Determinants of heterogeneity in European credit ratings

**Kurt Hornik · Rainer Jankowitsch ·
Manuel Lingo · Stefan Pichler · Gerhard Winkler**

Abstract In this study, we empirically analyze the determinants of heterogeneity in rating assessments across different segments of the European loan market. We conduct a benchmarking analysis using rating information on European corporate obligors from nine major Austrian banks that have a large share of foreign lending, particularly in the Central and Eastern European region. We provide evidence that, generally, overall heterogeneity among rating outcomes for foreign markets is higher than for domestic markets. Furthermore, we show that heterogeneity increases in transition economies and those markets where Austrian bank involvement is relatively low. Our evidence supports the hypothesis that heterogeneity in the assessment of credit risk is determined not only by the objective quality of information, which is deemed to be lower in transition economies, but also by the subjective access to information about obligors measured by the level of domestic bank involvement in the respective foreign market. Furthermore, we quantify potential effects on regulatory capital requirements.

K. Hornik
Institute for Statistics and Mathematics, WU (Vienna University of Economics and Business),
Augasse 2-6, 1090 Vienna, Austria

R. Jankowitsch (✉) · M. Lingo · S. Pichler · G. Winkler
Institute for Finance, Banking and Insurance, WU (Vienna University of Economics and Business),
Heiligenstädter Straße 46-48, 1190 Vienna, Austria
e-mail: rainer.jankowitsch@wu-wien.ac.at

S. Pichler
Vienna Graduate School of Finance, Vienna, Austria

M. Lingo · G. Winkler
Oesterreichische Nationalbank, Otto Wagner Platz 3, 1090 Vienna, Austria

1 Introduction

In this study, we analyze the determinants of heterogeneity in rating assessments across different segments of the European loan market. We aim at gaining information about the level of integration of European financial markets with respect to the availability, as well as the quality, of information essential for credit rating assessments. Following the literature on financial market integration (see, e.g., IMF 2000; Standard&Poor's 2003; European Central Bank 2004a), we generally expect more similar rating outcomes of obligors in the domestic market than in the foreign markets. Further, we expect that heterogeneity of rating outcomes is higher in transition economies compared to nontransition economies. Finally, we expect that heterogeneity in ratings is lower for those foreign markets that are more integrated with the domestic market than it is for their counterparts. This is particularly interesting in the context of the current efforts to improve European financial market integration (e.g., by means of the Financial Services Action Plan; see, e.g., European Commission 2005a, 2005b). In our empirical study, we use a unique data set stemming from nine internationally active Austrian banks. The data set is particularly useful because of the high exposure of Austrian banks, both in relative and absolute terms, to foreign markets, especially in Central and Eastern European transition countries.

Analyzing rating heterogeneity is of particular interest for banks and their supervisors because reliable assessment of the creditworthiness of obligors is an important precondition for the stability of a financial system. Inadequately high exposure to credit risk has been one of the leading causes of problems in financial institutions worldwide for many decades (Basel Committee on Banking Supervision 2000, 2004). As a consequence, analyzing the heterogeneity of bank ratings across different obligor groups has also become more important to academics (Carey 2001; Jacobson et al. 2006). Our paper offers interesting insights into this topic, as we quantify the heterogeneity for our sample and present an indication of the size of potential surcharges to capital requirements for groups with dissimilar ratings.

From a bank's point of view, analyzing the heterogeneity between its own internal ratings and those of its competitors is important as errors and inconsistencies in ratings may have substantial effects on pricing decisions, capital allocation, and business strategies such as starting a lending business in a new market segment. Thus, high degrees of dissimilarity ultimately influence the bank's competitive position (Carey 2001). As a consequence, banks have a natural interest in analyzing the heterogeneity of their ratings to ensure reliable credit assessments of their obligors and avoid the effects of adverse selection (e.g., Jankowitsch et al. 2007; Jordão and Stein 2003).

A regulatory supervisor will also have an interest in the degree of heterogeneity of ratings of certain obligors across banks and whether different types of rating systems

produce similar results (Jacobson et al. 2006). From the supervisor's perspective, investigating the consistency of rating information on an aggregate (macro) level allows an analysis of the extent to which individual banks in a financial market agree in their rating assessments of certain obligor groups. Conversely, such an analysis will aid the supervisor in identifying potential problem areas and hence risk clusters, i.e., obligors for which the financial system has no agreed opinion as regards their risk and thus fails to assess them consistently.

To date, there is not a great deal of empirical work on rating heterogeneity (see, e.g., Carey and Hrycay 2001; Carey 2001; Güttler 2005; Krämer and Güttler 2006; Jacobson et al. 2006). Carey (2001) provides evidence on the consistency of US banks' internal ratings. He conducts a comparison of ratings across lenders based on the Loan Pricing Corporation's Loan Loss Database. In a more recent paper, Jacobson et al. (2006) investigate the consistency of two large Swedish banks' estimates for credit risk parameters, in the process of which they also analyze the similarity of the banks' ratings and find substantial differences between them.

Most existing studies in this field measure the heterogeneity of rating outcomes by means of benchmarking. Generally, benchmarking can be defined as the comparison of ratings of certain obligors by one specific credit quality assessment source with the ratings of the same obligors by other sources. Thus, benchmarking methods make use of available multi-rater information, i.e., rating assignments for identical obligors made by different rating sources. The studies by Carey (2001) and Jacobson et al. (2006) focus on a descriptive analysis of the frequencies of rating deviations across lenders, which allows for only a qualitative assessment of rating heterogeneity. We extend this approach by employing a new methodological framework for benchmarking suggested by Hornik et al. (2007). This framework is based on appropriate measures of rating agreement and rating association and thus permits a meaningful quantification of rating heterogeneity, as well as statistical inference.

