Sharing more than dialects: Herding behavior in bank risk-taking explained by cultural vicinity

Barbara Seitz[1,2]
University of St. Gallen, Switzerland | Institute of Accounting, Control and Auditing

Andreas Rathgeber[3]
University of Augsburg, Germany | Institute of Materials Resource Management

Alfred Lameli[4]
Philipps-Universität Marburg, Germany | Research Center Deutscher Sprachatlas

Tobias Gaugler[5]
University of Augsburg, Germany | Institute of Materials Resource Management

Abstract:
We expand on research concerning the well-pronounced influence of geographical peer groups on human behavior. For this purpose, bank-specific risk-taking behavior and its relation to culturally close banks – measured by geographical as well as linguistic distance – is examined. We hypothesize that the level of risk taken by an distinct bank can be explained by the risk-taking behavior of other culturally close banks. Using a complete panel survey of all 1,111 separate and independent German cooperative banks from 2007 to 2010, we show with a high level of significance that banks adapt to the behavior of their culturally defined peer group. Interestingly, linguistic distance is superior to the geographical proxy. Results are robust after controlling for typical macroeconomic, bank specific, and – to eliminate unintentional herding – regional determinants. Our results are also robust to common econometrical and economic specifications. We amend existing literature on geographical herding firstly by a full census of German cooperative banks. As each cooperative bank is privileged with territorial exclusivity, our research is based on an intersect-free full coverage of the entire national territory. Secondly, we are able to refine research on geographical herding by measuring cultural vicinity via linguistic, i.e. dialect, proximity. Based on the evidence of banks selecting their peer group not by a "best-in-class" approach, but rather by dialectical proximity, we can show evidence of irrational herding resulting from psychosociological phenomena, such as mere exposure, as well as conformity effects.

EFM classification code / keywords:
120 - Behavioural Issues / behavioral finance, herding, bank risk-taking, linguistics, cultural area, German cooperative banks

---

[1] Tigerbergstr. 9 | CH-9000 St. Gallen | +41 78 824 33 34 | barbara.seitz@unisg.ch.
[2] Acknowledgements: Earlier versions of this paper were presented (1) at the 3rd CEQURA-Junior Research Workshop 2014 on "Advances in Financial and Insurance Risk Management" on September 30, 2014 organized by the Center for Quantitative Risk Analysis of Ludwig-Maximilians-Universität (LMU) Munich, (2) at the Annual Conference 2015 of the Multinational Finance Society (MFS) from June 28, 2015 to July 1, 2015 in Halkidiki, Greece, and (3) at the Annual Conference of the British Accounting & Finance Association (BAFA) from March 21 to 23, 2016 in Bath, UK. We thank our discussants Wendy Rotenberg, Vasileios Kallinterakis, and all conference participants for many helpful comments. We are grateful to Benedikt Gleich and Christian Stepanek for valuable support.
[3] Universitätsstr. 2 | D-86135 Augsburg | +49 821 598 3040 | andreas.rathgeber@mrm.uni-augsburg.de.
[4] Pilgrimstein 16 | D-35032 Marburg | +49 6421 28 22482 | lameli@uni-maburg.de.
[5] Universitätsstr. 2 | D-86135 Augsburg | +49 821 598 3050 | tobias.gaugler@mrm.uni-augsburg.de.

1. Introduction

We examine the influence of a bank's cultural peer group on its risk-taking behavior. We follow prior studies on (I) geographical herding and finance as well as on (II) cultural distance in financial decision-making. In particular, the role of geographic and linguistic proximity between different banks is key to our study. Examining geographical proximity in the context of finance, Malloy (2005) provides evidence that spatial proximity is positively related to the performance of analysts, Coval and Moskowitz (1999) find equity investors' preference for companies headquartered close to them. Becker, Ivkovich, and Weisbenner (2011) discover that local companies and financiers exploit the advantage of being neighbors by improving corporate governance, in order to overcome agency problems. Besides general studies on finance and geography, a more specific strand of literature uses geographical location as a proxy to describe the peer group- or herding behavior in financial decision-making. Aside from random parallel actions (unintended herding), we focus on intended herding behavior, where decision makers consciously make decisions analogous to other ones. Regarding herding behavior, two general venues can be distinguished: rational herding and irrational herding.

