



Centre for International
Finance and Regulation

Towards Financial System Integrity

Research Working Paper Series

Decomposing the smile: systematic credit risk in mortgage Portfolios

Yongwoong Lee

University of Technology, Sydney

Daniel Rosch

University of Regensburg, Germany

Harald Scheule

University of Technology, Sydney

WORKING PAPER NO. 13/2014

MARCH 2014

www.cifr.edu.au

This research was supported by the Centre for International Finance and Regulation (project number E001) which is funded by the Commonwealth and NSW Governments and supported by other Consortium members (see www.cifr.edu.au).



All rights reserved. Working papers are in draft form and are distributed for purposes of comment and discussion only and may not be reproduced without permission of the copyright holder.

The contents of this paper reflect the views of the author and do not represent the official views or policies of the Centre for International Finance and Regulation or any of their Consortium members. Information may be incomplete and may not be relied upon without seeking prior professional advice. The Centre for International Finance and Regulation and the Consortium partners exclude all liability arising directly or indirectly from use or reliance on the information contained in this publication.

Decomposing the smile: systematic credit risk in mortgage portfolios

Yongwoong Lee* Daniel Rösch† Harald Scheule‡§

March 7, 2014

Abstract

This study analyzes systematic and non-systematic credit risk in mortgage portfolios given US loan-level information by controlling for time-varying observable information in relation to the borrower, the collateral and the macro economy. The total risk in relation to rating class default rates is decomposed into systematic and class-specific non-systematic risk by a state space model. The paper finds that the total risk relates to credit quality in a smile-shaped pattern: systematic risk is negatively related and non systematic risk is positively related to average default rate levels. In addition, total risk increases during and after the Global Financial Crisis. The impact of the crisis on systematic risk is persistent whereas the impact on non-systematic risk appears to be temporary. The analysis of regulatory capital suggests that mortgage risk models in conjunction with periodic updating warrant a sufficient level of regulatory capital given the current regime. These findings are relevant to prudential regulators who are currently discussing the implementation of a monotone relationship between default probabilities and asset correlations under Basel III.

JEL classification: G20; G28; C51

Keywords: Asset correlation, Basel capital, Loss given default, Mortgage portfolio, Probability of default, State space model, Systematic risk.

*Finance Discipline Group, UTS Business School, University of Technology, Sydney, PO Box 123, Broadway NSW 2007, Australia, Phone: +61-2-9514-7770, Fax: +61-2-9514-7722, mailto: YongWoong.Lee@uts.edu.au.

†Department of Statistics and Risk Management, Faculty of Business, Economics and Business Information Systems, University of Regensburg, 93040 Regensburg, Germany, Phone: +49-941-943-2588, Fax : +49-941-943-4936, mailto: Daniel.Roesch@wiwi.uni-regensburg.de.

‡Finance Discipline Group, UTS Business School, University of Technology, Sydney, PO Box 123, Broadway NSW 2007, Australia, Phone: +61-2-9514-7724, Fax: +61-2-9514-7722, mailto: Harald.Scheule@uts.edu.au.

§We gladly acknowledge the valuable research assistance of Michael Tjendara. Many thanks to the participants of the seminars at the University of Edinburgh, the University of Hannover and the 2013 Credit Modelling Forum, Sydney. The support by the Centre for International Finance and Regulation (CIFR, project number E001) is gratefully acknowledged. CIFR is funded by the Commonwealth and NSW Governments and supported by other Consortium members (see www.cifr.edu.au).

1 Introduction

In March 2013, the Reserve Bank of New Zealand has proposed to increase the 15% Basel II/III asset correlation for mortgage loans with higher loan-to-value ratios (LTV) (see Irvine, 2013). The proposal includes an asset correlation of 20% for loans with an LTV between 80% and 90% and an asset correlation of 21% for loans with an LTV in excess of 90%. These changes aim to reflect a higher exposure to systematic risk for high LTV loans.

While this proposal applies to New Zealand, it is an interesting starting point for a global discussion as (i) in particular high risk loans (e.g., high LTV or sub-prime loans) were exposed to large credit losses in the US during the Global Financial Crisis (GFC), and (ii) our knowledge on the exposure of mortgage loans to systematic risk is limited.

The industry measures the exposure to systematic risk with the concept of asset correlations and the link between asset correlations and default probabilities is controversial. Prior literature has focused on corporate loans. Grundke (2008) provides an overview of empirical studies on asset correlations including data sources and estimation methods for corporate exposures and the relationship of correlations to default probabilities and firm size. Sironi and Zazzara (2003), Lopez (2004), and Das et al. (2007) find that asset correlations decrease with increasing probabilities of default (PD) and decreasing size of Italian and US corporate loans. Contrary to this, Dietsch and Petey (2004) find a positive relationship of asset correlation and default probability for European small and medium enterprises. In addition, some papers explain the time-varying co-movement of defaults based on macroeconomic variables and latent risk factors (see e.g., Duffie et al. (2009) who model latent factors conditional on observable macroeconomic information, i.e., frailty). Demchuk and Gibson (2006) use exclusively macroeconomic variables for modeling US corporate default rates. Koopman et al. (2011) use latent risk factors and macroeconomic variables explaining US corporate default rates.

Empirical evidence on systematic risk of mortgage loans is limited. Most research has focused on the prediction of mortgage defaults and loss rates using borrower, loan and macroeconomic characteristics. For instance, Elul et al. (2010) and Goodman et al. (2010) investigate the impact of negative equity, illiquidity and unemployment on predicting mortgage default.

Amromin and Paulson (2009) evaluate the relative impact of borrower, loan and macroeconomic characteristics on mortgage defaults and identify real estate prices as an important risk driver. Rajan et al. (2013) show that the deterioration of the accuracy of statistical default prediction model is triggered by the change in lender behavior as the level of securitization increases. Crook and Banasik (2012) forecast mortgage default rates based on the mortgage rate and the real house price index. Cowan and Cowan (2004) model the variance of default events for bank-internal credit scores, FICO scores and other mortgage information. Jimenez and Mencia (2009) propose a state space model linking the default rates with both macroeconomic variables and a frailty risk factor for retail loans.

The contributions of this paper relative to the literature are as follows. Firstly, we focus on mortgage default which is an important credit exposure class with limited research on systematic risk. The paper provides a multivariate predictive model for mortgage default probabilities by incorporating significant time-varying loan, borrower level as well as macroeconomic information. Scores provided by Fair Isaac and Co. (FICO), current LTV ratios and debt-to-income ratios are most predictive. We are first in kind to control for real estate prices by computing the current LTV ratio (CLTV). Prior studies are based on LTV at origination. We analyze the in-sample and out-of-time prediction quality of this predictive model for mortgage default. Secondly, our paper is first to measure total mortgage risk and decompose the total risk into systematic and non-systematic class-specific risk by controlling for the credit quality (i.e., the default threshold). The structural single factor model underlying the Basel framework is extended via a state space model with multiple risk factors and associated parameters estimated by the maximum likelihood method with Kalman filtering. The resulting parameter estimates may provide benchmark values for regulatory input parameters. The paper extends Cowan and Cowan (2004) who focus on the variance of default events which endogenously co-moves with credit quality but do not measure systematic risk conditional on the credit quality. The paper also builds on the paper by Jimenez and Mencia (2009) which analyzes macroeconomic time series information but does not control for mortgage-specific information, as well as interaction between such information and the economy (e.g., CLTV)

and the clustering of risk with regard to the credit quality. The decomposition of total risk into systematic and non-systematic risk is important as some banks are able to diversify the non-systematic risk of different classes while others are not able to diversify due to inefficient lending markets. Banks often have a business model that concentrates on lending to a particular risk segment such as prime or sub-prime loans, primary lending or refinancing. Thirdly, our study extends the literature in relation to the econometric framework as multiple latent variables are included in a model and auto correlations are modeled through first order autoregressive processes. This allows banks to quantify past realizations of systematic risk and include these into their predictive mortgage models. Fourthly, the paper analyzes the time series properties for all parameters and pre-crisis as well as crisis information.

The study is important as mortgage loans are the largest asset class in commercial banking and losses in relation to mortgage loans have triggered the failure of a large number of financial institutions in recent years. The understanding of the exposure to systematic and to non-systematic risks is central to the efficient allocation of capital in relation to minimum reserve and capital requirements. Our study is based on a large loan-level data set on US sub-prime mortgage loans.

The paper's findings are as follows. Firstly, a mortgage's total risk relates to the probability of default in a smile-shaped pattern: the total risk first decreases and then increases with increasing unconditional default risk. Note that the large changes in mortgage default rates which were observed during the GFC do not necessarily imply high exposures to systematic risk. This is due to the dependence of the default rate volatility on both the unconditional probability of default and systematic risk as shown in Gordy (2000). However, the argument holds for given unconditional PD level, which is why we analyze systematic risk and non-systematic risk by controlling for credit quality. Secondly, systematic risk increases as the credit rating of mortgage decreases, while non-systematic risk decreases as the credit rating of mortgage decreases. Lower rated mortgages are more sensitive to systematic risk than higher rated mortgages. Thirdly, total risk increases during and after the GFC. However, the increase of total risk results from non-systematic risk for the lowest risk class and from systematic risk

for the remainder of the risk classes. Fourthly, the impact of systematic risk is persistent whereas the impact on non-systematic risk is temporary. Finally, the regulatory capital given the current asset correlation regime of 15% is found to be sufficient for all risk classes and all observation periods.

The remainder of this paper is organized as follows. Section 2 develops a methodological approach to measure PD, systematic risk and non-systematic risks and their relationship. Section 3 introduces the data and provides descriptive statistics for mortgagee, mortgage collateral and economic characteristics and presents estimation results and empirical findings. PDs are estimated via a point-in-time Probit model using information available at origination and observation times for mortgages. Information relates to mortgagees, collateral and the economy. Point-in-time models consider time-varying information, since default probabilities change with the state of the economy.¹ Consistent with current industry practice, rating classes are assigned based on the estimated probabilities of default (or alternatively credit scores). Then the rating class-level historical default rates are simultaneously modeled by a state space model with multiple risk factors: systematic and class-specific risk factors for investigating systematic and non-systematic risks for mortgage exposures. Consistent with the data, all risk factors are assumed to be autoregressively correlated to capture the time dependence of historical default rates. The state space model is estimated for various time periods to examine the serial change of unconditional PD, systematic and non-systematic risks. Section 4 calculates the regulatory capital given the PD estimates and investigates the adequacy of regulatory capital for mortgage portfolios under the current regime and model-implied as well as data-implied exposures to total risk. Finally, Section 5 concludes and discusses implications for prudential regulation.

2 Methodology

The methodological procedure of this study is as follows. Section 2.1 estimates individual PDs of mortgages via a Probit model incorporating various borrower, loan, collateral and macroe-

¹Compare Loeffler (2004) for a comparison of default probability models.

conomic characteristics for different estimation periods. Section 2.2 explains the categorization of mortgage loans into ten risk classes based on the estimated PDs. In Section 2.3, the conditional PD is derived for autoregressively correlated risk factors. Section 2.4 establishes a state space model with multiple latent risk factors for estimating class-wide PDs, systematic risks and non-systematic risks.

2.1 Mortgagee default process and unconditional probability of default

Based on the theory of Merton (1974), a mortgage borrower defaults on a loan when the latent asset value falls below the nominal amount of debt at maturity (the default threshold). We follow the contributions by Gordy (2000, 2003) and Vasicek (1991, 1987) and assume that idiosyncratic risk is fully diversified in a bank portfolio which is infinitely granular.²

The asset return of mortgage borrower i ($i = 1, \dots, I$) in period t ($t = 1, \dots, T$) is driven by a common time-specific systematic risk factor F_t and an idiosyncratic (i.e., diversifiable) factor $\epsilon_{i,t}$:

$$V_{i,t} = \sqrt{\rho_i}F_t + \sqrt{1 - \rho_i}\epsilon_{i,t}, \quad (1)$$

where F_t and $\epsilon_{i,t}$ are standard normally distributed and independent from each other, with standardized weights $\sqrt{\rho_i}$ and $\sqrt{1 - \rho_i}$ with values between zero and one.³ Note that ρ_i is the exposure to systematic risk and known as the asset return correlation in the literature.