Theoretically, differences in the performance of rating systems across different obligor groups could be analyzed by means of backtesting (see, e.g., Krämer and Güttler 2006; Ryser and Denzler 2009). However, backtesting is often not possible due to lack of data on consistently created time-series of rating and default information.

This paper is organized as follows. In Sect. 2, we develop the methodology and formally describe the use of several measures of proximity between different rating systems. Section 3 contains a description of the data set, the rationale for our hypotheses on the expected similarities, and a presentation of the empirical results. Section 4 concludes.

2 A framework for benchmarking credit ratings

Using multi-rater panels for benchmarking rating systems makes it possible to quantify the similarity or dissimilarity, i.e., the *proximity*, between the raters (see Everitt and Rabe-Hesketh 1997). Hornik et al. (2007) show that there are several aspects of proximity that must be taken into account, most notably *agreement* and *association*. *Agreement* measures the raters' concordancy and discordancy for single objects

(obligors), whereas *association* measures the extent to which rating orders between pairs of obligors are preserved between raters. These general aspects of proximity are analyzed in other studies. For example, Carey (2001) provides descriptive statistics about the distribution of rating deviations among the internal ratings made by more than 20 US banks regarding a large corporate segment. Based on these statistics, he draws conclusions about agreement. Jacobson et al. (2006) compare the rating outcomes made by two Swedish banks for more than 17,000 common borrowers. Given two rating systems, each with a different number of rating classes, the authors cannot analyze agreement, but they do compute Spearman's rank correlation to assess the association between the two rating systems. This measure of association, however, is not consistent with important axiomatic requirements (see below) and should not be used to compare ordinal rankings with many ties (see Hornik et al. 2007).

Before we introduce appropriate measures for agreement and association, we note that ratings are typically provided on an ordinal scale (e.g., $AAA > AA+ > \cdots > D$). When measuring the proximity of ratings made by two banks, we look at their *co-ratings*, i.e., the obligors jointly rated. Given the discreteness of the ratings, one natural starting point is the contingency table $C = [c_{ij}]$ with the cross-tabulation of the co-ratings, i.e., the counts c_{ij} of obligors rated into classes i and j by the first and second rater, respectively. Suppose that both raters use the same structure of rating classes. Obviously, the ratings are in perfect agreement if all off-diagonal entries in the contingency table are zero. Otherwise, one can employ measures known as κ or, more generally, weighted κ, to assess the degree of agreement. These measures generally adjust for "agreement by chance," and, through the appropriate choice of weights w_{ij}, can quantify the closeness of agreement for co-ratings into classes i and j. For example, the same weights w_k could be used for co-ratings that are k classes apart.

Formally, the weighted κ measures are obtained as follows. Let $N_c = \sum_{i,j} c_{i,j}$ be the number of obligors rated by both raters, and $p_{i,j} = c_{i,j}/N_c$ be the joint proportion of obligors rated into classes i and j, respectively. Also, use $p_{i\cdot} = \sum_j p_{i,j}$ and $p_{\cdot j} = \sum_i p_{i,j}$ for the marginal proportions of obligors rated into class i by the first rater and into class j by the second rater, respectively. Then, weighted κ is defined as

$$\kappa_w = \frac{P_{o(w)} - P_{e(w)}}{1 - P_{e(w)}} \tag{1}$$

and compares the observed agreement $P_{o(w)} = \sum_{j=1}^{R} \sum_{i=1}^{R} w_{ij} p_{ij}$ to that expected if the ratings were independent (and hence $p_{ij} = p_{i\cdot}p_{\cdot j}$), given by $P_{e(w)} = \sum_{j=1}^{R} \sum_{i=1}^{R} w_{ij} p_{i\cdot}p_{\cdot j}$, where in general the weights w_{ij} satisfy $0 \le w_{ij} \le 1$ with $w_{ii} = 1$ and $w_{ij} = w_{ji}$, and R is the (common) number of rating classes employed. Hornik et al. (2007) propose to use the Cohen–Fleiss weights $w_{ij} = 1 - [(i - j)/(R - 1)]^2$, which measure disagreement as quadratic in the "difference" in rating classes.

Ratings on an ordinal scale naturally result in weak order relations (i.e., including ties) between the obligors rated. Association measures quantify the extent to which these order relations are in accordance. Kemeny and Snell (1962) show that there is a unique association (or distance) measure between order relations satisfying several natural axioms (basically, metricity, transitivity, and "betweenness"). We refer to this

measure as τ_x, by which notation we indicate that this measure is an extension of the well-known Kendall's τ that handles the case of ties more appropriately (see also Emond and Mason 2002). Note that other measures of rank correlation, such as Spearman's rank correlation or Somer's d, as suggested by Jacobson et al. (2006) and the Basel Committee of Banking Supervision (see Basel Committee on Banking Supervision 2005), respectively, do not fulfill all these axioms; in particular, they do not meet transitivity requirements. The Kemeny–Snell τ_x is based on the comparison of pairs of co-rated obligors: the association of ratings is related to the number of pairs for which the ordering is preserved between raters. Formally, if a_{uv} is the indicator of obligor u being rated higher than obligor v for the first rater, and b_{uv} is defined similarly for the second rater, then

$$\tau_x = 1 - \frac{\sum_{u=1}^{N_c} \sum_{v=1}^{N_c} |a_{uv} - b_{uv}|}{N_c(N_c - 1)}. \tag{2}$$

(Equivalently, τ_x corresponds to the "natural" symmetric difference metric between the order relations of the co-rated obligors induced by the ratings.) Note that the definition of τ_x does not assume that the ratings use the same ordinal scale.

