

Benchmarking Credit Rating Systems

Kurt Hornik^a, Rainer Jankowitsch^b, Manuel Lingo^b, Stefan Pichler^{bc}, Gerhard Winkler^d

^a *Department of Statistics and Mathematics, Vienna University of Economics and Business Administration, Augasse 2–6, A-1090 Vienna, Austria*

^b *Department of Finance and Accounting, Vienna University of Economics and Business Administration, Nordbergstrasse 15, A-1090 Vienna, Austria*

^c *Vienna Graduate School of Finance*

^d *Oesterreichische Nationalbank, Otto Wagner Platz 3, A-1090 Vienna, Austria*

Any views expressed represent those of the authors only and not necessarily those of Oesterreichische Nationalbank.

January 2007

Abstract

The validation of credit rating systems has recently attracted particular interest both from banks and their supervisors as well as from academic research. Whereas the main interest has been focused on backtesting methods, alternative approaches such as benchmarking are of growing importance. Benchmarking methods make use of available multi-rater information, i.e. rating assignments about identical obligors stemming from different rating sources. Employing a unique data set provided by the Austrian central bank with rating information on European corporate obligors by nine major Austrian banks, we conduct a benchmarking analysis based on the framework suggested by Hornik et al. (2005) on both market specific and bank specific levels. It turns out that overall similarity among rating systems is remarkably lower for foreign markets compared to the domestic market. Transition economies and markets with a generally low involvement of Austrian banks show particularly dissimilar results. Differences in the overall similarity of rating systems are partly explained by different performance of some individual rating systems in the domestic and foreign markets.

Key words: Rating validation, benchmarking, rating agreement, rating association, financial market integration, transition economies.

JEL classification: G20, C49

1 Introduction

The validation of credit rating systems has recently attracted particular interest both from banks and their supervisors as well as from academic research. Whereas the main interest has been focused on backtesting methods, alternative approaches such as benchmarking are of growing importance (Bank for International Settlements, 2005). Benchmarking methods make use of available multi-rater information, i.e. rating assignments about identical obligors stemming from different rating sources. Analyzing the proximity (similarity or dissimilarity) of rating information among a prespecified set of obligors may help to identify individual rating systems which show a remarkably lower degree of similarity and thus may be regarded as “outliers” (Hornik et al, 2005). On the other hand, this information can be used to examine potential differences in rating quality across various markets. In an extensive empirical study with rating information about European corporate obligors stemming from nine major Austrian banks we show the application of benchmarking for various analyses on the bank specific (“micro”) and on the market specific (“macro”) levels.

Clearly, accurate assessment of the creditworthiness of obligors is an important precondition for the stability of a financial system since inadequately high exposure to credit risk has been one of the leading sources for problems in financial institutions worldwide for many decades (Basel Committee on Banking Supervision, 2000). Thus the financial service industry and its supervisors have become more and more focused on credit risk measurement.

The continuously growing interest in credit risk models and rating systems has obtained a further boost by the new guidelines of Basel II which were issued by the Basel Committee on Banking and Supervision with the intention to promote international convergence of capital measurement and capital standards (Bank for International Settlements, 2004b). One of its general and most fundamental features is the closer linkage of a bank’s regulatory capital

requirements to the actual level of risk incurred. Concerning credit risk the exact alignment of regulatory capital with the risk of a bank's loan portfolio may even be based on banks' internal measures for key parameters of credit risk: Arranging for a so-called internal ratings-based approach the new framework allows banks to use their own estimates of key drivers of credit risk as main inputs in the computation of their minimum regulatory capital requirements. As such it ensures that regulatory capital is in line with a bank's own perception of the credit risk in its books.

Obviously, validation of banks' internal rating systems is crucial for both banks and their supervisors. As regards banks validation is not only motivated by the obligation under Basel II to assess the soundness and appropriateness of their internal risk measurement systems on an ongoing basis. In general, banks have a particular interest in validating their internal credit risk models to ensure accurate and reliable credit assessments of their obligors. In a traditional cut-off based lending framework banks need to know whether potential obligors will be able to meet their loan obligations in accordance with contractual agreements. In a risk-based lending framework where margins directly depend on the obligor's credit assessment the accuracy of rating systems is a key success factor (e.g., Jankowitsch et al., 2006, and Jordão and Stein, 2003).

Although validation is thus foremost the responsibility of banks themselves, of course their supervisors also have a core interest in rating validation. Supervisors have to review banks' internal validation processes and may in addition want to perform their own validation tasks. Supervisors are interested that each bank correctly quantifies the risk in all its undertakings and consequently aligns its capital base to its risk exposure. From a micro perspective supervisors ensure that the failure risk of individual banks is within a limited range. From a

macro perspective they foster the stability of the financial system as a whole. Furthermore they may help create a level playing field among the numerous competing credit institutions.

