DISCUSSION PAPER SERIES

No. 9171

BANK RATINGS: WHAT DETERMINES THEIR QUALITY?

Harald Hau, Sam Langfield and David Marqués Ibañez

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP9171.asp**

ISSN 0265-8003

BANK RATINGS: WHAT DETERMINES THEIR QUALITY?

Harald Hau, University of Geneva and CEPR
Sam Langfield, European Central Bank and Financial Services Authority (FSA)
David Marqués Ibañez, European Central Bank

Discussion Paper No. 9171
October 2012

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 9171

October 2012

ABSTRACT

Bank ratings: What determines their quality?*

This paper examines the quality of credit ratings assigned to banks in Europe and the United States by the three largest rating agencies over the past two decades. We interpret credit ratings as relative assessments of creditworthiness, and define a new ordinal metric of rating error based on banks' expected default frequencies. Our results suggest that rating agencies assign more positive ratings to large banks and to those institutions more likely to provide the rating agency with additional securities rating business (as indicated by private structured credit origination activity). These competitive distortions are economically significant and help perpetuate the existence of 'too-big-to-fail' banks. We also show that, overall, differential risk weights recommended by the Basel accords for investment grade banks bear no significant relationship to empirical default probabilities.

JEL Classification: E44, G21, G23 and G28
Keywords: conflicts of interest, credit ratings, prudential regulation, rating agencies and sovereign risk

Harald Hau
Professor of Economics & Finance
Swiss Finance Institute, Senior Chair
University of Geneva and GFRI
Unipignon
40, Boulevard du Pont d'Arve
CH 1211 Geneva 4
SWITZERLAND

Email: prof@haraldhau.com

Sam Langfield
European Systemic Risk Board
European Central Bank

Email: Sam.Langfield@ecb.int

David Marqués Ibañez
European Central Bank
Kaiserstrasse 29
60311 Frankfurt am Main
GERMANY

Email:
david.marques_ibanez@ecb.int

*Preliminary version of a paper prepared for the 56th panel meeting of *Economic Policy*, October 2012. Opinions expressed herein are those of the authors only. They do not necessarily reflect the views of, or involve any responsibility for, the institutions to which they are affiliated. Any errors are the fault of the authors. We are indebted to Thomas Drechsel and Matthias Efing for excellent research assistance and to Johannes Micheler and Antonia Simeonova for providing data and code. Allen Berger, Oliver Burkart, Jean-Pierre Danthine, Matthias Efing, Artus Galiay, Zijun Liu, Simone Manganelli, Jose Geli Manzano, John Muellbauer, Steven Ongena, Alex Popov, Ana Rita Ribeiro Mateus, Andrei Sarychev, Frank Smets, Balázs Zsámboki and four anonymous referees provided helpful comments. The Managing Editor in charge of this paper is Philip Lane.

Submitted 01 October 2012

1. INTRODUCTION

In the aftermath of the initial phase of the financial crisis in 2007–08, popular indignation often focused on credit ratings assigned to banks: most failing banks enjoyed investment grade status shortly before defaulting. Ratings of products sold by banks, such as securitised credit, were also found wanting. Ratings were subject to particularly sharp criticism since they are supposed to evaluate default risk over the economic cycle. These cumulative mistakes conveyed the impression that the entire rating system was flawed, along with large parts of the prudential regulation of banks, which relies heavily on credit ratings.

We pursue three objectives. First, we provide a comprehensive empirical measurement of the quality of banks' ratings over the past 20 years based on a new ordinal metric of rating error. Our method interprets bank credit ratings in a strictly ordinal manner: banks are ranked by their credit rating; and this ranking is then compared to a second ranking of expected default frequencies two years later. The difference between these two ranks is then defined as the *Ordinal Rating Quality Shortfall* (*ORQS*). The ranking procedure provides a good measure of rating quality, since it does not require measurements of bank risk to be correct in absolute terms. In particular, our ranking procedure is not affected by dramatic increases in cardinal measures of default probability observed during financial crises. Second, we use this non-parametric rating quality measure for a structural analysis into the determinants of rating quality. In particular, we examine the role of various bank characteristics on rating quality and rating bias in order to unveil their potential causes. Third, we discuss the policy conclusions of our evidence and outline the most promising policy option to improve bank rating quality.

Any analysis of rating quality faces the question, what is the meaning of a credit rating? Literature published by the rating agencies themselves is testimony to considerable confusion. Moody's *Rating Methodology* (1999) states that 'one of Moody's goals is to achieve stable expected default rates across rating categories', which suggests that ratings are absolute or *cardinal* measures of future default. By contrast, other documents characterise Moody's credit ratings as '*ordinal* measures of expected loss' (Moody's, 2006). Statements by other rating agencies are similarly contradictory about the meaning of credit ratings, and in particular about cardinal versus ordinal interpretation.

A cardinal rating for banks requires rating agencies to predict bank distress, including during generalised banking crises, whereas ordinal ratings only assess banks' relative creditworthiness. Our evaluation of bank rating quality adopts the weaker ordinal standard. Our intention is not to hold rating agencies to an unreasonable standard of absolute accuracy over time, but only to a much weaker requirement of cross-sectional consistency in their bank rankings.

Our analysis draws on a large and comprehensive dataset of bank ratings from the three major rating agencies. The data on credit ratings are combined with yearly accounting balance sheet information on rated banks and monthly expected defaults frequencies (*EDFs*) from those banks obtained from Moody's KMV. In total, our dataset has 38,753 bank-rating observations at quarterly frequency over the period ranging from 1990 to 2011. By using

EDFs calculated by Moody's as a measure of risk, we maintain methodological fairness by avoiding subjective risk modelling choices (see Section 4). *EDFs* capture perceptions of bank risk derived from a structural model incorporating expectations from equity markets. Moreover, unlike some other indicators of bank risk, *EDFs* are observed in relation to individual banks over a long time series.

To illustrate the advantage of an ordinal (non-parametric) analysis, consider the evolution of expected default frequencies (*EDFs*) for our sample banks depicted in Figure 1. The left-skewed distribution shows a spike at the high quantiles of bank credit risk from 2008. Short of predicting the financial crisis, credit ratings are unlikely to capture such enormous fluctuations in bank credit risk. Any cardinal measure of rating quality would therefore be strongly tainted by the unpredictability of the crisis itself. By contrast, our strictly rank-based measure of rating quality is not altered by a shift in the distribution of expected default frequencies, as long as the rank ordering remains unchanged.

[Insert Figure 1 here]

Our analysis provides a rich set of empirical insights into the structure and the determinants of credit rating quality. First, we find that ordinal rating quality is countercyclical. With the onset of a banking crisis, the (ordinal) information content of credit ratings increases. In normal times, bank credit ratings are informative about future expected default probabilities only for the 25% lowest-rated banks with ratings of BBB+ and below, but not for investment grades above BBB+. Unconditionally, our results suggest that an A-rated bank is as likely to become distressed as an AAA-rated institution.

Second, bank characteristics significantly influence bank rating quality. A traditional banking model with a large loan share increases the accuracy of the credit rating. At the same time, the accuracy of bank ratings decreases with bank size: larger banks feature upwardly biased ratings. This rating bias in favour of large banks is economically significant. An increase in the size of a bank by two standard deviations implies that the credit rating rank relative to the *EDF* rank is overestimated by 15 positions for every 100 banks in the sample. This corresponds, for example, to an unwarranted rating improvement from A- to A, which on average equates to a financing cost decrease of 40 basis points.

Third, our results suggest that there are conflicts of interest between banks and rating agencies that seem to alter the rating process. Using additional data on banks' agency-specific securitisation business, we find that rating agencies give systematically better ratings to banks that provide an agency with a large quantity of business in the form of rating asset-backed securities.

Fourth, multiple bank ratings by different rating agencies correlate with less favourable ratings relative to future *EDFs*. This finding casts some doubt on the assertion that rating competition fosters rating inflation through 'ratings shopping'.

These empirical insights lead us to a number of policy conclusions, which we summarise as follows:

1. The strong discrimination of credit risk within the investment grade category (as maintained under Basel II and Basel III) cannot be reconciled with our evidence on empirical bank default probabilities. Taken at face value, our results suggest that all investment grade bank ratings above A- deserve the same risk weight, at least with respect to bank ratings.
2. Rating agencies seem to systematically assign more favourable ratings to larger banks and to those institutions that provide the respective rating agency with additional rating business in the private structured credit markets. These results are in line with the 'too big to fail' problem and can lead to competitive distortions. As a result, an increase in supervisory intensity for large banks is warranted.
3. The generally low information content of bank ratings implies that punitive measures for (ex-post) rating failures cannot be translated into a workable policy framework. The hope that the incentives of rating agencies will change if investors pay directly for ratings seems similarly misplaced, in view of the evidence of investor collusion in ratings inflation over the pre-crisis period (Calomiris, 2009).
4. Given the strong negative externalities of bank opacity, a promising policy option lies in enhanced transparency of banks and rating agencies. Substantial improvement of banks' public disclosure and forensic examination of rating performance by an external party seem warranted. A related insight concerns heterogeneity in accounting practices across countries, which compounds incentive problems due to bank opacity, leading to costly delays in the recognition of banking problems.

The paper is organised as follows. The next section explains the motivation of our focus on bank ratings. Section 3 describes the literature on credit ratings, while Section 4 explains the data sources. Section 5 presents the methodology and Section 6 discusses the main hypotheses. Section 7 explains the regression results and section 8 robustness issues. The last section presents the main conclusions and policy implications.

2. WHY DO CREDIT RATINGS MATTER?

Investors' reliance on credit ratings has increased over the past 30 years. Financial transactions have grown in volume and complexity and finance has shifted from banks to capital markets, particularly in the US (Boot and Thakor, 2010). At the same time, deregulation and financial innovation – including securitisation and credit derivatives – have made the banking sector larger, more concentrated, more complex and more closely connected with capital markets.

Acquiring information is costly, particularly for fixed income investors, given collective action problems. Thus investors seek to outsource creditworthiness assessments to rating agencies. More than half of all corporate bonds are held by institutions subject to ratings-based investment restrictions (Bongaerts et al, 2011).

Bank ratings are a particularly important determinant of the issuance cost of senior unsecured debt. Senior unsecured debt remains the largest source of long-term funding for banks (Oliver Wyman, 2011). Since 2007, new issuance of unsecured debt as a share of total bank debt issuance has somewhat decreased, partly substituted by more deposits and secured debt. Secured debt accounted for less than 30% of total bank debt issuance in 2009; this figure had risen to 40% in the first half of 2012, according to data from Dealogic. In the case of European banks, total debt issuance amounted to approximately EUR1.1tn in 2009. Thus, despite recent marginal changes in business models, senior unsecured debt ratings remain an important assessment of bank creditworthiness.

But compared to other corporations, banks pose a particular challenge for external rating agencies. Banks are inherently opaque and exposed to a multiplicity of risks. Bank business is characterised to a significant extent by asymmetries of information and actual (and potential) regulatory interventions.[1] We may therefore consider that bank ratings provide a lower bound (or worst-case setting) for the quality of external ratings compared to other corporate ratings (Morgan, 2002).

At the same time, banks' central role in credit intermediation is important for efficient allocations of capital and risk, and thus for activity in the real economy. The collapse in credit supply during the financial crisis of 2008–09 led to a long-lasting reduction in the level of output relative to the pre-crisis trend (Reinhart and Rogoff, 2009; Campello et al, 2010). Publicly funded recapitalisation and guarantees on deposits and debt put pressure on the credibility of sovereigns' signatures. These considerations compound the economic importance of unbiased and efficient assessments of bank creditworthiness.

The particular role of credit ratings in the financial system is enshrined in policy. From 1936 onwards, regulatory authorities in the United States have, in many instances, delegated oversight of the credit quality of banks' portfolios to rating agencies (White, 2010). For instance, in exchange for liquidity, central banks require a minimum quality of collateral, defined in many cases by reference to credit ratings. In the realm on prudential banking regulation, the Basel II accord increased regulatory reliance on credit ratings. Under this agreement, minimum capital levels are specified as a proportion of risk-weighted assets, where risk weights may be calculated using credit ratings. Yet compared with the unweighted leverage ratio, there is no evidence to suggest that the risk-weighted capital ratio is a superior predictor of bank failure during crisis periods (Mariathasan and Merrouche, 2012). Moreover, anecdotal evidence suggests that although large banks sometimes use internal models as a substitute for credit ratings for their credit assessments, the internal models themselves often tend to rely heavily on ratings for actual or methodological input. The Basel III agreement expresses a broad intention to mitigate reliance on ratings of securitised loans, but introduces an additional role for credit ratings with respect to counterparty credit risk from over-the-counter derivatives (BCBS, 2010).

[1] This is best illustrated by the spectacular bankruptcies of Enron and WorldCom – both of which failed as 'financial conglomerates' rather than ordinary energy or telephone companies.

