Asset Correlation in RMBS Reference Portfolios

Marco Geidosch
Statement of Purpose
The Working Paper series of the UniCredit & Universities Foundation is designed to disseminate and to provide a platform for discussion of either work of UniCredit economists and researchers or outside contributors (such as the UniCredit & Universities scholars and fellows) on topics which are of special interest to UniCredit. To ensure the high quality of their content, the contributions are subjected to an international refereeing process conducted by the Scientific Committee members of the Foundation. The opinions are strictly those of the authors and do in no way commit the Foundation and UniCredit Group.

Scientific Committee
Franco Bruni (Chairman), Silvia Giannini, Tullio Jappelli, Levent Kockesen, Christian Laux, Catherine Lubochinsky, Massimo Motta, Giovanna Nicodano, Marco Pagano, Reinhard H. Schmidt, Branko Urosevic.

Editorial Board
Annalisa Aleati
Giannantonio De Roni

The Working Papers are also available on our website (http://www.unicreditanduniversities.eu)
Contents

Abstract 4

1. Introduction 5

2. Estimation Methodologies 7

3. Simulation Study 12

4. Data 16

5. Estimated Asset Correlation 21

6. Discussion and Interpretation of the Results 29
Asset Correlation in RMBS Reference Portfolios

Marco Geidosch*

UniCredit Bank AG, CEC4EC Economic Capital

Abstract

This paper contributes to the literature about estimating asset correlation in two ways. First, in simulation study, we compare the performance of different estimation approaches. By doing so, we provide knowledge about the behavior of the applied estimators which is an important precondition for the reliability and interpretation of the estimation process. Second, we present a novel data base to estimate asset correlation from, which is loss data of residential mortgage-backed securities (RMBS) transactions. Our data set consists to a large extent of the most toxic RMBS deals that sparked the subprime crisis. Contrary to the widely held view, our analysis reveals that asset correlation in the subprime market is surprisingly low (roughly 6%). By giving an intuitive and straight-forward explanation for these low values, we provide valuable insight into the mechanism and evolution of the subprime crisis in general, and into the risk characteristics of a credit portfolio in particular.

Keywords: credit portfolio risk; estimating asset correlation from default rates; subprime crisis; RMBS transactions

*I thank Gregor Dorfleitner, Michael Gordy and Alfred Hamerle for valuable comments and discussion.
1. Introduction

Broadly speaking, asset correlation measures the joint movement of obligors’ creditworthiness in a loan portfolio. It is one of the core input parameters for credit portfolio risk models. In the foundation internal ratings-based approach (IRBA) of Basel II, asset correlation is provided by the regulator, while in the advanced IRBA, banks can use their own estimates. Amongst others, this explains the abundance of empirical studies on estimating asset correlation. Grundke (2008) and Chernih et al. (2010) provide a comprehensive comparison of these studies.

The existing literature can be categorized into two streams, separated by the data from which asset correlation is inferred. For corporate portfolios, it can be derived from observable market data, like equity and asset returns or CDS spreads. In contrast, for retail or mortgage portfolios this data is not available, and asset correlation has to be derived from historical default rates. The main disadvantage of the latter approach is that observed defaults are in general very scarce, which negatively affects the validity of the estimation results. Usually, the approach based on observable market data comes up with higher correlation values than the default rate based method (roughly 20-30% versus 5-15%). While Frye (2008) gives an explanation for this discrepancy, Düllmann et al. (2010) argue that correlation estimates from equity returns are more efficient than those from default rates.

This study contributes to the second class of literature, which is based on default rates in two ways. First, the robustness of the estimation results w.r.t. applying different estimation techniques is investigated in a simulation study. We complement the work of Gordy and Heitfield (2010) and Frey and McNeil (2003) who analyze the performance of maximum likelihood and method of moments estimators, by adding two parametric approaches, proposed by Botha and van Vuuren (2010), to the evaluation process. A comparison of different estimation approaches is an important precondition for the reliability and interpretation of the estimation process. Low standard deviation and mean bias of five estimation approaches suggest that all estimators perform rather well and that asset correlation is robust against the choice of the estimator.

Second, we present a novel data source to estimate asset correlation from default rates, which is reference portfolios of residential mortgage-backed securities (RMBS) transactions. Put simply, the securities of an RMBS deal are backed by a pool of mortgages. Cashows from these mortgages are used to pay the financial claims of the RMBS investors. However, when some of the mortgages default, the RMBS investors incur losses.

Our data sample consists of all US RMBS deals available in Bloomberg that have been originated between 2004 and 2007 totaling for 4,4605 RMBS deals. After applying several data quality filters, our sample is reduced to 988 RMBS reference portfolios with the single portfolio consisting of 1,052 mortgage loans on average. Thus, our study is based on monthly loss data of over one million mortgages. Further, our sample includes the “most toxic” RMBS deals from the US subprime market which (amongst other things) caused the recent subprime crisis. The 2006 and 2007 subprime RMBS vintages for example, lost on average almost 25% of their original face value so far. The worst

---

1 Dorfleitner et al. (2012) study specification risk when using market data to calibrate a macroeconomic factor model.
reference portfolio of our sample lost nearly 80% of its original face value. Thereby, our study does not suffer from the typical problem of observing too few defaults and is one of the first studies to include the subprime crisis.

This study is related to Cowan and Cowan (2004) who investigate default correlation of subprime loans. However, there are three important distinctions. First, their definition of default is different to ours (payment 90+ days delinquent and foreclosure, respectively compared to actually booked losses). Second, while our data sample comes from a wide range of different lenders, theirs is restricted to a single subprime lender. And third, the Cowan and Cowan (2004) data covers the time from July 1995 through December 2001 and therefore stops well before the subprime boom and bust. In contrast, our loss sample start roughly at the end of 2007 and therefore perfectly coincide with the subprime meltdown.

Apart from the great number of mortgage loans, our data sample has two other implications. First, we can study asset correlation under realistic stress conditions, as the subprime crisis with its high default rates is considered the perfect storm scenario. This gives us important insight for stress testing issues, for example. And second, we can investigate the popular held view that an underestimation of asset correlation was one of the core reasons for the subprime crisis.²

In the following section, we present the estimation methodologies that have been proposed in the literature. Next, we show the results of a simulation study that compares standard deviation and mean bias of five estimators. Section 4 describes RMBS reference portfolios in general and our data sample in particular. Estimated asset correlation based on our data set are given in section 5, and in the final section we discuss the results in the light of the subprime crisis.

² For example, Salmon (2009) state that “naturally, default correlations were very low in [the years prior to the crisis]. But when the mortgage boom ended abruptly and home values started falling across the country, correlations soared.”
2. Estimation Methodologies

2.1 Single Factor Gaussian Copula Framework

All approaches that have been proposed in the literature to estimate asset correlation from default rates rely on a single factor Gaussian copula (SFGC) framework. In this framework, the latent variable or creditworthiness or asset value of an obligor \( i \) is modeled as a standard Gaussian random variable \( CW_i \), while the joint movement of many obligors’ creditworthiness is modeled with a multivariate Gaussian distribution.