The measures τ_x and κ introduced above are readily interpretable ways of quantifying the proximity of the co-ratings of obligors by two banks. As τ_x relies only on the ordinal ranking of obligors, it is related to the discriminatory power, whereas κ is related to the calibration quality of the rating system (see Basel Committee on Banking Supervision 2005; Oesterreichische Nationalbank 2004b).

When ratings from several banks are available, the overall structure of the ratings can be analyzed based on the pairwise proximities. Since there is no consistent method available for computing an aggregated "consensus" rating per obligor based on the observed ratings, all our analyses are based on pairwise comparisons rather than on comparisons to a consensus rating. In particular, we compute the average τ_x and the average κ for each bank with respect to all the other banks for all obligors and for specific submarkets (according to transitionality and/or financial market integration). Examining the average τ_x and κ over all banks in the whole data set and in the specific submarkets allows us to perform market-level analyses of the proximities of rating systems.

To check the robustness of our results we also present descriptive statistics of the distribution of rating deviations on a pooled level, similar to the procedure of Carey (2001). For this purpose, we simply count the number of pairwise rating assignments that fully agree (zero deviation), the assignments that deviate by exactly one rating class (regardless of the direction of deviation), those that deviate by two rating classes, and so forth. This distribution of rating deviations provides a quick intuitive overview of the structure of rating heterogeneity.

3 Heterogeneity in European credit ratings: empirical results

3.1 Data description

For our empirical analysis we use a unique data set on rating information derived from the Austrian Credit Register, which is operated by the Austrian National Bank

(for a review of the use of credit registers for Basel II purposes, see Basel Committee on Banking Supervision 2004). Under the Austrian Banking Act, all Austrian financial institutions are required to supply monthly information to this register regarding any changes in the status of their outstanding credits as well as information on new loans granted during the period. The Austrian Credit Register covers all major direct lending activities of Austrian financial institutions and thus contains information on all loans above a threshold of EUR 350,000. Approximately 67% of the total volume of all Austrian bank (direct) lending activities were included in credit register entries, totaling EUR 373 bn, at the end of 2004 (Oesterreichische Nationalbank 2004a). Each record of the register includes general information on each credit claim (e.g., economic situation, legal form, country) and the respective bank's own internal assessment of the individual obligor's creditworthiness as expressed by a rating. This information on credit risk, which has been collected since the beginning of 2003, makes the Austrian Credit Register a unique data source.

In this study, we use credit ratings of individual obligors, which are measured on a harmonized master scale. Based on very detailed information about the bank-specific rating systems (i.e., definition of default, time horizon, conditioning, point-in-time vs. through-the-cycle), the Austrian National Bank consistently maps the original rating information onto the master scale, which consists of eight buckets: four for investment-grade entities (AAA–BBB) and four for non-investment-grade obligors (BB–D). This master scale is very similar to the rating scales of the major rating agencies and its granularity has been optimized in order to minimize loss of information from the banks' individual rating scales.

For this study, we use a data set of rating information on European corporate obligors for the year 2005. Since our benchmarking methods are based on co-ratings of identical obligors by different banks, we need a certain minimum number of co-ratings. Thus, we select a sample based on the credit register reports made by nine large Austrian banks. Together, these banks' major loans on average account for 57% of the total volume of all major loans granted by banks and they comprise 69% of all pan-European foreign direct lending activities reported to the credit register by Austrian banks in 2005. These statistics suggest that the sample banks and their loan portfolios are highly representative of the Austrian banking industry. We exclude obligors rated by only a single bank and certain others so as to obtain a homogeneous sample for which there are ratings by external rating agencies. This yields a sample of 7,463 ratings on 2,858 obligors to be used in our benchmarking analyses.

In this data set we can distinguish between domestic and foreign obligors. We further differentiate between obligors from foreign markets in which Austrian banks have a strong presence (high-involvement countries) and obligors from markets in which Austrian banks play a minor role (low-involvement countries). This distinction is based on relative concentration. Using all the reports on European obligors to the credit register, we first compute the total exposure, defined as the aggregate outstanding amount of all types of credit based on book values, for each country. Then we rank the countries according to these total exposures. High-involvement countries are defined as those with the highest exposures of domestic banks whose cumulative share reaches 50% of all foreign lending. Note that we need a bank-independent criterion for the definition of high involvement since we want to compare the heterogeneity of the ratings of *all* banks on a certain submarket. These high-involvement

countries are: Croatia, Germany, Hungary, Italy, the Netherlands, and Poland. The low-involvement countries are: Albania, Andorra, Bosnia and Herzegovina, Belarus, Belgium, Bulgaria, the Czech Republic, Cyprus, Denmark, Estonia, Latvia, Lithuania, Finland, France, Great Britain, Greece, Ireland, Iceland, Liechtenstein, Luxembourg, Macedonia, Malta, Monaco, Norway, Portugal, Russia, Serbia and Montenegro, Slovakia, Slovenia, Spain, Sweden, Switzerland, and Turkey.