A default event occurs when and if the asset return $V_{i,t}$ on assets falls below threshold $h_{i,t-1}$:

$$V_{i,t} < h_{i,t-1} \quad (2)$$

with $h_{i,t-1} = \beta \mathbf{x}_{i,t-1}$, where $\mathbf{x}_{i,t-1} = (x_{i,t-1}^1, x_{i,t-1}^2, \dots, x_{i,t-1}^p)'$ is a information on the mortgage borrower, the mortgage loan, the loan collateral and the economy which is observable in time

²This model is consistent with the model for determining regulatory capital of banks applied in the Basel II and III regulation.

³Please note that a non-linear transformation of $V_{i,t}$ ensures that distributions of risk measures such as probability of default, default rate or loss rate are highly skewed. In other words, the propensity of large losses is much greater than a normal distribution would suggest.

t and before. β is a vector of the respective sensitivities.

The borrower default is modeled by the indicator

$$D_{it} = \begin{cases} 1 & \text{borrower } i \text{ defaults in period } t \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The unconditional default probability of the borrower is given by:

$$\mathbb{P}(D_{i,t} = 1) = \mathbb{P}(V_{i,t} < h_{it-1}) = \Phi(h_{i,t-1}) = \Phi(\beta \mathbf{x}_{i,t-1}). \quad (4)$$

2.2 Borrower categorization

It is econometrically challenging to estimate the exposure to systematic risk (ρ_i) on a borrower level due to (i) the unavailability of loan-level borrower asset values and (ii) the binary character of the dependent default variable. Like corporate borrowers, mortgagee asset values are unknown to the financial institution. Therefore, we categorize the borrowers into risk classes and estimate the exposure to systematic risk in a second step. Once the unconditional PD is estimated by the Probit model in Equation (4), mortgages are split into G risk classes. The smaller the index of a risk class, the lower the default probability. Mortgages in a given risk class are assumed to be homogeneous in terms of unconditional PD and their exposures to systematic and non-systematic risks.

2.3 Class level conditional probability of default

For an assigned mortgage borrower i in risk class g , the asset value return at time t is extended to a class-specific non-systematic risk component and an idiosyncratic error term:

$$V_{i,g,t} = \sqrt{\rho_g} F_t + \sqrt{1 - \rho_g} (\sqrt{\alpha_g} Z_{g,t} + \sqrt{1 - \alpha_g} \epsilon_{i,t}), \quad (5)$$

where ρ_g is a mortgage borrower i 's systematic sensitivity to a systematic risk factor F_t and α_g is the factor loading to a class-specific risk factor $Z_{g,t}$. Both parameters are defined between zero and one. Note that parameter sensitivity to the systematic risk factor is now group-

specific, i.e., $\sqrt{\rho_g}$. The idiosyncratic factor $\epsilon_{i,t} \sim N(0, 1)$ is independent from F_t and $Z_{g,t}$. All risk factors are assumed to be unobservable and their unconditional distributions are supposed to be standard normal. These assumptions are common in the credit risk models since the introduction of CreditMetrics (2007) and consistent with the model assumption in current prudential regulation as shown in McNeil et al. (2005) and Lando (2003). Suppose that $\tilde{V}_{i,g,t}$ and \tilde{h}_g are the unknown true asset return and default threshold of mortgage i in risk class g at time t , respectively. McNeil et al. (2005) show that the two threshold models $(V_{i,g,t}, h_g)$ and $(\tilde{V}_{i,g,t}, \tilde{h}_g)$ are equivalent if the default probabilities coincide, i.e. $\mathbb{P}(V_{i,g,t} \leq h_g) = \mathbb{P}(\tilde{V}_{i,g,t} \leq \tilde{h}_g)$, and if the two asset returns have the same multivariate dependence structure between mortgages (multivariate normal herein). Note that if $\alpha_g = 0$ then ρ_g represents the pairwise asset correlation of the asset value return process of two borrowers in class g which is consistent with the exposure to a single systematic risk factor underlying the Basel framework.

For the asset return process in Equation (5), the total risk of a mortgage borrower can be represented as the additive sum of systematic risk, class-specific non-systematic risk and idiosyncratic risk:

$$\begin{aligned} \sigma_g &= \text{var}(V_{i,g,t}) \\ &= \underbrace{\rho_g}_{\text{systematic risk}} + \underbrace{(1 - \rho_g)\alpha_g}_{\text{non-systematic risk}} + \underbrace{(1 - \rho_g)(1 - \alpha_g)}_{\text{idiosyncratic risk}} = 1. \end{aligned}$$

Mortgage credit portfolios consist of large numbers of loans with almost equal weights (given the same vintage). Therefore idiosyncratic risk $(1 - \rho_g)(1 - \alpha_g)$ is diversified away in such portfolios. Thus, the total risk (σ_g) of risk class is simply composed of systematic risk (ρ_g) and non-systematic risk ($\tilde{\alpha}_g$):

$$\sigma_g = \rho_g + \tilde{\alpha}_g \leq 1, \tag{6}$$

where $\tilde{\alpha}_g = (1 - \rho_g)\alpha_g$. The asset correlation between two borrowers i and j within the same risk class is equal to the total risk since

$$\text{corr}(V_{i,g,t}, V_{j,g,t}) = \rho_g + (1 - \rho_g) \alpha_g.$$

The asset correlation between two borrowers i and j in different risk classes is

$$\text{corr}(V_{i,g,t}, V_{j,g',t}) = \rho_g \rho_{g'}.$$

We denote $D_{i,g,t}$ as the default indicator for mortgage i at time t , taking either zero or one for a default threshold $h_{g,t-1}$, i.e.,

$$D_{i,g,t} = \begin{cases} 0, & \text{if } V_{i,g,t} \leq h_{g,t-1} \\ 1, & \text{otherwise.} \end{cases} \quad (7)$$

Then the unconditional PD of risk class g is

$$P(V_{i,g,t} \leq h_{g,t-1}) = \Phi(h_{g,t-1}),$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function. Note that the default is now class-specific due to our homogeneity assumption.

Conditional on $F_t = f_t$ and $Z_{g,t} = z_{g,t}$, the PD of mortgage i in the risk class g at time t is given by

$$\begin{aligned} p_g(f_t, z_{g,t}) &= \mathbb{P}[V_{i,g,t} \leq h_{g,t-1} | f_t, z_{g,t}] \\ &= \Phi\left(\frac{\Phi^{-1}(p_g)}{\sqrt{1 - \rho_g} \sqrt{1 - \alpha_g}} - \frac{\sqrt{\rho_g}}{\sqrt{1 - \rho_g} \sqrt{1 - \alpha_g}} f_t - \frac{\sqrt{\alpha_g}}{\sqrt{1 - \alpha_g}} z_{g,t}\right), \end{aligned} \quad (8)$$

where p_g is a class-wide unconditional PD of mortgages in the risk class g satisfying $p_g = \Phi(h_{g,t-1})$.

Given the total number of mortgages in the risk class g , $N_{g,t}$, the default rate of the risk class g at time t is

$$r_{g,N_{g,t}} = \frac{\sum_{i=1}^{N_{g,t}} D_{i,g,t}}{N_{g,t}}. \quad (9)$$

The default rate $r_{g,N_{g,t}}$ converges to the conditional default probability $p_g(f_t, z_t)$ as $N_{g,t}$ increases to infinity tby the law of large numbers, i.e.,

$$r_{g,t} = \lim_{N_{g,t} \rightarrow \infty} r_{g,N_{g,t}} \rightarrow p_g(f_t, z_{g,t}).$$

Then the default rate is

$$r_{g,t} = \Phi \left(\frac{\Phi^{-1}(p_g)}{\sqrt{1 - \rho_g \sqrt{1 - \alpha_g}}} - \frac{\sqrt{\rho_g}}{\sqrt{1 - \rho_g \sqrt{1 - \alpha_g}}} f_t - \frac{\sqrt{\alpha_g}}{\sqrt{1 - \alpha_g}} z_{g,t} \right). \quad (10)$$

For the time dependence of the default rates, we assume an AR(1) process for the systematic and class-specific risk factors:

$$f_t = \beta_f f_{t-1} + \sqrt{1 - \beta_f^2} \nu_{f,t} \quad (11)$$

$$z_{g,t} = \beta_g z_{g,t-1} + \sqrt{1 - \beta_g^2} \nu_{g,t}, \quad (12)$$

where $-1 < \beta_f < 1$, $-1 < \beta_g < 1$, $\nu_{f,t} \sim N(0, 1)$ and $\nu_{g,t} \sim N(0, 1)$ for all $g = 1, 2, \dots, G$. Note that the unconditional mean and variance of the systematic and class-specific risk factors are zero and one, respectively.

2.4 State space model for default rate

The conditional default rate in Equation (10) can be re-parametrized as the Measurement Equation for linearity and simplicity:

$$\Phi^{-1}(r_{g,t}) = \phi_{0,g} + \phi_{1,g} f_t + \phi_{2,g} z_{g,t} \quad (13)$$

for all $g = 1, 2, \dots, G$, where $\phi_{0,g} = \frac{\Phi^{-1}(p_g)}{\sqrt{1-\rho_g}\sqrt{1-\alpha_g}}$, $\phi_{1,g} = -\frac{\sqrt{\rho_g}}{\sqrt{1-\rho_g}\sqrt{1-\alpha_g}}$ and $\phi_{2,g} = -\frac{\sqrt{\alpha_g}}{\sqrt{1-\alpha_g}}$. The State Equation is constructed by the systematic and class-specific risk factors as

$$\boldsymbol{\xi}_t = \mathbf{F}\boldsymbol{\xi}_{t-1} + \boldsymbol{\nu}_t, \quad (14)$$

where $\boldsymbol{\xi}_t = (f_t, z_{1,t}, z_{2,t}, \dots, z_{G,t})'$, \mathbf{F} is the diagonal matrix of the auto regressive coefficients in Equation (11) and Equation (12), i.e., $\mathbf{F} = \text{diag}(\beta_f, \beta_1, \beta_2, \dots, \beta_G)$, and the vector $\boldsymbol{\nu}_t = (\nu_{f,t}, \nu_{1,t}, \nu_{2,t}, \dots, \nu_{G,t})'$ is a white noise vector such that

$$\mathbb{E}(\boldsymbol{\nu}_t \boldsymbol{\nu}_\tau') = \begin{cases} \mathbf{Q} & \text{for } t = \tau \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

In the above, $\mathbf{Q} = \mathbf{I}_{G+1} - \mathbf{F}^2$ and \mathbf{I}_{G+1} is the $((G+1) \times (G+1))$ identity matrix. Note that there is no measurement error in the Measurement Equation (13) since we assume that every risk class includes a large enough number of homogeneous mortgages.⁴ The state space model for default rate with the Measurement Equation (13) and the State Equation in (14) can be generally estimated by maximum likelihood estimation with Kalman filtering (see Hamilton (1994) for details).

3 Empirical Results

3.1 Data

The paper is based on loan level data on US securitized non-agency mortgage loans. The data was collected from monthly loan tapes for residential mortgage-backed securities by International Financial Research and matches the data from Rajan et al. (2013), who show that up to 90% of all US sub-prime mortgage loans are securitized.

The data set comprises 4,978,076 loans observed at quarterly intervals from 2000 to 2012.