To model the individual default risk, a structural or Merton (1974) type model is applied, which means that at a fixed time horizon obligor \( i \) defaults when its creditworthiness \( CW_i \) falls below a certain default threshold \( d_i \). Denoting the (unconditional) probability of default of obligor \( i \) with \( PD_i \) the default threshold can be calculated as

\[
PD_i = \Pr(\text{obligor i defaults}) = \Pr(CW_i \leq d_i) = \Phi(d_i) \Rightarrow d_i = \Phi^{-1}(PD_i)
\]

(1)

Dependence among borrowers is modeled with a factor model, that is the creditworthiness \( CW_i \) is driven by a single common systematic factor \( X \) and by an idiosyncratic factor \( \epsilon_i \), which are both assumed to be standard Gaussian and mutually independent:

\[
CW_i = \alpha_i X + \sqrt{1 - \alpha_i^2} \epsilon_i
\]

(2)

The factor loadings \( \alpha_i \) and \( \sqrt{1 - \alpha_i^2} \) ensure unit variance of the latent variable \( CW_i \). The estimation approaches described in the following sections are based on two crucial assumptions: first, probability of default is constant within the portfolio and second, the factor loading \( \alpha_i \) in the above equation (2) is also constant within the portfolio. Therefore, we suppress an index \( i \) for concise presentation whenever reasonable. Asset correlation \( \rho \) in the portfolio is then defined as the correlation between two obligors’ creditworthiness and is calculated as

\[
\rho = \text{corr}(CW_i, CW_j) = \alpha^2
\]

(3)

The probability that an obligor defaults conditional on the realization of the systematic risk factor \( X = x \) is

\[
PD_i | X = x = \Pr \left( \epsilon_i \leq \frac{d - ax}{\sqrt{1 - \alpha^2}} \right) = \Phi \left( \frac{d - ax}{\sqrt{1 - \alpha^2}} \right) \quad \text{with} \quad \epsilon_i \sim N(0,1)
\]

(4)
In this paper, for \( t = \{1, \ldots, T\} \) observations of data, \( D_t \) denotes the number of observed defaults in the time period \([t, t + 1]\) and \( N_t \) is the number of obligors at time \( t \).

### 2.2 Method of Moments Estimation

To derive two method of moments estimators, we rewrite the event of default with help of the indicator function \( Y \), that is

\[
Y_t = \begin{cases} 
1 & \text{if obligor } i \text{ defaults} \\
0 & \text{otherwise}
\end{cases}
\]

(5)

According to this notion, \( \mathbb{E}[Y^2] \) equals the probability of a simultaneous default of two obligors (also referred to as joint probability of default). Lucas (1995) propose to estimate \( \mathbb{E}[Y^2] \) as the weighted ratio of number of defaulted pairs to the number of all possible pairs in the sample:

\[
\hat{\mathbb{E}}[Y^2] = \sum_{t=1}^{T} w_t \frac{D_t(D_t - 1)}{N_t(N_t - 1)}
\]

(6)

where \( w_t \) is the weight of observation \( t \) w.r.t. to the number of obligors in the portfolio. Applied to real world data, one needs to adjust the above equation to avoid calculating negative values in cases with no observed defaults:

\[
\text{MM1}: \mathbb{E}[Y^2] = \sum_{t=1}^{T} w_t \frac{D_t^2}{N_t^2}
\]

(7)

An alternative method of moments estimator for \( \mathbb{E}[Y^2] \) is derived from the relation \( \mathbb{E}[Y^2] = \text{Var}[Y] + \mathbb{E}[Y^2] \)

While in our case \( \mathbb{E}(Y^2) \) is estimated as the squared average unconditional probability of default \( PD \), Gordy (2000) proposed to estimate \( \text{Var}[Y] \) as follows\(^3\)

\[^3\text{For a derivation see Appendix B of Gordy (2000).}\]
Combining both terms gives the second method of moments estimator

\[
\text{MM2: } \mathbb{E}[Y^2] = \frac{\text{Var}(\frac{D}{N^t}) - \mathbb{E}(\frac{1}{N^t})PD(1 - PD)}{1 - \mathbb{E}(\frac{1}{N^t})} + PD^2
\]  

\hspace{1cm} (9)

\(\text{MM1 and MM2 of equation (7) and (9) are two ways to estimate the joint probability of default } \mathbb{E}[Y^2]\) empirically from loss data. On the other side, we can calculate the joint probability of default analytically in the SFGC framework as

\[
\mathbb{E}[Y^2] = \Phi_2(d, d, \alpha^2)
\]  

\hspace{1cm} (10)

where \(\Phi_2(\cdot)\) denotes the bivariate Gaussian distribution function.\(^4\) Now, the idea is to back out asset correlation 2 from equation (10) numerically as the value that matches the empirically estimated joint probability of default, i.e. estimators \(\text{MM1 and MM2}\), respectively.

Method of moments estimator \(\text{MM1}\) is applied in Jobst and de Servigny (2005), Bandyopadhyay et al. (2007), de Servigny and Renault (2003), Nagpal and Bahar (2001) and Frey and McNeil (2003), while the \(\text{MM2}\) estimator is used in Dietsch and Petey (2002, 2004).

2.3 Maximum Likelihood Estimation

In the SFGC framework the conditional probability of default \(PD|x\) is given in equation (4). Recall that \(D_t\) is the time series of observed defaults and \(N_t\) is the corresponding number of borrowers in the portfolio. Conditioned on a realization of the systematic factor \(X = x\) defaults are assumed to be independent and therefore the number of defaults \(D_t\) follows a binomial distribution:

\[
\Pr(D_t|X = x) = \text{Bin}(D_t, N_t, PD|x) = \binom{N_t}{D_t} \cdot PD|x^{D_t} \cdot [1 - PD|x]^{N_t - D_t}
\]  

\hspace{1cm} (11)

Integrating over the risk factor \(X\) yields the unconditional distribution of the number of defaults:

\[
\Pr(D_t) = \int_{-\infty}^{\infty} \binom{N_t}{D_t} \cdot PD|x^{D_t} \cdot [1 - PD|x]^{N_t - D_t} \phi(x) dx
\]  

\hspace{1cm} (12)

\(^4\) \(\Phi_2(d, d, \alpha^2)\) gives the probability that two latent variables fall below the default threshold d given the model assumptions of a constant probability of default and constant factor loadings in the portfolio.
where \( \phi(x) \) denotes the density of the standard normal distribution.

Since the systematic risk factor \( X \) is assumed to be independent over time, the probability to observe the given default time series \( (D_1,\ldots,D_T) \) equals the product over the marginal densities \( Pr(D_t) \) from equation (12):

\[
Pr(D_1,\ldots,D_T) = \prod_{t=1}^{T} Pr(D_t)
\]

which finally gives the following log-likelihood function

\[
l(\alpha) = \sum_{t=1}^{T} \ln[Pr(D_t)] = \sum_{t=1}^{T} \ln \left[ \int_{-\infty}^{\infty} \frac{N_t}{D_t} \cdot PD(x^{d_t}) \cdot [1 - PD(x)]^{N_t-D_t} \phi(x) dx \right]
\]

The log-likelihood function \( l(\alpha) \) is maximized to get the asset correlation parameter \( \alpha \). We do not perform a simultaneous maximization over both parameters \( d \) and \( \alpha \) in the log-likelihood function (14) as this may lead to convergence problems of the integral.\(^5\) Instead, we estimate the default threshold \( d \) separately from the average probability of default and maximize the log-likelihood function subsequently. We checked that differences between the two-step and simultaneous estimation are immaterial.