The performance of credit rating agencies has faced heightened scrutiny since the onset of the financial crisis in 2007. The model of the credit rating agency industry – to take private information, and a fee, from an issuer, and publish a summary judgement in a rating, with special status conferred by public policy – has been heavily criticised (Pagano and Volpin, 2010; Financial Stability Board, 2010). High reliance on rating agencies increases the exposure of the financial system to the accuracy of ratings. Mistakes and biased forecasts have the potential to cause or exacerbate crises, rendering the financial system more vulnerable to cliff effects (Manso, 2011).[2]

3. LITERATURE

Credit ratings play a key role in the financial system, but determinants of their quality are poorly understood. There is scant empirical literature on bank credit ratings and the quality of such ratings. This is surprising, since credit ratings potentially contain information on banks' riskiness throughout the cycle that is not otherwise available to the market.

Agency and incentive problems are a central theme in the literature on credit ratings. These agency problems arise in different forms. The majority of the research focuses on the conflict of interest between the ratings consumer (i.e. the financial investor) and the issuer, who pays for the rating and has an incentive to lobby for positive bias from the rating agency. This conflict of interest arose in 1975, when credit rating agencies shifted from an 'investor pays' to an 'issuer pays' model (White, 2010; Pagano and Volpin, 2010). Under the latter model, issuers may credibly threaten to switch to a competing agency, which could lead to positive rating bias referred to as 'ratings shopping'. The larger the potential future business between rating agencies and their clients, the larger an agency's incentive to inflate ratings. Other research has focused on the power of rating agencies rather than that of their clients. Rating agencies may issue downside-biased unsolicited ratings for which no fee is charged, thus threatening credit issuers who do not solicit ratings (Partnoy, 2002; Fulghieri et al, 2010). According to Griffin and Tang (2011), rating teams that interact more closely with their clients produce more upwardly biased ratings than those teams in the supervisory unit. Other evidence points to additional upward bias in credit ratings whenever the value of the asset, and therefore the agency fee, is high (He et al, 2011).

A second and more perilous incentive conflict may arise from rating-contingent financial regulation of banks and other investors (i.e. the buy side) with agency problems of their own. As Calomiris (2009) highlights, rating inflation may arise from collusion between rating agencies and security investors in the pursuit of regulatory arbitrage, higher leverage and short-term profits. This could explain why such a large quantity of collateralised assets with inflated ratings turned out to be on bank balance sheets during the crisis. Opp et al (2012) show that rating-contingent regulation can significantly lower an agency's incentives

[2] In the case of AIG, over-the-counter derivatives contracts provided for margin calls in the event of a rating downgrade of the underwriter, precipitating a vicious circle.

to acquire costly information and to produce high-quality ratings. Investors do not scrutinise rated securities as they enjoy regulatory benefits from inflated ratings. In related work, Efing (2012) highlights that agencies may bias their ratings upwards even with access to free and full credit information, because they can share with the issuers the incremental revenues from selling rating-inflated debt to regulated banks seeking more leverage. The normative conclusion is that rating-contingent bank regulation might be very negative from a welfare perspective.[3]

Reputational capital is often seen to attenuate these agency problems (Cantor and Packer, 1995; Covitz and Harrison, 2003). Rating agencies have a long-term incentive to invest in their reputation for producing high-quality ratings that are unbiased assessments of creditworthiness. Yet a recent body of theoretical literature argues that the quality of credit ratings based on reputational concerns is likely to change over the business cycle as ratings' quality decreases during booms and increases during troughs (Bar-Isaac and Shapiro, 2011a). During periods of economic expansion, when fewer defaults occur, rating agencies' potential returns on reputational capital would be lower. Moreover, during these episodes it is more difficult for final investors to ascertain rating quality. The presence of 'naïve investors' would also strengthen the countercyclical nature of ratings quality (Bolton, Freixas and Shapiro, 2012).[4] Evidence of rapid and widespread downgrades of structured finance securities' ratings over 2007–08 is consistent with the hypothesis of counter-cyclical ratings quality (Benmelech and Dlugosz, 2009). Expansionary periods indeed coincide with higher revenues for rating agencies, but it is unclear whether this is due to cyclicality in the volume of rating business or cyclical rent extraction. Existing evidence suggests that credit ratings are a particularly good indicator of credit risk during crisis periods (Hilscher and Wilson, 2011).

Competition among rating agencies could also affect ratings quality through different channels with contrary predictions. Higher competition among rating agencies would reduce the benefit of good reputation, leading to lower rating quality (Camanho, Deb and Liu, 2010). Similarly, rating quality can be reduced if issuers shop for more favourable ratings (Bolton, Freixas and Shapiro, 2012). Becker and Milbourn (2010) assert that the entry of the rating agency Fitch in 1997 led to deterioration in ratings' quality. On the other hand, the industrial organization literature generally sees a positive role of competition for product quality – a finding that should also transfer to the market for credit ratings (Hörner, 2002).

Rating quality in the banking sector might be also affected by reasons unrelated to incentive problems. In particular, opacity and complexity might impair rating quality. Compared with other large corporations, big banks are extremely opaque in terms of their legal structure, risk exposures and value creation process. Such opacity makes it harder to predict financial distress for banks than for non-bank institutions. Rating disagreements across agencies occur more often in the case of banks' ratings than those of other industries

[3] For Efing (2012), this is the case when agencies can share with issuers the incremental revenues from selling rating-inflated debt to regulated banks that seek to arbitrage capital requirements.

[4] In other words 'ratings are more likely to be inflated when there is a larger fraction of naïve investors in the market who take ratings at face value' (Bolton et al, 2012). Note that this does not mean that asset managers (i.e. the agents of the ultimate investors) are naïve (Calomiris, 2009).

(Morgan, 2002). Structural changes in the banking sector have increased opacity in recent decades – thus rendering the assessment of bank creditworthiness even more complicated. Financial innovation has increased complexity in banking; more direct funding from financial markets and securitisation activity have formed part of a wider trend of innovation that has intensified credit risk transfer between intermediaries (Boot and Thakor, 2010). Costly observability of creditworthiness reduces the ability of market participants to screen noisy ratings and increases the cost to a rating agency of issuing informative forecasts (Bar-Isaac and Shapiro, 2011b). Generally, rating agencies might find it more profitable simply to issue lower-quality ratings rather than to confront increasing bank complexity. (Mathis et al, 2009; Skreta and Veldkamp, 2009; Opp et al, 2010).

If asset complexity is an important determinant of rating quality, then a bank's asset choice and business model should explain rating accuracy. A number of studies have focused on the impact of bank business models on bank risk and performance during the recent crisis. Beltratti and Stulz (2011), for example, found that banks with more Tier I capital and a higher ratio of loan to total assets performed better in the initial stages of the crisis. Berger and Bouwman (2011) show that during banking crises higher capital levels improve banks' performance, while a larger deposit base and more liquid assets are associated with higher returns. Cole and White (2012) show that higher levels of capital and stronger CAMEL ratings lower the likelihood of bank failure. Altunbas et al (2011) find that banks with higher risk are larger and have less capital, greater reliance on short-term market funding and aggressive credit growth. In light of this research, we hypothesise that a bank's business model is related to the accuracy of its credit rating.

4. DATA

We construct a comprehensive panel of banks' ratings from January 1990 to December 2011 based on rating data from Standard and Poor's, Moody's and Fitch. The ratings datasets record whenever a rating is changed, affirmed or withdrawn. We extract a time series by recording for each bank the most recent rating observation at the end of each quarter. Retained are all banks headquartered in the US or EU15. To avoid double-counting ratings within a single institution, we discard any bank that is junior in the organisational structure – for example, HSBC France is discarded; HSBC Holdings plc is retained.[5]

[Insert Table 1 here]

Ratings by the three rating agencies are translated into a numerical value from 1 to 24 according to Table 1, where the lowest rank number corresponds to the highest credit ranking. Summary statistics for the quarterly ratings data are provided in Table 2. We obtain an unbalanced panel with 38,753 quarterly bank ratings. Ratings are assigned to 369 banks,

[5] Data processing limitations do not allow us to extend the analysis to the bank subsidiary level.

which are each rated by between one and three agencies. Standard and Poor's provides the most complete coverage with 16,928 bank ratings at quarterly frequency, followed by Moody's (2,715) and Fitch (9,110). Rating coverage was relatively incomplete in the early 1990s, before widespread adoption of the Basel recommendations: 75% of all panel observations concern the second decade after January 2000.

[Insert Table 2 here]

The rating data are matched with annual accounting data from Bankscope. The matching process employs bank identifiers, a text-string matching algorithm (Winkler, 2006) and manual work. Most accounting data are available only after 1994 and feature varying degrees of reporting coverage. To account for data errors, we undertake some winsorizing of extreme observations on balance-sheet variables. For example, we impose that observations on leverage must lie between 0 and 100. Table 2 provides the definitions of the accounting variables we retain and their summary statistics.

Finally, we match the above panel with data on expected default frequencies (*EDFs*) as a measure of bank distress. *EDFs* are obtained from a structural model of corporate default and widely used to price corporate bond debt (Merton, 1974; Longstaff and Schwartz, 1995).[6] The main model inputs are the volatility of asset returns (which aims to capture business risk) and the difference between the market value of a bank's assets and the book value of its liabilities (accounting for leverage). Increases in volatility or leverage translate into higher *EDF* levels. Our analysis draws on *EDFs* calculated by the rating agency Moody's, contemporaneously to the rating process. Moody's calculations are undertaken monthly and draw on a large proprietary default database owned by KMV (Dwyer and Qu, 2007). It is possible to reconstruct proxy *EDFs* using only public data (Bloechlinger et al, 2012). But drawing on existing *EDF* data has the advantage that we do not need to make any parameter or calibration choices. Our measurement of rating errors is thus immune to any model selection or back-fitting criticism.

It is difficult to find the ideal indicator of actual bank risk. The *EDF* measure (as calculated by Moody's KMV) has a number of limitations. For instance, more elaborate structural models of credit risk have been shown to provide a better out of sample prediction of bank risk (Bharath and Shumway, 2008).[7] Also, *EDFs* incorporate bail-out expectations only indirectly due to the junior status of equity in banks' liability structures.

Our choice of the *EDF* indicator is justified by specific reasons linked to our research design. First, unlike other indicators of bank risk (such as spreads on credit default swaps), *EDFs* are available with a relatively long time series, facilitating panel analysis over

[6] More specifically the calculation of *EDF* builds on Vasicek and Kealhofer's extension of the Black-Scholes-Merton option-pricing framework to make it suitable for practical analysis.

[7] For our purpose our main assumption would be that its functional form is useful for forecasting defaults due to the relative nature of our variable and the short-term forecasting horizon for the EDF variable. We therefore do not assume that the Merton distance to default model used by KMV produces an optimal and sufficient statistic for the probability of default.

a longer time period. Second, *EDFs* are widely used by researchers, central banks, regulators, supervisors, financial institutions and investors (Garlappi et al, 2007; IMF, 2009). Third, compared with structural measures of credit risk such as CreditMetrics (created by JP Morgan), the *EDF* measure does not assume that all issuers are homogeneous within the same rating class. Fourth, *EDFs* represent a good approximation of default risk perceived by equity investors over a one-year horizon (Harada et al, 2010).[8] Even though defaults have occurred very suddenly over the recent financial crisis, *EDF* measures have predictive power in an ordinal sense: financial institutions that subsequently defaulted had high *EDF* measures relative to those of their peers (Munves et al, 2010).

5. METHODOLOGY

A very narrow definition of rating quality could focus on their ability to discriminate between banks that experience defaults and those that do not. But such an approach is problematic because of a small-sample problem. Outright corporate default is rare – especially for banks that typically benefit from (implicit) government guarantees of senior debtholders' claims. It is therefore more appropriate to consider bank ratings as a general assessment as to whether a bank is likely to experience financial distress in the future.[9] We therefore compare the credit ratings to *EDFs* measured *k* months forward in time. The latter approach moves the statistical problem away from predicting a very small default tail and broadens the analysis.

A second important issue concerns the interpretation of credit ratings. We prefer to interpret ratings as solely ordinal measures of default probabilities or financial distress. Moreover, long-term issuer ratings represent opinions on creditworthiness through the cycle, rather than short-term fluctuations in macroeconomic conditions (Moody's, 2006; Kiff, Kisser and Schumacher, 2012). Our own methodology accounts for this aspect by adopting a strictly ordinal interpretation of credit ratings by assigning a rank order to all credit ratings.

We rank-order the bank ratings of all three rating agencies in any given quarter. Banks rated AAA by an agency are given the lowest rank; AA the next lowest; etc. Rating agencies use 21-24 distinct rating buckets (see Table 1), resulting in some ties in our panel of 369 banks. In order to reduce the number of rating ties, we further subdivide the credit rating rank by the *rating outlook* as a second sorting criterion. Within a given credit rating category, banks with a positive outlook are given the lower rank; negative outlooks are given the higher rank.[10] A third and final sorting criterion is the *watchlist*. If more than one bank features the same credit rating and the same outlook, the banks 'on watch' receive a higher

[8] From a historical perspective, KMV analysed more than 2,000 companies that have defaulted or entered into bankruptcy over a 25-year period out of a comprehensive sample of listed companies from the Compustat database. The results show a sharp increase in the slope of the *EDF* between one and two years prior to default (Crouhy et al, 2000).

[9] Financial distress can be operationalized as the expected default frequency (*EDFs*) over a given time period.

[10] For example, consider four banks: banks A and B are rated AAA positive outlook; bank C is AAA stable; bank D is AAA negative; and bank E is AA+ positive. Here, we would assign rankings of 1.5 to bank A; 1.5 to bank B; 3 to bank C; 4 to bank D; and 5 to bank E.

