Bank ratings

SUMMARY

This paper examines the quality of credit ratings assigned to banks by the three largest rating agencies. 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 on average large banks receive more positive bank ratings, particularly from the agency to which the bank provides substantial securitization business. These competitive distortions are economically significant and contribute to 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.

— *Harald Hau, Sam Langfield and David Marques-Ibanez*

Economic Policy April 2013 Printed in Great Britain

Bank ratings: what determines their quality?

Harald Hau, Sam Langfield and David Marques-Ibanez

University of Geneva, Swiss Finance Institute and CEPR; European Systemic Risk Board Secretariat and UK Financial Services Authority; European Central Bank, Financial Research Division

1. INTRODUCTION

In the aftermath of the initial phase of the financial crisis in 2007–2008, 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 securitized credit, were also found wanting. Ratings were subject to particu-

Opinions expressed herein are those of the authors only. They do not necessarily reflect the views of, and do not involve any responsibility for, the institutions to which the authors are affiliated. Any errors are the fault of the authors. Harald Hau acknowledges financial support from the Geneva Financial Research Institute and the Swiss Finance Institute. The authors are indebted to Thomas Drechsel and Matthias Efing for excellent research assistance and to Johannes Micheler and Antonia Simeonova for providing data and code. We are grateful to our discussants who provided careful and insightful comments: Thorsten Beck and Isabel Schnabel at the 56th Panel meeting of Economic Policy in Cyprus; Ricardo Correa at the Global Research Forum organized by the European Central Bank and the New York Fed; and Lampros Kalyvas at a workshop at the European Banking Authority. Bo Becker, Allen Berger, Oliver Burkart, Jean-Pierre Danthine, Matthias Efing, Artus Galiay, Linda Goldberg, John Griffin, Iftekha Hasan, Janet Jackson, Zijun Liu, David Llewellyn, Simone Manganelli, Jose Geli Manzano, John Muellbauer, Steven Ongena, Alex Popov, Ana Rita Ribeiro Mateus, Andrei Sarychev, Frank Smets, Jeremy Stein, Balázs Zsámboki and four anonymous referees provided helpful comments.

The Managing Editor in charge of this paper was Philip Lane.

Economic Policy April 2013 pp. 289–333 Printed in Great Britain

larly 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, because it only requires ratings to be consistent over time. A higher credit rating must correspond to lower default risk, but not to any particular quantity of default risk. Thus, an ordinal rating metric may remain accurate even with the dramatic increases in cardinal default probabilities 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 characterize Moody's credit ratings as '*ordinal* measures' (Moody's, 2006). Statements by other rating agencies are similarly contradictory about the cardinal versus ordinal interpretation of credit ratings.

A cardinal rating for banks requires rating agencies to predict bank distress in normal times as well as during generalized 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 default frequencies (*EDF*s) 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 *EDF*s calculated by Moody's as a measure of risk, we maintain methodological fairness by avoiding subjective risk modelling choices (see Section 4). *EDF*s capture perceptions of bank risk derived from a structural model

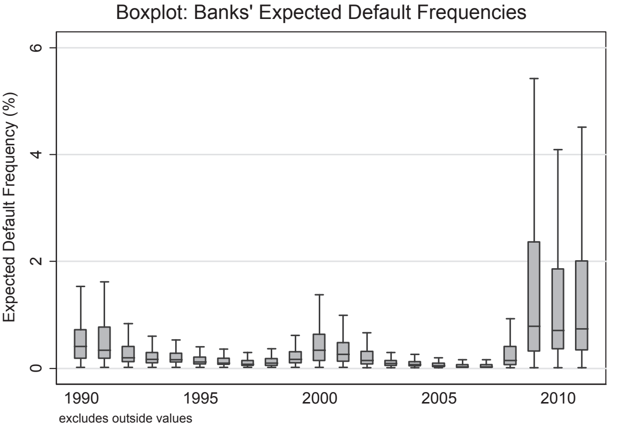

Figure 1. Moody's KMV one-year expected default frequencies (*EDF*™)

Notes: Graph shows the *EDF*s of the 369 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 *ED*Fs at the 25th and 75th percentiles. Adjacent values are the most extreme observations within 1.5 interquartile ranges of the nearest quartile. Values outside of these ranges are not shown in the graph.

incorporating expectations from equity markets. Moreover, unlike some other indicators of bank risk, *EDF*s 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 *EDF*s 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.

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. Bank size strongly correlates with more favourable 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 alter the rating process. Using additional data on banks' agency-specific securitization 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 *EDF*s. 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 summarize 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 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 buy-side investors' demand for inflated ratings (Calomiris, 2009).
4. Given the strong negative externalities of bank opacity, a promising policy option lies in enhanced transparency of banks. Substantial improvement of banks' public disclosure with granular reporting of risk positions seems 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 organized 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 securitization 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.*, 2012).

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. 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 US and EU15, total bank debt issuance amounted to approximately US$1,000 billion in 2011 – comprising US$442 billion of corporate bonds; US$134 billion of medium-term notes; US$116 billion of short-term debt and US$362 billion of covered bonds. Thus, despite recent marginal changes in funding 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 characterized 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 2008–2009 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 recapitalization 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.

[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 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 of 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 securitized loans, but introduces an additional role for credit ratings with respect to counterparty credit risk from over-the-counter derivatives (Basel Committee on Banking Supervision (BCBS), 2010).

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 criticized (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 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 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 sharpened in 1975, when credit rating agencies shifted from an 'investor pays' to an 'issuer pays' model (White, 2010; Pagano and Volpin,

[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.

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’. On average, the larger the potential future business between rating agencies and their clients, the larger an agency’s incentive to inflate ratings. Related analysis of structured debt ratings by Efing and Marqués-Ibáñez (2012) indicates that issuers that generate more rating business receive rating favours and benefit from lower yield spreads, and that this mechanism was strongest at the height of the credit cycle in 2004–2006. 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 of securities when the issuer is large, since issuer size is correlated with the agency fee (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 collateralized assets with inflated ratings turned out to be on bank balance sheets during the crisis. Opp *et al.* (2013) show that rating-contingent regulation can significantly lower an agency’s incentives to acquire costly information and to produce high-quality ratings. Investors do not scrutinize 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, 2013). 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 ascer-

[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.

tain rating quality. The presence of 'naïve investors' would also strengthen the counter-cyclical nature of ratings quality (Bolton *et al.*, 2012).[4] Evidence of rapid and widespread downgrades of structured finance securities' ratings over 2007–2008 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 *et al.*, 2010). Similarly, rating quality can be reduced if issuers shop for more favourable ratings (Bolton *et al.*, 2012). Becker and Milbourn (2011) 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 also be affected by reasons unrelated to incentive problems. In particular, opacity and complexity might impair rating quality. Compared with other large corporations, big banks are 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 between agencies occur more often in the case of banks' ratings than those of other industries (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 securitization activity has 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, 2011). 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.*, 2013).