The intuition behind the ML estimator is to find the correlation parameter \( \alpha \) for which occurrence of the observed default rate time series is most likely. The ML estimator and various enhancements are applied in Gordy and Heitfield (2010), Rösch and Scheule (2004), Hamerle et al. (2003), Demey et al. (2004) and Demey and Roncalli (2004).

### 2.4 Parametric Approach

The idea of the parametric approach is to fit two distributions to the observed default rate time series and to derive asset correlation from the fitted distributions. In particular, the Vasicek and the beta distribution are used to this end because they are both quite popular in the credit portfolio risk context.\(^6\)

\(^5\) See Gordy and Heitfield (2010), Appendix A for details.

\(^6\) For example, the beta distribution is used in the Basel II framework on securitization.
Vasicek (1987) derived a closed form solution for the probability distribution of the loan portfolio loss $L$. Let $PD$ be the constant probability of default in the portfolio and let $\rho$ be the constant pairwise asset correlation as defined in equation (3), then the probability that the portfolio loss $L$ is lower than a percentage $q$ of the total loan exposure is given by

$$\Pr(L \leq q) = \Phi\left(\frac{\sqrt{1 - \rho} \cdot \Phi^{-1}(q) - \Phi^{-1}(PD)}{\sqrt{\rho}}\right),$$

(15)

where the most prevalent loss - the mode - of the distribution is located at

$$L_{mode} = \Phi\left(\frac{\sqrt{1 - \rho}}{1 - 2\rho}\right) \cdot \Phi^{-1}(PD).$$

(16)

After calculating the average (unconditional) default rate $PD$ and the mode of the default rate time series $L_{mode}$, we can solve equation (16) for the asset correlation $\rho$. We denote this approach with $PA_V$.

The distance between the mode and the mean of a data sample is an indication of the skewness of the underlying distribution. That is the intuition behind this approach as the skewness of a credit portfolio’s loss distribution is closely related to the asset correlation. A practical problem of the $PA_V$ estimator concerns how many digits after the decimal point are used for the default rates. The mode of the default rate time series decreases as the number of digits increases, which in turn increases asset correlation. We found that using four digits after the decimal point yields the most sensible results.

The beta distribution is characterized by two parameters $\alpha$ and $\beta$ and its distribution function is given by

$$F_{\alpha,\beta}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \cdot \int_0^x (1 - t)^{\beta-1} \cdot t^{\alpha-1} dt, \quad 0 \leq x \leq 1, \quad \alpha, \beta > 0,$$

(17)

where $\Gamma(\cdot)$ is the gamma function. The beta distribution is fitted to the loss data by matching the first and second moment to the empirically estimated mean and standard deviation of the loss data, $PD$ and $sd_{PD}$, respectively. To be specific, $\alpha$ and $\beta$ of the beta distribution are calculated as follows

$$\hat{\alpha} = PD \cdot \left(\frac{PD \cdot (1 - PD)}{sd_{PD}^2} - 1 \right)$$

(18)

$$\hat{\beta} = \frac{\alpha}{PD} \cdot (1 - PD).$$

(19)
Based on the fitted beta distribution, the portfolio value at risk at a given confidence level $\text{VaR}_{\beta}(q)$ is calculated as

$$\text{VaR}_{\beta}(q) = \inf_x \{ F_{\tilde{\alpha}, \tilde{\beta}}(x) \geq q \}$$ (20)

Finally, asset correlation is derived as the value whose corresponding Basel II value at risk matches the value at risk based on the fitted beta distribution of equation (20). More formally, solving

$$\text{VaR}_{\beta}(q) = \Phi \left( \frac{\Phi^{-1}(\text{PD}) + \sqrt{\rho} \cdot \Phi^{-1}(q)}{\sqrt{1 - \rho}} \right)$$ (21)

for $\rho$, as the term on the right side is the Basel II value at risk. A very practical problem of the parametric beta estimator ($\text{PA}_{\beta}$) is to choose the confidence level $q$ in equation (20). We found that $q = 99.9\%$ yields the most reasonable results. Both parametric approaches are applied in Botha and van Vuuren (2010).

3. Simulation Study

In a simulation study, we compare the variation and bias of the estimators presented in the previous section depending on asset correlation $\rho$, probability of default $\text{PD}$ and the length of the time series $N$. To simulate default rates, we rely on the SFGC framework of section 2.1 and consider a fixed portfolio of 1,000 obligors. We assume a constant probability of default $\text{PD}$ and a constant factor loading $\alpha$ within the portfolio which is consistent with the estimation approaches.

We study the estimators’ performance in two different settings. First, we vary the input parameters on a quite general range to study the overall behavior of the estimators, while in the second setup we narrow the input parameter range to mimic the characteristics of the RMBS reference portfolios of our data sample. Effectively, the RMBS-like setting focuses on lower asset correlation and lower $\text{PD}$. In each setting, we take ten equidistant values from the parameter ranges given in Table 1. Thus, in total we have 1,000 input parameter triplets consisting of asset correlation $\rho$, probability of default $\text{PD}$ and the length of the time series $N$ in both settings. For each parameter triplet, we generate 50 independent default rate time series according to the SFGC framework, that is we draw $N$ samples from a 1,000 dimensional Gaussian distribution with correlation matrix $\Sigma = \sigma_{i,j} = \rho$ for $i \neq j$. A 1,000 dimensional sample represents the creditworthiness of 1,000 borrowers.
The number of defaults of a given sample equals how many observations fall below the default threshold \( d = \Phi(PD) \).

Figure 1 shows standard deviation and mean bias averaged over the 1,000 triplets in the general and the RMBS-like setting. The estimators’ variation pattern is rather consistent over both setups: except for the \( PA_V \) estimator, which has by far the highest standard deviation, the remaining estimators’ variations are in a very similar range.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Parameters of the simulation study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>General</td>
</tr>
<tr>
<td>( N ) (%)</td>
<td>[10, …, 100]</td>
</tr>
<tr>
<td>( \rho ) (%)</td>
<td>[0, …, 30]</td>
</tr>
<tr>
<td>( PD ) (%)</td>
<td>[0.5, …, 10]</td>
</tr>
</tbody>
</table>

Ten equidistant values are taken from the intervals in each setting of the simulation study.

However, the absolute variation level is lower in the RMBS-like than in the general setting (roughly 1.6% compared to 3.5%). As we will see in the following analysis, the reason for the higher level of variation in the general setup is the focus on higher asset correlation.