(lower) credit rating rank if the outlook is negative (positive).[11] Specifically, outlooks indicate the credit rating agency's opinion regarding the likely direction of an issuer's rating over the medium-term; *watchlist* indicates that a rating is under review for possible change in the short-term.

For each rating, we define a measure of rating error called the Ordinal Rating Quality Shortfall (*ORQS*). *ORQS* is the absolute difference between the rank of a bank's *i* credit rating by rating agency *a* among all bank ratings at time t and the corresponding rank of that bank's expected default frequency (*EDF*)[12] at time *t+k*, normalised by sample size. Formally, we define:

$$ORQS\ (a,i,t,k) = \frac{|EDF\ rank(i,t+k) - Credit\ Rating\ rank(a,i,t)|}{N-1}.$$

ORQS is bounded between 0 and 1, where 0 represents a perfect rating and 1 the maximum shortfall or error.[13] The metric allows for simple interpretation of the rating error. If a particular *ORQS* is for example 0.2 and the sample of all bank ratings at time *t* comprises 300 observations, this implies that the Credit Rating (CR*)* rank differs from the *EDF* rank by 60 observations. In other words, there are 60 bank-ratings for which the CR rank was lower (higher) and the later *EDF* rank higher (lower). We interpret positive error as rating optimism, whereas negative error implies rating pessimism.

Figure 3 provides two scatter plots where the *EDF* rank (scaled by 1/N) on the y-axis is plotted against the credit rating rank (also scaled by 1/N) on the x-axis. The scatter plot focuses on the case where *k* = 24 months. The left-hand graph depicts observations where the *EDF* is measured outside the financial crisis and the right-hand graph shows ratings for which the *EDF* (24 months later) falls within the financial crisis. The red and blue lines represent a kernel estimation of the mean and median of the scaled *EDF* rank, respectively. Full information in credit ratings would imply that the observations cluster along the 45 degree line. In this case the ranking of credit ratings would perfectly correspond to the ranking of *EDFs* 24 months later. The scatter plots show instead a more uniform dispersion of the observations over the entire quadrant, indicating low correlation between the credit rating and *EDF* ranks. For the non-crisis period (depicted in the left graph), the mean and median of the scaled *EDF* rank are approximately 0.5 for all of the 75% best rated banks (AAA to A-). Only for the crisis period (depicted in the right-hand graph) do we observe a small positive relationship between rating rank of the 75% best rated banks and the subsequent *EDF* rank.

[11] Outlook and watchlist are used by credit rating agencies as 'auxiliary signals about credit risk'. For more details see 'Moody's Rating Symbols & Definitions', Moody's Investors Services, June 2009.

[12] Similarly to the ranking procedure used for ratings, we also implement a subordinate second sort criterion for the purposes of calculating the final EDF rank. Specifically, if more than one bank has the same EDF, we implement a second sort criterion on the estimated distance-to-default. See section 6 for further explanation of Moody's KMV methodology.

[13] For a set of axioms similar to those of Kemeny and Snell (1962), the ORQS defines a distance metric for a pair of rankings. Compared to Cook, Kress and Seiford (1986), our distance measure does not consider partial rankings (pairs without ranking information) and normalize the minimum distance (Axiom 7) to 1/N instead of 1.

[Insert Figure 3 here]

Table 3 reports the Spearman (rank) correlation between both variables at different horizons for the *EDF* measurement (k={0, 12, 24, 36} months) for the full sample (Panel A), the pre-crisis (Panel B) and crisis (Panel C) periods. The Spearman correlation coefficient in the full sample moderately decreases from 0.283 to 0.176 as the horizon increases from k=0 to k=36 months. At 0.352, the Spearman correlation coefficient at k=24 is much larger in the bottom third of credit ratings than in the two other sample tiers. By contrast, the *top tier* and *middle tier* ratings provide no evidence for a statistically positive correlation between *EDF* and *CR* rank. As these two upper tiers correspond to *investment grade rating*, these tiers contain no information regarding future expected default frequencies. Such evidence is difficult to reconcile with current bank regulation, which stipulates large differences in risk weights between a 20% weight for grades AAA to AA- and a much larger 50% risk weight for credit risk in the A+ to A- range.

For *EDFs* calculated during the financial crisis this is visibly different as the positive correlation between *EDF* rank and credit rating rank extends to banks with a top tier rating (Figure 3, right-hand graph). The overall Spearman correlation for the crisis period rises to 0.321 for k=24 compared to only 0.178 outside the crisis. Credit ratings are therefore considerably more informative for banks' relative creditworthiness within a financial crisis than outside.

[Insert Table 3 here]

An important part of our analysis consists of explaining the determinants of rating errors, where *ORQS* becomes the dependent variable in a linear regression analysis. Its distribution is strongly positively skewed as shown in Figure 2. We therefore apply a Box-Cox transformation to *ORQS* and thereby create a rank-preserving new variable named *Transformed Ordinal Rating Quality Shortfall* (*TORQS*), which more suitable for regression analysis. The Box-Cox parameter of -0.224 brings the skewness exactly to zero.[14] The new *TORQS* features reduced kurtosis of 1.95 (relative to 2.55 for *ORQS*) and serves as the dependent variable for rating accuracy in all subsequent analysis. Its panel structure also allows us to explore the determinants of rating quality in the cross-section (across banks and rating agencies) in a linear framework:

$$ORQS\ (a,i,t,k) = X(a,i,t) \times \beta(k) + \epsilon,$$

where the explanatory variables

[14] A Box-Cox parameter of -1 corresponds to the log transformation. The latter scales down large ratings errors more strongly and is more discriminating for small rating errors.

$$X(a,i,t) = [BC(i,t), D(a), RV(a,i,t), D(t), D(i)]$$

include bank characteristics $BC(i,t)$, rating agency dummies or country dummies $D(a)$, bank-rating agency relationship variables $RV(a,i,t)$, time/crisis fixed effects $D(t)$ and country fixed effects $D(i)$. We can thus test a variety of economically interesting hypotheses regarding the determinants of ratings quality. These are elaborated in the next section.

The *ORQS* (and its transformation *TORQS*) treat positive and negative errors symmetrically. But some of our hypotheses relate to rating bias rather than error. The distinction between error and bias is elaborated in Calomiris (2009). Rating *error* arises from 'innocent' but 'flawed measures of underlying risk' (Calomiris, 2009), and is a function of the degree of complexity of the rated entity and the extent of the rating agency's investment in credit analysis. In contrast, rating *bias* generally refers to deliberate systematic over-rating, which might occur due to conflicts of interest arising from the issuer-pays model (Partnoy, 2006). As a proxy for rating bias, we capture a positive directional effect in the rating error by defining the *Directional Ordinal Rating Quality Shortfall* (*DORQS*) as:

$$DORQS\ (a,i,t,k) = \frac{EDF\ rank(i,t+k) - Credit\ Rating\ rank(a,i,t)}{N-1}.$$

The *DORQS* measure is sufficiently close to a normal distribution, enabling us to apply regression analysis directly without any further variable transformation. We also highlight that *DORQS* has by construction a zero cross-sectional mean and therefore does not detect any overall rating bias for all banks. Our analysis of bank rating bias is confined to rating distortion *within* the bank sample.

[Insert Figure 2 here]

6. HYPOTHESES ABOUT CREDIT RATING QUALITY

In this section we formulate and discuss five hypotheses about the determinants of rating quality. As suggested by Figure 3, the generally low information content of relative ratings for future relative expected default frequency (*EDF*) does not preclude that the rating error has a systematic structure, which should be explored separately.

H1: Ratings Quality during the Crisis and after Credit Booms

Ordinal ratings quality shortfall depends on the state of the financial system and the credit cycle. Bank credit ratings are more informative (in an ordinal sense) about bank distress when distress occurs during periods of financial crisis.

The Lehman bankruptcy and other prominent ratings failures have conveyed the misleading impression that bank ratings become more inaccurate during a financial crisis. However, this

is at odds with the summary statistics presented in the previous section. The Spearman rank correlation between *EDF* rank and the CR rank dramatically increases as the bank system entered the crisis. This suggests that *ORQS* has a cyclical component, particularly for the majority of banks rated A- or better. Expansionary credit cycles may also affect rating accuracy as they foreshadow later bank distress (Bar-Isaac and Shapiro, 2011a).

For the empirical part, we define a global financial crisis dummy. The dummy takes on the value of 1 when the mean *EDF* is greater than 2%. In our sample of banks, this occurs from 2008Q3 to 2010Q4, and again in 2011Q2 and 2011Q3. Importantly, our crisis dummy is contemporaneous with the observation on the *EDF* variable. For example, when *ORQS* is defined using a two-year gap between the credit rating and the *EDF*, the crisis dummy will equal 1 when the credit rating is measured two years prior to crisis (i.e., 2006Q3–2008Q4 and 2009Q2–2009Q3). As a measure of the credit cycle, we use private credit growth over the previous three years at country level. This second measure adds cross-sectional variation (across the 16 countries in our panel) unlike the crisis dummy, which features only time variation.

H2: Rating Quality across Rating Agencies and Countries

Ordinal ratings quality shortfall varies across rating agencies.

First, rating agencies may differ in their rating methodology and in the quality of their credit analysts. Differences between the ability of equity analysts have been documented by Bradshaw (2011) and Fang and Yasuda (2009). Second, rating agencies may also differ in their access to non-public bank information. Unfortunately, the incidence of unsolicited bank ratings is low, precluding exploration of the latter aspect in more detail. Third, agency and incentive problems may also differ across rating agencies and manifest themselves in certain rating biases. To explore cross-agency differences in rating accuracy and rating bias, we define dummy variables called *Moody's* and *S&P* which capture the average agency-specific rating shortfall relative to Fitch ratings.

H3: Rating Quality and Conflicts of Interest

Rating agencies provide better bank ratings to banks that are (i) larger and (ii) generate more securitisation business.

Large banks typically have many rated subsidiary entities, so that a large bank is in a much stronger client position. Bank size may therefore augment conflicts of interest for the rating agency. Moreover, asset securitisation provides a substantial income stream to both banks (as the asset originators) and to the rating agencies and may generate additional conflicts of interest. To explore these in more detail, we use the Dealogic database to identify 1,189 unique issuers of asset-backed securities with a total face value of US$6 trillion over 1990–2012. The securities comprise residential mortgage-backed securities RMBS (which make up 28% of the sample), other asset-backed securities (3.8%), collateralised mortgage-backed securities (3.4%), collateralised loan obligations (5%), other collateralised debt obligations

(5.9%) and home equity loans (53.8%). Importantly, the dataset records which of the three major rating agencies provided ratings when the security was issued. The supply of asset-backed securities concentrated among 200 issuers that account for 90% of the total market. We combine these 200 largest issuers with the 369 banks in our sample, obtaining 53 successful matches (which together account for 35% of the total market for asset-backed securities). The remaining 147 issuers are mostly non-bank issuers, such as Fannie Mae and Freddie Mac. Any bank outside the list of 200 top issuers is assumed not to issue any asset-backed securities. Even for the 53 most active banks, securitisation business is highly irregular over time, so that aggregation over the entire time period provides the best measurement of the overall securitisation business shared between a bank and a rating agency. As our proxy for conflicts of interest in bank ratings, we define a bank's *agency-specific securitisation business* (*ASSB*) as

$$ASSB(a,i) = log[1 + assets\ securitized\ by\ bank\ i\ and\ rated\ by\ ageny\ a].$$

The log transformation is appropriate because a large share of the securitisation business is concentrated among a relatively small number of banks, with the 10 largest banks accounting for roughly 65.7% of the asset origination of the 53 banks in our sample.

H4: Ratings Quality and Bank Characteristics

Ordinal ratings quality shortfall depends on key bank characteristics including size, capital structure, asset structure and funding structures.

The regulatory debate makes explicit reference to most of these bank characteristics. Large banks might be subject to more stringent regulation because of their systemic importance, while other regulatory proposals want to separate banks with trading income from those doing loan business only (Dodd-Frank Act, 2010; Independent Commission on Banking, 2011). Yet little is known about how these bank characteristics relate to the quality of bank ratings. For example, if large banks exhibit greater rating errors or benefit more from rating inflation, this provides an additional argument in favour of size-contingent bank regulation for bank capital (BCBS, 2011). Moreover, such findings would suggest that regulation of large (and systematically important) banks should be less reliant on rating agencies' assessments of creditworthiness.

We measure bank size by *Log assets* (natural log of the book value of assets). Large banks may generally be more complex and thus more difficult to rate, increasing both positive and negative rating errors. On the other hand, size often comes with revenue diversification and hence more stability, which suggests an offsetting effect on rating accuracy. However, unlike the conflict of interest mechanism, rating complexity and asset diversification should change the error variance, without creating a bias.

Capital structure is captured by *Leverage* defined as total assets divided by book equity multiplied by 100. Bank leverage has often been deemed excessive and conducive to more risk (Berger and Bouwman, 2011), hence our interest in whether it also contributes to

larger rating errors. Asset structure is proxied by two variables, namely the *Loan share* as total loans over total assets and *Trading share* defined as net profit on trading and derivatives also divided by total assets. The traditional bank loan has been deemed a more stable business model compared to financial market-making activity. Here we explore the impact of both business models on the accuracy of bank credit ratings. Funding structure is represented as *Short-term funding share* measured as deposit and short-term funding divided by total assets.