We conclude that the development of reliable validation techniques is a crucial issue for banks and their supervisors, for commercial and regulatory purposes, and from a micro but also a macro perspective. Validation methods may roughly be divided into *qualitative* and *quantitative* (or empirical) methods (Bank for International Settlements, 2005). Whereas the area of qualitative validation fulfills the primary task of ensuring the general applicability as well as the specific proper application of the risk measurement methods in practice, quantitative validation comprises all procedures in which statistical indicators about the accuracy or reliability of the estimated key parameters of credit risk (such as probability of default, PD) are calculated and interpreted on the basis of an empirical data set. Most quantitative validation techniques rely on *backtesting*, i.e., the comparison of estimated parameters with empirical realizations. The literature on the various backtesting methods is quite extensive and focuses to a large extent on the discriminatory power of PD estimates, i.e., the ability to distinguish ex-ante between defaulting and non-defaulting obligors (e.g., Oesterreichische Nationalbank, 2004b). In addition, other methods try to directly evaluate the accuracy of the PD estimates or calibration quality (e.g., Stein, 2002). For both, checking discriminatory power and even more for the testing of calibration quality, however, the requirements on the available data are rather high. Thus a major obstacle to the backtesting of PD often is the scarcity of data (especially on defaults) which in many cases prevents the examiner from making meaningful statistical inference.

As suggested by the Basel Committee on Banking Supervision *benchmarking* may be regarded as a reasonable and complementing alternative to backtesting (Bank for International Settlements, 2005). In its most general sense benchmarking may be defined as the comparison

of ratings of certain obligors stemming from one specific credit quality assessment source with ratings on the same set of obligors obtained from other sources. As benchmarking does not depend on the immediate availability of historical default data it is particularly helpful in situations with severe data restrictions when backtesting is not applicable (e.g., for small obligor populations with low default rates). Benchmarking may also be less costly than backtesting for supervisors, who are obliged to check the validity of numerous banks' internal rating systems almost simultaneously but are constrained by their limited human and financial resources. A supervisor might additionally be interested in the degree of heterogeneity of the internal ratings for certain obligors across banks and whether different types of rating systems produce similar results. Finally, a benchmarking analysis can be undertaken from both a micro perspective on an individual bank level as well as from a macro perspective on an aggregate, market-wide level. This enables the supervisor to analyze the extent to which individual banks in a financial market agree in their rating assessments of certain obligor groups. Conversely, the supervisor can identify 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 in a consistent way.

As such benchmarking may especially be of relevance for the validation tasks to be performed by supervisors. In many countries supervisors have access to banks' internal rating information. Either the data may be collected via on-site inspections or it has to be directly reported to the supervisor. Very often this sort of data is contained in credit registers managed by or accessible to supervisors. In this case supervisors can easily extract rating information on an overlapping set of obligors provided by different sources (e.g. banks reporting to the credit register) at the same time and use this data for benchmarking purposes. Thus credit registers possess an enormous potential as a key tool in the hands of supervisory authorities as has recently been thoroughly discussed (Bank for International Settlements, 2004a).

Against this background we use data on European corporate obligors of Austrian banks as recorded in the Austrian credit register to measure rating system performance by means of benchmarking. On the one hand we analyze performance on an individual bank (“micro”) level employing different measures of proximity based on the framework suggested by Hornik et al. (2005). This allows us to identify banks which show an apparently low level of overall similarity to other banks and may thus be regarded as potential outliers. On the other hand we show how this framework may be enhanced and applied from a macro perspective to measure aggregate proximity among all banks simultaneously on a market-wide level. Thus, a further contribution of our paper is the examination of potential differences in rating quality on an aggregate level across various sub-groups of obligors.

This analysis provides us with information to the extent to which European financial markets are fully integrated as regards 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, and European Central Bank, 2004b). Following the literature on financial market integration we can expect more similar rating outcomes on obligors in the domestic market than in the foreign markets. This hypothesis about aggregate proximity is motivated by the fact that credit markets are characterized by informational differences between companies wishing to finance their operations and often do not want to fully disclose all information about their undertakings (e.g., Apreda, 2002, and Moser and Venkataraman, 1996) and banks striving 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&Poors, 2003).

Further, the degree of similarity between rating outcomes can be used as an indicator of financial market integration of foreign markets in relation to the domestic market. An argument supporting this hypothesis is that a high level of integration with a foreign market usually coincides with a harmonization of disclosure principles and financial reporting practices (European Commission, 2000). Finally, soft facts may either be more easily accessible for banks on obligors residing in highly integrated markets or banks may simply take more effort to obtain them whereas they might rely more on purely quantitative (accounting) information on hard facts in less integrated markets.

As a consequence we can expect a higher degree of similarity for foreign markets being more integrated with the Austrian market than for others. To test this relationship we separate the foreign markets into two sub-markets due to 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.

Whereas financial market integration is an obvious factor for the availability and the quality of information needed as an input for rating systems, other factors can be considered as well. One of these factors which is very likely to explain differences in the quality of available data is the type of economy. In the context of European integration it is particularly interesting, whether an economy is in the status of transition from a former centrally planned socialist regime or not (e.g., Garibaldi et al., 2001, and IMF, 2000). Apart from macroeconomic stabilization policies, microeconomic elements are equally crucial for a successful economic conversion of transition economies. Sound and effective governance structures can be decisive for attracting portfolio investment (LaPorta et al., 1999, and Garibaldi et al., 2001) 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 or accounting standards may not be as established as in other economies and in case of potential deficiencies in transparency banks may have more difficulties in assessing the risk profile of a company located in a transition country (European Commission 2004).