In the general setting, the ML estimator appears unbiased, while all other approaches underestimate asset correlation. The \( PA_V \) and \( PA_{\beta_{\text{est}}} \) by almost 3% and 4% percentage points, respectively. The observations concerning the mean bias do not transfer from the general to the RMBS-like setup. The bias is almost reduced to zero for the \( PA_V \) estimator and it is reduced by almost
90% in the $P_{A_{\beta}}$ case. The behavior of the bias is reversed for the MM1 estimator: while it underestimates asset correlation by -0.8% on the global parameter range, it overestimates asset correlation by 0.8% for the RMBS-like portfolios. In the RMBS-like setting, MLE and MM2 estimator slightly underestimate asset correlation (-0.3% and -0.9%, respectively). These observations show that the estimators' performance strongly depends on the characteristics of the underlying portfolio.

Next, we study the estimation performance on a more granular basis, that is we split the averaged values of figure 1 to distill the stand-alone impact of asset correlation $\rho$, probability of default $PD$ and the length of the time series $N$, respectively. We do so only for the general setup, as for the RMBS-like parameters the conclusions are either redundant or meaningless due to the narrow parameter range.

We proceed in the following way. Averaged over 1,000 input parameter triplets the standard deviation of the MM2 estimator is 3.8% in figure 1, left graph. Asset correlation takes ten equidistant values in $[0,\ldots,30]$ and for each of these equidistant values we have 100 parameter triplets (remember there are 1,000 parameter triplets in total). Figure 2 a) now shows average (over 100 triplets) standard deviation for each of the ten equidistant asset correlation values per estimator.

Interestingly, the estimators' variation increases sharply as the underlying asset correlation increases. This pattern holds for all estimators, with the $PA_{\rho}$ behaving slightly different. That means, the estimation precision impairs as asset correlation increases which is particularly alarming as high correlation values can have important consequences for credit portfolio tail risks. On the other hand, the portfolio's credit quality does not influence the estimation precision as can be seen in figure 2 b). $PA_{\rho}$ is again the exception, as the estimator's variation increases when the credit quality impairs. The low impact of the probability of default is somewhat counterintuitive as one would expect the estimators' fluctuation to decrease when defaults tend to occur more frequently. Not surprisingly, the estimators' standard deviation decreases with growing sample size, see Figure 2 c).

We know from Figure 1 that MM1, MM2, $PA_{\rho}$ and $P_{A_{\beta}}$ on average underestimate asset correlation in the general setting. The downward bias increases with growing asset correlation, see Figure 2 d). While the absolute level of the downward bias is moderate for MM1 and MM2 (less than 3% in both cases), it is rather large for $PA_{\rho}$ and $P_{A_{\beta}}$. For example, if the true (but unknown) asset correlation is 30%, the beta approach will come up with an estimate of only 18.2%. Again, an underestimation of asset correlation of this magnitude will have severe consequences for the portfolio risk management. The mean bias of the ML estimator is comparatively independent of the underlying correlation value.

The mean bias of the MM1, MM2 and $PA_{\rho}$ estimator show the expected behavior as the probability of default increases. Their mean bias decreases. MLE moves from underestimating asset correlation to overestimation when the credit quality of portfolio impairs, both on a moderate level. The $P_{A_{\beta}}$ estimator's bias increases as defaults become more likely, which is a somewhat counterintuitive behavior. The MM1, MM2 and $P_{A_{\beta}}$ react quite similar to an increasing number of observations $N$, whereas the bias of the latter estimator is on a significant higher level. For $N = 100$ observations, the mean bias for MM1 and MM2 is lower than 0.5%, while it is 3.3% for the beta
estimator. The $PA_1$ estimator is rather independent from the number of observations. The MLE moves from underestimating to overestimation asset correlation when the number of observations is increased from $N = 10$ to $N = 100$, both on a moderate level.

We conclude from the simulation study that the estimators' performance is more or less similar. Most important, we find that asset correlation can be estimated rather robustly with respect to the estimator (with the exceptions discussed above). Neither does one estimator clearly outperform nor underperform, which is why we apply all estimators to our data sample and do not exclude any estimator.

**Figure 2**

Estimators' performance on granular level

This figure shows the impact of asset correlation (upper panel), probability of default (middle panel) and length of the time series (lower panel) on the standard deviation and mean bias of five estimators.
4. Data

Residential mortgage-backed securities (RMBS) are structured credit products which are created from a pool of mortgages. Principal and interest payments from a mortgage portfolio are used to pay the financial claims of the RMBS investors. The funds from the RMBS investors, on the other hand, are used to take off the mortgages from a bank’s balance sheet, which is one of the original motivations for securitization, known as regulatory capital relief. Similar structures may have student loans, credit card debt or auto loans as underlying pool of assets. The generic term for all these structures is asset-backed securities (ABS). However, we focus on RMBS as they are the most important asset class in terms of outstanding notional.

To execute an RMBS deal, a dummy company (special purpose vehicle, SPV) is founded in an offshore location which buys the mortgages (from a bank or mortgage lender) and sells securities to investors. To create confidence among the RMBS investors, the SPV is equipped with liquidity guarantees by its founding mother company. The securities of an RMBS transaction are split into tranches with different seniority levels. Cashflows from the mortgage portfolio are distributed to the tranches in a waterfall-like structure according to the securities’ seniority. Junior or equity tranches will receive payments only after the more senior tranches have been fully repaid, which is why they pay the highest yield and bear the greatest risk.

The example RMBS transaction of Figure 3 has a total face value of 10 euro which is evenly distributed to ten mortgages (= the reference portfolio). If two mortgages default, the remaining cashflow of 8 euro is paid to the tranches successively according to their seniority level. The senior and mezzanine tranche (face value of 4 and 3 euro, respectively) can be fully paid off. The residual cashflow of 1 euro is not enough for the equity tranche with a notional of 3 euro. It suffers a loss of 2 euros.

The cashflow waterfall acts as protection for the more senior tranches, as losses are absorbed by the subordinated tranches. A second credit enhancement comes from diversification in the reference portfolio as it was perceived very unlikely that a large part of the mortgages will not perform. Other credit enhancements come from over-collateralization, spread accounts or third party guarantees. As it was very hard to find investors for BBB rated tranches of RMBS deals, they were pooled and taken as assets itself in another round of securitization (ABS CDO or CDO-squared).

RMBS transactions played a central role in the subprime crisis 2007-08 and the subsequent financial crisis. Ultimately, pooling and tranching, no matter whether first or second round, was a tool to turn mediocre or subprime mortgages into tradable AAA securities. However, this mechanism turned out to be terribly wrong and sparked the subprime crisis with its well-known consequences.\footnote{For further information, see FCIC (2011), Barnett-Hart (2009), Crouhy et al. (2008) or Hull (2009).}

\footnote{Other reasons for securitization are spread arbitrage opportunities, funding and economic risk transfer.}

\footnote{Another name for a transaction described above is collateralized debt obligation (CDO). However, to keep the acronyms concise, we use RMBS deal in this paper.}
In this paper, we focus exclusively on the correlation within the reference portfolios of RMBS deals. Essentially, the detailed and often very complex structure of the liability side of an RMBS transaction is irrelevant for the scope of this paper. Consequently, our definition of correlation has nothing to do with correlation between RMBS tranches or RMBS deals. Rather, it describes the dependence between individual mortgage loans, which is important for any mortgage lender, any RMBS investor or originator or rating agency.