H5: Rating Quality and Competition

The Ordinal Rating Quality Shortfall is influenced by the level of competition in the market for bank ratings.

Rating competition may provide the rating agencies with incentives to improve their rating process and methodology in order to acquire a reputation for the most accurate ratings. For example, reputational effects appear to matter for equity analysts (Fang and Yasuda, 2009). But competition may also compromise the independence of the rating process if corporations can 'shop' for the best available rating (Becker and Milbourn, 2010). To explore the role of competition for rating quality, we measure market concentration based on the market share of the three largest rating agencies using the traditional Herfindahl-Hirschmann index (*HH index*). Figure 4 shows a decrease in market concentration from 1990 to 2011, which is mostly explained by an increased market share by the rating agency Fitch. Unfortunately, we do not have data on any of the smaller rating agencies; their market share is ignored, which implies some measurement error for the *HH index*. A better approach might therefore capture the market power of a rating agency at the bank level. Roughly 73% of all banks have multiple ratings, which should reduce the bank's dependency on a single rating. We create a *Multiple rating dummy*, which takes on the value of 1 (and 0 otherwise) whenever more than one rating agency has issued a bank rating. This measure captures cross-sectional variation in rating competition at the bank level, whereas the *HH index* features only time series variation.

7. EMPIRICAL ANALYSIS

7.1. Rating Quality during the Crisis and after Credit Booms

The distributional evidence in Figure 3, and the corresponding Spearman correlations in Panel B of Table 3, suggests that in normal times (when the *EDF* is observed outside of financial crisis) bank credit ratings contain information about future default risk only for speculative investment grades. For all investment grade ratings (corresponding approximately to a rating rank below the 66% percentile), the mean and median *EDF* rank do not vary substantially with the credit rating rank. The Spearman correlation between both variables is even slightly negative. This pattern changes if we restrict the sample to *EDFs* observed

during the financial crisis. Here, we find a positive Spearman correlation over the entire rating scale, with an overall rank correlation of 0.321 at the two-year horizon (k =24).

Columns (1)–(3) in Table 4 confirm this finding in panel regressions with *TORQS* as the dependent variable. We use a *Crisis* dummy to mark all ratings for which the *EDF* is reported at a moment of high global bank distress, namely for the quarters 2008Q4–2010Q4 and again 2011Q2–2011Q3. The coefficient estimate of -0.031 in column (1) implies that financial crisis reduced *TORQS* by 7.4% relative to its unconditional standard deviation of 0.4205. Column (2) adds *Credit growth* over a three-year period prior to the rating as additional controls for credit booms. At the end of a credit boom and in a financial crisis, the ordinal rating error decreases significantly. Hence ratings quality is counter-cyclical: bank ratings become more accurate during financial crises. This finding contradicts the frequently voiced criticism that bank ratings become less reliable when a financial crisis occurs.

The specification in Table 4, columns (1) and (2), uses country fixed effects, while column (3) reports coefficient estimates using bank fixed effects. Coefficients show little variation across these specifications. Given that ratings are measured at quarterly frequency, we expected considerable serial correlation in the error structure. The estimated serial correlation is indeed high at around 0.77. The reported standard errors are adjusted for this serial correlation.

We repeat these regressions with *DORQS* as the dependent variable in Table 4, columns (4)–(6). By construction, the *DORQS* as the difference between two rankings has a zero mean so that any time fixed effect or crisis dummy should also be zero (except for mission observations in the regression). Surprisingly, stronger past credit growth does not generate any statistically significant positive rating error bias for the *DORQS* beyond what is captured in country or bank fixed effects.

[Insert Table 4 here]

7.2. Rating Quality across Countries and Rating Agencies

Rating agencies may differ in their rating technology and the degree to which they have conflicts of interest with respect to revenue sources. As our data cover the three largest rating agencies, it is interesting to explore agency-specific differences in the accuracy of ratings. Here we also report and control for country fixed effects, as cross-country differences in accounting standards and regulatory supervision may also co-determine the rating precision.

Table 5 reports panel regressions with agency, country and time fixed effects; columns (1)–(3) focus on the non-directional rating error measured by *TORQS*, whereas columns (4)–(6) feature the directional error or rating bias *DORQS* as the dependent variable. The baseline specification in columns (1) and (4) controls for bank size measured by *Log assets* and reports all country fixed effects. The regression specification allows for serial correlation of the regression error and reports the adjusted standard errors. Bank size correlates strongly with both the non-directional and directional measure of rating error – a robust data feature discussed in more detail in the following section of this paper. The

country fixed effect in column (1) shows that banks headquartered in Austria, Ireland and the United Kingdom feature significantly higher ratings errors. The country coefficient for Ireland of 0.244 implies that Irish banks have on average 58% higher *TORQS* relative to its unconditional standard deviation of 0.4205. We also highlight that almost all European countries show strong positive bias for the rating errors of their respective banks relative to the ratings of US based banks. This finding appears to contradict the idea that the US rating agencies Moody's and S&P might have treated non-US banks unfavourably. Columns (2) and (5) introduce additional agency dummies for ratings issued by Moody's and S&P. Controlling for bank size, and time and country fixed effects, ratings by S&P are significantly more negative than those of Moody's and Fitch.[15] This implies that rating errors feature a systematic component that is related to an agency's overall rating policy.

[Insert Table 5 here]

7.3. Rating Quality and Agency Conflicts

A key regulatory concern relates to possible upward bias in ratings arising from conflicts of interest between a rating agency and a bank. A bank's power in relation to a rating agency is related to its size. Larger banks are more likely to have multiple and more comprehensive business relations with rating agencies. Often national bank subsidiaries might require additional ratings beyond the rating for the holding company. Another important revenue source for rating agencies concerns ratings of asset-backed securities. The larger the bilateral business volume measured by a bank's *agency-specific securitisation business* (ASSB), the more the quality of the bank rating might be compromised.

Regressions reported in Table 5, columns (3) and (5) confirm the important role of bank size as a determinant of rating accuracy and bias. Bank size measured as *Log assets* strongly and positively correlates with both non-directional rating error (*TORQS*) and the rating bias (*DORQS*). The regression coefficient of 0.042 for log assets in Table 5, column (6), implies that a bank size increase by two standard deviations translates into an inflated credit rating rank (relative to the EDF rank) by 15 positions for every 100 banks in the sample.[16] This corresponds to an undeserved rating improvement from A- to A. Based on yields to maturity on banks' bonds and medium-term notes issued to the primary market over 2002–12, a rating improvement from A- to A equates to a considerable reduction in funding costs of 40 basis points on average.

We highlight that conflicts of interest with big bank clients may not be the only interpretation of the strong size dependence of the bank rating error. An alternative interpretation could relate this bias to the 'too big to fail' privilege of big banks. Our analysis

[15] In 2012, Moody's finally undertook a general review of bank ratings and downgraded its ratings for a large group of banks with a six-month delay on its rivals Fitch and S&P (Economist, 11 June, 2012).

[16] Two standard deviations in log assets are 3.58 (see Table 2) so that we obtain a predicted change of 0.15 (=3.58 ×0.042) for DORQS. In Table 1, the difference between the average rank for an A and an A- rating is 0.162.

is based on 'all-in' ratings, which means that the rating agencies try to account for the ability of different sovereigns to bail out banks. Cross-country differences in governments' ability to bail out banks are captured by country fixed effects and should not affect our results. However, the implicit government support for banks might protect creditors of big banks more than those of small banks – something that the rating agency might account for in its rating process. If the bank's equity price and expected default frequency insufficiently accounts for this 'too big to fail' distortion, then the positive correlation between the *DORQS* and bank size can be predicted as the outcome of the rating agency's foresight rather than any conflict of interest. In this interpretation, the rating process just reflects substantial competitive distortion, rather than creates it.

Clearer inference on agency conflicts in bank ratings can be obtained from the agency-specific securitisation business (*ASSB*) variable. In Table 5, column (6), we find that this measure is related to a statistically significant upward bias in the rating.[17] Figure 5 captures this rating bias effect in a scatter plot of the directional rating error *DORQS* against bank size for all sample banks in January 2007. Grey dots and red circles distinguish banks with and without substantial securitisation business, respectively. The vertical line between the small blue circle and the grey dot depicts the predicted positive marginal change in the directional error due to the bank's agency-specific securitisation business. Figure 5 illustrates that most of the banks engaged in asset securitisation are large. Rating favours related to asset securitisation therefore come in addition to the general rating bias in favour of large banks. Again, we can quantify the economic magnitude of the rating bias in column (6). An increase in *ASSB* by two standard deviations (or 19.01) is associated with a rating improvement of 10 ranks for every 100 banks in the sample, which amounts to an economically substantial rating favour resulting from bilateral cooperation in security issuance.

Because bank size is largely exogenous and predetermined in the short run, the significant coefficient has a reasonable motivation, according to which large banks enjoy higher ratings relative to their future expected default frequencies. This should generally translate into more favourable financing conditions for large banks, generate larger asset growth and accentuate the 'too big to fail' problem. Reverse causality is a bigger issue for the *ASSB* variable, because banks may reward rating agencies for good bank ratings with more rating business related to asset securitisations. But from a policy point of view, this is no less problematic than if favourable bank ratings follow the securitisation business.

[Insert Table 6 here]

7.4. Rating Quality and Bank Characteristics

Banks differ not only in size, but also in profitability, capital structure, asset structure or business model and funding structure. How do these bank characteristics relate to rating

[17] This effect is also robust after inclusion of bank fixed effects.

accuracy and rating bias? Are previous findings robust if we control for these bank characteristics?

In the following extended regression analysis, it might not always be appropriate to consider these variables as exogenous to the rating error. Reverse causality is particularly plausible from the *level* of ratings to some bank characteristics such as profitability or funding structure. For example, banks with low ratings may face higher financing costs, seek shorter maturities on the liability side of their balance sheet or experience lower profitability. However, the dependent variable in our analysis is not the rating level, but rather *TORQS* or *DORQS*, which are less likely to have feedback effects on corporate decisions. We also note that a larger, but symmetric and transitory, rating error (like *TORQS*) should – to a first-order linear approximation – have no steady state effect on corporate decisions since its expected long-run impact is always zero. On the other hand, changes to the asset structure of a bank might involve considerable adjustment costs so that causal effect from the rating error (and particularly the *TORQS*) on bank asset choices are less plausible.

Table 6 extends the panel regressions in Table 5 by including additional bank characteristics. In columns (1)–(2) and (3)–(4), we use both country and time fixed effects. The reported standard errors are adjusted for the serial correlation at the bank level. Bank profitability (*RoA*) and capital structure (*Leverage*) show no evidence of any incremental correlation with either rating accuracy (*TORQS*) or rating bias (*DORQS*). By contrast, the asset structure is significantly related to rating accuracy, but not to the rating bias. The negative coefficient on the *Loan share* variable in columns (1)–(3) suggests that a traditional lending-based banking model is associated with higher rating accuracy. Surprisingly, a high *Trading share* also correlates (weakly) with higher rating accuracy. This could be explained by the strong countercyclical nature of trading revenues. The average correlation of bank trading revenue with the VIX index of equity market volatility is relatively important at 0.18. Market-making and proprietary trading appears to deliver revenue stabilizing income in times of financial crisis when market volatility is high. Our finding of a significant negative coefficient on the *Trading share* variable in columns (1)–(3) implies that credit rating agencies systematically underestimate the countercyclical effect of trading activity on bank creditworthiness. The *Short-term funding share* variable correlates with a smaller ratings bias in columns (4)–(6). This variable not only measures the degree of maturity transformation, but also the size of the deposit base of a bank. This means that banks with a large depositor base tend to be systematically underrated relative to their future expected default frequencies.

Inclusion of these various bank characteristics does not change the coefficient estimates from Table 5 for the bank size variable (*Log assets*) and the bank's agency-specific securitisation business (*ASSB*). Even conditional on bank characteristics, bank size and bank securitisation activity with a rating agency remain highly correlated with the rating bias.

7.5. Ratings Quality and Competition

Finally, we explore the role of competition in the market for bank ratings. After 2000, competition in the rating market increased as Fitch became a more important competitor

through acquisitions of smaller rating agencies and a general expansion of its rating business (Becker and Milbourn, 2010). Based on the number of bank ratings generated by the three major rating agencies, we construct a Herfindahl-Hirschmann index (*HH index*) of industry concentration, which shows decreasing industry concentration after 2000. A separate *Multiple rating dummy* captures cross-sectional variation in the market power of rating agencies vis-à-vis individual banks.

In Table 6, columns (3) and (6), we report the panel regressions with both measures of competition for TORQS and DORQS, respectively. Since the *HH index* represents a pure time series, neither specification features time fixed effects. The *Multiple rating dummy* shows a strong negative correlation with *DORQS*. The 73% of banks ratings classified as multiple ratings were on average less favourable (by on average three ranks for every 100 banks) than those for which only one rating was issued. This evidence contradicts the hypothesis that rating competition implied rating inflation. The *HH index* itself is insignificant in each specification, which could reflect measurement problems with respect to the correct market share of all relevant rating agencies.