We further distinguish between foreign obligors residing in transition countries and foreign obligors located in nontransition countries. Following the extensive literature on transition economies, we define a transition country as one that has switched from a centrally planned socialist regime to a market economy (e.g., Garibaldi et al. 2001; IMF 2000). According to this definition, the transition countries are: Albania, Bosnia and Herzegovina, Belarus, Bulgaria, Croatia, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Macedonia, Poland, Russia, Serbia and Montenegro, Slovakia, and Slovenia; every other European country is in the group of nontransition countries.

In a last step, we split these samples—differentiating between high-involvement nontransition, high-involvement transition, low-involvement nontransition, and low-involvement transition countries—to study joint effects.

Table 1 provides descriptive statistics for the entire data set and for the various subsamples described above.

Note that on the overall level, even the smallest bank has at least 102 obligors in common with one or more other institutions. Furthermore, we infer from the number of foreign obligors and the respective ratings, that our sample is well suited for analyses of proximity in different submarkets of obligors. The nine banks in our sample are very active at the pan-European level, with a strong presence in the Central and

Table 1 Description of the data set. The first column shows the total number of ratings per obligor group. The second column shows the total number of obligors for each group. Columns 3, 4, and 5 show the bank-specific minimum, median, and maximum numbers of obligors with at least one co-rating

Obligor Group	Total Number of Ratings	Total Number of Obligors	Number of Obligors per Bank for which at Least One Co-Rating Exists		
			Minimum	Median	Maximum
Overall	7463	2858	102	647	2183
Domestic Market	5899	2230	46	509	1761
Foreign Market	1564	628	36	175	344
High Involvement	725	293	11	78	162
Low Involvement	839	335	25	87	182
Nontransition Economies	922	377	16	96	213
Transition Economies	642	251	20	61	131
Nontransition & High Inv.	511	209	5	61	120
Transition & High Inv.	214	84	6	19	42
Nontransition & Low Inv.	411	168	11	35	93
Transition & Low Inv.	428	167	14	42	89

Eastern European transition countries. Three of these countries, Croatia, Hungary, and Poland, also fall into the high-involvement category.

We also compute the number of co-ratings at the obligor level. For the entire sample, the minimum and maximum numbers of co-ratings are 2 and 9, respectively. The mean and median numbers of co-ratings are 2.71 and 2, respectively, indicating that most corporate customers have business relations with only a small number of banks. These values do not vary much across the different subsets.

3.2 Hypotheses

Using the framework presented in Sect. 2, we measure the aggregate proximity among all banks on market-wide levels. These analyses provide us with information about the determinants of rating heterogeneity. First, according to the literature on financial market integration, we can expect more similar rating outcomes of domestic market obligors than for foreign market obligors. This hypothesis about aggregate proximity is motivated by the fact that credit markets are characterized by informational asymmetries between the loan-seeking companies, which often do not want to disclose too much information about their undertakings (e.g., Apreda 2002; Moser and Venkataraman 1996), and the loan-granting banks, which are trying to earn adequate interest on the capital invested while protecting themselves from the risk of default. This information asymmetry should be lower for domestic obligors than for foreign obligors as still existing differences in national accounting standards, definitions, and regulations make cross-border deals more difficult and complicate the assessment of a foreign obligor's creditworthiness (Standard&Poor's 2003). In addition to these differences in the *objective* quality of information, there are differences in the *subjective* access to soft information, meaning knowledge not necessarily based on facts and numbers, but instead having to do with things like quality of management or business outlook. Obviously, banks will have more of this type of information for domestic obligors than they will for foreign market obligors.

In the following, we focus on the determinants of rating heterogeneity across foreign markets. Again, we analyze the influence of the objective quality of information and the subjective access to information about obligors. The availability and objective quality of accounting information needed as an input for rating systems is expected to be most relevant determinant of the objective quality of rating inputs (Krahnen and Weber 2001). One factor that is very likely to explain differences in the quality of available accounting data is the type of economy the obligor is operating in, i.e., whether or not the economy is in a state of transition (e.g., Garibaldi et al. 2001; IMF 2000). In addition to macroeconomic stabilization policies, microeconomic elements are equally crucial for a successful transition to a market economy. Sound and effective governance structures can be decisive in attracting portfolio investment (LaPorta et al. 1999; Garibaldi et al. 2001; Andres et al. 2007) and advancing domestic growth and prosperity (e.g., IMF 2000). In an ongoing process of massive deregulation and liberalization, as well as privatization of formerly state-owned industries, disclosure rules and accounting standards may not be as well established as in other economies and such an environment can make it difficult for banks to accurately assess the risk profile of a company located in

a transition country (European Commission 2004). Thus, we expect that the overall similarity of rating outcomes will be lower in transition economies compared to non-transition economies. We examine this relationship by splitting our foreign sample into transition and nontransition economies.

The degree of heterogeneity between rating outcomes is also likely to depend on the access to the soft facts relevant as inputs in a rating system. We presume that access to subjective information is easier in foreign markets that are highly integrated with the domestic market than in markets with a low degree of integration. Soft facts about obligors may either be more easily accessible when the obligors operate in highly integrated markets, or banks may simply expend effort to obtain this sort of information in a highly-integrated market, whereas they might rely more on purely quantitative (accounting) information in less integrated markets. As a consequence, we expect a higher degree of similarity in ratings for obligors operating in foreign markets that are more integrated with the Austrian market than for their counterparts. To test this relationship, we separate the foreign markets into two submarkets based on their level of integration with the Austrian market. We use the involvement of Austrian banks, measured as shares of their total exposures, as a proxy for financial market integration.