Our data is from the US RMBS market and contains all deals available in Bloomberg that have been originated between 2004 and 2007, which are 4,605 deals in total. All tranches of an RMBS deal are backed by the same mortgage portfolio, i.e. there is one reference portfolio per RMBS deal. In Bloomberg, RMBS transactions are divided into three loan categories:

- **Residential B/C (ResBC)** deals consist of a majority (>50%) of subprime or B and C rated loans. Subprime mortgages are characterized by loans under which one or more previous payments were 30+ days delinquent.

- **Whole Loan (WL)** describes a MBS that is too large to be issued or guaranteed by Ginnie Mae, Freddie Mac or Fannie Mae. That is, a whole loan MBS represents jumbo mortgages and characterizes loans with a participation of one or more lenders.

---

Essentially, a fourth category is _Agency_, whose underlying mortgages are issued by a US government-sponsored agency like Fannie Mae or Freddie Mac. However, for this category there are no loss data available.
Alternative A-paper (AltA) is a type of US mortgage that, for various reasons, is considered riskier than prime and less risky than subprime. Typically AltA mortgages are characterized by borrowers with less than full documentation, lower credit scores, higher loan-to-values and more investment properties. AltA mortgages may have excellent credit but may not meet underwriting criteria for other reasons.

For 2,576 deals (56%) loss data is not available,\(^\text{11}\) for 881 deals (19%) the time series are incomplete, for 69 deals (1%) the time series is shorter than 12 month, for 69 deals (1%) the ending point of the time series is not consistent and 22 deals (0.5%) have been removed because of implausible data or errors.

Characteristics of the final data sample which consists of 988 reference portfolios of RMBS transactions are described in table 2. Reported values are averages weighted according to the following scheme. The 94 reference portfolios of the AltA 2006 vintage, for example, differ in how many mortgages are bundled into each portfolio (1,059 on average though). For a fixed year-category tuple, the values of table 2 are weighted according to the number of loans in the reference portfolios, ensuring that bigger portfolios get more weight. Ultimately, we want the averages to reflect the size of the underlying portfolios (in terms of number of mortgages). Weighted averages of the last column (WAVG) are calculated yet slightly different. As table 2 shows, the year-category data differ in the number of deals and in loans per deal, which ultimately means that the total number of mortgages per year-category tuple differs. For example, over 50% of all mortgage loans of our sample belong to the 2005, 2006 or 2007 ResBC vintage. The averages of the last column take these differences into account by weighting the values according loans per year-category tuple (= LpD • Deals). For obvious reasons the averages of Deals and LpD in the last column are not weighted. We apply this weighting scheme again when we report estimated asset correlation in table 3.

A reference portfolio consists on average of 1,052 mortgages (simple average), with an (weighted) average face value (FV) of 214,531 USD for the single mortgage. On average, we have 45.8 monthly observations per deal. As we have collected the data in January 2012, our sample covers roughly the years 2008-2011. Loss data prior to 2008 is not available, but this time period accounts for only 12.6% of the losses on average. This means in turn, that our sample contains over 87% of the total losses. Weighted average monthly default rate (MDR) is 0.9% with the year of origination having a significant impact on the default rate. For all three categories, the monthly default rate goes up sharply reaching its peak in the 2006 vintage. Cumulative total loss (CTL) measures the total loss incurred to date in relation to the original face value of the RMBS transaction. Especially the 2006 and 2007 ResBC deals with a total cumulative loss of almost a quarter consists of very low quality mortgages. Average FICO score clearly marks the ResBC category as subprime, while the other categories consist on average of low prime quality mortgages.\(^\text{12}\) For the 2004 ResBC deals, the FICO score is only available for four transactions with an average of 626. The FICO data is available

\(^{11}\) It is actually quite alarming that for over 50% of the RMBS deals loss data is not available.

\(^{12}\) There is no official definition of subprime, but in general a FICO score over 700 is considered as prime, while a FICO of under 650 is deemed subprime. The remainder is called midprime. The FICO score ranges between 300 and 850.
only for approximately 70% of the 988 reference portfolios. However, as it is not required to calculate asset correlation, a missing FICO score is no criterion for exclusion.

Finally, note that an overall weighted average monthly default rate of 0.9% in our sample implies very low credit quality portfolios.\(^{13}\) From a statistical perspective, a high default rate improves the validity of our estimation results. As noted earlier, the usual drawback of estimating asset correlation from default rates is that the default event is very rare.

The number of defaulted obligors per month is not directly reported in Bloomberg. Instead, we have monthly data on the number of loans per deal, current deal face value, incurred loss and loss severity for each RMBS reference portfolio. Reported loss data are net losses, i.e. we divide these net losses by the corresponding loss severity to get the gross loss. The default rate of a given month is then gross loss in relation to the current deal face value. In case an estimator needs to proceed with the number of defaulted obligors, we multiply the default rate with current number of loans. We calculate monthly default rates separately for each of the 988 reference portfolios.

\(^{13}\) This is especially alarming in the light of the maturity of the transactions which is in general well beyond 2030.
Table 2  
Characteristics of the 988 RMBS reference portfolios

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Deals</td>
<td>75</td>
<td>96</td>
<td>164</td>
<td>59</td>
<td>38</td>
<td>106</td>
<td>94</td>
<td>57</td>
<td>31</td>
<td>77</td>
<td>84</td>
<td>107</td>
<td>82.3</td>
</tr>
<tr>
<td>LpD</td>
<td>746</td>
<td>1,250</td>
<td>2,078</td>
<td>2,524</td>
<td>663</td>
<td>930</td>
<td>1,059</td>
<td>1,097</td>
<td>279</td>
<td>496</td>
<td>756</td>
<td>747</td>
<td>1,052</td>
</tr>
<tr>
<td>FV (×10³)</td>
<td>13.4</td>
<td>14.9</td>
<td>12.1</td>
<td>13.3</td>
<td>17.8</td>
<td>23.6</td>
<td>27.9</td>
<td>34.5</td>
<td>33.1</td>
<td>34.5</td>
<td>38.5</td>
<td>51.1</td>
<td>21.5</td>
</tr>
<tr>
<td>Obs</td>
<td>40.0</td>
<td>42.0</td>
<td>47.5</td>
<td>47.4</td>
<td>43.3</td>
<td>45.4</td>
<td>46.4</td>
<td>46.9</td>
<td>42.8</td>
<td>43.9</td>
<td>44.5</td>
<td>47.4</td>
<td>45.8</td>
</tr>
<tr>
<td>MDR</td>
<td>0.5</td>
<td>1.0</td>
<td>1.3</td>
<td>1.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
<td>1.1</td>
<td>0.5</td>
<td>0.9</td>
</tr>
<tr>
<td>CTL</td>
<td>4.0</td>
<td>10.1</td>
<td>23.6</td>
<td>25.0</td>
<td>1.3</td>
<td>3.8</td>
<td>9.6</td>
<td>9.2</td>
<td>1.9</td>
<td>6.4</td>
<td>15.7</td>
<td>9.1</td>
<td>15.0</td>
</tr>
<tr>
<td>FICO</td>
<td>n.a.</td>
<td>618</td>
<td>634</td>
<td>624</td>
<td>706</td>
<td>711</td>
<td>706</td>
<td>712</td>
<td>710</td>
<td>721</td>
<td>719</td>
<td>731</td>
<td>656</td>
</tr>
</tbody>
</table>

This table presents averages of the following categories: number of deals, number of loans per deal (LpD), face value of the single loan (FV), number of observations (Obs), monthly default rate (MDR), cumulative total loss (CTL) measured in percentage of the original deal face value and FICO score. Reported averages in the last column are weighted with respect to the number of loans per year category (= Deals × LpD), except for Deals and LpD which are simple averages for obvious reasons.
5. Estimated Asset Correlation

As seen in the simulation study, neither did we exclude nor did we select any estimator. Instead, we apply all estimators separately to the loss data of the 988 RMBS reference portfolios. We choose not to fix the underlying time periods across portfolios, instead we took the time series independently per portfolio which allows us to exploit the full data availability. Otherwise, we had to cut all time series to the length of the shortest one in the sample.