8. ROBUSTNESS

The analysis to this point was based on the two year lag (*k=8* quarters) between the credit rating and the *EDF* measurement. Next we show that our results are robust to other forecast horizons, for example one year or three years. Table 7 repeats the full specification in Table 6, columns (3) and (6), for different measurement lags of *k=0, 4, 12* quarters. Interestingly, the coefficients for the rating bias related to bank size and securitisation business with the rating agency remain more or less stable for different lags. This implies that our main findings – concerning the rating privilege of large banks as well as the rating bias related to securitisation business – are robust to the lag between observations on the rating and *EDF*.

Somewhat less robust are the coefficients on variables characterizing the bank's asset structure. For example, at the three-year horizon (*k=12*), the negative correlation between *Loan share* and *TORQS* drops to a 10% significance level (column (3)), whereas the bank's *Trading share* is now negatively related to *DORQS* at the 1% level. Banks with a high trading income (relative to their assets) on average deserved a better rating at this three-year horizon relative to what was in fact assigned. The *Multiple ratings dummy* is still associated with lower bank ratings error *DORQS* at the one-year horizon, but becomes statistically insignificant at the three-year forecast horizon. The latter affect might be caused by the reduced sample size at the longer horizon.

Finally, we explore whether the rating bias in favour of large banks may simply reflect larger implicit government guarantees for the debt of 'too big to fail' banks. Ratings used so far refer to the creditworthiness of banks' senior unsecured debt. These are so-called 'all-in' ratings because they incorporate the likelihood that a government bails out creditors. As discussed in section 7.3, the rating bias in favour of larger banks might instead reflect agencies' rational assessment of bank size dependent government support.

We test the robustness of our results to the effect of government support by exploiting a different type of credit rating. Fitch Ratings publishes 'bank individual ratings', which assess banks' creditworthiness as independent stand-alone entities, absent reliance on government guarantees. The rank difference between the 'all-in' and 'stand-alone' ratings describes the credit improvement due to implicit government support. We define the rank improvement from the 'stand-alone' to the 'all-in' credit rank (by the same bank) as *Rank difference*. Its positive correlation with bank size, measured by *Log assets*, of 0.51 could potentially account for the rating bias in favour for large banks.

Table 8 explores the explanatory power of *Rank difference* for both *TORQS* and *DORQS*. 'Stand-alone' ratings are available to us only from Fitch, so that we can only use a reduced sample of 7,980 bank-rating observations, 31% of the full sample. Specifications (1) and (3) repeat the analysis from Table 6, column (3) and (6) for all Fitch rated banks. Specifications (2) and (4) use these same specifications and Fitch sample plus the *Rank difference* control. Even in the reduced sample of Fitch ratings in specification (1), the *Log assets* variable shows a significant coefficient of 0.02 for rating error *TORQS*. The result is stronger in column (3), with an estimated coefficient of 0.06 for rating bias *DORQS*.

Including the *Rank difference* in specifications (2) and (4) slightly reduces the regression coefficient for *Log assets* to 0.020 and 0.056 respectively. But statistical and economic significance, particularly in column (4), remains high. We note that the coefficient for *Rank difference* has the expected positive sign and is also statistically significant. We conclude for the Fitch ratings that only an economically small part of the substantial rating bias in favour of large banks can be attributed to implicit government guarantees.

9. CONCLUSIONS AND POLICY IMPLICATIONS

The ongoing financial and banking crises have shifted rating agencies and the quality of their opinions into the centre of the policy arena. The issue of rating quality is closely connected to a larger debate about bank regulation, which is often founded on rating-contingent bank capital requirements. To inform this debate, the current paper contributes with a number of stylized empirical facts about the quality of bank ratings.

We ground our analysis on the premise that it is inherently difficult to predict the timing and intensity of a systemic banking crisis. This insight informs our strictly ordinal definition of rating quality. In our analysis, it is not the absolute (cardinal) level of default risk that matters, but rather the rank-order of default risk among all banks. We then apply this ordinal approach to a large database on bank ratings issued by the three major rating agencies over the period 1990 to 2011. The corresponding measure of bank distress is the expected default frequency (*EDF*) measured by the widely used Merton model of corporate default. We draw our *EDF* measures directly from Moody's in order to avoid any parameter choices that might bias the rating quality metric against a finding of high rating quality.

Our first insight concerns overall ratings quality. We show that bank ratings *in the upper investment grade range* bear no substantial ordinal relationship to expected default

probabilities two years later. The Spearman rank correlation between the credit rating rank and the *EDF* rank is even slightly negative when *EDF*s are measured outside crises. This finding runs contrary to risk-weights applied in the standardized approach to credit risk under the first pillar of the Basel II accord. Under these Basel recommendations, exposures to financial institutions are assigned a 20% risk-weight if the external credit rating is between AAA and AA-; a 50% risk-weight if the external rating is between A+ and A-; and a 100% risk-weight for the lowest investment grade ratings from BBB+ to BBB-. These risk-weights are used by national bank regulators to determine whether banks meet minimum regulatory capital requirements. But such large step-changes in risk weights cannot be reconciled with our evidence that the AAA to AA- bucket is statistically indistinguishable from the A+ to A- bucket in terms of predicting future *EDF* rankings. This discrepancy is likely to generate substantial distortions. To the extent that minimum regulatory capital requirements bind, we expect banks to hold more exposure to other banks rated AAA to AA- compared with banks rated A+ to A-. These Basel II risk-weights thus distort interbank markets and entrench the market position of banks rated AA- and above. We also highlight the countercyclical nature of rating quality. The information content of ratings increases during a financial crisis. If the expected default risk is measured during a crisis period, even bank ratings in the investment grade range become somewhat informative. The Spearman correlation between the credit rating rank and *EDF* rank is 14% for the top third of rating observations. In an ordinal (rather than cardinal) sense, credit ratings become more meaningful at the onset of a financial crisis.

Second, our analysis reveals systematic relationships between the direction (bias) of the rating error and bank size: large banks obtain systematically more favourable credit ratings relative to their expected default risk measured two years later. This bias is most likely related to the agency conflicts between the rating agency and the bank which increase in the size and economic power of a bank. At the extreme, large banks with economic power might become ‘too big to downgrade’ for the rating agency. In small part, the distortion in large banks’ ratings can be attributed to more substantial government guarantees for large banks. But results presented in Table 8 indicate that, at least for the subsample of Fitch-rated banks, the finding of rating error and rating bias in favour of large banks is robust to the inclusion of government guarantees. Overall, the consequences for interbank competition are deplorable, because the rating bias distorts the financing costs of large banks and reinforces the creation of ‘too big to fail banks’ devoid of economic rationale.

Third, new information from the Thomson Reuters Dealogic database is used to map the bilateral business relations in securitisation issuance between banks and the three major rating agencies over the period 1990–2012. We define a bank’s *agency-specific securitisation business* (*ASSB*) and show that it has significant explanatory power for the rating bias even after controlling for many bank characteristics. In other words: the more a bank used a particular rating agency for rating its asset-backed securities at issuance, the more this agency rewarded the bank with a better bank credit rating. We consider that this represents clear evidence that conflicts of interest in the securitisation business compromised the quality of bank credit ratings.

In light of the shortcomings in the current rating process, public policy should encourage alternative sources of credit rating information. Recent work by Bloechlinger, Leippold and Maire (2012) shows that one can produce corporate credit rating measures at par or superior to those of the credit rating agencies at almost no cost, using public information only. The latter suggests that the three largest rating agencies owe their predominance in the market for corporate ratings more to regulatory privilege than information advantage. With the Dodd-Frank Act in the US, which aims to reduce regulatory reliance on rating agencies, some segments of the rating market might become low-cost commodities in the future, dominated by not-for-profit organizations.

In order to reduce the cost of processing bank accounting information, banks' public reporting requirements should be vastly enhanced to facilitate cheaper and better credit analysis. Those reporting requirements are still heterogeneous across countries. A number of countries do not require quarterly financial statements for non-listed banks and provide significant room for manoeuvre to allocate certain items to the trading or banking book (Huizinga and Laeven, 2010). In most countries, bank regulators protect their privileged data access, and do not share crucial bank data publicly (or even with other bank regulators) in a narrow pursuit of their own agency power and to shield themselves from accountability. Future bank regulation therefore needs to create an entirely new information environment for external credit analysis. Better public information and more bank reporting is the best strategy to reduce the power and exorbitant influence of rating agencies in the current system.

REFERENCES

Altunbas, Y., S. Manganelli and D. Marques-Ibanez (2011). 'Bank risk during the financial crisis: do business models matter?', European Central Bank Working Paper 1394.

Bar-Isaac, H. and J. Shapiro (2011a). 'Ratings quality over the business cycle', working paper.

Bar-Isaac, H. and J. Shapiro (2011b). 'Credit ratings accuracy and analyst incentives', *The American Economic Review: Papers and Proceedings*, 101(3), 120-124.

Basel Committee on Banking Supervision (BCBS) (2010). 'Basel III: A global regulatory framework for more resilient banks and banking systems', Bank for International Settlements.

Basel Committee on Banking Supervision (BCBS) (2011). 'Global systemically important banks: assessment methodology and the additional loss absorbency requirement', Bank for International Settlements.

Becker, B., and T. Milbourn (2010). 'How did increased competition affect credit ratings?', Harvard Business School Finance Working Paper 09-051.

Beltratti, A. and R.M. Stultz (2012). 'Why did some banks perform better during the credit crisis? A cross-country study of the impact of governance and regulation', *Journal of Financial Economics*, forthcoming.

Benmelech, E. and J. Dlugosz (2009). 'The credit rating crisis', NBER Macro Annual, 161-207.

Berger, A. and C. Bouwman (2011). 'How does capital affect bank performance during financial crises?', Wharton Financial Institutions Center Working Paper 11-22.

Bharath, S.T. and T. Shumway (2008). 'Forecasting default with the Merton distance to default model', *The Review of Financial Studies*, 21(3), 1339-1369.

Bloechlinger, A., M. Leippold and M. Basile (2012). 'Are ratings the worst form of credit assessment apart from all the others?', Swiss Finance Institute Research Paper 12-09.

Bolton, P., X. Freixas and J. Shapiro (2012). 'The credit ratings game', *Journal of Finance*, 67(1), 85-112.

Bongaerts, D., K.J. Martijn Cremers and W. Goetzmann (2011). 'Tiebreaker: certification and multiple credit ratings', National Bureau of Economic Research Working Paper 15331.

Boot, A. and A.V. Thakor (2010). 'The accelerating integration of banks and markets and its implications for regulation', in A. Berger, P. Molyneux and J. Wilson (eds.), The Oxford Handbook of Banking, 58-90.

Bradshaw, M.T. (2011). 'Analysts' forecasts: what do we know after decades of work?', mimeo.

Calomiris, C.W. (2009). 'A recipe for ratings reform', The Economists' Voice Berkeley Electronic Press, 6(11), 0-5.

Camanho, N., P. Deb and Z. Liu (2010). 'Credit rating and competition', working paper, mimeo.

Campello, M., J. Graham and C. Harvey (2010). 'The real effects of financial constraints: Evidence from a financial crisis', *Journal of Financial Economics*, 470-487.

Cantor, R. and F. Packer (1995). 'The credit rating industry', *Journal of Fixed Income*, 5(3), 10-34.

Cole, R.A. and L.J. White (2012). 'Déjà vu all over again: The causes of US commercial bank failures this time around', *Journal of Financial Services Research*, Forthcoming.

Cook, W.D., M. Kress, and L.M. Seiford (1986). 'Information and preferences in partial orders: A bimatrix representation', *Psychometrika*, 51(2), 197-207.

Covitz, D., and P. Harrison (2003). 'Testing conflicts of interest at bond rating agencies with market anticipation: evidence that reputation incentives dominate,' Board of Governors of the Federal Reserve System Finance and Economics Discussion Series 2003-68.

Crouhy, M., D. Galai and M. Robert (2000). 'A comparative analysis of current credit risk models', *Journal of Banking and Finance*, 241, 59-117.

Demirgüc-Kunt, A. and Huizinga, H.P. (2011). 'Do we need big banks? Evidence on performance, strategy and market discipline', Discussion Paper 2011-021, Tilburg University, Center for Economic Research.

Dodd-Frank Wall Street Reform and Consumer Protection Act (2010).

Dwyer, D. and S. Qu (2007). 'EDF™ 8.0 Model Enhancements', Moody's KMV.

Efing, M. (2012). 'Bank capital regulation with an opportunistic rating agency', working paper, mimeo.

Fang, L. and A. Yasuda (2009). 'The effectiveness of reputation as a disciplinary mechanism in sell-side research' , *Review of Financial Studies*, 22(9), 3735-3777.