Thus, we conclude that the overall similarity of rating outcomes can be expected to be lower in transition economies compared to non-transition economies. From the opposite point of view similarity of rating outcomes can be used as an indicator of market transitionality. We examine this relationship by splitting our foreign sample in transition and non-transition economies.

As for Austrian banks, markets in the Central and Eastern European region play a dominant role, we are in the fortunate situation to find markets which both exhibit a high level of integration with the Austrian market and are transition economies. This enables us to examine the question whether financial market integration or transitionality plays a more dominant role in explaining similarity or dissimilarity of rating outcomes.

This paper is organized as follows. Section 2 revisits the methodology suggested by Hornik et al. (2005) and formally describes the use of several measures of proximity between different rating systems. Section 3 contains a description of the data set and the empirical results on bank and market-wide levels. Section 4 concludes the paper.

2 Rating Validation based on Proximity Measures

When using multi-rater panels for the benchmarking of rating systems, it is clearly imperative that one is able to quantify the similarity or dissimilarity, i.e., the *proximity*, between the raters. Hornik et al. (2005) show that there are several aspects of proximity to be taken into account, most notably *agreement* and *association*. The former is based on comparisons of the numbers of objects (obligors) rated concordantly and discordantly, respectively; the latter measures the extent to which rating orders between pairs of obligors are preserved between raters.

To make this more precise, let us first notice that ratings are typically given on an ordinal scale (e.g., $AAA > AA+ > \dots > D$). When measuring the proximity of the ratings given 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 bank, respectively. Suppose that both banks use the same 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 amount 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 which 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 banks, and $p_{i,j} = c_{i,j} / N_c$ be the joint proportion of obligors rated

into classes i and j , respectively. Also, write $p_{i\cdot} = \sum_j p_{i,j}$ and $p_{\cdot j} = \sum_i p_{i,j}$ for the marginal proportions of obligors rated into class i by the first bank and into class j by the second bank, respectively. Then weighted κ is defined as

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

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 $0 \leq w_{ij} \leq 1$ with $w_{ii} = 1$ and $w_{ij} = w_{ji}$, and R is the (common) number of rating classes employed. Hornik et al (2005) propose to use the so-called Cohen-Fleiss weights $w_{ij} = 1 - (i - j)^2 / (R - 1)$, which measure disagreement as quadratic in the “difference” in rating classes.

Ratings on an ordinal scale naturally give rise to (weak) order relations on the obligors rated. Association measures quantify the extent to which these order relations are in accordance. Kemeny and Snell (1962) have shown that there is a unique association (or equivalently, distance) measure between order relations satisfying several natural axioms (basically, metricity and “betweenness”). Hereafter, we refer to this measure as τ_x ; this notation indicates that it is an extension of the well-known Kendall’s τ , which handles the case of ties more appropriately than τ (see also Emond and Mason, 2002). 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 bank, and b_{uv} is defined similarly for the second bank, then

$$\tau_x(A, B) = 1 - \frac{\sum_{u=1}^{N_c} \sum_{v=1}^{N_c} |a_{uv} - b_{uv}|}{N_c(N_c - 1)/2} \quad (2)$$

(Equivalently, τ_x corresponds to the “natural” symmetric difference metric between the order relations on 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 provide readily interpretable means of quantifying the proximity of the co-ratings of obligors between two banks. When ratings from several banks are available, their overall structure can be analyzed based on the pair-wise proximities. In addition, proximities can be computed for sub-groups of obligors, e.g., according to economic sector and/or economic region. This makes it possible to perform both bank level and market level analyses of the proximities of rating systems. Methodological details of this approach are given in Section 3.2.

3 Benchmarking Rating Systems: Empirical Applications

3.1 Data Description

In our empirical analysis we use a unique data set on rating information stemming from the Austrian credit register which is operated by the Austrian National Bank. Credit registers indeed provide a unique opportunity for supervisors to face the challenges of Basel II (Bank for International Settlements, 2004a). In particular a credit register may act as a supplementary tool allowing the supervisor to address many of the validation issues that Basel II entails such as a benchmarking of rating estimates. According to the Austrian Banking Act, all Austrian financial institutions have to supply monthly information to this register on any changes in the status of their outstanding credits as well as information on the new loans

granted during the period. In detail, the Austrian credit register covers all major direct lending activities of Austrian credit and financial institutions as well as contract insurance companies and thus contains information on all loans above a threshold of EUR 350,000. Taken together approximately 67% of the total volume of all Austrian banks' (direct) lending activities were covered by the credit registers entries with a total amount of EUR 373 bn at the end of 2004 (Oesterreichische Nationalbank, 2004a). Each record of the register comprises information on each credit claim and its obligor including the name, the address, the legal form, and the economic sector affiliation of each obligor, the type of financial instrument used for refinancing, and the outstanding as well as the undrawn amounts of credit commitments. In addition to this very general sort of information on bank debt which may be shared among institutions, each record in the Austrian credit register further contains obligor specific information on credit risk which must not be granted to any institution but only be used for supervisory purposes. It consists of obligor specific information on provisions for loan losses, on the value of collateral the respective obligor has pledged to the reporting institution, and – most important for this paper - on the respective institution's own internal assessment of the individual obligor's creditworthiness as expressed by a rating. The information on credit risk which has been recorded since the beginning of 2003 makes the Austrian credit register unique and enables the supervisor to closely monitor credit risk in the entire system.