Table 3 presents average asset correlation for the applied estimator, year of origination and loan category. We report average values according to the weighting scheme of table 2, that is for a fixed year-category combination, the averages are weighted according to the number of loans per reference portfolio and the last column is weighted according to mortgages per year-category combination.

Estimated asset correlation is quite robust against the year of origination and against the category. Given a fixed estimator, the asset correlation values lie in a very similar range across the year-category tuples. Maybe the only exceptions are the 2004 Whole Loan vintage with somewhat higher values (9.6% averaged across estimator, see bottom row) and the 2004, 2005 ResBC vintages with slightly lower values (3.5% and 3.4%, see bottom row). Comparing simple averages of the five estimators further demonstrates the homogenous correlation values across the year-category tuples. Apart from the exceptions noted above, the correlation values range between 5.0% and 6.8% (see last row). Further, the applied estimator has only a small effect on asset correlation. Averaged and weighted across the year-category tuples, the values lie between 5.0% and 6.4% (see last column). The MM2 and the $PA_{beta}$ estimator tend to come up with slightly lower values on average (5.2% and 5.0%, respectively) than the remaining estimators. Given a fixed year-category combination, the estimated correlation values are quite homogenous across the estimators. Averaged over the five estimators and weighted across the 988 portfolios, estimated asset correlation is 5.7% (see bottom right value).

We conclude that, in general, neither the year of origination nor the category nor the applied estimator has significant impact on asset correlation and that it can therefore be estimated in a quite robust way. From a practical point of view, the differences between estimation methodologies become even more negligible. There is not much difference between an asset correlation of 5.0% and 6.4% in terms of economic capital, especially when measured at a very high confidence level.
Table 3
Average asset correlation of the 988 RMBS reference portfolios

<table>
<thead>
<tr>
<th></th>
<th>ResBC</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MM1</td>
<td>4.4</td>
<td>3.7</td>
<td>5.8</td>
<td>7.1</td>
<td>9.4</td>
<td>6.6</td>
<td>5.1</td>
<td>6.3</td>
<td>12.3</td>
<td>8.0</td>
<td>5.8</td>
<td>7.8</td>
<td>6.1</td>
<td></td>
</tr>
<tr>
<td>MM2</td>
<td>1.6</td>
<td>2.4</td>
<td>2.5</td>
<td>7.1</td>
<td>5.4</td>
<td>4.9</td>
<td>4.7</td>
<td>6.0</td>
<td>7.3</td>
<td>4.9</td>
<td>4.8</td>
<td>7.2</td>
<td>5.2</td>
<td></td>
</tr>
<tr>
<td>MLE</td>
<td>2.2</td>
<td>2.8</td>
<td>8.2</td>
<td>8.3</td>
<td>5.5</td>
<td>4.6</td>
<td>5.3</td>
<td>8.3</td>
<td>9.0</td>
<td>4.9</td>
<td>4.7</td>
<td>8.2</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>PA\text{v}</td>
<td>5.9</td>
<td>5.0</td>
<td>6.7</td>
<td>5.1</td>
<td>5.5</td>
<td>5.8</td>
<td>6.2</td>
<td>4.2</td>
<td>8.7</td>
<td>5.2</td>
<td>4.9</td>
<td>4.1</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>PA\text{leta}</td>
<td>3.6</td>
<td>3.0</td>
<td>4.7</td>
<td>5.8</td>
<td>8.0</td>
<td>5.4</td>
<td>4.1</td>
<td>5.1</td>
<td>10.5</td>
<td>6.5</td>
<td>4.8</td>
<td>6.5</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>3.5</td>
<td>3.4</td>
<td>6.2</td>
<td>6.7</td>
<td>6.8</td>
<td>5.4</td>
<td>5.1</td>
<td>6.0</td>
<td>9.5</td>
<td>5.9</td>
<td>5.0</td>
<td>6.8</td>
<td>5.7</td>
<td></td>
</tr>
</tbody>
</table>

This table presents average asset correlation for the applied estimator, year of origination and category. The bottom row are averages over the applied estimators, while the last column are weighted according to the number of loans per year-category pair.
In addition to average values, we want to compare asset correlation across estimators in greater detail. Figure 4 displays boxplots of the 988 estimated asset correlation values for the five estimators. Most apparent difference between the estimation methodologies is their tendency to produce negative estimates. Only for the MM2 and the $PA_Y$ estimator do we get negative asset correlation values. While the MM1 estimator can theoretically come up with negative values, the estimates of the beta approach are greater than zero for analytical reasons. The ML estimation is in fact symmetrically, which means that a positive estimate is equally probable as the corresponding negative value but for obvious reasons we report positive values.

Further, the MM1, the ML and the beta estimator are very similar with clearly right skewed distributed values. In contrast, the values of the MM2 and the $PA_Y$ estimator are distributed more symmetrically. The Vasicek estimator differs somewhat to the other estimators when the maximum value and the general dispersion of the values is considered. Actually, the wider dispersion of the Vasi estimator is consistent with the higher standard deviation in the Monte Carlo simulation (see table 1). Concerning the outliers in the boxplots, note that asset correlation above 35-40% are exceptional high values compared to the default rate based literature. Apart from these slight differences, the estimators' results are rather similar and we cannot observe any significant differences between the estimation methodologies. Again, we find that asset correlation is more or less robust against the choice of the estimator.

$14$ Asset correlation in the beta approach is a solution of an equation which is quadratic in $\frac{1}{\sqrt{s}}$ see section 2.4.

$15$ This fact is due to the integration over the (symmetric) Gaussian risk factor $X$ in equation (14) which makes positive and negative values equally likely.
The above correlation values are based on separately evaluating 988 mortgage portfolios. Another approach is to aggregate the granular loss data of different reference portfolios to the single default rate of a super portfolio. We do so for each year-category combination, that is we form twelve super portfolios (three categories times four years of origination) and calculate the corresponding default rate time series. We keep the super portfolio separated by the year-category dimension, as the estimation approaches assume constant probability of default within a portfolio and the monthly default rate clearly differs across the year-category tuples (see the monthly default rate row of table 2).