Fitch (2012). '*Fitch Ratings – Definitions of Ratings and Other Forms of Opinion'* - April 2012.
Financial Stability Board (2010). 'Principles for reducing reliance on CRA ratings'.
Fulghieri, P., G. Strobl and H. Xia (2010). 'The economics of solicited and unsolicited credit ratings', mimeo.
Garlappi L., R. Uppal, and T. Wang (2007). 'Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach', *Review of Financial Studies*, 20, (1), 41-81.
Griffin, J. and D. Tang (2011). 'Did credit rating agencies make unbiased assumptions on CDOs?', *American Economic Review: Papers & Proceedings*, 101(3), 125-130.
He, J., J. Qian, P.E. Strahan (2011). 'Credit ratings and the evolution of the mortgage-backed securities market', *American Economic Review: Papers & Proceedings*, 101(3), 131-135.
Hilscher, J. and M. Wilson (2011). 'Credit ratings and credit risk', working paper, mimeo.
Hörner, J. (2002), 'Reputation and competition', *The American Economic Review*, 92(3), 644-663.
Huizinga, H. and L. Laeven (2009). 'Accounting discretion of banks during a financial crisis' IMF Working Papers 09/207.
Independent Commission on Banking (2011). *Final Report Recommendations.*
International Monetary Fund (2009). *Global Financial Stability Review*, April.
Kemeny, J.G., and L.J. Snell (1962). Preference ranking: An axiomatic apprach. In: J.G. Kemeny, and L.J. Snell (Eds.), *Mathematical models in the social sciences* (pp. 9-23). New York: Ginn.
Kiff, J., M. Kisser, and L. Schumacher (2012). 'An inspection of the through-the-cycle rating methodology,' *IMF Working Paper*, *forthcoming*.
Longstaff, F. and E. Schwartz (1995). 'Valuing risky debt: A new approach', *Journal of Finance* 50, 789-820.
Manso, G. (2011). 'Feedback effects of credit ratings', working paper.Mariathasan, M. and G. Merrouche, 'The manipulation of Basel risk-weights', Oxford University working paper 621.
Mathis, J., J. McAndrews and J.C. Rochet (2009). 'Rating the raters: Are reputation concerns powerful enough to discipline rating agencies?' *Journal of Monetary Economics*, 56, 657-74.
Merton, R. (1974). 'On the pricing of corporate debt: the risk structure of interest rates', *The Journal of Finance*, 29,(2), 449-470.
Moody's Investors Services (1999). 'Rating methodology: The evolving meanings of Moody's bond ratings'.
Moody's (2006). Moody's credit rating prediction model, Special Note.
Morgan, D. (2002). 'Rating banks: risk and uncertainty in an opaque industry', *The American Economic Review*, 92, 874-888.
Munves, D., D. Hamilton and A. Smith (2010). 'Banks and their EDF measures now and through the credit crisis: Too high, too low, or just about right?', Moody's Analytics, June.
Oliver Wyman (2011). 'The state of European bank funding'.
Opp, C., M. Marcus and M. Harris (2012). 'Rating agencies in the face of regulation', mimeo.
Packer, F. and N. Tarashev (2011). 'Rating methodologies for banks', BIS Quarterly Review.
Pagano, M. and P. Volpin (2010). 'Credit ratings failures and policy options', *Economic Policy*, 62, 401-431.
Partnoy, F. (2002). 'The paradox of credit ratings', in *Ratings, Rating Agencies, and the Global Financial System*, ed. R. Levich, C. Reinhart, and G. Majnoni, Boston: Kluwer 65-84.
Partnoy, F. (2006). 'How and why credit rating agencies are not like other gatekeepers', in *Financial Gatekeepers: Can They Protect Investors?*, ed. Y. Fuchita and R. Litan, Brookings Institution Press and Nomura Institute of Capital Markets Research, 59-102.
Reinhart, C. and K. Rogoff (2009). 'The aftermath of financial crises', *American Economic Review*, 99(2), 466-472.
Skreta, V. and L. Veldkamp (2009). 'Ratings shopping and asset complexity: A theory of ratings inflation', *Journal of Monetary Economics*, 56, 678-695.
Standard and Poor's (2009). 'Understanding Standard & Poor's rating definitions', *Standard & Poor's Ratings Direct*, June 3.
White, L. (2010). 'Markets: The credit rating agencies', *Journal of Economic Perspectives*, 24(2), 211-26.
Winkler, W. (2006). 'Overview of record linkage and current research directions', Statistical Research Division, US Census Bureau.

Table 1: Transformation of credit ratings data

Category	All				S&P		Moody's		Fitch	
	Rating	Ave rank	Obs	% equal or better	Rating	Obs	Rating	Obs	Rating	Obs
Prime	1	0.007	383	0.99%	AAA	236	Aaa	113	AAA	34
	2	0.029	1,080	3.78%	AA+	131	Aa1	482	AA+	467
High grade	3	0.072	2,217	9.50%	AA	573	Aa2	858	AA	786
	4	0.174	5,802	24.47%	AA-	1,920	Aa3	2,243	AA-	1,639
	5	0.324	6,748	41.88%	A+	2,559	A1	2,735	A+	1,454
Upper medium grade	6	0.497	7,178	60.40%	A	3,355	A2	2,367	A	1,456
	7	0.659	5,297	74.07%	A-	2,523	A3	1,454	A-	1,320
	8	0.769	3,015	81.85%	BBB+	1,738	Baa1	718	BBB+	559
Lower medium grade	9	0.846	2,836	89.17%	BBB	1,558	Baa2	589	BBB	689
	10	0.907	1,753	93.69%	BBB-	1,059	Baa3	375	BBB-	319
	11	0.947	609	95.26%	BB+	306	Ba1	211	BB+	92
Non-investment grade speculative	12	0.948	377	96.24%	BB	184	Ba2	80	BB	113
	13	0.963	585	97.75%	BB-	256	Ba3	267	BB-	62
	14	0.974	267	98.44%	B+	173	B1	59	B+	35
Highly speculative	15	0.980	208	98.97%	B	124	B2	54	B	30
	16	0.985	172	99.42%	B-	106	B3	43	B-	23
Substantial risks	17	0.978	10	99.44%	CCC+	2	Caa1	8	CCC+	0
Extremely speculative	18	0.982	41	99.55%	CCC	20	Caa2	13	CCC	8
In default with little prospect of recovery	19	1.000	9	99.57%	CCC-	8	Caa3	1	CCC-	0
	20	0.988	31	99.65%	CC	11	Ca	4	CC	16
	21	0.990	48	99.78%	C	0	C	41	C	7
In default	22		0	99.78%	R	0				
	23		0	99.78%	SD	0			RD	0
	24	0.997	87	100.00%	D	86			D	1

Note: Ratings given by each agency are transformed to a universal rating. For example, an AAA rating given by S&P is deemed equivalent to an Aaa rating given by Moody's; both are transformed to a rating of 1 for the purposes of our ranking procedure. The table also indicates the total number of observations on each rating by each agency. The column "% equal or better" conveys an impression of the distribution of these ratings.

Table 2: Summary Statistics

	Description	Source	Obs	Mean	Median	Std. Dev.	Min	Max
Ratings variables:								
Credit rating rank (simple)	Fractional rank of credit ratings	CRAs	21131	149.5	132.0	96.83	1.00	367.0
Credit rating rank (with outlook and watchlist)	Fractional rank of credit ratings using the outlook and watchlist	CRAs	21131	149.5	134.5	97.27	1.00	367.0
EDF	One-year expected default frequency	Moody's KMV	25572	0.83	0.14	3.01	0.01	35.00
EDF rank	Fractional rank of EDF	Moody's KMV	21131	149.5	135.0	97.65	1.00	367.0
ORQS	Ordinal Rating Quality Shortfall (ORQS) with 8-quarter forward EDF	Authors' calculations	21131	0.29	0.26	0.21	0.00	0.99
DORQS	Directional Ordinal Rating Quality Shortfall (ORQS)	Authors' calculations	21131	0.00	-0.01	0.36	-0.97	0.99
TORQS	Box-Cox Tranformation of Ordinal Rating Quality Shortfall (ORQS) with 8-quarter forward EDF	Authors' calculations	21131	-0.75	-0.73	0.42	-1.50	0.19
Rank difference: 'all-in' minus 'stand-alone'	Difference between the rank of a bank's Fitch all-in rating and the rank of a bank's Fitch stand-alone rating, normalized by sample size	Authors' calculations	7980	0.00	-0.02	0.26	-0.61	0.92
Dummy variables:								
Crisis	Dummy (=1) if 8-quarter forward EDF falls into crisis period with high average EDF	Authors' calculations	38753	0.19	0.00	0.39	0.00	1.00
Multiple rating dummy	Dummy (=1) if bank is rated by more than one agency.	Authors' calculations	38753	0.73	1.00	0.45	0.00	1.00
Macroeconomic variables:								
Credit growth	Change in country-level private credit stock on 12 quarters previous	Statistical Offices; Datastream	35157	0.16	0.18	0.16	-0.42	1.09
Bank balance sheet variables:								
Log assets	Natural log of a bank's on balance-sheet assets in USD	Bankscope	23975	10.75	10.73	1.79	2.36	15.15
RoA	Return on average assets	Bankscope	23304	0.82	0.84	1.49	-22.43	26.77
Leverage	Assets divided by equity multiplied by 100	Bankscope	23402	17.87	13.39	14.08	1.01	100.00
Loans share	Total loans divided by total assets	Bankscope	22785	57.24	61.05	18.16	0.02	98.22
Trading share	Net profits on trading and derivatives divided by total assets	Bankscope	38753	0.00	0.00	0.00	0.00	0.03
Short-term funding share	Deposits and short-term funding divided by total assets	Bankscope	23009	0.68	0.73	0.18	0.00	0.96
Agency specific securitization business	Ratio of securitization business between a bank and rating agency relative to agency's total securitization rating business	Authors' calculations	38753	4.87	0.00	9.51	0.00	25.96
Herfindahl-Hirschmann index	Measure of concentration in the market for bank ratings	Authors' calculations	38753	0.36	0.35	0.02	0.35	0.42

Source: Authors' calculations.

Table 3: Rating Quality and Rank Correlation

Panel A: Full Sample

Spearman Correlation between	Subsamples			Full sample
Rating Rank and EDF Rank	Top Tier	Middle Tier	Bottom Tier	
k=0	0.031 ***	0.023 **	0.417 ***	0.283 ***
k=12	-0.004	-0.016	0.378 ***	0.238 ***
k=24	-0.009	-0.036 ***	0.352 ***	0.205 ***
k=36	-0.017	-0.026 **	0.342 ***	0.176 ***

Panel B: Non-Crisis Period

Spearman Correlation between	Subsamples			Full sample
Rating Rank and EDF Rank	Top Tier	Middle Tier	Bottom Tier	
k=0	0.021 *	0.004	0.400 ***	0.259 ***
k=12	-0.027 **	-0.036 ***	0.366 ***	0.206 ***
k=24	-0.043 ***	-0.049 ***	0.378 ***	0.178 ***
k=36	-0.046 **	-0.039 ***	0.380 ***	0.165 ***

Panel C: Crisis Period

Spearman Correlation between	Subsamples			Full sample
Rating Rank and EDF Rank	Top Tier	Middle Tier	Bottom Tier	
k=0	0.086 ***	0.129 ***	0.511 ***	0.401 ***
k=12	0.119 ***	0.098 ***	0.449 ***	0.384 ***
k=24	0.139 ***	0.042	0.271 ***	0.321 ***
k=36	0.125 ***	0.043	0.164 ***	0.233 ***

Note: Parameter *k* denotes the time lag (in months) for the *EDF* measurement. The symbols *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. Top tier ratings comprise (mostly) ratings from AAA to AA-, middle tier ratings those from A+ to A-, and bottom tier ratings those from BBB+ to C. The tiers are constructed by dividing the sample into three.

Table 4: Credit Ratings During Crisis and after Credit Booms

Dependent Variable	Non-Directional Error: TORQS			Directional Error: DORQS		
	(1)	(2)	(3)	(4)	(5)	(6)
Crisis Dummy	-0.031 ***	-0.026 ***	-0.025 ***	0.003	0.003	0.003
	(0.006)	(0.007)	(0.007)	(0.003)	(0.003)	(0.003)
Credit growth		-0.211 ***	-0.201 ***		0.033	0.032
		(0.040)	(0.040)		(0.024)	(0.024)
Av serial correlation	0.777	0.768	0.768	0.777	0.768	0.768
Country fixed effects	Yes	Yes	No	Yes	Yes	No
Bank fixed effects	No	No	Yes	No	No	Yes
Time fixed effects	No	No	No	No	No	No
No. of observations	21,131	18,218	18,218	21,131	18,218	18,218

Note: The panel regressions allow for an AR(1) serial correlation structure and random effects. Symbols represent: Crisis = dummy for a crisis eight quarters forward, with crisis defined as the period from 2008Q4:2010Q4 and 2011Q2:2011Q3; Credit growth = change in country-level private credit stock on 12 quarters previous. Coefficients for country and bank fixed effects are not reported. The symbols *, **, and *** represent significance levels of 10 per cent, 5 per cent, and 1 per cent respectively.

Source: Authors' calculations.