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). It is closely related to the rating scales of the

major rating agencies and its granularity has been optimized in order to minimize the loss of information in the mapping of banks' individual rating scales onto it.

The data set considered here consists of rating information on European corporate obligors within 2005. Since our benchmarking methods are based on co-ratings of identical obligors by different banks we have to ensure a minimum amount of co-ratings. Thus, we select a sample based on the reportings of nine large Austrian banks to the credit register. Taken together their volume of major loans on average accounted for 57% of the total volume of all major loans granted by banks and they amounted to 69% of all pan-European foreign direct lending activities reported to the credit register by Austrian banks within 2005. Both ratios suggest that the sample banks and their loan portfolios can be considered to be highly representative for the purposes of our study. We exclude obligors only rated by a single bank, or, in order to obtain a homogeneous sample, for which external ratings by rating agencies are available. This yields an overall sample consisting of 7,882 ratings on 3,026 obligors which are used in our benchmarking analyses.

This sample allows us to study the proximity of ratings not only on a bank specific (micro) but also on a market-specific (macro) level. Thus we divide our overall sample into several sub-samples to be able to measure proximity in different obligor groups. To test the hypothesis whether overall proximity among banks is higher for domestic obligors than for foreign obligors we divide our overall sample into one sub-sample for domestic and one for foreign obligors. In a next step we further split up the sub-sample of foreign obligors to analyze variations in proximity in greater detail.

On the one hand we differentiate between obligors from foreign markets in which Austrian banks show a high presence (hereafter high involvement countries) and obligors from markets

where Austrian banks play a minor role (low involvement countries) to test whether overall proximity is higher for high involvement countries than it is for low involvement countries. This distinction is based on relative concentration: Using all the reportings on European obligors to the credit register we first compute the exposure per country. Then we identify those European countries which show the highest involvement of domestic banks and whose cumulative share reaches 50% of all foreign lending. These high involvement countries are (in alphabetic order): Croatia, Germany, Hungary, Italy, The Netherlands, and Poland. The remaining low involvement countries consist of: Albania, Andorra, Bosnia and Herzegovina, Belarus, Belgium, Bulgaria, Czech Republic, Cyprus, Denmark, Estonia, Latvia, Lithuania, Finland, France, Great Britain, Greece, Ireland, Island, Liechtenstein, Luxembourg, Macedonia, Malta, Monaco, Norway, Portugal, Russia, Serbia and Montenegro, Slovakia, Slovenia, Spain, Sweden, Switzerland, and Turkey.

On the other hand we distinguish foreign obligors residing in transition countries from foreign obligors located in non-transition countries to study whether transition still matters and whether overall proximity is lower among banks for this subset than for obligors from non-transition countries. Following the extensive literature on transition economies we define a transition country as one which has switched from centrally planned socialist regime to a market economy (e.g., Garibaldi et al. 2001, and IMF 2000). As such the countries of transition are Albania, Bosnia and Herzegovina, Belarus, Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Macedonia, Poland, Russia, Serbia and Montenegro, Slovakia, and Slovenia whereas the sub-set of non-transition countries comprises all the other European countries.

In a third and final step we further split these samples and differentiate between high involvement non-transition, high involvement transition, low involvement non-transition, and

low involvement transition countries to study joint effects of involvement and transitionality on the levels of proximity in the respective segments.

Table 1 shows descriptive statistics for the composition of our overall data set and the various sub-samples.

- insert Table 1-

Note that on the overall level even the smallest bank has at least 102 obligors in common with one or more of the other institutions. Furthermore we infer from the number of foreign obligors and the respective ratings that our sample seems to be well suited for a macro-analysis of proximity in different sub-markets of obligors. The nine banks in our sample are very active on a pan-European level with a strong presence in the Central and Eastern European countries of transition. Three of these countries, namely Croatia, Hungary, and Poland are even among the high involvement countries.

In addition to the figures given in Table 1 we also compute the number of co-ratings on an obligor level. For the overall sample the minimum and maximum numbers 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 to only a small number of banks. These values do not vary much across the different subsets.

3.2 Analysis on a Bank Specific Level

Based on the data set described in the previous section we can analyze the performance of credit rating systems. By making use of the measures for association (τ_x) and agreement (κ) presented in Section 2 we assess the level of similarity between the nine individual banks. Using aggregates of these bivariate comparisons allows us to identify banks with lower levels of similarity, typically to be considered as outliers. We conduct this analysis for different obligor groups allowing deeper insight into which market segments potentially cause lower levels of similarity for certain banks.