To form the twelve super portfolios, we have to fix a consistent time period of 40 observations and remove all portfolios with less observations. Characteristics of the twelve super portfolios and corresponding asset correlation values are displayed in table 4. Our initial sample is reduced to 705 deals (sum of the first row). There is a big discrepancy between the super portfolios with respect to the size measured in number of loans. The 2004 WL super portfolio is the smallest with 5,273 loans, while the 2006 ResBC super portfolio is the biggest with 292,487 mortgages on average. Total number of mortgages is 865,496 (sum of second row). On average (weighted according to number of loans) we observe 1,853 monthly defaults, which again shows that defaults in our sample are not scarce (see last column).

Turning now to the estimated asset correlation values, the most apparent observation is that asset correlation is drastically reduced in the super portfolios. Numbers in parentheses denote changes compared to evaluating the reference portfolios separately (see table 3). There is not a single case in which correlation in the super portfolios is higher than in the separately evaluated portfolios. Maximum asset correlation reduction averaged over the estimators, occurs in the 2004 WL case with a decrease of 8.2 percentage points (from 9.6% to 1.4%) and minimum reduction in the 2004 ResBC case with 2.8 percentage points. Weighted across the year-category tuple and averaged over the five estimators, asset correlation is reduced by 4.6 percentage points from 5.7% to 1.1% (compare bottom right values of table 3 and 4). Notably, there are five super portfolios with negative asset correlation (for different estimation approaches). Our other findings, which is that asset correlation is robust against the year-category dimension and the estimator, still hold (however now on a lower correlation level).
Table 4
Asset correlation based on 12 super portfolios

<table>
<thead>
<tr>
<th></th>
<th>ResBC</th>
<th>AHA</th>
<th>WL</th>
<th>WAVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loans</td>
<td>35,805</td>
<td>89,566</td>
<td>292,487</td>
<td>124,930</td>
</tr>
<tr>
<td></td>
<td>35,805</td>
<td>89,566</td>
<td>292,487</td>
<td>124,930</td>
</tr>
<tr>
<td></td>
<td>9,640</td>
<td>53,976</td>
<td>60,360</td>
<td>33,325</td>
</tr>
<tr>
<td></td>
<td>5,273</td>
<td>28,287</td>
<td>56,144</td>
<td>71,093</td>
</tr>
<tr>
<td>AMD</td>
<td>137</td>
<td>916</td>
<td>4,120</td>
<td>1,533</td>
</tr>
<tr>
<td></td>
<td>137</td>
<td>916</td>
<td>4,120</td>
<td>1,533</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>209</td>
<td>419</td>
<td>255</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>223</td>
<td>641</td>
<td>407</td>
</tr>
<tr>
<td>MM1</td>
<td>.7 (-3.7)</td>
<td>.6 (-5.0)</td>
<td>.9 (-4.9)</td>
<td>.8 (-6.5)</td>
</tr>
<tr>
<td></td>
<td>.7 (-3.7)</td>
<td>.6 (-5.0)</td>
<td>.9 (-4.9)</td>
<td>.8 (-6.5)</td>
</tr>
<tr>
<td></td>
<td>.7 (-3.7)</td>
<td>.6 (-5.0)</td>
<td>.9 (-4.9)</td>
<td>.8 (-6.5)</td>
</tr>
<tr>
<td></td>
<td>.7 (-3.7)</td>
<td>.6 (-5.0)</td>
<td>.9 (-4.9)</td>
<td>.8 (-6.5)</td>
</tr>
<tr>
<td>MM2</td>
<td>.2 (-1.5)</td>
<td>.0 (-2.3)</td>
<td>.1 (-5.6)</td>
<td>.4 (-5.8)</td>
</tr>
<tr>
<td></td>
<td>.2 (-1.5)</td>
<td>.0 (-2.3)</td>
<td>.1 (-5.6)</td>
<td>.4 (-5.8)</td>
</tr>
<tr>
<td></td>
<td>.2 (-1.5)</td>
<td>.0 (-2.3)</td>
<td>.1 (-5.6)</td>
<td>.4 (-5.8)</td>
</tr>
<tr>
<td></td>
<td>.2 (-1.5)</td>
<td>.0 (-2.3)</td>
<td>.1 (-5.6)</td>
<td>.4 (-5.8)</td>
</tr>
<tr>
<td>MLE</td>
<td>.6 (-1.6)</td>
<td>.6 (-2.2)</td>
<td>1.3 (-6.9)</td>
<td>.8 (-7.6)</td>
</tr>
<tr>
<td></td>
<td>.6 (-1.6)</td>
<td>.6 (-2.2)</td>
<td>1.3 (-6.9)</td>
<td>.8 (-7.6)</td>
</tr>
<tr>
<td></td>
<td>.6 (-1.6)</td>
<td>.6 (-2.2)</td>
<td>1.3 (-6.9)</td>
<td>.8 (-7.6)</td>
</tr>
<tr>
<td></td>
<td>.6 (-1.6)</td>
<td>.6 (-2.2)</td>
<td>1.3 (-6.9)</td>
<td>.8 (-7.6)</td>
</tr>
<tr>
<td>PAV</td>
<td>1.4 (-4.4)</td>
<td>.2 (-5.2)</td>
<td>3.7 (-3.0)</td>
<td>.5 (-4.6)</td>
</tr>
<tr>
<td></td>
<td>1.4 (-4.4)</td>
<td>.2 (-5.2)</td>
<td>3.7 (-3.0)</td>
<td>.5 (-4.6)</td>
</tr>
<tr>
<td></td>
<td>1.4 (-4.4)</td>
<td>.2 (-5.2)</td>
<td>3.7 (-3.0)</td>
<td>.5 (-4.6)</td>
</tr>
<tr>
<td></td>
<td>1.4 (-4.4)</td>
<td>.2 (-5.2)</td>
<td>3.7 (-3.0)</td>
<td>.5 (-4.6)</td>
</tr>
<tr>
<td>PAV_{na}</td>
<td>.6 (-3.0)</td>
<td>.5 (-2.5)</td>
<td>.9 (-3.9)</td>
<td>.7 (-5.1)</td>
</tr>
<tr>
<td></td>
<td>.6 (-3.0)</td>
<td>.5 (-2.5)</td>
<td>.9 (-3.9)</td>
<td>.7 (-5.1)</td>
</tr>
<tr>
<td></td>
<td>.6 (-3.0)</td>
<td>.5 (-2.5)</td>
<td>.9 (-3.9)</td>
<td>.7 (-5.1)</td>
</tr>
<tr>
<td></td>
<td>.6 (-3.0)</td>
<td>.5 (-2.5)</td>
<td>.9 (-3.9)</td>
<td>.7 (-5.1)</td>
</tr>
<tr>
<td>mean</td>
<td>.7 (-2.8)</td>
<td>.3 (-5.0)</td>
<td>1.3 (-4.0)</td>
<td>.7 (-6.0)</td>
</tr>
<tr>
<td></td>
<td>.7 (-2.8)</td>
<td>.3 (-5.0)</td>
<td>1.3 (-4.0)</td>
<td>.7 (-6.0)</td>
</tr>
<tr>
<td></td>
<td>.7 (-2.8)</td>
<td>.3 (-5.0)</td>
<td>1.3 (-4.0)</td>
<td>.7 (-6.0)</td>
</tr>
<tr>
<td></td>
<td>.7 (-2.8)</td>
<td>.3 (-5.0)</td>
<td>1.3 (-4.0)</td>
<td>.7 (-6.0)</td>
</tr>
</tbody>
</table>