Table 5: Rating Quality by Bank Size, Agency-Specific Securitization Business and Rating Agency

Variable	Non-Directional Error: TORQS			Directional Error: DORQS		
	(1)	(2)	(3)	(4)	(5)	(6)
Size						
Log assets	0.013 **	0.013 **	0.019 ***	0.051 ***	0.052 ***	0.042 ***
	(0.006)	(0.006)	(0.007)	(0.005)	(0.005)	(0.006)
Securitization						
ASSB			-0.002 **			0.005 ***
			(0.001)			(0.001)
Agency Dummies						
Moody's		-0.014	-0.017		0.040	0.046 *
		(0.026)	(0.026)		(0.025)	(0.025)
S&P		-0.004	-0.006		-0.087 ***	-0.083 ***
		(0.025)	(0.025)		(0.024)	(0.024)
Country fixed effects:						
Austria	0.218 ** (0.091)			0.347 *** (0.088)		
Belgium	0.006 (0.090)			0.199 ** (0.087)		
Cyprus	-0.319 * (0.178)			0.282 (0.172)		
Denmark	-0.073 (0.089)			0.250 *** (0.087)		
Finland	0.002 (0.121)			0.367 *** (0.119)		
France	0.104 ** (0.051)			0.331 *** (0.051)		
Germany	-0.016 (0.045)			0.232 *** (0.043)		
Greece	-0.101 * (0.058)			0.138 ** (0.056)		
Ireland	0.244 *** (0.076)			0.242 *** (0.074)		
Italy	-0.009 (0.034)			0.040 (0.034)		
Netherlands	-0.199 (0.159)			0.132 (0.160)		
Portugal	0.144 * (0.078)			0.159 ** (0.077)		
Spain	0.001 (0.046)			0.135 *** (0.045)		
Sweden	-0.109 (0.067)			0.194 *** (0.068)		
United Kingdom	0.171 *** (0.052)			0.266 *** (0.050)		
Country fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
observations	17,226	17,226	17,226	17,226	17,226	17,226

Note: Reported are panel regressions with bank level random or fixed effects. The regression allows for serial AR(1) correlation of the error. Symbols represent: Log assets = natural log of a bank's on balance-sheet assets in USD; ASSB = agency specific securitization business (business volume between agency and bank measured in logs); Moody's = dummy for a Moody's rating; S&P = dummy for an S&P rating. Coefficients for time fixed effects are not reported; coefficients for country fixed effects are reported only in column (1), although the fixed effects are included in all regressions. The symbols *, **, and *** represent significance levels of 10 per cent, 5 per cent, and 1 per cent respectively.
Source: Authors' calculations.

Table 6: Rating Quality and Additional Bank Characteristics

Dependent Variable	Non-Directional Error: TORQS			Directional Error: DORQS		
	(1)	(2)	(3)	(4)	(5)	(6)
Size						
Log assets	0.014 *	0.014 *	0.008	0.046 ***	0.046 ***	0.042 ***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Securitisation						
ASSB	-0.003 **	-0.003 **	-0.002 *	0.004 ***	0.004 ***	0.004 ***
	0.001	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Agency dummies						
Moody's		-0.009	-0.003		0.048 *	0.047 *
		0.027	(0.027)		(0.026)	(0.026)
S&P		0.001	0.006		-0.088 ***	-0.089 ***
		0.025	(0.025)		(0.024)	(0.024)
Profitability						
RoA	0.000	0.000	-0.002	0.003	0.003	0.003
	(0.005)	(0.005)	(0.004)	(0.003)	(0.003)	(0.002)
Capital structure						
Leverage	0.000	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Asset structure						
Loans share	-0.002 ***	-0.002 ***	-0.002 ***	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Trading share	-4.355 *	-4.341 *	-5.261 **	0.616	0.528	0.218
	(2.257)	(2.257)	(2.255)	(1.301)	(1.299)	(1.291)
Funding structure						
Short-term funding share	-0.008	-0.008	0.039	-0.065 **	-0.072 **	-0.058 *
	(0.050)	(0.050)	(0.048)	(0.033)	(0.033)	(0.032)
Rating Competition						
Multiple rating dummy			0.001			-0.030 ***
			(0.018)			(0.011)
HH index			0.455			-0.140
			(0.420)			(0.249)
Av. serial correlation	0.761	0.761	0.763	0.859	0.859	0.859
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	No	Yes	Yes	No
No. of observations	15,426	15,426	15,426	15,426	15,426	15,426

Note: The independent variables are: Log assets = natural log of a bank's on balance-sheet assets in USD; ASSB = agency specific securitization business (business volume between agency and bank measured in logs); Agency dummies for Moody's and S&P are 1 if the rating is from the respective agency and 0 otherwise; RoA = return on average assets; Leverage = assets divided by equity multiplied by 100; Loans share = total loans divided by total assets; Trading share = net profits on trading and derivatives divided by total assets; Short term funding share = deposits and short-term funding divided by total assets; Multiple rating dummy = dummy taking the value 1 if a bank is rated by more than one agency, 0 otherwise; HH index = Herfindahl Hirschmann index for concentration in the market for bank ratings. Coefficients for country and time fixed effects are not reported. The symbols *, **, and *** represent significance levels of 10 per cent, 5 per cent, and 1 per cent respectively.

Source: Authors' calculations.

Table 7: Robustness Check for Different Lags Between EDF and Credit Rating

Dependent Variable	Non-Directional Error: TORQS			Directional Error: DORQS		
Lag (in quarters)	0	4	12	0	4	12
	(1)	(2)	(3)	(4)	(5)	(6)
Size						
Log assets	-0.007	0.001	0.005	0.022 ***	0.040 ***	0.036 ***
	(0.007)	(0.007)	(0.008)	(0.005)	(0.005)	(0.006)
Securitisation						
ASSB	-0.001	-0.001	-0.002	0.004 ***	0.005 ***	0.005 ***
	0.001	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Agency dummies						
Moody's	0.010	0.018	-0.021	0.050 **	0.043 *	0.056 **
	0.028	0.026	(0.029)	(0.023)	(0.024)	(0.027)
S&P	0.020	0.016	-0.008	-0.074 ***	-0.090 ***	-0.086 ***
	0.027	0.025	(0.027)	(0.022)	(0.023)	(0.026)
Profitability						
RoA	0.021 ***	0.005	-0.003	-0.010 ***	-0.004 *	0.002
	(0.004)	(0.004)	(0.005)	(0.002)	(0.002)	(0.003)
Capital structure						
Leverage	0.003 ***	0.001	0.001	0.001 ***	0.001 **	0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Asset structure						
Loans share	-0.001 ***	-0.001 ***	-0.001 *	-0.001 ***	0.000 *	0.000
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Trading share	6.050 ***	-4.053 **	-3.569	0.882	-1.096	-4.322 ***
	(2.063)	(2.009)	(2.673)	(1.000)	(1.073)	(1.507)
Funding structure						
Short-term funding share	0.011	0.033	-0.009	-0.026	0.003	-0.103 ***
	(0.051)	(0.047)	(0.053)	(0.029)	(0.030)	(0.035)
Rating Competition						
Multiple rating dummy	0.010	-0.014	0.002	-0.025 **	-0.034 ***	-0.017
	(0.019)	(0.018)	(0.019)	(0.011)	(0.011)	(0.012)
HH index	0.008	-0.341	1.017 **	0.144	0.131	-0.141
	(0.461)	(0.425)	(0.437)	(0.239)	(0.245)	(0.251)
Av. serial correlation	0.748	0.755	0.769	0.851	0.859	0.861
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	No	No	No	No	No	No
No. of observations	18,615	17,274	13,578	18,615	15,426	13,578

Note: The independent variables are: Log assets = natural log of a bank's on balance-sheet assets in USD; ASSB = agency specific securitization business (business volume between agency and bank measured in logs); Agency dummies for Moody's and S&P are 1 if the rating is from the respective agency and 0 otherwise; RoA = return on average assets; Leverage = assets divided by equity multiplied by 100; Loans share = total loans divided by total assets; Trading share = net profits on trading and derivatives divided by total assets; Short term funding share = deposits and short-term funding divided by total assets; Multiple rating dummy = dummy taking the value 1 if a bank is rated by more than one agency, 0 otherwise; HH index = Herfindahl Hirschmann index for concentration in the market for bank ratings. Coefficients for country and time fixed effects are not reported. The symbols *, **, and *** represent significance levels of 10 per cent, 5 per cent, and 1 per cent respectively.

Source: Authors' calculations.

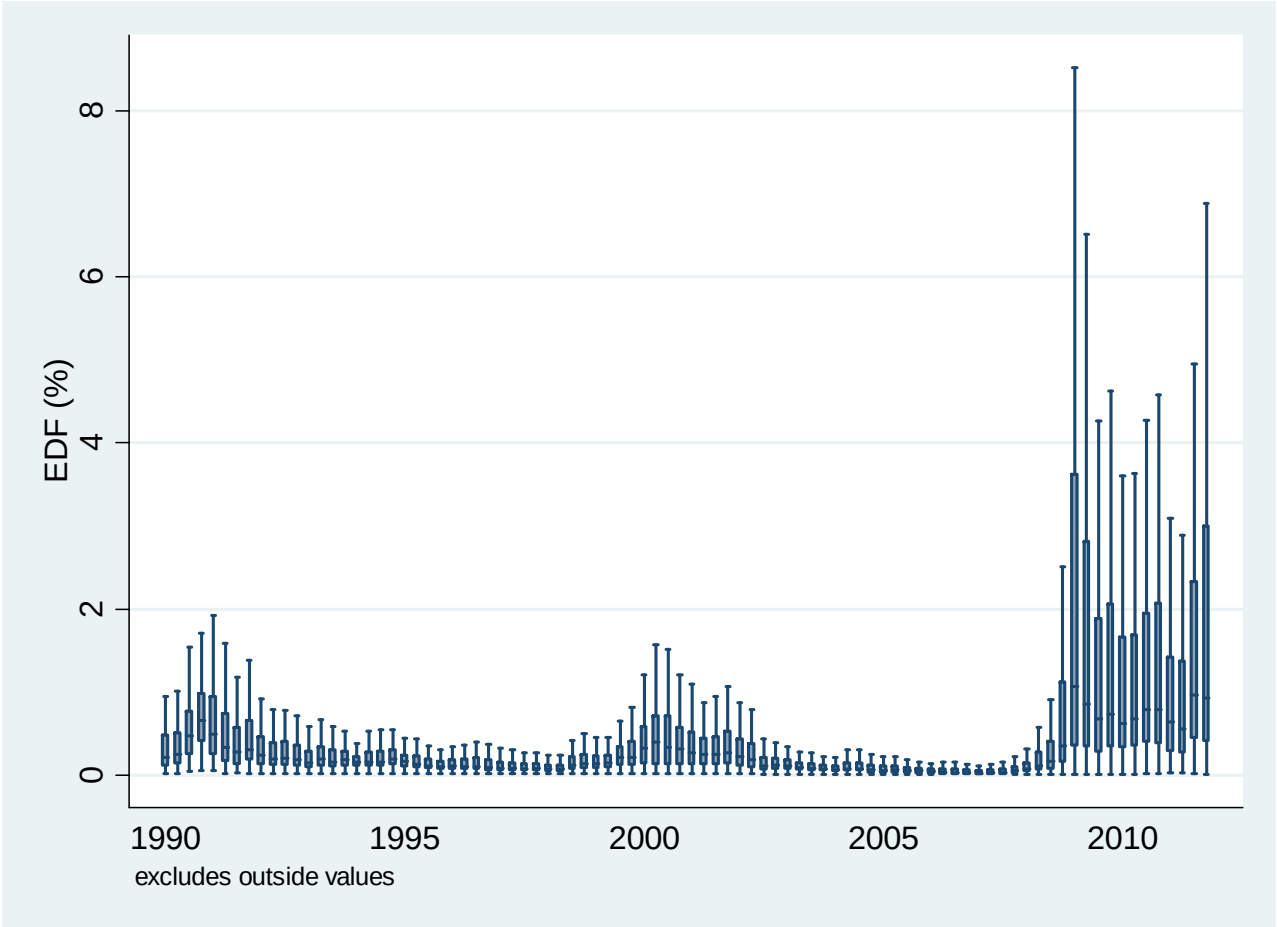
Table 8: Robustness Check for the Effect of Government Support: Fitch Ratings Only

Dependent Variable	Non-Directional Error: TORQS		Directional Error: DORQS	
	(1)	(2)	(3)	(4)
Size				
Log assets	0.024 *	0.020	0.061 ***	0.056 ***
	(0.013)	(0.013)	(0.010)	(0.010)
Securitisation				
ASSB	-0.003	-0.003	0.003	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Government support				
Rank difference: 'all-in' minus 'stand-alone'		0.102 *		0.169 ***
		(0.056)		(0.033)
Profitability				
RoA	-0.008	-0.008	0.008 *	0.008 *
	(0.008)	(0.008)	(0.004)	(0.004)
Capital structure				
Leverage	-0.001	-0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Asset structure				
Loans share	-0.002 ***	-0.002 **	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Trading share	-6.259 *	-5.913	-1.092	-0.288
	(3.610)	(3.614)	(2.054)	(2.053)
Funding structure				
Short-term funding share	0.049	0.059	0.044	0.059
	(0.087)	(0.088)	(0.058)	(0.058)
Rating Competition				
Multiple rating dummy	-0.037	-0.036	-0.016	-0.017
	(0.035)	(0.035)	(0.021)	(0.021)
HH index	0.620	0.785	-0.181	0.150
	(0.875)	(0.879)	(0.513)	(0.515)
Av. serial correlation	0.874	0.782	0.874	0.873
Country fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	No	No	No	No
No. of observations	4,480	4,480	4,480	4,480

Note: The independent variables are: Log assets = natural log of a bank's on balance-sheet assets in USD; ASSB = agency specific securitization business (business volume between agency and bank measured in logs); Directional rank difference = difference between the rank of a bank's senior unsecured debt rating and the rank of a bank's individual (stand-alone) rating, normalized by sample size; RoA = return on average assets; Leverage = assets divided by equity multiplied by 100; Loans share = total loans divided by total assets; Trading share = net profits on trading and derivatives divided by total assets; Short term funding share = deposits and short-term funding divided by total assets; Multiple rating dummy = dummy taking the value 1 if a bank is rated by more than one agency, 0 otherwise; HH index = Herfindahl Hirschmann index for concentration in the market for bank ratings. Coefficients for country and time fixed effects are not reported. The symbols *, **, and *** represent significance levels of 10 per cent, 5 per cent, and 1 per cent respectively.