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 (2012), for example, found that banks with more

[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).

Tier I capital and a higher ratio of loan to total assets performed better in the initial stages of the crisis. Berger and Bouwman (2013) 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 hypothesize that a bank's business model is related to the accuracy of its credit rating.

4. DATA

We construct a comprehensive panel of US and EU15 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. Our benchmark analysis concerns banks' long-term issuer ratings, which refer to the probability of repayment of senior unsecured credit obligations. In an extended analysis, we also scrutinize bank financial strength ratings, which assess banks' creditworthiness as independent stand-alone entities, absent reliance on government guarantees.

To focus on group-wide financial distress and avoid double-counting ratings within a single institution, we discard any bank that is junior in the organizational structure – for example, HSBC Holdings plc is retained as the bank holding company; junior entities within this group, such as HSBC France SA, are discarded. More practically, *EDF*s are mostly available at the level of the listed entity, which generally corresponds to the bank holding company or most senior banking entity within a group, rather than individual subsidiaries.

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, 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 period after January 2000.

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

Table 1. Transformation of credit ratings data

	All				S&P		Moody's		Fitch	
Category	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 examp le, 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 all-in rating and the rank of a bank's standalone rating, normalized by sample size	Authors' calculations; CRAs	16135	0.00	0.01	0.27	0.71	0.96
Dummy variables:								
Crisis	Dummy (=1) if 8-quarter forward EDF falls into crisis p eriod 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 Authors' 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

Table 2. (Continued)

	Description	Source	Obs	Mean	Median	Std. Dev.	Min	Max
Leverage	Assets divided by equity all divided by 100	Bankscope	23402	0.18	0.13	0.14	0.01	1.00
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
ASSB (Agency-specific securitization business)	Log of the securitization volume between a bank and a rating agency	Authors' calculations; Dealogic	38753	4.87	0.00	9.51	0.00	25.96
ASSB ex-guarantee	Same as ASSB, excluding deals which are guaranteed either by the issuing bank or a third party	Authors' calculations; Dealogic	38753	4.85	0.00	9.48	0.00	25.95
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

some winsorizing of extreme observations on balance-sheet variables. For example, we impose that observations on leverage must lie between 0 and 1. Table 2 provides the definitions of the accounting variables we retain and their summary statistics.

Finally, we match the above panel with data on *EDF*s as a measure of bank distress. *EDF*s are obtained from a structural model of corporate default and widely used to price corporate bond debt (Merton, 1974; Longstaff and Schwartz, 1995).[5] 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 *EDF*s calculated by a division of the rating agency Moody's, contemporaneously to the rating process. Moody's calculations are undertaken monthly and draw on a large proprietary default database (Dwyer and Qu, 2007). It is possible to reconstruct proxy *EDF*s 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 impossible to find the ideal indicator of bank risk. The *EDF* measure (as calculated by Moody's KMV) has a number of limitations. First, *EDF*s refer to the probability of default on all credit obligations, regardless of seniority. In contrast, long-term issuer ratings refer to the probability of repayment of senior unsecured credit obligations. We abstract from this problem by noting that senior unsecured credit is the most common type of credit in banks' liability structures. Second, *EDF*s incorporate expectations of creditor bail-out only indirectly due to the junior status of equity in banks' liability structures. Third, 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).[6]

Notwithstanding its limitations, our choice of the *EDF* indicator is justified by specific reasons linked to our research design. First, *EDF*s attempt to measure the probability of default on obligations to creditors, and are therefore comparable with ratings.[7] Unlike other indicators of bank risk (such as spreads on credit default

[5] 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.

[6] 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.

[7] Both S&P and Fitch assign credit ratings based solely on the probability of default on obligations to creditors. However, Moody's credit assessment criteria are more complex: credit ratings represent 'ordinal measures of *expected loss*' (Moody's, 2006), where expected loss can be interpreted as the product of the probability of default and loss given default. Elsewhere, Moody's states that 'ratings reflect both the likelihood of default and any financial loss suffered in the event of default'. In this paper, we abstract from Moody's conflation of probability of default and loss given default, and treat Moody's issuer ratings as equivalent to S&P and Fitch ratings. Nevertheless, any structural between-group variation in the ratings process would be captured by a Moody's rating dummy, which is reported in most regressions.

swaps), *EDF*s are available with a relatively long time series, facilitating more robust panel analysis. Comparability between ratings and *EDF*s is further facilitated by the continuous nature of the *EDF* variable, which allows bank risk to vary within a rating category. Other structural measures of credit risk, such as Credit-Metrics (created by JP Morgan), assume that issuers are homogeneous within the same rating class.

Second, *EDF*s represent a good approximation of default risk perceived by equity investors over a one-year horizon (Crouhy *et al.*, 2000). 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).

Third, perhaps most importantly, any residual noise in *EDF* observations is unlikely to have any structure related to the hypotheses examined in this paper. In particular, as a mechanical measure based on equity prices, *EDF* noise is unlikely to be correlated with variables related to our 'conflict of interest' hypothesis. Only ratings, which depend on human judgement, can plausibly have a structure consistent with the conflict of interest hypothesis.

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 general assessments of a bank's probability of future financial distress, which we operationalise as the forward *EDF*. We therefore compare the credit ratings to *EDF*s 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. Moreover, long-term issuer ratings represent opinions on creditworthiness through the cycle, rather than short-term fluctuations in macroeconomic conditions (Moody's, 2006; Kiff *et al.*, 2013). 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 between 21 and 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.[8] 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 (lower) credit rating rank if the outlook is negative (positive).[9] 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 *EDF*[10] at time $t + k$, normalized 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 (see Table 2 and Figure 3b).[11] 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.

The interpretation of *ORQS* as an indicator of rating quality requires that the *EDF* rank is an unbiased measure of banks' relative credit risk. The noisier our *EDF* benchmark, the more problematic the interpretation of *ORQS* as an indicator of rating quality, as opposed to measurement error reflected in the *EDF*. To increase the precision of the *EDF* rank as a useful benchmark for rating quality, we measure the *EDF* at k periods ahead. Our benchmark regressions take k to be two years, but results are robust to different values of k (see Table 7). Taking forward *EDF* observations incorporates additional equity price information. By construction, this forward observation makes the *EDF* rank more informative about relative bank risk at time $t + k$ than the

[8] For example, consider five banks: banks A and B are rated AAA stable outlook; bank C is AAA negative outlook; bank D is AAA negative outlook and on watch; bank E is AA+ positive outlook. Here, we would assign rankings of 1.5 to bank A; 1.5 to bank B; 3 to bank C; 4 to bank D; 5 to bank E. Each rank is then normalized by the sample size: in this case, 5.

[9] 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.

[10] We also implement a subordinate second sort criterion for the purposes of calculating the final *EDF* rank, in a similar manner to the ranking procedure used for ratings. 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.