⁴In the empirical analysis, a state space model with measurement errors was also applied as a robustness check. The estimated variance of the measurement errors has no effect on the estimates of other parameters, although we do not present the outcomes in the paper. Thus, measurement error terms are not further discussed. The empirical results are available upon request.

We record 1,143,228 loan foreclosure events (many during the GFC in 2007 and thereafter). The total number of loan-quarter observations is 56,946,616. We define the default event as loan foreclosure. Observations with missing values in key variables were omitted from the analysis.

Table 1 summarizes the number of observations per origination and observation year. Consistent with Demyanyk and Hemert (2011) origination years immediately prior to the GFC and observation years during the GFC have a larger proportion of default events.

[Table 1 about here.]

Figure 1 plots the quarter-by-quarter total default rates from 2000 to 2012. The figure shows the default rate and average loss rate given default over time. The effect of the sub-prime crisis is extremely strong as the default rate increases during the crisis from 0.52% at 2004:Q2 to 3.90% at 2009:Q2. The grey bars indicate quarters which include a period of economic downturn as indicated by the National Bureau of Economic Research.

[Figure 1 about here.]

Loss rates given default (LGDs) are recorded at default for individual mortgages.⁵ The aggregated loss ratio given default in Figure 1 is calculated as the ratio of all losses over a given quarter to the outstanding account balance at the time of default for the quarter. The gap between the two lines is proportional to the recovery. The default rates and loss rates given default increase during economic downturns.

Mortgage information at origination

Mortgage-specific variables at origination include origination year (OY), ARM indicator (ARM), FICO score (FICO), original balance, original appraisal value of the collateral/property (OAV), original loan-to-value ratio, owner occupancy type (OWNOCCP) and dwelling type (DWLTYPE).

Table 2 and Table 3 describe these variables for default and non default observations.

[Table 2 about here.]

[Table 3 about here.]

The FICO score measures the credit worthiness of a borrower using information from

⁵Section 4 includes LGDs in the computation of Basel capital and implied capital.

borrower credit files. A higher FICO score implies a higher credit quality. The average FICO score of defaulted mortgages is 30 points lower than for non-defaulted mortgages. The average original balance is 3.6% larger for defaulted than for non-defaulted mortgages. The mean of appraisal values of the collateral (i.e., the financed real estate) is \$40,000 smaller for defaulted mortgages than for non-defaulted mortgage. The mean of original LTV ratio is 8% higher for defaulted mortgages than non-defaulted mortgages. The ARM indicator denotes whether or not a mortgage rate may be adjusted after it is issued. Generally speaking, the mortgage rate is linked to a reference rate such as LIBOR for adjustable rate mortgages. We assign a code of 1 to adjustable rate mortgages and a code of 0 to fixed rate mortgages. The default rate of adjustable rate mortgages is 1.7% higher than for fixed rate mortgages in Table 3. The purpose of owner occupancy is classified into three types: residence (code=1), investment (code=2), and others (code=0). The code for residence includes both primary and secondary residence, the code for others contains tenant, vacant and unknown. Mortgages for residence and investment are 0.6% less risky in terms of the default rate than mortgages for other purposes. Dwelling type includes single family homes (code=1), planned urban developments (code=2), and condominiums (code=3), and others (code=0). Mortgages for planned urban developments and condominiums are 0.1% riskier than other purposes in terms of the default rate.

Mortgage information at observation time

Mortgage-level variables at observation time t are current balance (CB), and current loan-to-value ratio (CLTV). The current balance of defaulted mortgages is on average \$27,000 larger than non-defaulted mortgages.

The current loan-to-value ratio of mortgages is calculated using the Case-Shiller Index for 20 major metropolitan statistical areas in the US as follows. We first compute the distance of individual mortgage properties to 20 US cities for which a Case-Shiller home price index is available (via a mortgage's property zip-code). If a zip-code is not available, then we use the zip code of the largest city in the state of the property in terms of population. Then, we select the closest Case-Shiller index for a mortgage property and approximate the current

loan-to-value ratio at observation time t as

$$CLTV_{i,t} = \frac{CB_{i,t}}{CAV_{i,t}}.$$

The current appraisal value (CAV) is modeled by the following approximation:

$$CAV_{i,t} = OAV_i \cdot \frac{CSI_{i,t}}{CSI_{i,t_0}},$$

where CSI_{i,t_0} and $CSI_{i,t}$ denote the Case-Shiller Index at the origination time t_0 and the observation time t , respectively. For mortgages where the property zip code and state are unavailable, we use the 10-city composite Case-Shiller Index (CSI).⁶ Figure 2 plots the CSI and shows the spectacular decline during the GFC. CLTV is 27% higher for defaulted mortgages than for non-defaulted mortgages in Table 2.

[Figure 2 about here.]

Macro economic information

Next to the Case-Shiller indices, we use the country-level debt-to-income ratio (DTI) from the Federal Reserve Bank. DTI is an estimate of the ratio of debt payments to disposable personal income. Debt payments consist of the estimated required payments on outstanding mortgage and consumer debt. We chose DTI as it is common in banking and reflects economic information such as growth as well as employment levels. Figure 2 shows the quarterly reported country-level DTI which is correlated with the level of default rate in Figure 1.

3.2 Probit model

Based on the literature, we tested a wide range of mortgage variables and included the following variables in the Probit model: CLTV (positive effect from Amromin and Paulson (2009), Elul et al. (2010) and Goodman et al. (2010)), ARM (positive effect from Amromin and Paulson (2009)), DWLTYPE (mixed effect from Elul et al. (2010) and Foote et al. (2008)) and OWNOCPP (mixed effect from Amromin and Paulson (2009)) for loan characteristics and

⁶We use the 10-city composite Case-Shiller Index rather than the 20-city composite Case-Shiller Index as the latter is only available for years after 2000.

FICO score (negative effect from Elul et al. (2010)) for borrower characteristics, and DTI (positive effect from Amromin and Paulson (2009) and Goodman et al. (2010)) for macroeconomic characteristics.

Furthermore, we include dummy variables for the mortgage origination year which capture vintage effect and year-specific economic and market conditions. Demyanyk and Hemert (2011) show that mortgage loans which were originated just before the GFC have higher default rates than loans originated in earlier years.

Table 4 presents the results from the point-in-time Probit model for the various estimation periods from 2000:Q2 to 2006:Q4 for the first model to 2000 (Q2) to 2012 (Q4) for the last model.

[Table 4 about here.]

Our results in Table 4 are consistent with the empirical findings in the literature. Most prominently, PDs increase with CLTV and DTI and decrease with FICO score. The marginal effect of those variables on actual default is highly significant over time. PDs also vary with dwelling and owner occupancy types. As expected, investment mortgage loans are riskier than residential mortgage loans and single family mortgages are less risky than others.

The goodness-of-fit statistics (Area Under the Receiver Operating Characteristics curve (AUROC), R^2 and pseudo R^2) generally decrease as the estimation time horizon expands towards 2012. This decline in accuracy of the Probit model is consistent with the finding by Rajan et al. (2013) who find that default prediction errors increase over time and hypothesize that this is due to the degree of securitization on mortgages and lending standards. In addition, the trend persists after the GFC. Note that other explanations such as population changes in terms of number of mortgages or risk characteristics are also plausible.

The measure AUROC is based on the area under receiver operating characteristic curve (ROC). ROC curves are a common performance measure for ordinal rating systems. Figure 3 displays the true positives rate on the y-axis (sensitivity, which measures how well the model identifies default) and the false positives rate on the x-axis (one minus the specificity, which measures how well a model identifies non-default).

[Figure 3 about here.]

The better a ranking/rating system is able to attribute non-impairment outcomes with lower ranks and impairment outcomes with higher ranks, the larger the area under the ROC curve is (AUROC, see Agresti, 1984). The areas under the ROC curve are reported in Table 7 for the estimation and the prediction years.

These measures are common in the literature and can also be transformed into Gini coefficients and therefore be expressed as cumulative impairment frequencies for ratings, which are sorted from high to low risk classes (compare, e.g., Vassalou and Xing (2004) and Chava and Jarrow (2004)).

3.3 Default rates of risk classes

Based on the PD estimates from the Probit model, we construct ten risk classes with a width of 30 basis points for the first nine classes. Thus mortgages with PDs between 0 and 0.3% are assigned to the first class while mortgages with PDs greater than 2.7% are assigned to the tenth (i.e., the last) class.

The PD boundaries are selected to ensure that all classes have large observation and default numbers. We calculate the actual default rate for the risk class g at time t as in Equation (9) for all classes.

Figure 4(a) plots the quarter-by-quarter default rates by risk class at the last estimation time horizon, i.e., from 2000:Q2 to 2012:Q4.

[Figure 4 about here.]

Figure 4(a) shows that riskier classes have higher default rate levels with higher volatilities over the entire time horizon. During the sub-prime crisis, default rates rise and absolute change rates are monotone in unconditional PD levels and therefore risk classes. This observation may imply that mortgages with higher PD are more vulnerable to external economic shocks and thereby more sensitive to the systematic risk, than mortgages with lower PDs. This is the working hypothesis for our study.

3.4 Results of state space model

The parameters of the state space model with the Measurement Equation (13) and the State Equation (14) are estimated by the maximum likelihood method with Kalman filtering. The original parameters are computed by reversing the re-parametrization in relation to the Measurement Equation (13). Standard errors are calculated based on the delta method.

[Table 5 about here.]

Table 5 presents the parameter estimates of the state space model for the last estimation period from 2000 to 2012. Most of the parameters are highly significant. In Panel A, $\phi_{0,g}$ indicates the unconditional mean of the Measurement Equation and is proportional to the default threshold for risk class g . As expected, the default threshold increases as the credit quality decreases from class 1 to class 10. The parameter $\phi_{1,g}$ relates to the sensitivity of class-wide default to the systematic risk factor f_t . The worse the rating class, the higher the sensitivity to the systematic risk factor. The estimates of $\phi_{2,g}$ show that the worse the rating class, the less the sensitivity to the class-specific risk factor. These findings imply that the better rating classes are more exposed to the class-specific risk and the worse rating classes are more exposed to economic downturns due to their high sensitivity to the systematic risk factor. The estimated auto regressive parameters for the risk factors in Panel B provide evidence for a strong time dependence of risk factors and thereby, the auto regressive dependence of systematic risk realizations.

3.4.1 Unconditional PD and total risk

The parameters of asset value returns in Equation (5) are calculated in Table 6 by applying the associated re-parametrization in the Measurement Equation (13) given the estimates of state space model in Panel A of Table 5. The unconditional PD (p_g) and the systematic risk measure (ρ_g) monotonically increase as the credit qualities of risk classes deteriorate, while the total risk (σ_g) and the non-systematic risk (α_g and $\tilde{\alpha}_g$) decline with the unconditional PD.

[Table 6 about here.]

For validation of our results, the estimated unconditional PD, the total risk and the sys-

tematic risk from the single risk factor model in Vasicek (2002) are compared. Vasicek (2002) is equivalent to our model with $\alpha_g = 0$ for all g in Equation (5), i.e.,

$$V_{i,g,t} = \sqrt{\tilde{\rho}_g} F_t + \sqrt{1 - \tilde{\rho}_g} \epsilon_{i,t}, \quad (15)$$

where the systematic risk factor F_t and idiosyncratic factor $\epsilon_{i,t}$ follow independent standard normal distributions. Note that there is no class-specific risk and that the idiosyncratic risk should be diversified away for homogeneous portfolios in Equation (15). As such, the total risk is the same as the systematic risk, i.e. $\sigma_g = \tilde{\rho}_g$ for the single factor model.

Figure 5 shows the total risk and the historical average of default rate of the ten risk classes.

[Figure 5 about here.]