One-Dimensional Identification of Outliers. τ_x and κ provide us with bivariate measures to compare association and agreement for ratings of two banks based on their set of overlapping obligors. For the nine banks this amounts to 36 bivariate comparisons for each measure. Therefore the individual observations for each bank have to be aggregated in order to measure the overall proximity of each bank compared to all other banks. Following Hornik et al. (2005) we use arithmetic averages as a one-dimensional representation of the bivariate measures. First we analyze association and agreement of the banks based on all available ratings. Then we compare the rating proximity for the domestic and foreign obligor segments separately allowing us to interpret the overall results.

- insert Table 2-

Table 2 shows the average proximity measures for each bank and obligor segment. When using all obligor ratings the highest average τ_x and κ are 0.544 and 0.666 whereas the lowest average τ_x and κ are 0.422 and 0.482, respectively, showing a remarkable difference in the similarity levels of individual banks. However, there is no defined cut-off level for the

average measures to identify banks with significantly deviating rating systems. For illustrational purposes we consider the three banks with the lowest averages as potential outliers. Using our one-dimensional representation we would therefore identify the banks 4, 7, and 9 as outliers on the basis of their average τ_x and κ . In this case the outlier-banks are the same for both proximity measures. In general, this will not be the case, as for example indicated by bank 7 which has the second lowest average τ_x but the third lowest average κ . Thus it is important to measure both aspects of proximity.

To gain deeper insight into possible reasons for overall market dissimilarity we split our obligor sample into two market segments (see Table 2). The first segment is represented by all Austrian obligors, i.e., the banks' domestic market, whereas the second segment consists of all foreign obligors. Focusing on the outliers in the overall sample (i.e., banks 4, 7 and 9) we find that bank 4 is identified as an outlier for both measures in the domestic market but for neither in the foreign market. These results indicate that the overall low level of similarity for this bank is mainly caused by its performance in the domestic market. Bank 7 certainly has a rating system of lower similarity for the foreign market. In the domestic market, only association is low. Bank 9 is identified as outlier in the domestic market by both measures. In the foreign market, again only agreement is low.

Multi-Dimensional Identification of Outliers

Hornik et al. (2005) introduce a multivariate MDS/MST method as an alternative approach to analyzing proximities in multi-rater panels and, in particular, identifying outliers. A two-dimensional map of the banks obtained by metric multi-dimensional scaling (MDS) of their pairwise proximities is annotated by the minimal spanning tree (MST) of the proximities. The MST indicates higher-dimensional effects when nearby points in the map are not joined by an edge of the tree. Based on the MDS/MST combination, outliers could be seen as banks (i.e.,

points in the map) which are “far” from all other banks and have only “few” edges connecting them to the others.

-insert Figure1-

Figure 1 shows the resulting MDS maps including MST for τ_x and κ . Considering the results for the overall dataset the MDS/MST exactly confirms the one-dimensional results for both measures. Most banks are located in the center and are well connected to each other representing rather similar rating systems, whereas bank 4, 7, and 9 can be found on the edges with only one connection. Thus, as shown in Hornik et al. (2005) the MDS/MST representation provides a useful tool to assure the one-dimensional results including the choice of the number of outliers.

Considering the two sub-markets we find for bank 4 and 9 that the MDS/MST implies exactly the same results as the one-dimensional measures, providing further support for the interpretation of the previous analysis. The situation for bank 7 is different: here, MDS/MST supports that bank 7 is an outlier in the foreign market, but there is no indication that this is the case in the domestic market. Based on the results of the one-dimensional aggregation and MDS/MST for the two sub-markets we can provide further interpretation for the overall outliers. Two banks have low levels of similarity for only one of the two sub-markets: bank 4 in the domestic market and bank 9 in the foreign market. The third bank shows deviations in both markets.

3.3 Analysis on Market Specific Level

In this section we use the proximity measures to analyze association and agreement in different sub-markets. The most basic separation is between domestic and foreign markets as presented in Section 3.2. Comparing the τ_x and κ of the individual banks for these two markets shows lower levels of similarity for the foreign market (see Table 2). With our data set we can separate obligors into further sub-markets based on the type of economy and the degree of financial market integration. Employing the information argument presented in Section 1 we expect higher levels of similarity for non transition as well as for high involvement economies as in these better or more information, respectively is available. Here we test the implied hypotheses. First we separate into different groups of obligors by using either the type of economy or the degree of financial market integration. In a further analysis we build subsets based on combinations of these two attributes.

To analyze these effects we use the bank specific bivariate measures of each bank for different sub-markets to find potential differences. The high number of possible combinations, however, apparently makes it straightforward to use aggregated measures to analyze the different groups where we use overall averages of the bivariate measures for each sub-market. Additionally we test for significance differences in the level of similarity for different sub-markets based on the bivariate measures employing a conditional test based on Wilcoxon's rank sum statistic (see Good, 2000 or 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.