[11] 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 *et al.* (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.

credit rating rank observed at time t. While we concede that the *ORQS* picks up some residual measurement error related to shortcomings of the *EDF* as structural measure of default risk, it is not plausible that such measurement error correlates with the variables used for testing our hypotheses. Valid inference on the structure credit rating errors only requires that the *EDF* measurement error is small and not correlated with our explanatory variables.

A further critique of the *ORQS* concerns the conceptual difference between an *EDF* and a credit rating. While the *EDF* provides a strictly probabilistic assessment of the default event, a credit rating might in addition account for the magnitude of creditor loss in the event of default. However, this conceptual difference should be of limited relevance for our analysis. The small number of actual bank defaults makes it difficult to incorporate precise information on expected credit loss into bank ratings. Both *EDFs* and credit ratings are essentially probabilistic assessments of future default. Figure 2 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 dashed and solid lines represent kernel estimations 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 *EDF*s 24 months later. The scatter plots show instead a more uniform dispersion of the observations over the entire quad-

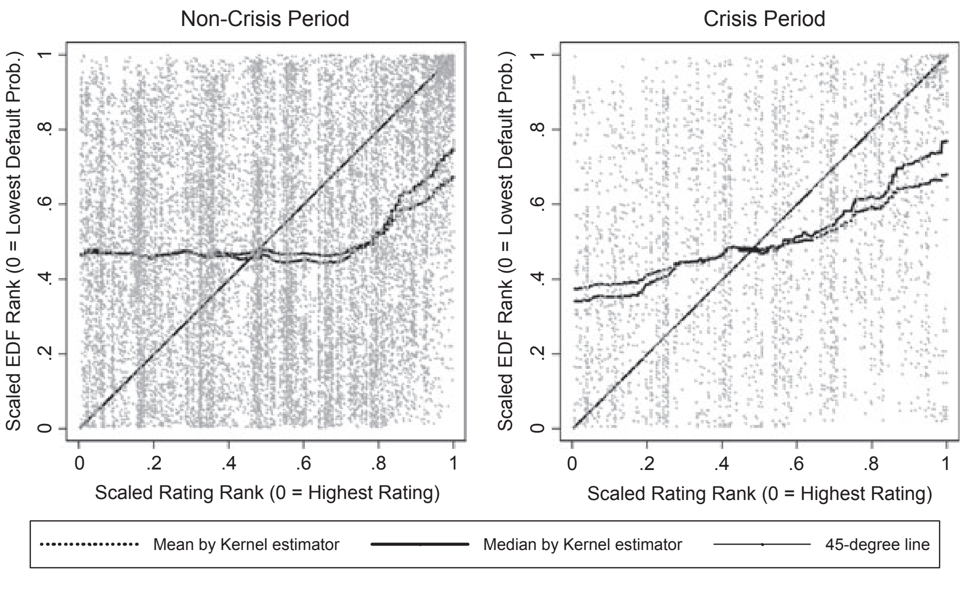

Figure 2. Rating quality in crisis versus non-crisis times

rant, 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.

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 *EDF*s 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 2, 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 rat-

Table 3. Rating quality and rank correlation

Panel A: Full sample Spearman correlation between	Subsamples			
rating rank and EDF rank	Top tier	Middle tier	Bottom tier	Full sample
k = 0	0.034***	0.022**	0.418***	0.281***
k = 12	−0.004	−0.015	0.381***	0.236***
k = 24	−0.008	−0.032***	0.357***	0.203***
k = 36	−0.016	−0.029**	0.342***	0.175***
Panel B: Non-crisis period				
k = 0	0.025**	0.003	0.402***	0.257***
k = 12	−0.027**	−0.036***	0.371***	0.205***
k = 24	−0.039***	−0.049***	0.380***	0.177***
k = 36	−0.044***	−0.034**	0.382***	0.164***
Panel C: crisis period				
k = 0	0.082***	0.134***	0.508***	0.399***
k = 12	0.116***	0.102***	0.446***	0.382***
k = 24	0.131***	0.061	0.277***	0.319***
k = 36	0.121***	0.039	0.158***	0.230***

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.

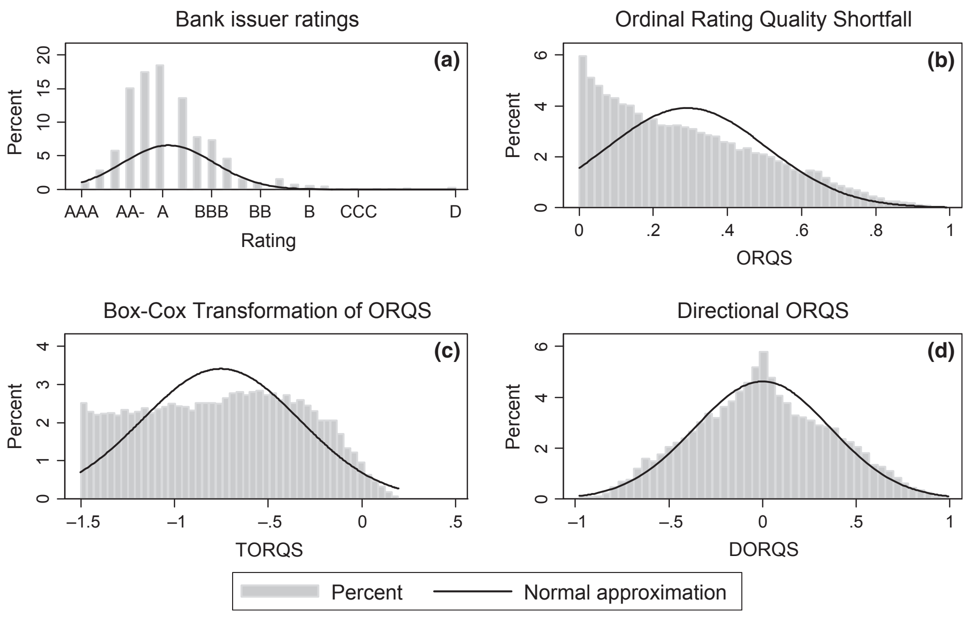

Figure 3. Distributions of ratings and measures of ratings quality

ings are therefore considerably more informative for banks' relative creditworthiness within a financial crisis than outside.

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 3b. 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 is more suitable for regression analysis. The Box–Cox parameter of –0.224 brings the skewness exactly to zero (Figure 3c).[12] 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

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

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

[12] 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.

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 (Figure 3d), enabling us to apply regression analysis directly without any further variable transformation. We also highlight that *DORQS* has by construction a near-zero cross-sectional mean (Table 2) 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.

6. HYPOTHESES ABOUT CREDIT RATING QUALITY

In this section we formulate and discuss four hypotheses about the determinants of rating quality. As suggested by Figure 2, 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, 2013).

