR(+), NODJ(+)
				(UPB-min(original appraised value, original purchase price))/UPB	0.09–0.14*	$\Delta r(-)$, PS(−), SRR(+), NODJ(+), DIL(−), RTC(+), LPI to foreclosure (+)
Pennington-Cross (2003)	16,272	1995–1999 (foreclosed)	GSEs	(UPB-sale price)/UPB	0.95–0.96	CLTV(+), CLTV × subprime (+), PS(−), NODJ(+), Age(−), size(−), size^2 (+)
Calem and LaCour-Little (2004)	120,289	1989–1997 (originated)	GSEs	(UPB-Gross sale proceeds)/UPB	0.25*	CLTV(+), CLTV ⩾ 90(−), LTV (−), LTV ⩾ 80 (−), size (−), size^2 (+), RELINC (−), RELINC^2 (+), Age (+)

Δr: Increase in interest rate in the year of termination relative to origination; ΔU: rise in unemployment rate; HPA: house price appreciation; PS: power-of-sale method of foreclosure (non-judicial); SRR: statutory right of redemption; NODJ: deficiency judgment prohibited; RELINC: relative median income in the property zip code; DIL: deed in lieu indicator; RTC: '1' if the Resolution Trust Corporation disposed the REO property, '0' otherwise; LPI to foreclosure: months from last paid installments to foreclosure; *: indicates R^2.

[25] *Sources*: Mortgage Bankers Association and MICA.

Appendix B

Variable definitions

AGE : loan age, the number of months between the origination date and the foreclosure date (or the settlement date if the foreclosure date is missing). Define *AGE* indicator variables for the following ranges: (, 24] = '*AGE24*'; (24, 48] = '*AGE48*'; (48,84] = '*AGE84*'.

BOVVAL : broker's opinion of value, as-is, at default, observed before foreclosure.

CLTV : current loan-to-value ratio or loan-to-value ratio at time of default (t), defined as $\mathrm{CLTV}_{it} = 100 * \mathrm{CUPB}_{it}/[(\mathrm{HPI}_t/\mathrm{HPI}_{t_f})\mathrm{BOVVAL}_{it_f}]$, where t_f is foreclosure date. Define CLTV indicators for the following ranges: (,80] = 'CLTV080'; (80,90] = 'CLTV090'; (90,95] = 'CLTV095'; (95,100] = 'CLTV100'; (100,110] = 'CLTV110'; (110,120] = 'CLTV120'; (120,) = 'CLTV120+'.

CUPB : unpaid balance at default.

HPI : quarterly house price index, reported by the Office of Federal Housing Enterprise and Oversight (OFHEO).

HPR : house price ratio, defined as $\mathrm{HPR}_{it} = 100\mathrm{HPI}_{it}/\mathrm{HPI}_{i(t-18months)}$, HPI_{it} is the house price index of the state where the ith property is located at time t. Define HPR indicators for the following ranges: (,100] = 'HPR100'; (100,105] = 'HPR105'; (105,110] = 'HPR110'; (110,) = 'HPR110+'.

JUDICIAL : '1' indicates that the state in which a property is located has a judicial foreclosure process (as opposed to non-judicial or power-of-sale method of foreclosure).

LGD : loss given default, defined as $100 \cdot (\mathrm{CUPB} + \mathrm{ACRINT} + \mathrm{FCLEXP} + \mathrm{PROEXP} - \mathrm{NETREC})/\mathrm{CUPB}$, where ACRINT is accrued interest, FCLEXP is foreclosure expenses, PROEXP is property maintenance expenses, NETREC is the net recovery. All cash flows are discounted at the 1-year LIBOR plus 3% from the foreclosure or settlement date to the time of default.

LNPURP : loan purpose indicator, '1P' for purchase, and '2R' for refinance.

LNSZN : loan amount relative to area median home price at origination, i.e., LOANAMT/MEDPRC. Define LNSZN indicators for the following ranges: (, 0.6] = 'LNSZN060'; (0.6,0.8] = 'LNSZN080'; (0.8,1.1] = 'LNSZN110'; (1.1,) = 'LNSZN110+'.

LOANAMT : original loan amount.

LTV : original loan-to-value ratio, defined as $\mathrm{LTV}_i = 100\mathrm{LOANAMT}_i/\mathrm{ORIGVAL}_i$. Define LTV indicator variables for the following ranges: (, 80] = ' LTV080'; (80, 90] = 'LTV090'; (90,) = 'LTV90+'.

MEDPRC : median home price at loan origination in the metropolitan statistic area, or state median price at origination if the former is not available.

NETSALVAGE : salvage value net of sales cost and repairs. It is the actual sale price if known or regression-adjusted BOVVAL.

NODJ : '1' indicates in the state where a property locates deficiency judgments are prohibited.

OCCUP : intended occupancy at origination indicator, '1O' for owner occupies.

ORIGVAL : original property value (lesser of purchase price or appraised value).

PRESALE : property sold prior to foreclosure indicator, '1Y' for yes, and '2N' for no, otherwise unknown.