Source: Authors' calculations.

Figure 1: Moody's KMV one-year Expected Default Frequencies (EDF™)



Notes: Graph shows the *EDF*s of the 397 banks in our unbalanced panel. In the boxplot, the median *EDF* is given by the horizontal line inside the box. The box contains observations on *EDFs* between the 25th and 75th percentiles. Adjacent values (the 'whiskers' of the plot) are the most extreme observations within 1.5 interquartile ranges of the nearest quartile. Outside values are not shown.

Figure 2: Distribution of ORQS and its Box-Cox transformation

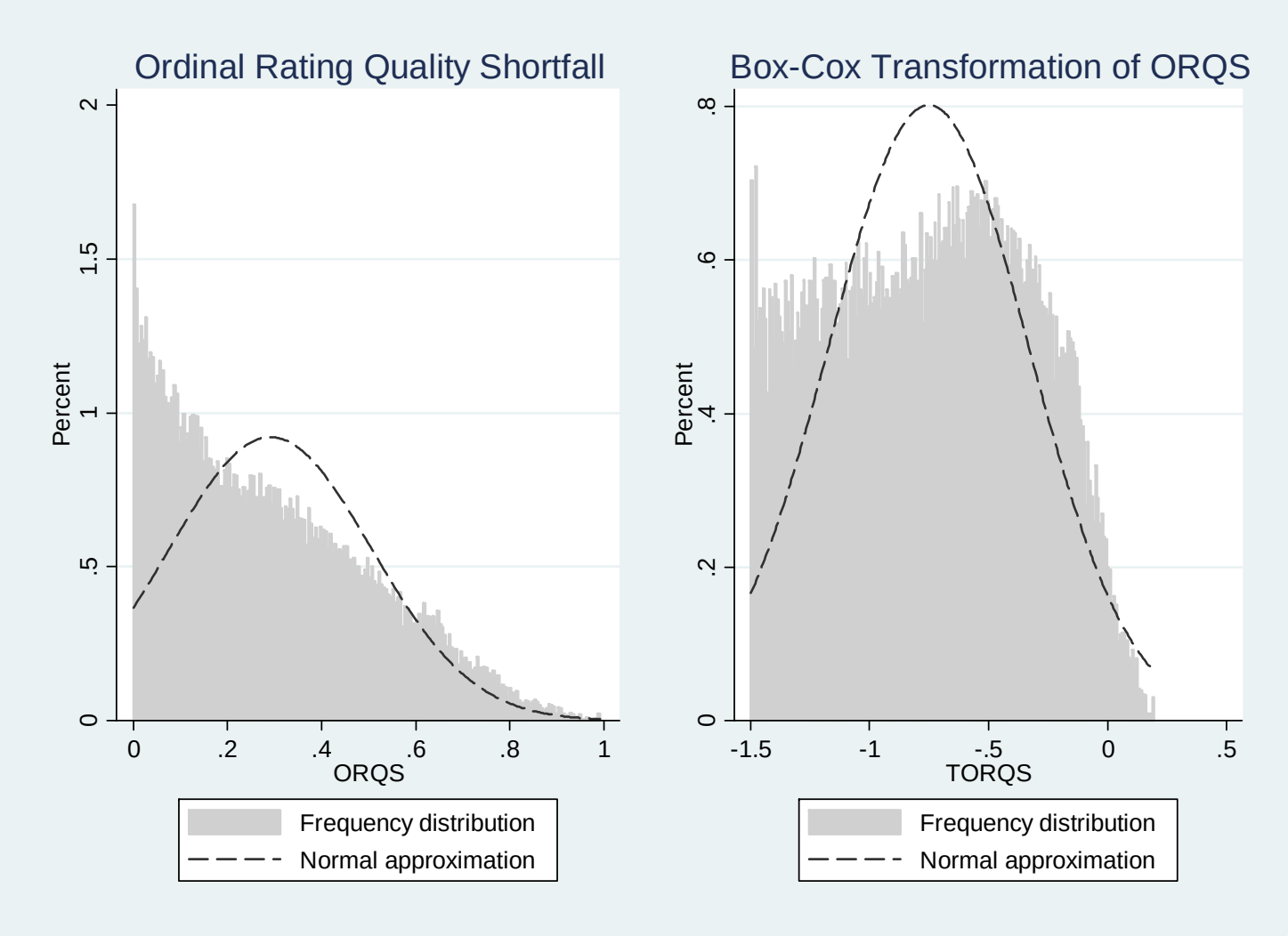


Figure 3: Rating Quality in Crisis versus Non-crisis times

Figure 4: Market Concentration for Bank Ratings

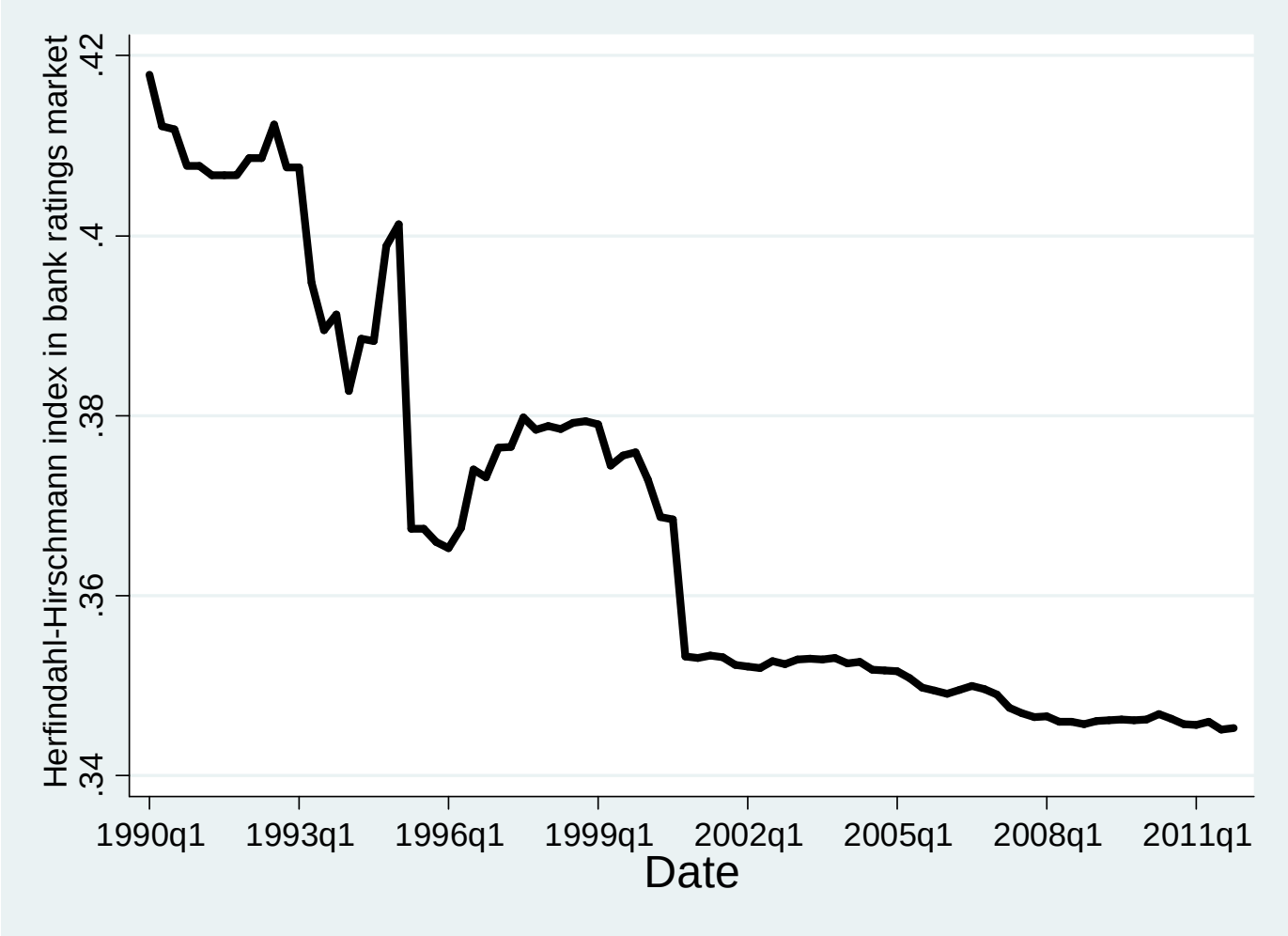
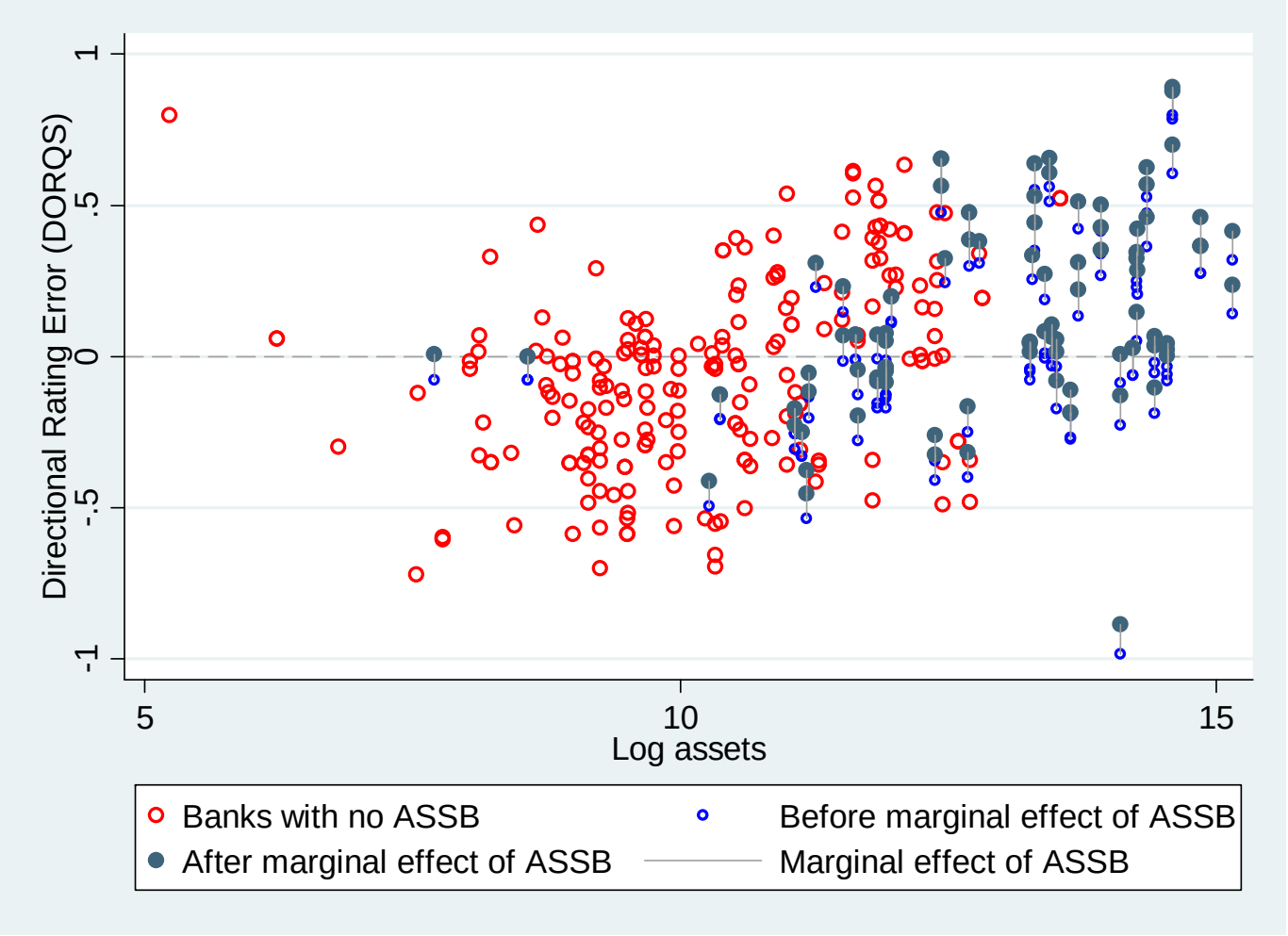


Figure 5: Marginal effect of Agency-Specific Securitization issuance on directional rating error



Notes: ASSB stands for agency-specific securitization business. Graph shows cross-section of banks on January 2007.