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 securitization 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 securitization 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 (which make up 28% of the sample), other asset-backed securities (3.8%), commercial mortgage-backed securities (3.4%), collateralized loan obligations (5%), other collateralized debt obligations (5.9%) and home equity loans (53.8%). Importantly, we observe 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, securitization business is highly irregular over time, so that aggregation over the entire time period provides the best measurement of the overall securitization 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 securitization business* (*ASSB*) as

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

The log transformation is appropriate because a large share of the securitization 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 Wall Street Reform and Consumer Protection 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 all divided by 100. Bank leverage has often been deemed excessive and conducive to more risk (Berger and Bouwman, 2013), hence our interest in whether it also contributes to larger rating errors. Asset structure is proxied by two variables, namely the *Loan share* (total loans over total assets) and *Trading share* (net profit on trading and derivatives divided by total assets). 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.

An extensive literature concerns the nexus between competition and rating quality. Competition may foster reputational effects, which appear to matter for the reporting quality of equity analysts (Fang and Yasuda, 2009). Others have argued that competition may also compromise rating quality if corporations can 'shop' for the best available rating (Becker and Milbourn, 2011). To control for variation in the level of competition over time, we measure market share of the three largest rating agencies using the Herfindahl–Hirschmann index of concentration (HH index). Market concentration decreased over the 1990s due to an increase in market share of 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 second control variable considers the rating conditions at the bank level. Roughly 73% of all banks have multiple ratings, which should reduce the absolute importance of any single rating of this bank. We create a *Multiple rating dummy*, which takes on the value of 1 whenever more than one rating agency has issued a bank rating (and 0 otherwise). Banks have some control over how many ratings they sequentially solicit. Thus banks with only one rating are most likely to enjoy a more favourable one.

7. EMPIRICAL ANALYSIS

7.1. Rating quality during the crisis and after credit booms

The distributional evidence in Figure 2, 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

Table 4. Credit ratings during crisis and after credit booms

	Non-directional error: TORQS			Directional error: DORQS		
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
Crisis dummy	-0.030^{***}	-0.024^{***}	-0.024^{***}	0.003	0.003	0.003
	(0.006)	(0.007)	(0.007)	(0.003)	(0.003)	(0.003)
Credit growth		-0.210^{***}	-0.200^{***}		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: Reported are panel regressions with bank level random effects. The regression allows for serial AR(1) correlation of the error. The panel regressions allow for an AR(1) serial correlation structure and random effects. Symbols represent: Crisis = dummy for a crisis 8 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%, 5%, and 1% respectively.
Source: Authors' calculations.

and in a financial crisis, the ordinal rating error decreases significantly. Hence ratings quality is counter-cyclical: bank ratings are a better predictor of credit risk during a financial crisis than in normal times. This finding contradicts the frequently voiced criticism that bank ratings are less reliable during financial crises.

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.

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.[13] 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 next section.

The country fixed effect in column (1) shows that banks headquartered in Austria, Ireland and the United Kingdom, and to a lesser extent France and Portugal, feature significantly higher ratings errors. The results are more apparent in column (4) for the directional error: the coefficient on every EU country dummy is positive, suggesting that European banks enjoy strong positive rating bias relative to the ratings of US-based banks. This result is robust when the regression is run on the subsample of Moody's and S&P ratings, thus contradicting the assertion that these US-based rating agencies might have been prejudiced against non-US banks.

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, S&P ratings are significantly more negative than those of Moody's and Fitch. This implies that rating errors feature a systematic component that is related to an agency's broad rating policy.

The analysis in the following sections retains agency and country dummies as control variables. However, the country dummies are not reported separately as they are very similar to those provided in Table 5, columns (1) and (4), respectively.

7.3. Rating quality and bank size

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.

Regressions reported in Table 5, columns (2) and (5), confirm the important role of bank size as a determinant of rating accuracy and bias. Bank size measured as Log

[13] The reported standard errors are adjusted for serial correlation of the errors using the *xtregar* command in Stata. Quantitatively very similar results are obtained if we correct the standard errors for clustering at the bank level.

Table 5. Rating quality by bank size, agency-specific securitization business and rating agency

	Non-directional error: TORQS			Directional error: DORQS		
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
Size						
Log assets	0.013** (0.006)	0.019 *** (0.007)	0.019 *** (0.007)	0.051*** (0.005)	0.042 *** (0.006)	0.042 *** (0.006)
Securitization						
ASSB		−0.002 ** (0.001)			0.005 *** (0.001)	
ASSB ex−guarantee			−0.002 ** (0.001)			0.005 *** (0.001)
Agency dummies						
Moody's		−0.017 (0.026)	−0.017 (0.026)		0.046 * (0.025)	0.046 * (0.025)
S&P		−0.006 (0.025)	−0.006 (0.025)		−0.083 *** (0.024)	−0.083 *** (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 effects	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	17,226	17,226	17,226	17,226	17,226	17,226

Note: Reported are panel regressions with bank level random 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); ASSB ex-guarantee = same as ASSB, excluding deals which are guaranteed either by the issuing bank or a third party; 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%, 5%, and 1% respectively. Source*:* Authors' calculations.

assets strongly and positively correlates with both non-directional rating error (*TORQS*) and rating bias (*DORQS*). The regression coefficient of 0.042 for Log *assets* in Table 5, column (5), 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.[14] This corresponds, for example, 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 average reduction in funding costs of 40 basis points.

We highlight that conflicts of interest between rating agencies and large banks may not be the only interpretation of the rating error and positive bias arising from bank size. An alternative interpretation could relate this bias to the 'too big to fail' privilege of big banks. Our analysis is based on 'all-in' ratings, which account for the ability of banks' home sovereign 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. Section 8 probes this alternative explanation further by using so-called 'financial strength' or 'stand-alone' ratings, which explicitly ignore sovereign support for banks. The rank difference between stand-alone ratings and conventional 'all-in' issuer ratings, which incorporate the conditional probability of government support, captures the 'too big to fail' privilege of systemically important banks in the form of a rating uplift. Including this control does not eliminate the considerable bias in favour of large banks, which casts doubt on this alternative interpretation.

A third interpretation of bias in favour of large banks could be that rating agencies collectively misjudged the relative fragility of large banks in a financial crisis. For example, the crisis revealed enormous potential losses for large dealer banks related to their over-the-counter product exposure (Duffie, 2010). The two-year-ahead measurement of banks' expected default frequency might introduce a hindsight bias which is particularly pronounced for large banks. To eliminate this potential for hindsight bias, Table 7 repeats the regression in Table 6 for the special case with $k = 0$ (instead of $k = 24$ months) so that the *EDF* and credit rating are observed contemporaneously. The magnitude of the bank size coefficient drops by half, but remains highly significant. At

[14] 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.

best, delayed learning about riskiness resulting from bank size can therefore explain about half of the rating bias in favour of large banks.

It is also important to highlight the strength of the rating error attributed to large banks at the onset of the crisis in January 2007. Figure 4, panel A, divides all directional rating errors into quintiles and plots the mean and median bank size within each quintile. The average bank size within each quintile increases almost tenfold, from a mean asset value of US$74 billion for the 20% most underrated banks to US $713 billion for the 20% most overrated banks. Similarly, median bank size jumps from US$16 billion in assets to US$261 billion.