PROPTYPE : property type indicator, '1SFD' for single family, '2CON' for condo, '3DUP' for 2–4 units.

PS : '1' indicates in the state where a property is located lenders may choose a power-of-sale foreclosure process, as opposed to judicial states where lenders must follow a judicial foreclosure process.

SALVAGEPCT : net salvage value/original property value.

SRR : '1' indicates that in the state where a property is located there is statutory right of redemption.

STRESS : economic downturn indicator, '1' if HPR < 100 in a state.

Appendix C

Exclusion criteria

The following exclusion criteria are applied for data cleaning and exclusion to eliminate potential biases and data errors

LGD $\leqslant -50$, or LGD $\geqslant 100$	LOANAMT $\leqslant 0$	NETSALVAGE $\leqslant 0$
FCTIME < DLQTIME, or $\leqslant 0$	SALVAGEPCT $\leqslant 0$	STLTIME $\leqslant$ 0
CUPB $\leqslant$ 10,000, or >1.2LOANAMT		ORIGVAL $\leqslant$ 10,000
BOVVAL $\leqslant$ 5000, or >3ORIGVAL, or < 0.5ORIGVAL		

CLTV, LTV, and SALVAGEPCT that are three standard deviations away from their respective means

References

Abaffy, J., Bertocchi, M., Dupa-ová, J., Moriggia, V., Consigli, G., 2007. Pricing nondiversitiable credit risk in the corporate Eurobond market. Journal of Banking and Finance 31 (8), 2233–2263.

Altman, E., Resti, A., Sironi, A., 2005a. Default recovery rates in credit risk modeling: a review of the literature and recent evidence. Journal of Finance Literature 1, 21–45.

Altman, E., Resti, A., Sironi, A. (Eds.), 2005b. Recovery Risk: The Next Challenge in Credit Risk Management. Risk Books, London.

Ambrose, B.W., Buttimer Jr., R.J., Capone, C.A., 1997. Pricing mortgage default and foreclosure delay. Journal of Money, Credit, and Banking 29 (3), 314–325. Why 'Basel II' may need a leverage ratio restriction.

Calem, P.S., 2003. Loss severity calculation for residential mortgages. Unpublished manuscript, Board of Governors of the Federal Reserve System.

Calem, P.S., LaCour-Little, M., 2004. Risk-based capital requirements for mortgage loans. Journal of Banking and Finance 28, 647–672.

Clauretie, T.M., Herzog, T., 1990. The effect of state foreclosure laws on loan losses: evidence from the mortgage insurance industry. Journal of Money, Credit, and Banking 22 (2), 221–233.

Crawford, G.W., Rosenblatt, E., 1995. Efficient mortgage default option exercise: evidence from loss severity. The Journal of Real Estate Research 10 (5), 543–555. Bank loan losses-given-default: A case study.

Dermine, J., Neto de Carvalho, C., 2006. Bank loan losses-given-default: a case study. Journal of Banking and Finance 30 (4), 1219–1243.

Lekkas, V., Quigley, J.M., Van Order, R., 1993. Loan loss severity and optimal mortgage default. Journal of the American Real Estate and Urban Economics Association 21 (4), 353–371.

Maddala, G.S., 2001. Introduction to Econometrics, 3rd ed. Wiley, Chichester, England. pp. 160–161.

Nyström, K., Skoglund, J., 2006. A credit risk model for large dimensional portfolios with application to economic capital. Journal of Banking and Finance 30 (8), 2163–2197.

Pence, K.M., 2006. Foreclosing on opportunities: state laws and mortgage credit. Review of Economics and Statistics 88, 177–182.

Pennington-Cross, A., 2003. Subprime and prime mortgages: Loss distributions. Working paper, Office of Federal Housing Enterprise Oversight.

Pennington-Cross, A., 2006. The duration of foreclosure in the subprime mortgage market: a competing risks model with mixing. Working paper, Federal Reserve Bank of St. Louis.

Qi, M., 2005. Survey of research on downturn LGD. Work paper, Basel Accord Implementation Group.

US Basel IA Advanced Notice of Proposed Rulemaking, 2005. Risk-based capital guidelines; Capital adequacy guidelines; capital maintenance: domestic capital modifications. Federal Register 70 (202) (Thursday, October 20).

US Basel IA Notice of Proposed Rulemaking, 2006. Risk-based capital guidelines; capital adequacy guidelines; capital maintenance: domestic capital modifications; proposed rules and notice. Federal Register 71 (247) (Tuesday, December 26).

US Basel II Final Rule, 2007. Risk-based capital standards: advanced capital adequacy framework – Basel II; final rule. Federal Register 72 (235) (Friday, December 7).

US Basel II Notice of Proposed Rulemaking, 2006. Risk-based capital standards: advanced capital adequacy framework and market risk; proposed rules and notices. Federal Register 71 (185) (Thursday, September 25).

VanHoose, D., 2007. Theories of bank behavior under capital regulation. Journal of Banking and Finance 31 (12), 3680–3697.