-insert Table 3-

As presented, 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 sound rating system is more easily achieved for domestic markets.

Concerning the type of economy we obtain an average association (τ_x) of 0.45 for non-transition economies in contrast to 0.36 for transition economies. A similar result can be observed for agreement: 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 which implies that information does not have the same standard as in non-transition countries or cannot be collected easily. Thus the type of economy certainly has an important impact on the overall similarity of the rating systems, at least in our data set.

For financial market integration we compare high to low involvement observing a difference in association (τ_x) of 0.08 (0.46 vs. 0.38) which is not statistically significant whereas for agreement we find a difference of 0.17 (0.58 vs. 0.41) which is significant. These results indicate that the available information in markets with low involvement might not be sufficient for slotting obligors into the correct rating class while it still might be sufficient for ranking them relative to each other.

Using combinations of the involvement and the transition attribute allows a more detailed analysis. When dividing the obligors into transition and non-transition economies for each level of involvement we find that for low involvement this separation causes significant differences for agreement and association. Interestingly, for high involvement the differences are only significant for association. This indicates that the availability of transparent and standardized information might be less diverse for transition and non-transition economies

with high involvement. When dividing the obligors into high and low involvement for each type of economy the previous results for involvement are confirmed.

Analysing combinations where the type of economy and the level of involvement are different for both groups we find the obvious result that obligors of non-transition economies with high involvement have a significantly higher level of similarity than transition economies with low involvement. Interestingly the comparison between non-transition countries with low involvement and transition economies with high involvement does not yield a significant difference for neither association nor agreement indicating a comparable availability of transparent and standardized information.

4 Conclusion

This paper demonstrates the application of benchmarking methods to validate banks' internal credit rating systems. Along the lines of Hornik et al. (2005) we analyze the proximity, i.e., similarity or dissimilarity, between outcomes of different rating systems. We enhance this framework to allow for the quantitative analysis of rating assessment variation across markets and the inclusion of several attributes potentially explaining these variations.

Provided with rating information about European corporate obligors stemming from nine major Austrian banks we obtain measures of association and agreement as important aspects of proximity between the rating outcomes of the banks in different markets. In addition to the comparison of averages of bivariate proximity measures we employ a multi-dimensional scaling technique combined with minimal spanning trees to detect potential "outlier" banks, i.e. banks whose rating outcomes apparently differ from those of other banks. On a bank specific ("micro") level we show that the dissimilarity of rating outcomes for some banks can

be related to specific segments of either domestic or foreign obligors whereas for some banks the dissimilarity exists for all market segments.

On the market-specific (“macro”) level it turns out that overall similarity among rating systems is remarkably lower for foreign markets compared to the domestic market. Transition economies and markets with a generally low involvement of Austrian banks show particularly dissimilar results. Our results suggest that the type of the economy, i.e. being in transition or not, has a stronger effect on the overall proximity of rating outcomes than financial sector integration, as proxied by the involvement of domestic banks in the foreign markets. As one could expect a completion of transition processes for most of the European economies in our sample our approach could be used for future research or deriving supervisory policies to examine a potential “phasing out” of transitionality effects.

References

R. Apreda. How corporate governance and globalization can run afoul of the law and good practices in business: the Enron's disgraceful affair. Working paper No. 225, Universidad del Cema , 2002. URL <http://www.cema.edu.ar/publicaciones/download/documentos/225.pdf>.

Bank for International Settlements. A review of credit registers and their use for Basel II. Working Paper, Financial Stability Institute, Bank for International Settlements, CH-4002 Basel, Switzerland, 2004a. URL <http://www.bis.org/fsi/awp2004.pdf>.

Bank for International Settlements. International convergence of capital measurement and capital standards: A revised framework, 2004b. URL <http://www.bis.org/publ/bcbs107.htm>.

Bank for International Settlements. Studies on the validation of internal rating systems (revised), 2005. URL http://www.bis.org/publ/bcbs_wp14.pdf.

Basel Committee on Banking Supervision. Principles for the management of credit risk, Basel, Switzerland, 2000. URL <http://www.bis.org/publ/bcbs75.pdf>.

E. J. Emond and D. W. Mason. A new rank correlation coefficient with applications to the consensus ranking problem. *Journal of Multicriteria Decision Analysis*, 11:17-28, 2002.

European Central Bank. Measuring financial integration in the Euro area. Occasional Paper No. 14, Frankfurt am Main, 2004a. URL www.ecb.int/pub/pdf/scpops/ecbocp14.pdf.

European Central Bank. Developing statistical indicators of the integration of the Euro area banking system. Working Paper No. 300, Frankfurt am Main, 2004b.

URL <http://www.ecb.int/pub/pdf/scpwps/ecbwp300.pdf>.

European Commission. Communication from the Commission to the Council and the European Parliament. EU financial reporting strategy : the way forward, Brussels, 2000.

URL http://europa.eu.int/eur-lex/lex/LexUriServ/site/en/com/2000/com2000_0359en01.pdf.