This pronounced bias in favour of large banks is apparent also if rating error is measured not in terms of the future *EDF* rank, but simply by the change in the rating rank over two years. Figure 4, panel D, reports the mean and median bank size for quintiles of rating changes from the 20% largest upgrades to the 20% largest downgrades between January 2007 and January 2009. The ratings which were most inflated in January 2007 – that is, the ratings which were subsequently downgraded the most over the crisis – concern disproportionately the largest banks. The 20% most downgraded banks dominate other quintiles in terms of size, with mean (median) assets of US$713 billion (US$262 billion) in January 2007. Independent of the definition of rating error, large banks enjoyed on average overoptimistic credit ratings before the financial crisis.

7.4. Rating quality and securitization business

An 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 securitization business* (ASSB), the more the quality of the bank rating might be compromised. Unlike the bank size variable, the *ASSB* variable allows us to infer how different agencies rated the same bank as a function of their specific business relationship in rating structured products. This permits clearer inference on conflicts of interest in bank ratings.

In Table 5, column (5), we find that the *ASSB* measure is related to a statistically significant upward bias in the rating.[15] 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 and black circles distinguish banks with and without substantial securitization business, respectively. The vertical lines between small and large black circles depict the predicted marginal change in the directional error due to the bank's agency-specific securitization business. Figure 5 also illustrates that most of the banks engaged in asset securitization are large. Rating favours related to agency-specific securitization business therefore occur in addition to the general rating bias in favour of large banks. Again, we can quantify the economic magnitude of the rating bias in

[15] This effect is also robust to the inclusion of bank fixed effects.

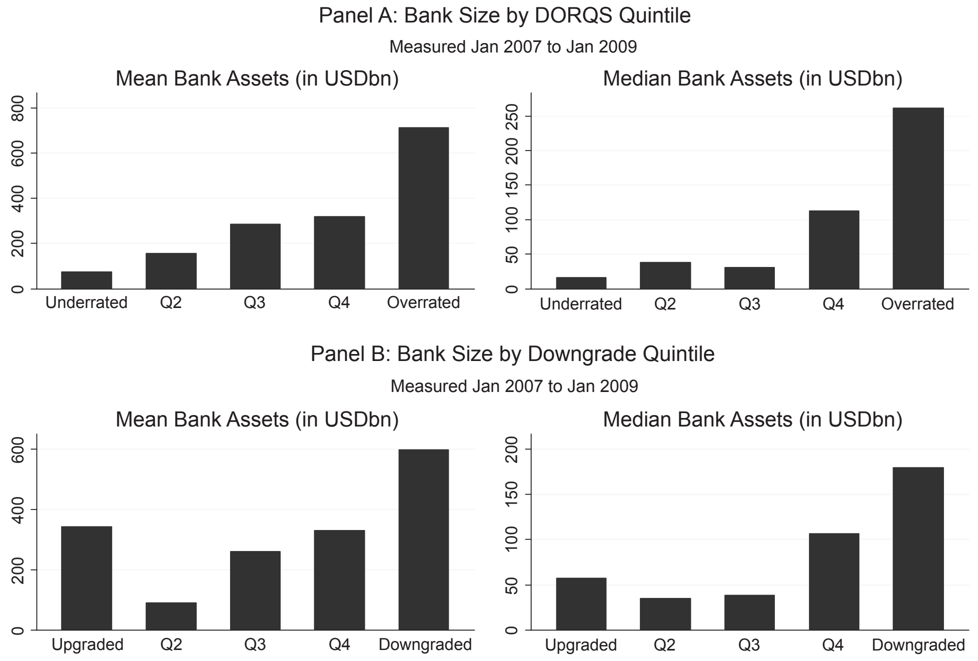

Figure 4. Bank size by quintiles of directional rating error and downgrade magnitude over 2007–2009

Notes: Panel A groups banks into quintiles according to their directional rating error (DORQS) in January 2007 ranging from most the most underrated (Q1) to the most overrated banks (Q5) and reports the mean and median bank size for each quintile. Panel B sorts the same banks into quintiles according to their rating revision between January 2007 and January 2009 ranging from the most upgraded (Q1) to the most downgraded banks (Q5) and reports again the quintile mean and median for bank size.

column (5). 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. This marginal effect amounts to an economically substantial rating favour resulting from bilateral business related to securitization.

A more benign interpretation for the significantly positive ASSB coefficient could be a correlated rating error between the bank rating and the ratings in structured products. Occasionally, an issuing bank provides credit enhancing guarantees so that credit risk of the structured product becomes correlated with the bank credit rating. An overoptimistic bank rating by any credit rating agency might thus jointly occur with a more favourable rating for the bank's structured products. In turn, this may generate more rating business for the agency if the issuing bank chooses agencies based on the best available rating. In order to control for this channel, we set to zero all business volume where the security has a guarantor. Such guarantees concern 3.7% of all securitization deals in our sample. The new variable *ASSB ex-guarantee* should be more robust to the reverse causality based on correlated errors between bank and securitization ratings. Table 5, column (6) shows that this alternative 'conflict of interest proxy' produces identical regression results. Explicit credit guarantees for some of the securitization volume do not explain why more favourable bank ratings correlate with agency-specific securitization business.

7.5. 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 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.

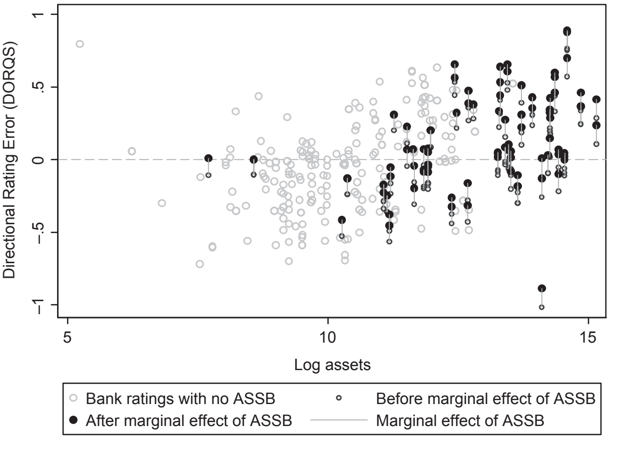

Figure 5. Marginal effect of agency-specific securitization issuance on directional rating error

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

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, but no directional rating bias. This suggests that bank asset complexity or insufficient disclosure represent important obstacles to rating quality.[16]

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.[17] 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 securitization business (*ASSB*). Even conditional on bank characteristics, bank size and bank securitization activity with a rating agency remain highly correlated with the rating bias.

7.6. 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, 2011). Based on the number of bank ratings generated by the three major rating agencies, we construct a Herfindahl–Hirschmann

[16] One representative of a rating agency highlighted the frequent lack of disaggregate data on bank assets as an important informational shortcoming which also extends to bank management.