This analysis has the additional benefit of providing us with information about the extent to which European financial markets are fully integrated in regard to the availability as well as the quality of information essential for credit assessments (for traditional means of measuring financial market integration, see, e.g., European Central Bank 2004a, 2004b). Since for Austrian banks, Central and Eastern European markets are key, we are in the fortunate situation of being able to study markets that are not only located in transition economies, but also are highly integrated with the Austrian market. Examining different submarkets defined with respect to *combinations* of their properties enables us to answer the question of whether financial market integration or transitionality plays the more dominant role in explaining heterogeneity of rating outcomes.

3.3 Benchmarking analysis

In this section, we provide results on rating proximity within different European submarkets. The most basic separation is between domestic and foreign markets. As described in Sect. 3.2, we further divide obligors into other submarkets based on type of economy and degree of financial market integration. Along the lines of Carey (2001), we first compute the relative frequency of rating deviations among all co-ratings within a specific submarket. Table 2 shows these rating deviations for the different submarkets, as well as for combinations of these submarkets.

Interpretation of the descriptive statistics presented in Table 2 is straightforward. If all rating systems had perfectly similar outcomes, we would observe 100% in the column for a rating deviation of zero. A higher share of observations for deviations of one or more rating classes (and a higher share of greater deviations) indicates more heterogeneous rating systems. Comparing domestic and foreign markets, we find that rating heterogeneity is lower in the domestic market. In the domestic market, rating deviations of zero or one rating grade are more frequent than in the foreign market,

Table 2 Marginal frequencies of rating deviations for different groups of obligors. Note that the maximum rating disagreement observed in the sample is six rating classes, although seven rating classes is the theoretical maximum deviation for our eight-grade master scale

Obligor Group	Rating Deviations						
	0	1	2	3	4	5	6
All Obligors	41.63	41.87	11.99	3.71	0.70	0.07	0.02
Domestic Market	41.45	43.11	11.24	3.43	0.69	0.10	0.00
Foreign Market	40.99	39.27	15.13	4.06	0.48	0.00	0.07
Nontransition	44.87	36.92	12.35	5.13	0.61	0.00	0.12
Transition	36.01	42.30	18.71	2.67	0.31	0.00	0.00
High Involvement	47.10	40.42	9.66	2.67	0.15	0.00	0.00
Low Involvement	35.72	38.28	19.85	5.25	0.77	0.00	0.13
Transition & High Inv.	42.33	41.86	13.49	2.33	0.00	0.00	0.00
Transition & Low Inv.	32.78	42.52	21.38	2.85	0.48	0.00	0.00
Nontransition & High Inv.	49.34	39.74	7.86	2.84	0.22	0.00	0.00
Nontransition & Low Inv.	39.17	33.33	18.06	8.06	1.11	0.00	0.28

whereas deviations of two or three grades are less frequent. When we compare the submarkets of the foreign market, we find support for our hypothesis that the average deviation is higher for obligors from transition economies than it is for obligors from nontransition economies. The comparison of obligors operating in low-involvement economies to those operating in high-involvement economies also supports our hypothesis that rating heterogeneity is lower for markets with a high level of integration with the domestic market. These conclusions can also be made for submarkets defined by combining type of economy with level of involvement.

Interestingly, these overall results are very similar in magnitude to Carey's (2001) findings. In his study, Carey also uses rating information mapped onto a scale that is very similar to the scales used by external rating agencies. His master scale is akin to the one employed by the Austrian Central Bank in terms of the number of rating grades and the level of risk associated with each rating grade. Carey's (2001) results are thus more or less directly comparable to our results on all obligors of Austrian banks, as shown in the first row of Table 2. Carey finds relative frequencies of 44, 39, 12.1, 3.3, and 1.6% for a rating disagreement of 0, 1, 2, 3, and 4 or more rating classes, respectively. We thus conclude that the overall degree of rating heterogeneity across Austrian banks is similar to the rating heterogeneity across US banks.

These descriptive statistics of the distributional characteristics of rating deviations permit an important first interpretation of the results. However, they do not allow for a more detailed analysis of rating heterogeneity since it is not possible to disentangle effects of agreement and association. Consider a case where many deviations of one or more rating classes are observed. This clearly indicates a low degree of agreement, but it is still possible that the association between the rating outcomes (the structure of relative orderings) is reasonably good. On the other hand, in a case with 50% deviations of zero and 50% deviations of one rating class, we can conclude that there is a reasonably high degree of agreement. Association, however, could be very poor.

Table 3 Comparison of different groups of obligors based on association (τ_x) and agreement (κ). The first and second columns show the groups' overall level of association (τ_x), whereas the fourth and fifth columns show the groups' overall level of agreement (κ). The third and sixth columns show the exact p-values of conditional tests based on Wilcoxon's signed rank statistic, indicating the significance of differences between the obligor groups

Obligor Group	Comparison Based on τ_x			Comparison Based on κ		
	Average 1st group	Average 2nd group	p-value	Average 1st group	Average 2nd group	p-value
Domestic Market vs. Foreign Market	0.5116	0.4300	**0.0050**	0.6069	0.5122	**0.0144**
Nontransition Economies vs. Transition Economies	0.4456	0.3550	**0.0305**	0.5112	0.3331	**0.0028**
High Involvement vs. Low Involvement	0.4612	0.3841	**0.1349**	0.5808	0.4144	**0.0010**
Nontransition & High Inv. vs. Transition & High Inv.	0.4537	0.3753	**0.6021**	0.6005	0.3997	**0.0324**
Nontransition & Low Inv. vs. Transition & Low Inv.	0.4144	0.2893	**0.0361**	0.4146	0.2389	**0.0098**
Nontransition & High Inv. vs. Nontransition & Low Inv.	0.4537	0.4144	**0.4091**	0.6005	0.4146	**0.0102**
Transition & High Inv. vs. Transition & Low Inv.	0.3753	0.2893	**0.0609**	0.3997	0.2389	**0.0142**
Nontransition & High Inv. vs. Transition & Low Inv.	0.4537	0.2893	**0.0018**	0.6005	0.2389	**0.0000**
Nontransition & Low Inv. vs. Transition & High Inv.	0.3753	0.4144	**0.9707**	0.3997	0.4146	**0.8256**