The total risk of the state space model and the Vasicek model decrease as the credit qualities of risk classes deteriorate. On the other hand, Figure 4(a) presents that the worse rated classes experienced more volatile default rates than the better rated classes. Note that the volatility of defaults depends on both the unconditional PD (p_g) and the total risk ($\tilde{\rho}_g$) under the single factor model in Equation (15) as shown in Gordy (2000),⁷ i.e.,

$$\text{var}(p_g(f_t) | p_g, \tilde{\rho}_g) = \Phi_2(\Phi^{-1}(p_g), \Phi^{-1}(p_g); \tilde{\rho}_g) - p_g^2, \quad (16)$$

where $\Phi_2(\cdot, \cdot; \tilde{\rho})$ is the cumulative density function of the bi-variate normal distribution with the correlation $\tilde{\rho}_g$ which is the total risk or asset correlation of the risk class g and p_g is the unconditional default probability.

Cowan and Cowan (2004) model the variance of the conditional default probability, or alternatively the variance of the default rate or number of defaults. Note that this measure generally depends on the systematic risk and the level of credit quality. Figure 6 plots the estimated default rate volatility of the state space model and Vasicek model obtained by Equation (16) against the historical average of default rate of the risk classes. The volatility of the default rate depends on both systematic risk and the level of the credit quality, i.e. the

⁷Note that systematic risk ($\tilde{\rho}_g$) and total risk (σ_g) are the same in a single factor model as mentioned earlier.

unconditional probability of default. The volatility increases as the credit quality of the risk class decreases while the total risk decreases as the credit quality of the risk class decreases. The high volatility of high risk classes is related to the unconditional PD rather than to the total risk. The total risk plays a relatively larger role for the volatility of the default rate for the low risk classes which have small PDs but high total risk.

[Figure 6 about here.]

3.4.2 Systematic and non-systematic risks

Figure 7 plots the estimated systematic and non-systematic risks against the average of default rate for the ten risk classes. The term non-systematic risk implies that banks are able to diversify this risk in efficient lending markets. However, banks often have a business model that concentrates on lending to a particular risk segment (and therefore unconditional PD levels) such as prime or sub-prime loans, primary lending or refinancing. Such business models may imply that banks find it difficult to diversify their loan portfolios across risk classes and the class level non-systematic risk may actually be interpreted as a systematic risk for such banks. Figure 7 shows that the systematic risk increases and the non-systematic risk decreases when the credit quality of risk class deteriorates.

[Figure 7 about here.]

These observations suggest that the simultaneous influence of mortgage characteristics and systematic risk is not additive due to the interactions between them. In other words, the effect of systematic risk on default differs depending on the unconditional PD, i.e., the risk class. Economic downturns may increase unconditional PDs and therefore the exposure to systematic risk. This may in turn result in a fortification of the conditional PDs.

Figure 5 presents the decreasing relationship between systematic risk and the historical average of class-wide unconditional PDs according to the risk classes under the single factor model. The single factor ignores the class-specific behavior of default rate for mortgage exposures and may result in a misleading interpretation.

3.4.3 Serial change of risks during the GFC and thereafter

Consistent with this paper, Dietsch and Petey (2004) discuss the effect of the length of the time series of data on estimated asset correlations. They point out that historical data should cover at least one complete cycle of the economy in order to avoid a downward estimation bias of asset correlation. In addition, Lucas (1995) suggests that estimated asset correlations over short estimation horizons are generally low since defaults may occur due to idiosyncratic risk.

[Figure 8 about here.]

Figure 8 plots the estimated unconditional PD and total risk per estimation period. Figure 8(a) shows that the effect of estimation time horizon is much stronger on the estimation of total risk than the expected loss (the unconditional PD). Thus, the length of estimation time horizon is more important for estimating risk measures. These results are consistent with Lucas (1995) and Dietsch and Petey (2004).

[Figure 9 about here.]

Figure 9 shows the evolution of systematic and non-systematic risks of ten risk classes per estimation period. During the crisis, the systematic risk gradually increases across all risk classes except the lowest risk class.

The non-systematic risk also increases during the crisis but reverts to lower levels except for the lowest risk class. Figure 8(b) implies that total risk increases due to the crisis and stays at a high level thereafter for all risk classes. Thus, these observations imply that the increase of total risk stems from both systematic and non-systematic risks during the crisis and to a larger degree from systematic risk after the crisis.

3.5 Robustness checks

3.5.1 Prediction accuracy of the Probit model

This paper has used the results from the Probit models as a basis to measure unconditional risk with regard to systematic and non-systematic risk. The model has been assessed in terms of in-sample fit statistics (see Table 4). We predict the PD of mortgages at the prediction year by providing the Probit model at the estimation time with changed borrower and loan

characteristics of mortgages at a prediction year to assess the model accuracy out-of-time. We then calculate the AUROC.

Table 7 presents these out-of time AUROCs for different combinations of estimation and prediction years.

[Table 7 about here.]

Like the in-sample AUROC in Table 4, Table 7 is consistent with Rajan et al. (2013) who point out the deterioration of accuracy of a default prediction model over time as securitization of mortgage increases. Our default prediction model extends Rajan et al. (2013) by incorporating additional mortgage characteristics into their default prediction model, which analyses FICO scores and LTVs. In addition, our paper relates to an extended post-crisis observation period.

Figure 10 compares the state space model-implied unconditional PDs with the realized default rates per risk class and model. The lower the total risk of a class, the closer to the scatters to the diagonal, which implies that default rate predictions are closer to default rate realizations.

[Figure 10 about here.]

Both the systematic and unsystematic risk are modeled by AR(1) processes to account for the intrinsic auto correlation. Figure 11 shows the conditional PDs, conditional on the time-lagged and the contemporary error terms for forecasting at 2008:Q1. Controlling for these random terms, the fit relative to the diagonal improves considerably. The conditional PDs are identical to the diagonal if both time-lagged and contemporary error terms are included. In other words, the predictive quality of the default probability model increases if the autoregressive structure of the systematic and non-systematic risk factors are taken into account when estimating the conditional probability of default. The unconditional model in 2008:Q1 deviate to a larger degree from the diagonal than the model conditional on the realization of the time lagged systematic and non-systematic risk factors: The maximum absolute deviation for the unconditional 2008 model is 2.19% for class 10. The maximum absolute deviation for the 2008 model conditional on the realization of the time lagged systematic and non-systematic

risk factors is 1.35% for class 10.

[Figure 11 about here.]

3.5.2 Risk classification

We checked the impact of class boundaries and widths and therefore different observations and default numbers on the empirical results. The systematic risk factor (f_t) and class-specific risk factors ($z_{g,t}$) are induced by equally-weighted default rates of ten risk classes in the state space model. Thus, we test our empirical outcomes for an alternative risk classification with an equal number of mortgages in every risk class, i.e., $\sum_{t=1}^T N_{g,t} = N$ for all g . The results of this alternative risk classification are in essence the same and are available upon request.

4 Implication on regulatory capital

This section assesses the capital adequacy of the current regulatory regime and a regime which takes into account asset correlations based on our (lower) measures of total risk. As we measure default risk on a quarterly basis these sensitivities are not directly comparable to the Basel asset correlations and we proceed as follows:

Basel capital

We use the Probit model from the previous section to analyze the adequacy of regulatory capital by forecasting the default probabilities of future periods. We predict PDs for future periods one year ahead. In other words, we apply the latest Probit model for a period and apply future borrower, loan and macroeconomic information.⁸ The annual PD is inserted into the formula of the Basel Internal Ratings-Based approach. The regulatory capital requirement for mortgage i at time t over a year risk horizon can be calculated by

$$C_{i,t} = EAD_{i,t} \cdot LGD_{g,t} \cdot \left\{ \Phi \left[\frac{\Phi^{-1}(p_{i,t}^A) + \sqrt{\rho_{Basel}} \Phi^{-1}(0.999)}{\sqrt{1 - \rho_{Basel}}} \right] - p_{i,t}^A \right\},$$

⁸Note that for the quarters prior to 2006, the 2006 model is applied in-sample.

where $EAD_{i,t}$ is the exposure-at-default, which is equal to variable CB, $LGD_{g,t}$ is the downturn LGD of class g . According to Basel (2006), the downturn LGD at time t is calculated as the maximum of the exposure-weighted average loss rate given default during the estimation period subject to a Basel floor of 10%. $p_{i,t}^A$ is the annualized unconditional PD of mortgage i at time t , respectively and ρ_{Basel} is given as 15%. Then the regulatory capital ratio for the risk class g at time t over a year risk horizon is:

$$C_{g,t} = \frac{\sum_{i=1}^{N_{g,t}} C_{i,t}}{\sum_{i=1}^{N_{g,t}} EAD_{i,t}}. \quad (17)$$

Implied capital - unconditional PD

For mortgage i in the risk class g , the implied capital based on the predicted PD ($p_{i,t}$) and the estimated total risk (σ_g) is obtained by

$$\tilde{C}_{i,t} = EAD_{i,t} \cdot LGD_{g,t} \cdot 4 \cdot \left\{ \Phi \left[\frac{\Phi^{-1}(p_{i,t}) + \sqrt{\sigma_g} \Phi^{-1}(0.999)}{\sqrt{1 - \sigma_g}} \right] - p_{i,t} \right\},$$

where the conditional PD is multiplied by the annualizing factor 4 since $p_{i,t}$ and σ_g are on a quarterly basis.

The implied capital ratio for the risk class g at time t over a year risk horizon is:

$$\tilde{C}_{g,t} = \frac{\sum_{i=1}^{N_{g,t}} \tilde{C}_{i,t}}{\sum_{i=1}^{N_{g,t}} EAD_{i,t}}. \quad (18)$$

Implied capital - conditional forecast PD

Conditional on f_{t-1} and $z_{g,t-1}$, the implied capital based on the forecast PD and the estimated systematic risk exposures (ρ_g) and non-systematic risk exposures (α_g) is obtained by

$$\hat{C}_{i,t} = EAD_{i,t} \cdot LGD_{g,t} \cdot 4 \cdot \left\{ \Phi \left[\frac{\Phi^{-1}(p_{i,t})}{\sqrt{1 - \rho_g} \sqrt{1 - \alpha_g}} - \frac{\sqrt{\rho_g}}{\sqrt{1 - \rho_g} \sqrt{1 - \alpha_g}} \beta_f f_{t-1} \right. \right. \\ \left. \left. - \frac{\sqrt{\alpha_g}}{\sqrt{1 - \alpha_g}} \beta_g z_{g,t-1} + \sqrt{\tilde{\sigma}_g} \Phi^{-1}(0.999) \right] - p_{i,t} \right\},$$

where

$$\tilde{\sigma}_g = \frac{\rho_g (1 - \beta_f^2) + \alpha_g (1 - \rho_g) (1 - \beta_g^2)}{(1 - \rho_g) (1 - \alpha_g)}$$

and the conditional forecast of PD is multiplied by the annualizing factor 4 since the parameters are on a quarterly basis.

The implied capital ratio for the risk class g at time t over a year risk horizon is:

$$\hat{C}_{g,t} = \frac{\sum_{i=1}^{N_{g,t}} \hat{C}_{i,t}}{\sum_{i=1}^{N_{g,t}} EAD_{i,t}}. \quad (19)$$

Loss ratio

We calculate the loss ratio given by

$$L_g = 4 \cdot \frac{\sum_{i=1}^{N_{g,t}} EAD_{i,t} LGD_{g,t} \cdot D_{i,t}}{\sum_{i=1}^{N_{g,t}} EAD_{i,t}}, \quad (20)$$

where $D_{i,t}$ is the default indicator in Equation (7). Note that a quarterly loss is annualized by multiplying the quarterly loss ratio by the annualizing factor 4.