[17] For evidence of bank trading profitability with respect to exchange rate volatility, see for example Hau (1998).

Table 6. Rating quality and additional bank characteristics

Dependent variable	Non-Directional error: TORQS (1)	(2)	(3)	Directional error: DORQS (4)	(5)	(6)
Size						
Log assets	0.014*	0.007	0.013*	0.046***	0.042***	0.046***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Securitization						
ASSB	−0.003**	−0.002*		0.004***	0.004***	
	(0.001)	(0.001)		(0.001)	(0.001)	
			−0.003**			0.004***
ASSB ex-guarantee			(0.001)			(0.001)
Agency dummies						
Moody's	−0.008	−0.003	−0.008	0.047*	0.047*	0.047*
	(0.027)	(0.027)	0.027	(0.026)	(0.026)	(0.026)
S&P	0.001	0.006	0.001	−0.089***	−0.089***	−0.089***
	(0.025)	(0.025)	0.025	(0.024)	(0.024)	(0.024)
Profitability						
RoA	0.000	−0.002	0.000	0.004	0.003	0.003
	(0.005)	(0.004)	(0.005)	(0.003)	(0.002)	(0.003)
Capital structure						
Leverage	0.009	−0.018	0.009	−0.019	−0.005	−0.019
	(0.071)	(0.070)	(0.071)	(0.043)	(0.042)	(0.043)
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	−4.322*	−5.261**	−4.323*	0.506	0.218	0.528
	(2.256)	(2.255)	(2.256)	(1.299)	(1.291)	(1.299)
Funding structure						
Short-term funding share	−0.012	0.039	−0.012	−0.072**	−0.058*	−0.072**
	(0.050)	(0.048)	(0.050)	(0.033)	(0.032)	(0.033)
Rating Competition						
Multiple rating dummy		0.001			−0.029***	
		(0.018)			(0.011)	
HH index		0.455			−0.145	
		(0.420)			(0.249)	
Av. serial correlation	0.762	0.763	0.761	0.859	0.859	0.859
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	No	Yes	Yes	No	No
No. of observations	15,426	15,426	15,426	15,426	15,426	15,426

Note: Reported are panel regressions with bank level random effects. The regression allows for serial AR(1) correlation of the error. 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); ASSB ex-guarantee = same as ASSB, excluding deals which are guaranteed either by the issuing bank or a third party; 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 all divided 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%, 5%, and 1% respectively.
Source: Authors' calculations.

index (*HH index*) of industry concentration, which shows decreasing industry concentration after 2000. A separate *Multiple rating dummy* captures cross-sectional variation in the number of ratings for individual banks.

In Table 6, columns (2) and (5), 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. Unlike the *HH index*, the *Multiple rating dummy* shows strong negative correlation with *DORQS*. The 73% of banks ratings classified as multiple ratings were less favourable (by on average three ranks for every 100 banks) than those for which only one rating was issued. If banks acquire ratings sequentially, then they should have stronger incentives to solicit additional ratings if the first rating is unfavourable. Such strategic 'shopping for better ratings' predicts that the multiple rating dummy should be associated with less average overrating. This prediction is borne out by the data. Importantly, inclusion of both control variables in Table 6 does not change the qualitative evidence discussed in the previous sections.

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 (2) and (5), for different measurement lags of $k = 0$, 4, 12 quarters. Interestingly, the coefficients for the rating bias related to bank size and securitization 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 securitization 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 effect 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 the likelihood of government support, which depends partly on bank size.

Table 7. Robustness check for different lags between EDF and credit rating

	Non-Directional error: TORQS			Directional error: DORQS		
Dependent variable	0	4	12	0	4	12
Lag (in quarters)	(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)
Securitization						
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	0.260***	0.064	0.069	0.111***	0.083**	0.069
	(0.070)	(0.066)	(0.077)	(0.036)	(0.038)	(0.045)
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: Reported are panel regressions with bank level random effects. The regressions allows for serial AR(1) correlation of the error. 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 all divided 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%, 5%, and 1% respectively.
Source: Authors' calculations.

We test the robustness of our results to the effect of government support by exploiting a different type of credit rating. Fitch and Moody's publish 'bank financial strength 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 improvement in creditworthiness due to

Table 8. Robustness check for the effect of government support

	Non-directional error: TORQS		Directional error: DORQS	
Dependent variable	(1)	(2)	(3)	(4)
Size				
Log assets	0.013	0.009	0.048***	0.044***
	(0.010)	(0.010)	(0.008)	(0.008)
Securitization				
ASSB	−0.002	−0.002	0.004**	0.004**
	(0.002)	(0.002)	(0.002)	(0.001)
Government support				
Rank difference: 'all-in' minus 'stand-alone'		0.181***		0.326***
		(0.039)		(0.024)
Profitability				
RoA	−0.008	−0.008	0.004	0.004
	(0.007)	(0.007)	(0.004)	(0.004)
Capital structure				
Leverage	−0.071	−0.076	0.009	0.022
	(0.094)	(0.094)	(0.058)	(0.057)
Asset structure				
Loans share	−0.002***	−0.002***	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)
Trading share	−6.424**	−5.831**	−1.516	−0.108
	(2.974)	(2.973)	(1.747)	(1.725)
Funding structure				
Short-term funding share	0.020	0.033	0.006	0.025
	(0.069)	(0.069)	(0.047)	(0.045)
Rating competition				
Multiple rating dummy	−0.019	−0.020	−0.012	−0.014
	(0.028)	(0.028)	(0.018)	(0.018)
HH index	−0.158	−0.021	−0.539	−0.098
	(0.723)	(0.723)	(0.450)	(0.444)
Av. serial correlation	0.758	0.782	0.855	0.855
Country fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	No	No	No	No
No. of observations	7,488	7,488	7,488	7,488

Note: Reported are panel regressions with bank level random effects. The regression allows for serial AR(1) correlation of the error. This table excludes S&P ratings, since S&P did not produce bank financial strength ratings until 2011. 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 all divided 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 %, 5 %, and 1 % respectively.

Source: Authors' calculations.

implicit government support.[18] We define the improvement from the 'stand-alone' to the 'all-in' credit rating rank (of the same bank) as *Rank difference*. The *Rank difference* variable is positive when a bank's stand-alone rating rank is better than its all-in rating rank. In general, we would expect that large banks are more likely to benefit from conditional government support, since large banks are more likely to be perceived as systemically important. We observe as such in our data: there is a positive correlation of the *Rank difference* variable with bank size, measured by Log *assets*, of 0.31.

Table 8 explores the alternative hypothesis that our finding of positive bias in favour of large banks might be caused by omission of the likelihood of government support. 'Stand-alone' ratings are available to us from Moody's and Fitch on a sub-sample of banks, such that the regressions reported in Table 8 are run on a reduced sample of 7,488 bank-rating observations, 49% of the full sample. Columns (1) and (3) repeat the analysis from Table 6, columns (2) and (5) for all Fitch and Moody's rated banks. Columns (2) and (4) use these same specifications plus the *Rank difference* variable. In the reduced sample of Moody's and Fitch ratings in specification (3), the Log *assets* variable shows a significant coefficient of 0.048 for rating error *DORQS*, compared with 0.042 in Table 6, column (5).