Therefore, in a more formal analysis we utilize the measures of rating proximity presented in Sect. 2. The measurement of agreement and association is based on the bank-specific bivariate comparisons between all possible combinations of banks for each submarket. We then use the overall averages of the bivariate measures of each individual bank for each submarket. Additionally, we test for significance of the differences in the levels of similarity in different submarkets employing a conditional test based on Wilcoxon's rank sum statistic (see Good 2000; Ernst 2004).

Table 3 shows the overall average levels of association (τ_x) and agreement (κ) for each comparison of different groups of obligors as well as the corresponding exact conditional p-values.

As expected, the overall levels of agreement and association are higher in the domestic market than in the foreign market. This difference is also statistically significant, indicating that the calibration of a sound rating system is more easily achieved for the domestic market. With respect to the type of economy, we obtain an average association (τ_x) of 0.45 for nontransition economies in contrast to 0.36 for transition economies. A similar result can be observed for agreement (κ), where we observe values of 0.51 vs. 0.33, respectively. These differences are statistically significant, indicating that the rating systems used in transition economies potentially are of lower quality and that differences in the availability and the objective quality of information are an important determinant of rating heterogeneity.

For financial market integration, we compare submarkets with high to submarkets with low involvement of domestic banks, observing a difference in association (τ_x) of 0.46 vs. 0.38, which is not statistically significant, whereas for agreement, we find a difference of 0.58 vs. 0.41, which is significant. These results support the hypothesis that access to subjective information, which is expected to better in markets with high involvement, plays an important role in explaining rating heterogeneity.

Using combinations of the involvement and transition attributes allows a more detailed analysis. The rating heterogeneity of nontransition markets having a high involvement of Austrian banks (i.e., Germany, Italy, and the Netherlands) is significantly lower than for the other submarkets, and very similar to that of the domestic market. Thus, we find that rating heterogeneity is lower in markets where both the objective quality of information and access to subjective rating information are better. On the other hand, we find a significantly higher degree of rating heterogeneity for transition markets with low involvement of domestic banks. Again, this supports the hypothesis that the quality of and the access to relevant rating information is the major determinant of rating heterogeneity.

Finally, we examine the question concerning which potential source of rating heterogeneity has the stronger effect. Comparing nontransition markets with low involvement (better objective quality of information but worse access to subjective soft facts) to transition markets with high involvement of Austrian banks (worse objective quality of information but better access to subjective soft facts), we cannot observe a significant difference for either measure of proximity. We conclude that for transition markets with a high degree of integration with the domestic market (as in Croatia, Hungary, and Poland for Austria), the potential lack of quality-accounting information is outweighed by better access to subjective soft facts.

Our results have important practical implications for banks and their supervisors. The quality of rating systems is heavily dependent on the information available in a specific foreign market. Reliable outcomes of a rating system can be expected in nontransition markets with a high level of market integration. More heterogeneous and thus less reliable results are to be expected in markets that are either in transition or have a low level of market integration. Special care must be taken with respect to rating systems operating in transition markets with a low level of integration with the domestic market.

3.4 Regulatory capital requirements

In this section, we translate the results of our benchmarking analysis into potential effects on regulatory capital requirements. For this analysis, we use a simulation framework that relies on the setup presented in Jankowitsch et al. (2007). In principle, low-quality rating systems have two important negative effects on a bank's financial stability. First, return on the portfolio will be lower than expected as rating errors have asymmetric consequences on loan volumes depending on the price elasticity of the borrowers (see, e.g., Broecker 1990). If a borrower receives too favorable a rating, he or she will most likely accept the loan offer, whereas rejection is more likely in the case of a rating error in the opposite direction, i.e., the bank will suffer from adverse selection. Therefore, rating errors have a negative impact on the portfolio return, as the expected loss is higher than implied by the ratings. Second, due to

the adverse selection effect the calculated regulatory capital held by the bank is too low. As more borrowers with too favorable ratings are in the portfolio, the resulting regulatory capital based on these ratings may be inadequate.

Jankowitsch et al. (2007) show that capital requirements estimated with a low-quality rating system can be more than two percentage points too low compared to the actual risk. This result is driven by the magnitude of the *rating error*. In their framework, the rating error is defined as a white-noise error term that is added to the logit-transformation of an obligor's default probability. This error term is normally distributed with mean zero and a standard deviation σ. Thus, the parameter σ drives the magnitude of potential estimation errors in the PD and in the rating.

We use this framework to estimate potential effects on banks' capital holdings. We calibrate the parameter σ such that a certain κ-value for banks with identical σ results, assuming that the portfolio the banks rate resembles the portfolio of a typical bank in our sample. We calibrate the parameter σ using κ-values as this measure of agreement is more closely related to the calibration quality of a rating system, whereas τ_x as a measure of association is more related to the discriminatory power, as argued in Hornik et al. (2007).