Characteristics of the 12 super portfolios are given, together with the corresponding asset correlation for the five estimators. AMD denotes average number of monthly observed defaults. Numbers in parenthesis are changes compared to separately evaluated portfolios. The bottom row are averages over the applied estimators, while the last column is weighted according to the number of loans per year-category pair.
Why is asset correlation reduced in the super portfolios? Looking at typical default rate time series from the simulation study of section 3 will yield a simple and intuitive explanation. The upper two graphs of Figure 5 show default rates of a portfolio with zero asset correlation but different probabilities of default (1.6% and 10.0%, respectively). Same for the lower panel, however the asset correlation is increased to 27%. We clearly see how the look of default rates changes if i) the probability of default increases (left side compared to right side) or ii) the asset correlation increases (upper panel compared to lower panel). In the former case, the stable default rate behavior is maintained, but the curve is shifted upwards. In the second case, the (mean) location of the curve stays unchanged but the style of the curve goes from a smooth path to a heavily fluctuation line.

In both right hand side portfolios of Figure 5, on average one out of ten loans will default at any given point in time. However, the way how the default rate of 10% is realized is fundamentally different. In the upper case the default rate does not exceed 13% while in the lower graph there are five observations at which the default rate exceeds 25% percent (marked in red). These outliers (and variation, of course) are a clear sign of high asset correlation.
The economic intuition of this effect is that default rates follow the cyclic behavior of economy. Transferred to the model setup, a strong connection to the macroeconomy means a high exposure to the systematic risk factor which corresponds to high asset correlation (see equation (2) and (3)). On the other hand, if there is no exposure to the systematic risk factor, default rates should be very stable over time. Thus, we have a good chance of guessing the asset correlation from merely looking at the default rate time series; at least, we can discriminate between high and low asset correlation. The greater the variation in the default rates, the greater the asset correlation in the portfolio.

Next, we check for variation and outliers in the default rate time series of the RMBS reference portfolios to get an idea of the underlying asset correlation. Figure 6 compares default rate curves of 20 randomly selected RMBS reference portfolios (upper graph) with 20 simulated default rates with similar probability of default but for two different asset correlation values. The default rate curves of the RMBS portfolios are very smooth and homogeneous with hardly any outliers indicating low asset correlation. The appearance of the RMBS default rates resembles more closely the simulated default rates with asset correlation equal to 4.4% (middle graph) than the heavily fluctuating default rates of the 13% asset correlation portfolio (bottom graph). Again, we have strong evidence of very low correlation in the RMBS reference portfolios. However, this evidence is not based on any model framework and estimation methodologies but rather common sense observation and maybe this way of deriving conclusion is even more reliable and robust.

So, how does the above discussion help to interpret the lower correlation in the super portfolios? We know that spikes in the default rate curve are signs of high correlation. It clearly takes a more extreme scenario to produce an outliers default rate in a 300,000 loan portfolio than in a 2,000 loan portfolio (compare the sizes of the 2006 ResBC portfolios in table 2 and 4). Combining many mortgage portfolios to a single portfolio acts like smoothing the default rate because the high default numbers in some parts of the portfolio are absorbed by the large number of loans. In fact, if variation and outliers in the default rate time series disappear if a single large portfolio is formed, that means that variation and outliers are not due to systematic effects but rather a coincidental result of arbitrarily cut portfolios. If variation and outliers had a systematic reason, they should not vanish in the super portfolios.
Figure 8
Simulated default rates compared to RMBS reference portfolios’ default rates

20 ResBC 2006 deals, PD=1.5%, ρ=7%

20 Samples, PD=1.7%, ρ=4.4%

20 Samples, PD=1.6, ρ=13%
6. Discussion and Interpretation of the Results

In general, the estimated asset correlation in this study is consistent with the literature that use default rates. But given the nature of our data set, they are surprisingly low. Remember that especially the 2006 and 2007 ResBC deals consist of typical subprime loans with extremely high default rates. And according to the commonly held view, the underestimation of asset correlation was one of the core reasons for the subprime crisis. However, an estimated asset correlation of roughly six percent (or even one percent based on the super portfolios) in the RMBS reference portfolios hardly supports this view. So the question is how an average asset correlation of six percent in RMBS reference portfolios could spark a financial crisis that brought the global financial system to collapse.

The answer lies not in correlation between single-name mortgages but in steadily and homogeneously increasing loss rates of the RMBS reference portfolios. Figure 7 shows the evolution of the incurred loss measured in percentage of the original deal face value of ten randomly selected RMBS reference portfolios. There are hardly any irregularities or variation in the curves. The losses are realized in a very stable and homogeneous way. In fact, there seems to be hardly any uncertainty in the cumulative loss rates (which is consistent with the stable default rates of Figure 6, upper graph). Further, the similarity of the cumulative loss rates is remarkable. The RMBS reference portfolios behave almost identically in terms of how and when losses occur. These highly synchronous and stable cumulative loss rates have a very important implication for the default behavior of tranches of RMBS transactions and in particular for the mezzanine tranches.

![Figure 7](cumulative_loss_rates.png)

*Cumulative loss rates of 10 RMBS reference portfolios*

Cumulative loss is measured in percentage of the original deal face value. The RMBS deals are from the 2006 ResBC vintage.
If we assume an average monthly default rate of roughly 1.0% (see last column of table 2) and a very low asset correlation in the corresponding reference portfolio, we can very reliably predict the time when any given tranche is hit (remember that low asset correlation means no uncertainty in the cumulative loss). If we further assume attachment points of 5% and 15% for the mezzanine tranche, it will be completely wiped out after roughly 15 month. And if all reference portfolios behave similarly, this effect will be similar for all mezzanine tranches which implies a high default correlation among the mezzanine tranches (not to be mistaken for correlation among individual mortgage loans).\footnote{For an evaluation of the risk of CDO transactions see Hamerle and Plank (2008).} Eventually, low asset correlation in reference portfolios turns into high default correlation of mezzanine tranches of RMBS deals, which explains the low asset correlation values in the light of the subprime crisis.

To put the importance of asset correlation into perspective, from a bank's view on risk capital, asset correlation is very important.\footnote{Risk capital is the amount of capital needed to absorb potential losses of a portfolio.} However, for the evaluation of a single RMBS deal, asset correlation is rather meaningless. To see why, return to Figure 5. In both right-hand side portfolios on average one out of ten loans will default. But the upper portfolio is clearly the superior choice, when it comes to minimizing risk capital, as it seems highly unlikely that the default rate will exceed 20% at any point in time. On the other side, for the tranches of an RMBS deal, it is only the cumulative loss that matters and in this regard, the upper and lower portfolios are equivalent. After a sufficient amount of time, 10% loans will have defaulted in both portfolios. So, for an RMBS deal the primarily focus is on the cumulative loss rather than asset correlation. And these arguments show how the low asset correlation values of this study can be interpreted in the light of the subprime crisis.
References