European Commission. Issues in corporate governance. Economic Papers No. 200, 2004.

URL http://europa.eu.int/comm/economy_finance/publications/economic_papers/2004/ecp200en.pdf.

M. D. Ernst. Permutation methods: Basis for exact inference. *Statistical Science*, 19/4: 676-685, 2004

B. Everitt and S. Rabe-Hesketh. *The Analysis of Proximity Data*. Edward Arnold Publishers Ltd, 1997.

P. Garibaldi, N. Mora, R. Sahay, and J. Zettelmeyer. What moves capital to transition economies? IMF staff papers, Vol 48, 2001.

URL <http://www.imf.org/external/pubs/ft/staffp/2001/04/pdf/garibald.pdf>.

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

K. Hornik, R. Jankowitsch, M. Lingo, S. Pichler and G. Winkler. Validation of credit rating systems using multi-rater information. Working Paper, 2005.

URL <http://ssrn.com/abstract=871213>.

International Monetary Fund. Institutions matter in transition, but so do policies. Working Paper No. 70, 2000. URL <http://www.imf.org/external/pubs/ft/wp/2000/wp0070.pdf>.

R. Jankowitsch, S. Pichler, W.S.A. Schwaiger, Modelling the economic value of credit rating systems, forthcoming Journal of Banking and Finance, 2006.

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

URL: http://www.moodykmv.com/research/whitepaper/MorePowerfulModel_TR030124.pdf.

J. G. Kemeny and J. L. Snell. Mathematical Models in the Social Sciences, chapter Preference Rankings: An Axiomatic Approach. MIT Press, Cambridge, 1962.

R. LaPorta, F. Lopez-de-Silanes, A. Shleifer, R. Vishny. Investor protection and corporate valuation, NBER Working Paper No. 7403, Cambridge, 1999.

J.T. Moser and S. Venkataraman. The economic consequences of disclosure requirements for derivatives. Chicago Fed Letter, Federal Reserve Bank of Chicago, No. 110, 1996.

Oesterreichische Nationalbank. Statistisches Monatsheft 12/2004. Wien, 2004a.

Oesterreichische Nationalbank. Guidelines on Credit Risk Management. Rating Models and Validation, Wien, 2004b. URL http://www.oenb.at/en/img/rating_models_tcm16-22933.pdf.

Standard&Poors. Report published on differences in bank asset quality accounting across Europe, 2003.

R. Stein. Benchmarking default prediction models: Pitfalls and remedies in model validation. Technical Report #020305, Moody's KMV, 2002.
URL: http://riskcalc.moodysrms.com/us/research/crm/Validation_Tech_Report_020305.pdf.

APPENDIX

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
<i>Overall</i>	7882	3026	102	647	2183
<i>Domestic Market</i>	5900	2230	46	509	1761
<i>Foreign Market</i>	1564	628	36	175	344
<i>High Involvement</i>	725	293	11	78	162
<i>Low Involvement</i>	840	335	25	87	182
<i>Non-Transition Economies</i>	922	377	16	96	213
<i>Transition Economies</i>	642	251	20	61	131
<i>Non-Transition & High Inv.</i>	214	84	6	19	42
<i>Transition & High Inv.</i>	511	209	5	61	120
<i>Non-transition & Low Inv.</i>	214	84	14	42	89
<i>Transition & Low Inv.</i>	412	168	11	35	93

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.

Bank	Proximity Measure					
	All Obligors		Domestic Obligors		Foreign Obligors	
	τ_x	κ	τ_x	κ	τ_x	κ
1	0.5305	0.6588	0.6135	0.6972	0.3430	0.4417
2	0.5049	0.6349	0.5149	0.6350	0.4229	0.5518
3	0.4899	0.6371	0.5092	0.6451	0.4688	0.5380
4	0.4218	0.4818	0.3755	0.4105	0.5785	0.5797
5	0.5367	0.6114	0.5460	0.6209	0.3388	0.4595
6	0.5439	0.6656	0.5536	0.6629	0.4822	0.6480
7	0.4701	0.6099	0.5031	0.6228	0.3826	0.4316
8	0.5089	0.6302	0.5068	0.6435	0.4464	0.5036
9	0.4792	0.5346	0.4814	0.5240	0.4065	0.4561
Overall Average	0.4984	0.6072	0.5116	0.6069	0.4300	0.5122

Table 2: Average bank specific association (τ_x) and agreement (κ) measures for different groups of rated obligors. The three banks with the lowest degree of similarity are highlighted. The last line shows the overall average similarity per obligor group.