Since the observed κ-values of the benchmarking analysis lie between 0.2 and 0.6, we calibrated σ-values for κ-values within this interval (see Table 4). For these calibrated values of σ, we simulate the underestimation of the capital holdings by using the base case values presented in Jankowitsch et al. (2007) for all other model parameters, i.e., loss given default (LGD) is 45%, borrowers price elasticity (α) is 500, and the risk-free interest rate is 3%.

Table 4 sets out the results of this analysis. First, we compare the resulting capital holdings for the group of borrowers in transition economies with low involvement of the banks (i.e., the borrower group with the lowest rating agreement, a κ of approximately 0.2) to the group of domestic borrowers (i.e., the borrower group with the highest rating agreement, a κ of approximately 0.6). We find that for a κ-value of 0.2, the regulatory capital holdings are underestimated by 2.86 percentage points compared to a κ-value of 0.6. Thus, if we assume that the margin of conservatism included in the regulatory capital calculation is calibrated to the quality of the rating system for domestic borrowers, the regulatory capital is clearly not sufficient for borrower groups with lower rating agreement. Interestingly, there is a nearly linear relationship

Table 4 Comparison of regulatory capital requirements for different values of κ. Capital requirements for κ-values lower than 0.6 are compared to the case with a κ-value of 0.6. This analysis presents the underestimation of capital requirements by low-quality rating systems compared to a high-quality rating system measured in percentage points. The corresponding values for σ and τ_x are also shown

κ	τ_x	σ	Underestimation of Capital Requirements
0.6	0.46	0.49	–
0.5	0.43	0.63	0.60%
0.4	0.41	0.80	1.36%
0.3	0.36	1.02	1.99%
0.2	0.30	1.35	2.86%

between underestimation of the capital and the κ-values, i.e., for a κ-value of 0.4 (e.g., nontransition, low-involvement group or transition, high-involvement group), we find an underestimation of the capital requirement by 1.36 percentage points. Table 4 also shows the resulting τ_x-values for the estimated σ-values. Here, we find that the resulting τ_x-values are in line with the values we find in our benchmarking analysis.

These results are important for banks, as well as for supervisors, in assuring an adequate level of regulatory capital. Banks may incorporate these results into their internal capital adequacy assessment processes by increasing the economic capital for transactions in markets with a high expected degree of rating heterogeneity. Supervisors may consider charging an additional capital requirement for transactions in those markets.

4 Conclusions

Using multi-rater information from nine major Austrian banks, we conduct a benchmarking analysis to investigate the determinants of heterogeneity in rating assessments across different segments of the European loan market. We show that heterogeneity in ratings is generally lower in domestic than in foreign markets. Furthermore, we identify the type of economy (transition vs. nontransition) and the level of involvement of domestic banks as the essential determinants of heterogeneity within the foreign market.

We find the lowest degree of rating heterogeneity in those foreign markets where both the objective quality of information and access to subjective rating information are relatively good (i.e., nontransition markets with high involvement of Austrian banks). On the other hand, we find the highest degree of rating heterogeneity in those foreign markets where both the objective quality of information and access to subjective rating information are relatively poor (i.e., transition markets with low involvement of Austrian banks). Finally, we conclude that for transition markets with a high degree of integration with the domestic market, the problem of lack of quality-accounting information is compensated for by better access to subjective soft facts.

Our results have important practical implications for banks and their supervisors as the quality of rating systems is heavily dependent on the information available in a specific foreign market. We present an indication of the size of potential surcharges to capital requirements for groups with dissimilar ratings. On the one hand, banks may incorporate these results in their internal capital adequacy assessment processes by increasing the economic capital for transactions in markets with a high expected degree of rating heterogeneity. On the other hand, supervisors may consider charging an additional capital requirement for transactions in those markets. Furthermore, our analysis can be the foundation for future research. When time-series data on rating information become available, future studies could investigate the evolvement of financial market integration over the course of time by tracking rating heterogeneities.

References

Andres, C., Betzer, A., Weir, C.: Shareholder wealth gains through better corporate governance—The case of European LBO-transactions. Financ. Mark. Portf. Manag. **21**, 403–424 (2007)

Apreda, R.: How corporate governance and globalization can run afoul of the law and good practices in business: The Enron's disgraceful affair. Working Paper 225, Universidad del Cema (2002)

Basel Committee on Banking Supervision: Principles for the management of credit risk, Basel, Switzerland (2000)

Basel Committee on Banking Supervision: A review of credit registers and their use for Basel II. Working Paper, Financial Stability Institute, Basel, Switzerland (2004)

Basel Committee on Banking Supervision: Studies on the validation of internal rating systems (revised) (2005)

Broecker, T.: Credit-worthiness test and interbank competition. Econometrica **58**, 429–452 (1990)

Carey, M.: Some evidence on the consistency of bank's internal credit ratings. Working paper, Federal Reserve Board (2001)

Carey, M., Hrycay, M.: Parameterizing credit risk models with rating data. J. Bank. Finance **25**, 197–270 (2001)

Emond, E.J., Mason, D.W.: A new rank correlation coefficient with applications to the consensus ranking problem. J. Multicriteria Decis. Anal. **11**, 17–28 (2002)

Ernst, M.D.: Permutation methods: Basis for exact inference. Stat. Sci. **19**(4), 676–685 (2004)

European Central Bank: Measuring financial integration in the Euro area. Occasional Paper 14, Frankfurt am Main (2004a)

European Central Bank: Developing statistical indicators of the integration of the Euro area banking system. Working Paper 300, Frankfurt am Main (2004b)