Figure 12 plots the Basel regulatory capital ratio, implied capital ratios and the loss ratio per risk class. The Basel capital exceeds the realized losses for all risk classes and all years. Thus, the regulatory assumption of 15% for the asset correlation is sufficient to cover unexpected losses. However the regulatory approach may not be efficient. The empirical results may also support lower asset correlations, which are increasing with risk. Insufficient capital allocations are possible for sub-segments, or alternatively, if the PD model is not updated over time. Our model-implied capital ratios were able to provide a sufficient level of capital except for selected

periods during the GFC when the increase in loss risk outweighed past experience.

[Figure 12 about here.]

5 Conclusion and implications on prudential regulation

Credit portfolio losses and therefore bank capital are substantially affected by the exposure to systematic and non-systematic risks. This paper estimates PDs, total risk, systematic risk and non-systematic risk of mortgage exposures and analyzes the empirical relationship between these measures.

We find that the sum of systematic risk and non-systematic risk follows a smile shape. The decomposition of total risk into systematic and non-systematic risk shows that systematic risk increases and non-systematic risk decreases as the unconditional PD of mortgages increases. This implies a higher sensitivity of high PD loans to changes in the economy.

By extending the estimation period from 2006 to 2012, this paper also investigates the serial changes of total, systematic and non-systematic risks during the GFC and thereafter. The sub-prime crisis raises the total risk for all mortgages. However, the increase of total risk is mainly caused by non-systematic risk for mortgages in the lowest risk class and systematic risk for mortgages in other risk classes. The exposure to systematic risk increases permanently while non-systematic risk increases temporarily. Finally, this paper evaluates regulatory capital adequacy for mortgage exposures given the current regime of an asset correlation of 15% and finds that the regulatory assumption on asset correlation is sufficient to cover unexpected loss. A model-implied (and data-implied) regulatory capital approach may warrant a more effective capital allocation provided that PD models are periodically updated and that the time-lagged realizations of systematic risk and non-systematic risk are taken into account.

This paper also emphasizes an important role of the length of historical data for regulatory capital estimation. Basel (2006) proposes at least 5 years of data to estimate loss characteristics such as EAD, PD and LGD. We start with a slightly longer time series and demonstrate the change of parameter estimates during and after the GFC by sequentially updating our models for extended time periods. In particular, risk measures are more sensitive than unconditional

PDs to the duration of data in the presence of business cycles.

References

- Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. Wiley, New York.
- Amromin, G. and Paulson, A. (2009). Comparing patterns of default among prime and subprime mortgages. *Economic Perspectives*, 33(2).
- Basel (2006). *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. Bank for International Settlements, Basel Committee on Banking Supervision, Basel.
- Chava, S. and Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, 8(4):537–569.
- Cowan, A. M. and Cowan, C. D. (2004). Default correlation: An empirical investigation of a subprime lender. *Journal of Banking & Finance*, 28:753–771.
- CreditMetrics (2007). *CreditMetrics (TM)-technical document*. RiskMetrics group.
- Crook, J. and Banasik, J. (2012). Forecasting and explaining aggregate consumer credit delinquency behaviour. *International Journal of Forecasting*, 28(1):145–160.
- Das, S. R., Duffie, D., Kapadia, N., and Saita, L. (2007). Common failings: How corporate defaults are correlated. *Journal of Finance*, 62:93–117.
- Demchuk, A. and Gibson, R. (2006). Stock market performance and the term structure of credit spreads. *Journal of Financial and Quantitative Analysis*, 41(4):863–887.
- Demyanyk, Y. and Hemert, O. V. (2011). Understanding the Subprime Mortgage Crisis. *Review of Financial Studies*, 24(6):1848–1880.
- Dietsch, M. and Petey, J. (2004). Should SME exposures be treated as retail or corporate exposures? A comparative analysis of default probabilities and asset correlations in French and German SMEs. *Journal of Banking & Finance*, 28:773–788.

- Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009). Frailty correlated default. *Journal of Finance*, LXIV(5):2089–2123.
- Elul, R., Souleles, N. S., Chomsisengphet, S., Glennon, D., and Hunt, R. (2010). What "triggers" mortgage default? *American Economic Review*, 100(2):490–494.
- Foote, C. L., Gerardi, K., and Willen, P. S. (2008). Negative equity and foreclosure: Theory and evidence. *Journal of Urban Economics*, 64(2):234–245.
- Goodman, L. S., Ashworth, R., Landy, B., and Yin, K. (2010). Negative equity trumps unemployment in predicting defaults. *Journal of Fixed Income*, 19(4):67–72.
- Gordy, M. B. (2000). A comparative anatomy of credit risk models. *Journal of Banking & Finance*, 24(1):119–149.
- Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, 12(3):199–232.
- Grundke, P. (2008). Regulatory treatment of the double default effect under the New Basel Accord: how conservative is it? *Review of Managerial Science*, 2:37–59.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, New Jersey.
- Irvine, S. (2013). Consultation paper: Review of bank capital adequacy requirements for housing loans (Stage One). *Reserve Bank of New Zealand*.
- Jimenez, G. and Mencia, J. (2009). Modelling the distribution of credit losses with observable and latent factors. *Journal of Empirical Finance*, 16:235–253.
- Koopman, S. J., Lucas, A., and Schwaab, B. (2011). Modeling frailty-correlated defaults using many macroeconomic covariates. *Journal of Econometrics*, 162:312–325.
- Lando, D. (2003). *Credit Risk Modelling - Theory and Applications*. Princeton University Press, Princeton.

- Loeffler, G. (2004). An Anatomy of Rating Through the Cycle. *Journal of Banking and Finance*, 28:695–720.
- Lopez, J. A. (2004). The empirical relationship between average asset correlation, firm probability of default and asset size. *Journal of Financial Intermediation*, 13(2):265–283.
- Lucas, D. J. (1995). Default correlation and credit analysis. *Journal of Fixed Income*, 4(4):76–87.
- McNeil, J. A., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton.
- Merton, R. C. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 7:141–183.
- Rajan, U., Seru, A., and Vig, V. (2013). The failure of models that predict failure: distance, incentives and defaults. *Journal of Financial Economics (forthcoming)*.
- Sironi, A. and Zazzara, C. (2003). The Basel Committee proposals for a new capital accord: Implications for Italian banks. *Review of Financial Economics*, 12(1):99–126.
- Vasicek, O. (1987). Probability of loss on loan portfolio. *Working Paper, KMV Corporation*.
- Vasicek, O. (1991). Limiting loan loss probability distribution. Working Paper. KMV Corporation.
- Vasicek, O. (2002). Loan portfolio value. *Risk*, December:160–162.
- Vassalou, M. and Xing, Y. (2004). Default risk in equity returns. *The Journal of Finance*, 59(2):831–868.

Tables

Table 1: Frequency counts of origination and observation year

This table shows the number of mortgages (N), the number of defaults (D) and the default rates (D/N) per origination and observation year. The origination year of mortgages issued before 1990 is replaced by 1990 for mapping to macroeconomic variables such as the Case-Shiller Index and debt-to-income ratio.

Year	Origination Year			Observation Year		
	N	D	D/N	N	D	D/N
1990	82,309	364	0.0044			
1991	4,840	40	0.0083			
1992	9,729	61	0.0063			
1993	24,089	166	0.0069			
1994	26,713	197	0.0074			
1995	45,714	415	0.0091			
1996	65,106	467	0.0072			
1997	527,899	1,335	0.0025			
1998	1,173,231	3,060	0.0026			
1999	934,544	6,109	0.0065			
2000	779,823	14,843	0.0190	140,200	2,189	0.0156
2001	633,570	9,403	0.0148	696,465	5,227	0.0075
2002	1,528,595	21,587	0.0141	695,448	5,757	0.0083
2003	4,910,205	34,887	0.0071	935,247	7,527	0.0080
2004	10,061,555	102,124	0.0101	2,077,628	12,955	0.0062
2005	14,871,698	263,433	0.0177	4,533,915	31,160	0.0069
2006	16,748,786	532,468	0.0318	7,783,329	77,033	0.0099
2007	4,129,991	143,669	0.0348	10,281,593	193,035	0.0188
2008	381,460	8,505	0.0223	9,046,203	282,476	0.0312
2009	2,906	61	0.0210	6,946,571	247,213	0.0356
2010	382	-	0.0000	5,487,139	130,573	0.0238
2011	1,352	-	0.0000	4,551,755	92,789	0.0204
2012	2,119	34	0.0160	3,771,123	55,294	0.0147
Total	56,946,616	1,143,228	0.0201	56,946,616	1,143,228	0.0201

Table 2: Descriptive statistics of metric variables

This table presents the descriptive statistics of mortgage-level metric variables at both origination and observation times for all observations and per default status. Original balance, original appraisal value and actual current balance are in \$1,000. Debt-to-income ratio is in percentage.

Variable	Non-default		Default		Total						
	Mean	STD	Mean	STD	Mean	STD	1%	Q1	Median	Q3	99%
Origination time											
FICO Score	683	70	653	67	682	70	512	636	687	738	802
Original Balance (in \$1,000)	248	224	257	197	248	223	18	95	180	347	1000
Original Appraisal Value (in \$1,000)	376	392	336	296	375	391	43	148	264	490	1718
Original LTV Ratio	0.70	0.24	0.78	0.17	0.70	0.24	0.00	0.65	0.80	0.80	1.00
Observation time											
Actual Current Balance (in \$1,000)	231	218	258	212	232	218	0	82	165	329	991
Current LTV Ratio	0.72	0.34	0.99	0.31	0.73	0.34	0.00	0.53	0.75	0.90	1.59
Macroeconomic											
Case-Shiller Index (CSI10)					165.36	34.34	104.01	147.22	157.92	191.00	226.17
Debt-to-income ratio					9.97	0.86	8.67	9.15	9.87	10.84	11.34

Table 3: Frequency counts of categorical variables

This table shows the number of mortgages (N), the number of defaults (D) and the default rates (D/N) according to categorical mortgage-level variables. The assigned codes are ARM (1) and FRM (0) for ARM indicator, residence (1), investment (2) and the others (0) for owner occupancy types and single family (1), planned urban developments (2) condominium (3) and the others (0) for dwelling types.

Variable	Code = 0			Code = 1			Code = 2			Code = 3		
	N	D	D/N	N	D	D/N	N	D	D/N	N	D	D/N
ARM Indicator	27,035,478	308,796	0.0114	29,911,138	834,432	0.0279						
Occupancy	1,196,042	30,940	0.0259	47,872,521	961,321	0.0201	7,878,053	150,967	0.0192			
Dwelling	10,480,778	203,891	0.0195	35,972,668	718,260	0.0200	6,578,788	139,258	0.0212	3,914,382	81,819	0.0210

Table 4: Parameter estimates for point-in-time Probit model

This table presents the parameter estimates of the Probit model for mortgage default given as

$$P(V_{i,t} \leq h) = \Phi \left(\beta_0 + \beta_1 FICO_i + \beta_2 ARM_i + \beta_3 DWLTYPE1_i + \beta_4 DWLTYPE2_i + \beta_5 DWLTYPE3_i + \beta_6 OWNOCPP1 + \beta_7 CLTV_{i,t} + \beta_8 DTI_t + \sum_{t'=9}^{30} \beta_{t'} OY_{t'} \right).$$

The coefficients of origin year (OY) dummies are skipped due to simplicity. These dummies are consistent with Dempanyk and Hemert (2011). One, two and three asterisks indicate significance at the 5%, 1% and 0.1% confidence levels, respectively. The assigned codes are ARM (1) and FRM (0) for ARM indicator, residence (1), investment (2) and others (0) for owner occupancy types and single family (1), planned urban developments (2) condominium (3) and the others (0) for dwelling types. Standard errors are in the parentheses.