Customer 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
<i>Domestic Market vs. Foreign Market</i>	0.5116	0.4300	0.0050	0.6069	0.5122	0.0144
<i>Non-Transition Economies vs. Transition Economies</i>	0.4456	0.3550	0.0305	0.5112	0.3331	0.0028
<i>High Involvement vs. Low Involvement</i>	0.4612	0.3841	0.1349	0.5808	0.4144	0.0010
<i>Non-Transition & High Inv. vs. Transition & High Inv.</i>	0.4537	0.3753	0.6021	0.6005	0.3997	0.0324
<i>Non-Transition & Low Inv. vs. Transition & Low Inv.</i>	0.4144	0.2893	0.0361	0.4146	0.2389	0.0098
<i>Non-Transition & High Inv. vs. Non-Transition & Low Inv.</i>	0.4537	0.4144	0.4091	0.6005	0.4146	0.0102
<i>Transition & High Inv. vs. Transition & Low Inv.</i>	0.3753	0.2893	0.0609	0.3997	0.2389	0.0142
<i>Non-Transition & High Inv. vs. Transition & Low Inv.</i>	0.4537	0.2893	0.0018	0.6005	0.2389	0.0000
<i>Non-Transition & Low Inv. vs. Transition & High Inv.</i>	0.3753	0.4144	0.9707	0.3997	0.4146	0.8256

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.

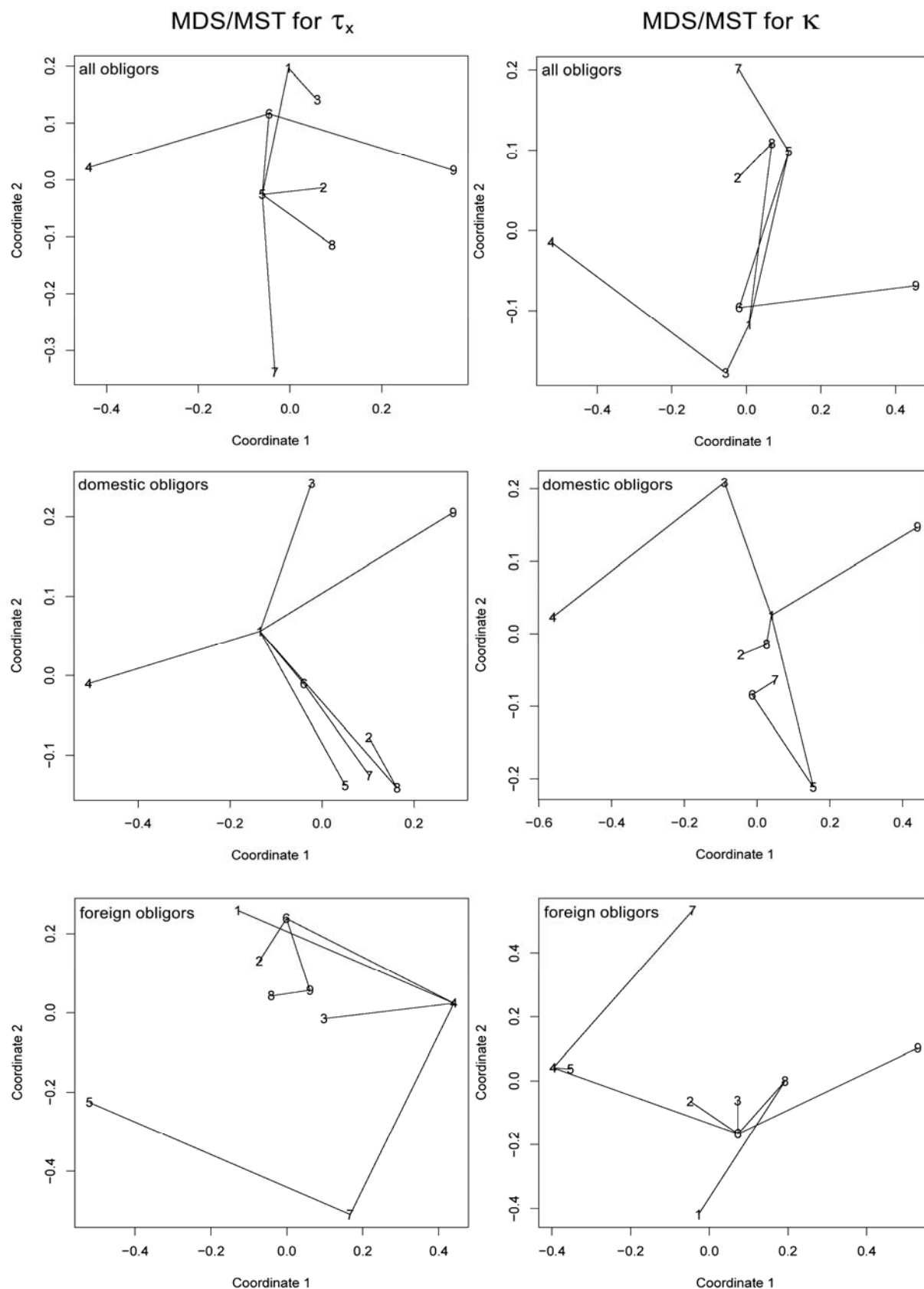


Figure 1: Multidimensional scaling plots in combination with minimal spanning trees based on bank specific association (τ_x) and agreement (κ) measures. The numbers represent the nine banks where the distances and the numbers of connections between these banks are related to the underlying proximities. One could identify banks with rating systems of potentially different quality by large distances from all other banks and a low number of connections.