European Commission: Issues in corporate governance. Economic Papers 200 (2004)

European Commission: White Paper on financial services policy 2005–2010, Brussels (2005a)

European Commission: Green Paper on financial services policy (2005–2010), Brussels (2005b)

Everitt, B., Rabe-Hesketh, S.: The Analysis of Proximity Data. Edward Arnold, Sevenoaks (1997)

Garibaldi, P., Mora, N., Sahay, R., Zettelmeyer, J.: What moves capital to transition economies? IMF Staff Papers 48 (2001)

Good, P.I.: Permutation Tests: Practical Guide to Resampling Methods for Testing. Springer, New York (2000)

Güttler, M.: Using a bootstrap approach to rate the raters. Financ. Mark. Portf. Manag. **19**, 277–295 (2005)

Hornik, K., Jankowitsch, R., Lingo, M., Pichler, S., Winkler, G.: Validation of credit rating systems using multi-rater information. J. Credit Risk **3**, 1–27 (2007)

International Monetary Fund (IMF): Institutions matter in transition, but so do policies. Working Paper 70 (2000)

Jacobson, T., Lindé, J., Roszbach, K.: Internal rating systems, implied credit risk and the consistency of banks' risk classification policies. J. Bank. Finance **30**, 1899–1926 (2006)

Jankowitsch, R., Pichler, S., Schwaiger, W.S.A.: Modelling the economic value of credit rating systems. J. Bank. Finance **31**, 181–198 (2007)

Jordão, F., Stein, R.: What is a more powerful model worth. Technical Report 030124, Moody's KMV (2003)

Kemeny, J.G., Snell, J.L.: Preference rankings: An axiomatic approach. In: Mathematical Models in the Social Sciences. MIT Press, Cambridge (1962)

Krahnen, J.P., Weber, M.: Generally accepted rating principles: A primer. J. Bank. Finance **25**, 3–23 (2001)

Krämer, W., Güttler, A.: On comparing the accuracy of default predictions in the rating industry. Working Paper, University of Dortmund, October 2006

LaPorta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R.: Investor protection and corporate valuation. NBER Working Paper 7403, Cambridge (1999)

Moser, J.T., Venkataraman, S.: The economic consequences of disclosure requirements for derivatives. Chicago Fed. Letter, Federal Reserve Bank of Chicago, 110 (1996)

Oesterreichische Nationalbank: Statistisches Monatsheft 12/2004, Wien (2004a)

Oesterreichische Nationalbank: Guidelines on credit risk management. Rating models and validation, Wien (2004b)

Ryser, M., Denzler, S.: Selecting credit rating models: A cross-validation based comparison of discriminatory power. Financ. Mark. Portf. Manag. **23**, 187–203 (2009)

Standard&Poor's: Report published on differences in bank asset quality accounting across Europe (2003)

Kurt Hornik is the Head of the Research Institute for Computational Methods and the Chair of the Institute for Statistics and Mathematics at the WU (Vienna University of Economics and Business). He completed his doctoral research and habilitation at the Vienna University of Technology. His research includes statistical computing, statistical graphics, statistical and machine learning, data mining, and a variety of application domains for quantitative data analysis, in particular quantitative risk management. Dr. Hornik has co-authored about 200 publications in refereed journals and conference proceedings, is among the ISI 100 most highly cited researchers in category "Engineering," and holds the Decoration of Merit in Gold of the Republic of Austria for his scientific achievements.


Rainer Jankowitsch is Associate Professor of Finance at the WU (Vienna University of Economics and Business). He completed his doctoral research at the University of Vienna and recently finished his habilitation. His research is focused on credit and liquidity risk, banking, risk management, and financial markets. He published in various finance journals in the last five years. Dr. Jankowitsch received the Best Paper Award of the German Finance Association in 2008 for his work on liquidity risk, which is an output from a research cooperation with the New York University. His current research is focused on the financial crisis.

Manuel Lingo is analyst at Oesterreichische Nationalbank where he currently is responsible for the development of operations of the In-house Credit Assessment System (ICAS) used for Eurosystem monetary operations. Before joining Oesterreichische Nationalbank he worked as research assistant at the WU (Vienna University of Economics and Business) and as freelance consultant for Pricewaterhouse-Coopers. He publishes in journals related to credit risk (Journal of Credit Risk and Journal of Risk Model Validation). His current research focuses on rating system development and validation. Manuel Lingo holds a Ph.D. in Finance from the WU.

Stefan Pichler is Professor of Banking and Finance at the WU (Vienna University of Economics and Business) and deputy speaker of the Vienna Graduate School of Finance. He completed his doctoral research at the University of Graz and his habilitation at the Vienna University of Technology. His research focus is on fixed income markets, financial risk management, and banking regulation. He has co-authored numerous publications in refereed journals.
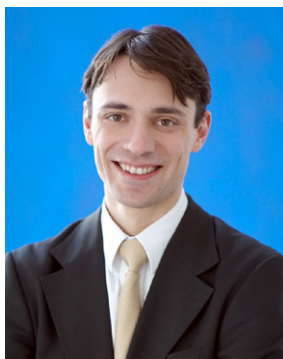
Gerhard Winkler currently holds the position of Deputy Head in Oesterreichische Nationalbank's Credit Division. As Associate Professor (non-tenured) at the Department of Finance, Accounting and Statistics, WU (Vienna University of Economics and Business), he has been active in research and teaching for many years. He completed his post-doctoral qualification (habilitation) in 2008. His main research interests are credit risk measurement, fixed income analysis, and bank efficiency.