Variable	Code	2006:Q4	2007:Q4	2008:Q4	2009:Q4	2010:Q4	2011:Q4	2012:Q4
Intercept		-2.561*** (0.0302)	-3.5635*** (0.0227)	-2.8395*** (0.0196)	-2.6811*** (0.0193)	-2.639*** (0.0146)	-2.465*** (0.0113)	-2.6347*** (0.1535)
FICO Score		-0.0045*** (0.0000)	-0.004*** (0.0000)	0.0035*** (0.0000)	-0.0032*** (0.0000)	-0.003*** (0.0000)	-0.0029*** (0.0000)	-0.0029*** (0.0000)
ARM	1	0.0739*** (0.0013)	0.0984*** (0.0009)	0.1*** (0.0007)	0.1045*** (0.0006)	0.0993*** (0.0005)	0.0954*** (0.0005)	0.0929*** (0.0005)
Dwelling Type	1	0.0411*** (0.0020)	0.01*** (0.0013)	-0.0041*** (0.0009)	-0.0056*** (0.0008)	-0.0058*** (0.0007)	-0.0048*** (0.0007)	-0.004*** (0.0007)
	2	-0.0366*** (0.0033)	-0.002 (0.0019)	0.0083*** (0.0014)	0.0068*** (0.0012)	0.0048*** (0.0011)	0.0032** (0.0010)	0.0018 (0.0010)
	3	-0.0637*** (0.0041)	-0.0439*** (0.0024)	-0.0281*** (0.0017)	-0.0243*** (0.0014)	-0.0217*** (-0.0013)	-0.0206*** (0.0013)	-0.0201*** (0.0012)
Owner Occupancy	1	-0.0521*** (0.0029)	-0.0617*** (0.0017)	-0.0663*** (0.0013)	-0.0573*** (0.0012)	-0.0574*** (0.0011)	-0.055*** (0.0010)	-0.0532*** (0.0010)
	2	0.0095** (0.0035)	0.0412*** (0.0021)	0.0613*** (0.0016)	0.0753*** (0.0014)	0.0757*** (0.0013)	0.071*** (0.0012)	0.0655*** (0.0012)
Current LTV Ratio		0.8338*** (0.0066)	1.0095*** (0.0046)	1.3316*** (0.0031)	1.1337*** (0.0019)	1.0587*** (0.0017)	1.0024*** (0.0016)	0.9786*** (0.0015)
DTI		25.0302*** (0.2487)	30.9402*** (0.1866)	20.1795*** (0.1630)	17.5458*** (0.1539)	16.709*** (0.1153)	14.8133*** (0.0810)	15.1758*** (0.0650)
Orig. Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Goodness-of-fit								
AUROC		0.8014	0.7889	0.7944	0.795	0.7861	0.7797	0.7773
Rescaled R-Square		0.113	0.1091	0.1226	0.1262	0.1182	0.1122	0.1099
R-Square		0.0105	0.0136	0.0195	0.0225	0.0214	0.0203	0.0196
No. of obs		16,862,158	27,143,702	36,189,859	43,136,404	48,623,533	53,175,281	56,946,401

Table 5: Parameter estimates of state space model, estimation period 2000-2012

This table presents the parameter estimates for the state space model for the estimation period from 2000 to 2012. Panel A presents the parameter estimates of the Measurement Equation: for risk class $g = 1, 2, \dots, 10$:

$$\Phi^{-1}(r_{g,t}) = \phi_{0,g} + \phi_{1,g}f_t + \phi_{2,g}z_{g,t}$$

and Panel B shows the parameter estimates of the State Equation:

$$\xi_t = \mathbf{F}\xi_{t-1} + \nu_t,$$

where $\xi_t = (f_t, z_{1,t}, z_{2,t}, \dots, z_{G,t})'$ and $\mathbf{F} = \text{diag}(\beta_f, \beta_1, \beta_2, \dots, \beta_G)$. One, two and three asterisks indicate significance at 5%, 1% and 0.1%, respectively. Standard errors are in the parentheses.

Panel A: The estimates of Measurement Equation

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10
$\phi_{0,g}$	-3.1245*** (0.1115)	-2.7258*** (0.0490)	-2.5129*** (0.0762)	-2.3514*** (0.0518)	-2.2255*** (0.0444)	-2.1403*** (0.0411)	-2.0596*** (0.0428)	-1.9921*** (0.0409)	-1.9377*** (0.0394)	-1.7889*** (0.0673)
$\phi_{1,g}$	-0.0408* (0.0239)	-0.0831*** (0.0239)	-0.0940*** (0.0259)	-0.1056*** (0.0256)	-0.1152*** (0.0276)	-0.1260*** (0.0288)	-0.1303*** (0.0298)	-0.1207*** (0.0287)	-0.1253*** (0.0301)	-0.1340*** (0.0326)
$\phi_{2,g}$	-0.2251*** (0.0496)	-0.1221*** (0.0175)	-0.1184*** (0.0301)	-0.0818*** (0.0189)	-0.0595*** (0.0121)	-0.0346*** (0.0045)	-0.0248*** (0.0038)	-0.0285*** (0.0039)	-0.0407*** (0.0044)	-0.1035 (0.0320)

Panel B: The estimates of State Equation

	$z_{1,t}$	$z_{2,t}$	$z_{3,t}$	$z_{4,t}$	$z_{5,t}$	$z_{6,t}$	$z_{7,t}$	$z_{8,t}$	$z_{9,t}$	$z_{10,t}$
β	0.6682*** (0.1955)	0.8504*** (0.0678)	0.7210*** (0.0675)	0.9030*** (0.0487)	0.8547*** (0.0686)	0.7856*** (0.0880)	0.4422*** (0.1142)	-0.1556 (0.2022)	0.3137** (0.1221)	0.1852 (0.1248)
									0.1852 (0.1248)	0.8769*** (0.0831)

Table 6: Parameter estimates of the asset value return, estimation period 2000-2012

This table exhibits the parameter estimates of the asset value return in Equation (5) for the estimation period 2000 to 2012. The estimates are reversed back from the parameter estimates in Table 5 using the associated parametrization to the Measurement Equation (13) of the state space model. The standard errors in the parentheses are calculated by the delta method.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10
p_g	0.0012 (0.0004)	0.0035 (0.0005)	0.0065 (0.0014)	0.0099 (0.0014)	0.0137 (0.0016)	0.0169 (0.0017)	0.0206 (0.0021)	0.0240 (0.0023)	0.0274 (0.0025)	0.0389 (0.0056)
ρ_g	0.0016 (0.0018)	0.0068 (0.0039)	0.0086 (0.0047)	0.0110 (0.0053)	0.0130 (0.0062)	0.0156 (0.0070)	0.0167 (0.0075)	0.0144 (0.0067)	0.0154 (0.0073)	0.0175 (0.0083)
α_g	0.0482 (0.0097)	0.0147 (0.0057)	0.0138 (0.0060)	0.0066 (0.0041)	0.0035 (0.0033)	0.0012 (0.0020)	0.0006 (0.0015)	0.0008 (0.0016)	0.0017 (0.0024)	0.0106 (0.0066)
σ_g	0.0497 (0.0018)	0.0213 (0.0039)	0.0223 (0.0047)	0.0175 (0.0053)	0.0165 (0.0062)	0.0168 (0.0070)	0.0173 (0.0075)	0.0152 (0.0067)	0.0171 (0.0073)	0.0279 (0.0083)
$\tilde{\alpha}_g$	0.0482 (0.0097)	0.0146 (0.0056)	0.0137 (0.0059)	0.0066 (0.0041)	0.0035 (0.0032)	0.0012 (0.0020)	0.0006 (0.0015)	0.0008 (0.0016)	0.0016 (0.0024)	0.0104 (0.0065)

Table 7: Effectiveness of default prediction

This table presents the Area Under the Receiver Operating Characteristics curve (AUROC) of the Probit model at the estimation time applied to mortgages at the prediction years to measure the prediction accuracy of defaults. For example, 70.92% of AUROC in the estimation year 2006 and the prediction year 2007 is obtained by applying the Probit model at 2006:Q4 in Table 4 to mortgages in the prediction year 2007 with updated observation time variables. The higher the AUROC, the more accurate the default prediction model.

Estimation Year	Prediction Year					
	2007	2008	2009	2010	2011	2012
2006	0.7092	0.6736	0.6272	0.6151	0.6096	0.6258
2007	.	0.6958	0.6523	0.6309	0.6237	0.6446
2008	.	.	0.6578	0.6307	0.6227	0.6311
2009	.	.	.	0.6337	0.6251	0.6328
2010	0.6278	0.6325
2011	0.6363

Figures

Figure 1: Default rate and loss ratio

This figure plots the aggregate default rate of the total portfolio and the loss ratio given default calculated as the aggregated loss over the aggregate outstanding balance at the observation time, i.e.,

$$\frac{\sum_{i=1}^{N_t} CB_{i,t} \cdot D_{i,t} \cdot LGD_{i,t}}{\sum_{i=1}^{N_t} CB_i},$$

where $CB_{i,t}$, $D_{i,t}$ and $LGD_{i,t}$ denote the current balance, the default indicator and the loss-given-default ratio for the mortgage i at time t , respectively. The grey bars indicate years which include a period of economic downturn as indicated by the National Bureau of Economic Research.

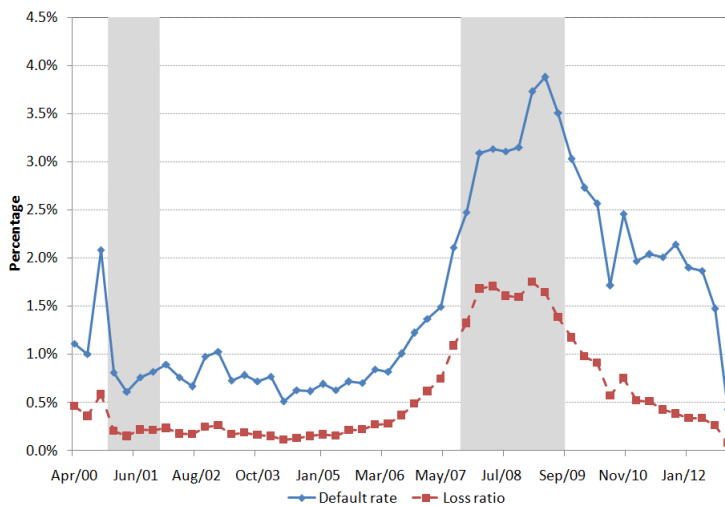


Figure 2: Case-Shiller index and debt-to-income ratio

This figure shows the historical the US 10-city composite Case-Shiller Index on the left vertical axis and the country-level debt-to-income ratio from the Federal Reserve Bank on the right vertical axis. The higher the Case-Shiller Index, the higher average house price.

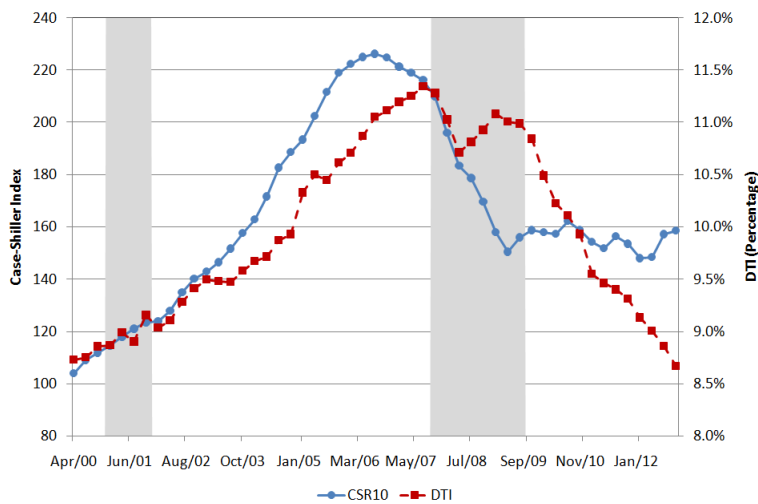


Figure 3: Receiver Operating Characteristic curve per estimation period

This figure shows the receiver operating characteristic (ROC) curve. The ROC curve is a common performance measure for ordinal rating systems. The plot displays the true positives (sensitivity) and the false positives (one minus the specificity). The better a ranking/rating system is able to attribute non-default outcomes with lower ranks and default outcomes with higher ranks, the larger the area under the curve (AUROC, see Agresti, 1984).