Including the *Rank difference* in specifications (2) and (4) slightly reduces the regression coefficient for Log *assets* to 0.013 and 0.044 respectively. But statistical and economic significance in column (4) remains high. The coefficient for *Rank difference* has the expected positive sign and is also statistically significant. We conclude for the Fitch and Moody's 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 meas-

[18] Stand-alone ratings use somewhat different notation to their all-in counterparts. Moody's rates banks A, A–, B+ B, B–, etc. through C and D. Fitch rates banks A, A/B, B, B/C etc. through C, D and F. For the purposes of the ranking, we judged, for example, that A– rated banks were relatively safer than A/B rated banks which were relatively safer than B– rated banks.

ure 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 rating agencies' conflict of interest, which increases with bank size. 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 and Moody's 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 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 securitization issuance between banks and the three major rating agencies over the period 1990–2012. We define a bank's *agency-specific securitization 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 evidence suggesting that conflicts of interest in the securitization 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 *et al.* (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, 2009). 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 exorbitant influence of rating agencies in the current system.

Discussion

Isabel Schnabel

Johannes Gutenberg University Mainz

In the recent financial crisis, credit rating agencies were heavily criticized for having failed to recognize the build-up of risks in the banking sector. This paper asks whether rating agencies at least predicted the *ordering* of banks' riskiness correctly, even if they failed to predict the crisis itself. To this end, the authors compare the ranking of banks' credit ratings by the three major rating agencies with the ranking of their expected default frequencies two years later. They also analyse which factors determine the size of rating errors (deviations in either direction) and rating biases (deviations in a particular direction) by developing ordinal measures of rating errors and biases, and by regressing these measures on bank- and country-specific variables.

The authors find that ratings poorly reflect the ordering of banks' creditworthiness, especially in non-crisis times and for investment grade banks. Moreover, rating errors and biases are found to be higher for large banks, and rating biases (but not errors) are higher for banks with a larger securitization volume at a given rating agency. Both results are interpreted as evidence of conflicts of interest in rating agencies.

Ordinal versus cardinal ratings

The paper's emphasis on the ordinal aspect of credit ratings is a welcome addition to the literature. As mentioned in the paper, rating agencies themselves do not seem to have a clear perception of whether ratings should be considered ordinal or cardinal.[19] In reality, they are likely to be both. For example, a rating of AAA does not only mean that a bank belongs to the group of best banks, but it also suggests high creditworthiness in an absolute sense. Standard & Poor's describe a company rated AAA as having an 'extremely strong capacity to meet financial commitments', which clearly is an absolute, rather than a relative statement about a firm's creditworthiness. If there were no absolute meaning to ratings, current banking regulation, which binds regulatory capital requirements to ratings, would be entirely inappropriate. While acknowledging that the relative (ordinal) aspect of ratings is important, absolute (cardinal) aspects cannot be fully discarded.

One interesting finding in the paper is the better ability of rating agencies to predict the ordering of banks in times of crisis and the ordering of below-investment-grade banks. While this result is surprising at first sight, it may partly be driven by the larger heterogeneity of banks in the respective time period or bank group, which makes it much easier to generate a reliable ranking. If banks are homogenous, small errors in the measurement of creditworthiness can translate into relatively large ranking deviations. It is not clear whether the same result would be found if a cardinal metric of rating errors were used.

Conflicts of interest versus bad judgment

One argument of the authors for focusing on ordinal ratings is that financial crises are hard to predict, whereas ordinal rankings are less affected by the occurrence of crises. While the first statement is certainly true, the second is not. A financial crisis does not affect all banks to the same degree. A bank's vulnerability to systemic risk (e.g., its interconnectedness or its degree of correlation with other banks) may be just as difficult to predict as the crisis itself. Similarly, regulatory interventions in a crisis do not

[19] Following the authors, we call ratings cardinal if they can be mapped into absolute levels of credit risk.

affect all banks in the same way and cannot easily be predicted. The different treatment of Bear Stearns and Lehman Brothers is a case in point. Hence, the ordering of banks is likely to change in a financial crisis, which may partly explain observed rating biases. Thus, the overrating of large banks could also be explained by an underestimation of large banks' exposures to systemic risk arising from their interconnectedness in interbank markets or their strong reliance on market funding. The overvaluation of banks with high volumes of securitization could have been due to an insufficient appreciation of the risks from this type of business. Even though the paper uses agency-specific securitization volumes, it does not control for overall securitization volumes. Therefore, it cannot isolate the effect of conflicts of interest. Hence, the analysis cannot identify whether the observed overvaluation of certain types of banks was a deliberate action of rating agencies due to conflicts of interest, or whether it was simply bad judgment. The paper strongly pushes the first interpretation, but the second one cannot be ruled out.

This problem arises because the employed measures of rating errors mix the measurement aspect of ratings (i.e., the relationship between ratings and *currently* expected default frequencies) and the ability of ratings to forecast future developments, which is inherently difficult in times of crises, not least due to potential structural breaks in the time series. Table 7 gives some insights into the implications of this distinction by considering different forecasting horizons. As expected, rating biases are significantly smaller if one considers a horizon of 0, but they are still there. A further investigation of this issue would be useful.

Overall, it seems difficult to draw clear policy conclusions from the analysis. The suitability of credit ratings for regulatory capital requirements would have to be analysed using a cardinal rating metric since the reliance on ratings in regulation can only be justified if ratings have some absolute meaning, that is, if they can be mapped into absolute levels of credit risk. Regarding rating agencies' conflicts of interest, the analysis does not clearly distinguish between deliberate biases and bad judgment. Nevertheless, the emphasis of the paper on the ordinal aspects of ratings and the distinction between rating errors and biases improve our understanding of ratings and will certainly stimulate further research on rating accuracy and the question whether regulation should continue to rely so strongly on rating agencies' judgments.

Thorsten Beck

CentER, Tilburg University

This paper shows that (i) credit ratings of banks are only able to predict expected default frequency during crisis but not normal times, (ii) bank ratings in the upper investment grade range are not able to predict bank fragility, (iii) large banks receive more favorable credit ratings relative to their expected default frequency, and (iv) banks with a large securitization business with a specific rating agency receive more favourable credit ratings from this agency relative to their expected default frequency.

It is important to stress that these last two results clearly reflect biases and not simply prediction errors. This is a very policy relevant and timely paper as we are still in the midst of the regulatory reform debate. The role of credit rating agencies, especially their regulatory role in determining risk weights, is subject to heavy debate and this paper adds to this policy discussion.

Let me first offer some comments on the findings. The results that there is no high correlation between credit ratings and EDF among investment grade banks and in non-crisis times does not seem surprising, for several reasons. First, as the authors point out themselves, bank failure is a rare event, especially in non-crisis times. And as the authors focus on relative ratings, the distance between contiguous risk categories will be less accurately measured in non-crisis times when the average risk of bank failure is miniscule. Second, shocks might come from unexpected corners unrelated to factors that make it into the risk ratings. Only after these shocks have been realized, will they be reflected in risk ratings and the latter be updated accordingly.

The policy recommendations of the authors make sense. While rating agencies and their regulatory privilege was for a long time seen as part of the market discipline agenda, their shortcomings, both in risk modelling *per se* and in their incentive structure, have become clear. Even if the incentive problems can be addressed, the lack of predictability of shocks remains a problem, so that reduced reliance on rating agencies is called for. More transparency in banks' financial statements, as advocated by the authors, can help broaden the number of informed observers.

A more general conclusion one can draw is that very sophisticated models of capturing risk and reflecting them in capital risk weights is the wrong approach, which might give the illusion of improved regulation without actually providing it, a point recently also made by Andrew Haldane. Focusing on a set of simpler rules including an incentive compatible bank resolution framework seems a more promising route to financial stability than ever more refined risk weights for capital requirements. And addressing especially the too-big-to-fail phenomenon head-on through living wills and special resolution regime for SIFIs can also help alleviate the rating bias reported by the authors.

Panel discussion

Fabiano Schivardi asked if the orthogonality between CR rankings and future EDF rankings could be due to regulatory institutions acting (specifically between years t and t+2) on the information provided by bank CR rankings today. Johannes Spinnewijn was surprised by the use of the 45 degree line as the benchmark in the analysis of the correlation between CR and future EDF rankings given that it ignores the role played by uncertainty in the discrepancy between the two. He maintained that it is never possible to perfectly correlate an *ex-ante* risk assessment with the *ex-post* realization.

Clemens Fuest noted that if rating agencies were abusing their powers and leading banks into bankruptcies, then one could argue the opposite of what the graphs of future EDF versus CR ratings are indicating. In the presence of this endogeneity, he thought that it would be better to focus on the contemporaneous relationship between the two sets of rankings.

Harald Hau agreed that a cardinal approach would yield more information. However, by adopting the minimalistic ordinal metric, they were erring on the side of caution. In particular, doing so results in more robust conclusions and not having to make parameter assumptions. Regarding structural breaks, Hau clarified that any shift in the distribution would be captured by the time fixed effects. On the other hand, according to Hau a problem arises if these breaks are in some way interacting with bank characteristics. Hau said that in this instance they could identify the breaks and look at subsamples, though he would not expect to find qualitatively different results. He admitted that obtaining evidence in support of breaks generating the effects would be troubling. David Marques-Ibanez agreed with Schnabel about forming conclusions on the basis of the 'conflicts of interest' argument. He added that the banks may have been securitizing more before the crisis which in turn could have made it more difficult for the agencies to assess the quality of the banks. With respect to size though, he claimed that the results are particularly robust. Summing up, he thought that the issue of endogeneity is not very relevant in their analysis given that they are concerned with forecasting.

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 (2011). 'Credit ratings accuracy and analyst incentives', *The American Economic Review: Papers and Proceedings*, 101(3), 120–24.

—— (2013). 'Ratings quality over the business cycle', *Journal of Financial Economics*, forthcoming.

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

—— (2011). '*Global systemically important banks: assessment methodology and the additional loss absorbency requirement*', Bank for International Settlements.

Becker, B. and T. Milbourn (2011). 'How did increased competition affect credit ratings?' *Journal of Financial Economics*, 101(3), 493–514.

Beltratti, A. and R.M. Stultz (2012). 'The credit crisis around the globe: why did some banks perform better?', *Journal of Financial Economics*, 105(1), 1–17.

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

Berger, A. and C. Bouwman (2013). 'How does capital affect bank performance during financial crises?', *Journal of Financial Economics*, forthcoming.

Bharath, S.T. and T. Shumway (2008). 'Forecasting default with the Merton distance to default model', *The Review of Financial Studies*, 21(3), 1339–69.

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 (2012). 'Tiebreaker: certification and multiple credit ratings', *Journal of Finance*, 67(1), 113–52.

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* (pp. 58–90), Oxford University Press, Oxford.

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 Economist's Voice*, 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*, 97, 470–87.

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*, 45, 5–29.

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.

Duffie, D. (2010), *How Big Banks Fail and What to Do about It*, Princeton University Press, Princeton, NJ.

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', SFI Working Paper.

Efing, M. and D. Marqués-Ibáñez (2012). '*Structured debt ratings: evidence on conflicts of interest*', University of Geneva, 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), 3735–77.

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.

Griffin, J. and D. Tang (2011). 'Did credit rating agencies make unbiased assumptions on CDOs?' *American Economic Review: Papers and Proceedings*, 101(3), 125–30.

Hau, H. (1998). 'Competitive entry and endogenous risk in the foreign exchange market', *Review of Financial Studies*, 11(4), 757–87.

He, J., J. Qian and P.E. Strahan (2011). 'Credit ratings and the evolution of the mortgage-backed securities market', *American Economic Review: Papers and Proceedings*, 101(3), 131–35.

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), 644–63.

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 approach', in J.G. Kemeny and L.J. Snell (eds.), *Mathematical Models in the Social Sciences* (pp. 9–23), Ginn, New York.

Kiff, J., M. Kisser and L. Schumacher (2013). '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 (2012). '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), 449–70.
Moody's Investors Services (1999). 'Rating methodology: the evolving meanings of Moody's bond ratings'.
Moody's (2006). 'Moody's credit rating prediction model', mimeo.
Morgan, D. (2002). 'Rating banks: risk and uncertainty in an opaque industry', *The American Economic Review*, 92, 874–88.
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.
Opp, C., M. Marcus and M. Harris (2013). 'Rating agencies in the face of regulation', *Journal of Financial Economics*, forthcoming.
Pagano, M. and P. Volpin (2010). 'Credit ratings failures and policy options', *Economic Policy*, 62, 401–31.
Partnoy, F. (2002). 'The paradox of credit ratings', in R. Levich, C. Reinhart and G. Majnoni (eds.), *Ratings, Rating Agencies, and the Global Financial System* (pp. 65–84), Kluwer, Boston.
—— (2006). 'How and why credit rating agencies are not like other gatekeepers', in Y. Fuchita and R. Litan (eds.), *Financial Gatekeepers: Can They Protect Investors?* (pp. 59–102), Brookings Institution Press and Nomura Institute of Capital Markets Research.
Reinhart, C. and K. Rogoff (2009). 'The aftermath of financial crises', *American Economic Review*, 99(2), 466–72.
Skreta, V. and L. Veldkamp (2009). 'Ratings shopping and asset complexity: a theory of ratings inflation', *Journal of Monetary Economics*, 56, 678–95.
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.