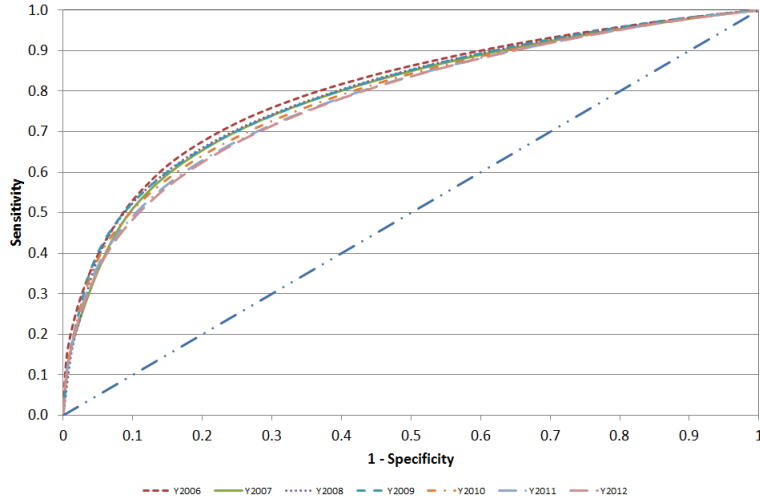
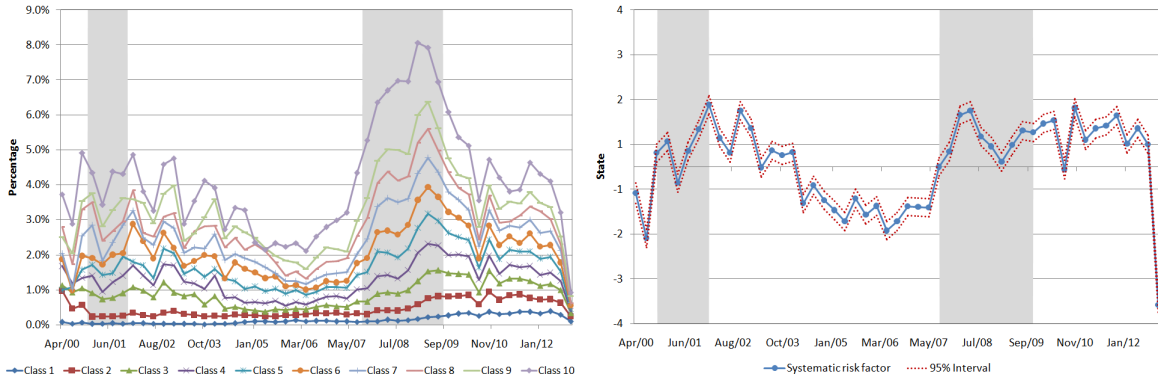


Figure 4: Default rate per risk class and the estimated systematic risk factor, estimation period 2000-2012

Figure (a) displays the historical default rates of ten risk classes based on the estimated PD from the Probit model with the estimation period spanning from 2000 to 2012. Figure (b) shows the estimated systematic risk factor and its 95% confidence interval from Kalman filtering given on the parameter estimates in Table 5. Note that the systematic risk factor in Figure (a) was multiplied by minus one to show the co-movement with default rates of Figure (a).



(a) Class-wide default rate

(b) Estimated systematic risk factor

Figure 5: Total risk per risk class, estimation period 2000-2012

This figure displays the estimated total risk from the state space model in Table 6 and the Vasicek model. The default rate per risk class is on the x-axis.

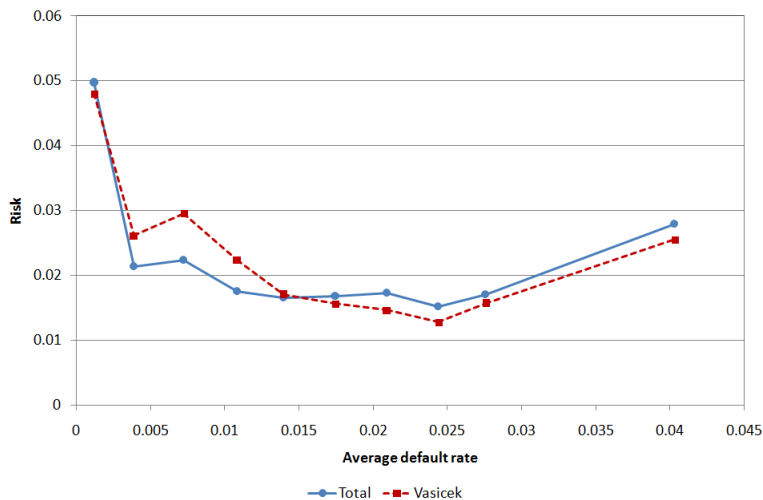


Figure 6: The volatility of default rate per risk class based on Gordy (2000), estimation period 2000-2012

The figure presents the variance of default probability in Equation (16) as shown in Gordy (2000). The variance is obtained by $var(p_g(f_t) | p_g, \sigma_g)$ for the state space model and $var(p_g(f_t) | p_g, \tilde{\rho}_g)$ for the Vasicek model. The default rate per risk class is on the x-axis.

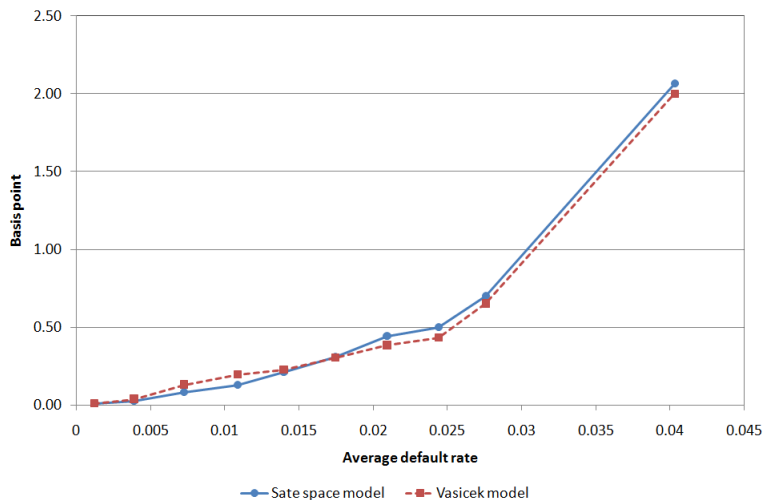


Figure 7: Systematic and non-systematic risks per risk class, estimation period 2000-2012

This figure presents the estimated systematic (ρ_g) and non-systematic ($\tilde{\alpha}_g$) risks from the state space model per risk class. The average default rate per risk class is on the x-axis.

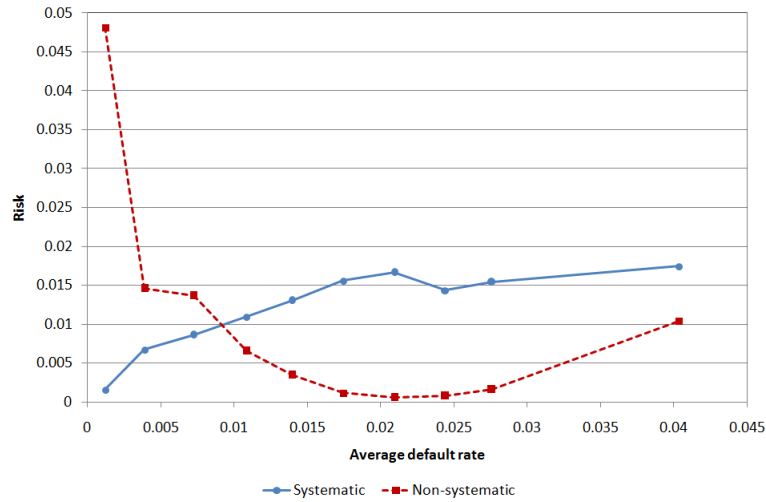
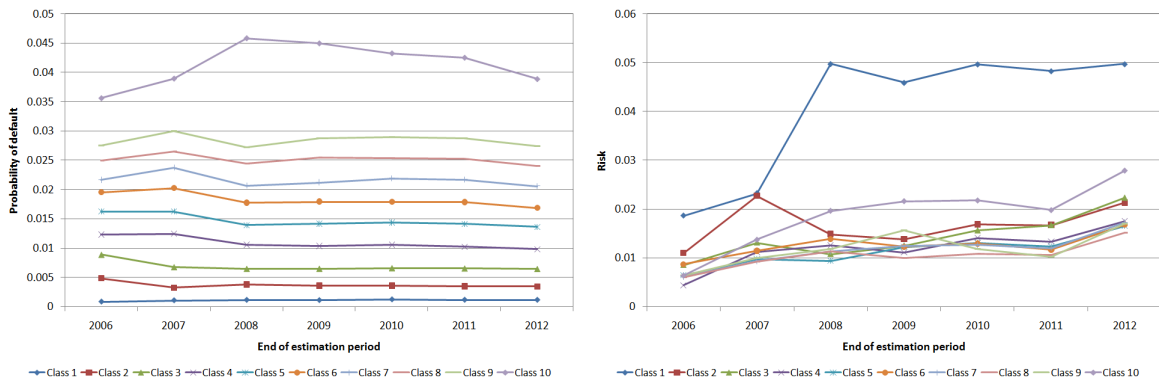


Figure 8: Serial change of unconditional PD and total risk per risk class

Figure (a) plots the estimated unconditional PD and Figure (b) plots the total risk (σ_g) as the end of estimation period extends from 2006 to 2012 per risk class. The end year of the estimation periods is on the x-axis.



(a) Unconditional PD

(b) Total risk

Figure 9: Serial change of systematic and non-systematic risks per risk class

Figure (a) plots the estimated systematic (ρ_g) and Figure (b) plots the non-systematic ($\tilde{\alpha}_g$) risks as the end of estimation period extends from 2006 to 2012 per risk class. The end year of the estimation periods is on the x-axis.

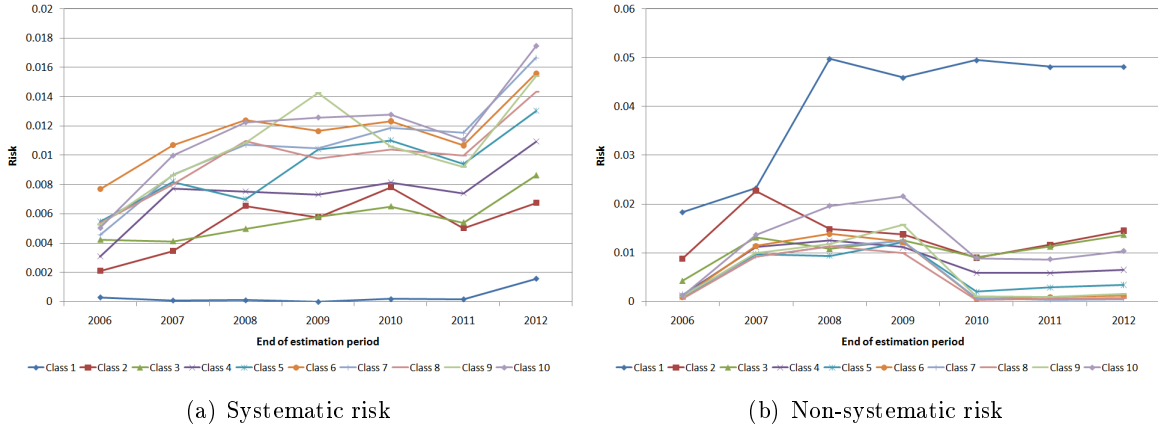


Figure 10: Realized default rate vs. predicted unconditional PD per risk class and estimation period

This figure compares the average default rate (y-axis) with the class-wide unconditional PD estimates ($\Phi(\phi_{0,g})$) from the state space model (x-axis) for different estimation periods. For example, the unconditional PDs for year 2006 are obtained by the state space model with the estimation period from 2000 to 2006. Each line represents the pairs of the average default rate and the class-wide unconditional PD for ten risk classes at the first quarter of the last year in each estimation period.

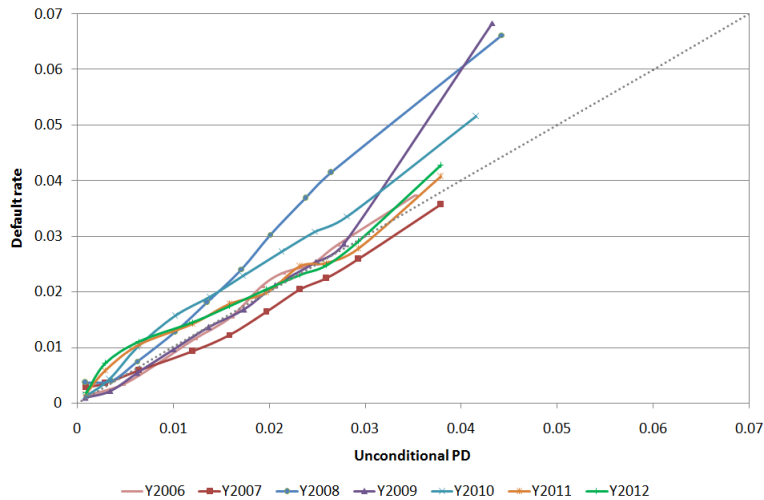


Figure 11: Realized default rate vs. predicted PD, estimation period 2000-2008

This figure plots the quarter-ahead predicted class-wide PDs using the state space model at the estimation time horizon spanning from 2000:Q2 to 2008:Q4. The PDs are predicted for 2008:Q1 at 2007:Q4 across ten risk classes as an example. The unconditional PD denotes the unconditional mean of the state space model ($\Phi(\phi_{0,g})$) as in Figure 10, the conditional forecast PD is obtained by

$$\hat{r}_{g,t+1|t} = \Phi\left(\phi_{0,g} + \phi_{1,g}\beta_f \hat{f}_{t-1} + \phi_{2,g}\beta_g \hat{z}_{g,t-1}\right)$$

for $g = 1, 2, \dots, 10$, and the conditional PD is calculated by

$$\hat{r}_{g,t|t} = \Phi\left(\phi_{0,g} + \phi_{1,g}\hat{f}_t + \phi_{2,g}\hat{z}_{g,t}\right),$$

where \hat{f}_{t-1} , \hat{f}_t , $\hat{z}_{g,t-1}$ and $\hat{z}_{g,t}$ are the Kalman filtered risk factors at time $t-1$ and t , respectively. The average default rate per risk class is on the y-axis.

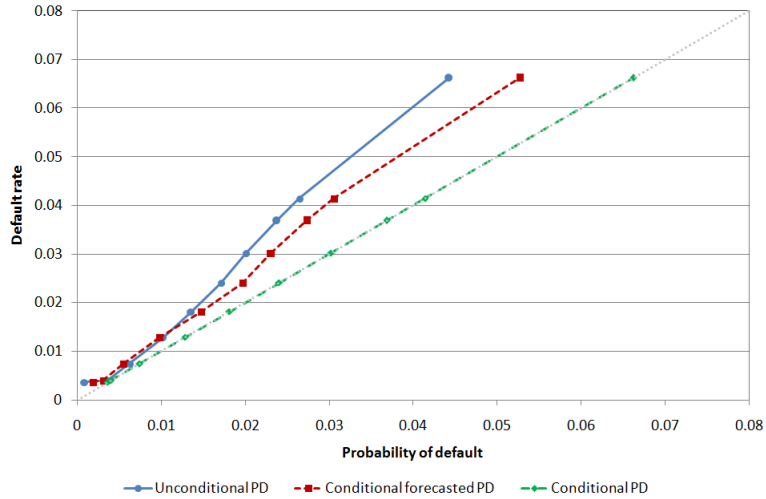
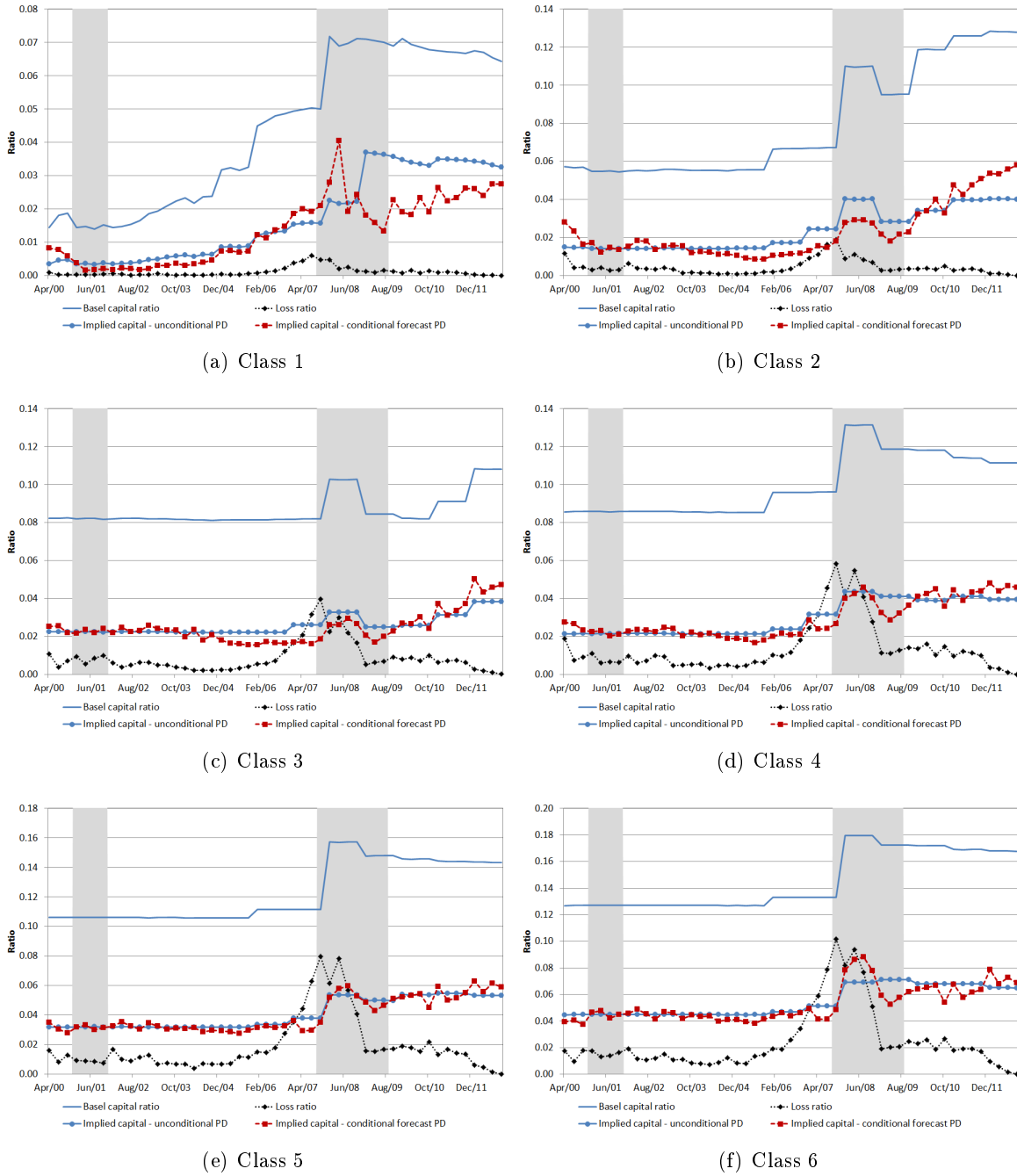
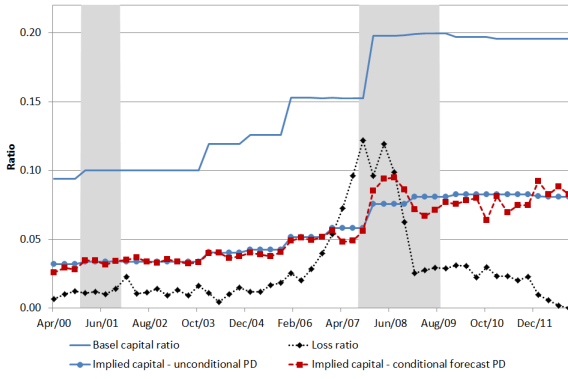


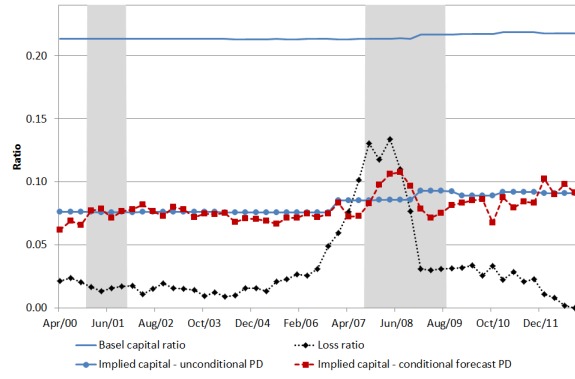
Figure 12: Basel capital ratio, implied capital ratios and actual loss ratio

The figures present the actual loss ratio in (20), the Basel capital ratio(17) based on the assumption of asset correlation 15% and the implied capital ratios in (18) and (19).

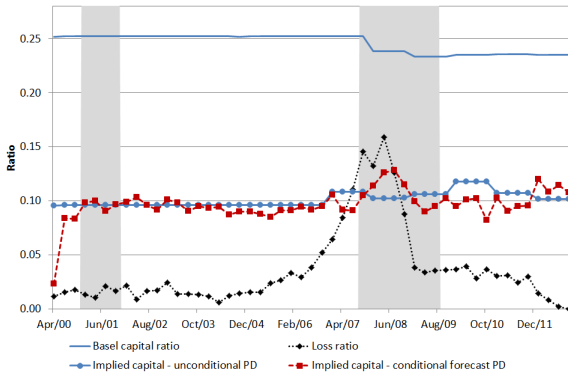




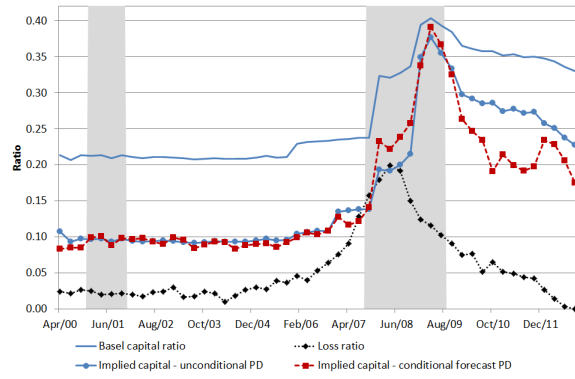
(g) Class 7



(h) Class 8



(i) Class 9



(j) Class 10