We align our study to both strands of literature and examine them in detail. According to the first, *rational herding* helps to reduce costs by following a neighbor's financial decision. There are several models, like those of Chevalier and Scharfstein (1996), Devenow and Welch (1996), Zeckhauser, Patel, and Hendricks (1991), in which herding is a result of rational behavior by managers due to less costly information collection or lower cost of insolvency in economic downturns. This is especially likely when other firms in the industry are perceived as having greater expertise (Bikhchandani, Hirshleifer, and Welch 1998) or when blame is sharable (Scharfstein and Stein 1990). Taking a step further, Zwiebel (1995) shows that corporate managers' employee types are inferred from their relative performance, because managers perceived to be below a cutoff type are fired. The second idea, *irrational herding*, is deeply rooted in psychological and sociological theories of human behavior. Following a neighbor's decision is considered a valid strategy, due to several behavioral effects. Among psychological effects, e.g. the anchoring effects, the most prominent is to explain herding by the effect of mere exposure (Heath and Tversky 1991, Huberman 2001, Hirshleifer 2001). In a sociological view, the conformity effect manifests in the actions of investors, who learn by observing the actions of others (Hirshleifer 2001). However – rational as well as irrational – herding needs a peer group definition. Several authors use a geographic proximal peer group. However, the peer group typically includes an individual's family, neighbors, friends, or peers, shown by different sociological studies in the spirit of social interaction theory (Mizruchi 1989, Marquis, Glynn, and Davis 2007).

In our study, we define a peer group based out out of a culturally similar region and analyze both geographical and cultural distances. Compared to prior studies, this setting allows us to distinguish between rational and irrational herding. Using a balanced panel of German cooperatives from 2007 to

2010, we address two research questions: *(1) Do cooperative banks switch in their bank risk-taking behavior following a peer group of culturally close banks? (2) Is herding even more present among culturally close banks compared to geographical neighbors?* Based on these questions, we hypothesize that the more similar two banks are from a cultural (i.e. linguistic) perspective, the more similar are their respective risk-taking behaviors. By differentiating between a geographic and a linguistic measurement of distance, we are able to obtain deeper insights into geographical herding and, in addition, whether such herding is irrationally or rationally driven. Unlike the vast majority of related studies, where investigations are conducted across different sectors, we present a comparison only between banks. Since we only focus on companies from one and the same industry, more specific results can be expected. Empirical work on geographical herding focusses on neighborhood effects – meaning similar decision-making in a certain region.

In the case of capital markets, several studies concentrate on neighborhood effects (herding) in the behavior of institutional or retail investors. Hong, Kubik, and Stein (2005) find that mutual fund managers are more likely to buy (or sell) a particular stock in any quarter, if other managers in the same city are buying (or selling) that same stock. Ivkovich and Weisbrenner (2007) and Brown et al. (2008) investigate the importance of geography in explaining equity market participation, concluding that the word of mouth is the major driver. By disentangling social connection from community effect, Pool, Stoffman, and Yonker (2015) report that socially connected fund managers have more similar holdings and trades. Most importantly, they are able to differentiate between informal social communication from formal communication using additional ethnic variables.

Turning to corporate finance, a handful of studies observe also neighborhood effects in investment decision and financial policy. While Dougal, Parsons, and Titman (2015) empirically examine that a firm's investment is highly sensitive to the investments of other firms head-quartered nearby, Gao, Ng, and Wang (2011) provide comprehensive evidence on the local effects of corporate financial policies. Especially the first study predicts that the investment expenditures of neighboring firms move together in response to the up and downs of local economic conditions though networks (rational herding). Similar to the study of Gao, Ng, and Wang (2011), Addoum et al. (2015) use bankruptcy filings as a proxy and discover that companies adjust their finance and investment policies after a bankruptcy occurs in their geographical neighborhood; they thereby conclude the existence of behavioral herding. One study concentrates on neighborhood effects in corporate payout policies as an example of special corporate finance policy. In a setting of German public banks, Rathgeber and Wallmeier (2011) observe that banks are more likely to pay dividends, if the neighboring banks pay dividends and vice versa (neo-institutional framework). Looking at special issues of corporate finance vis-à-vis the role of management, Kedia and Rajgopal (2008) find that the location of a firm's headquarters can explain variation in broad-based option grants. In a similar setting Bouwman (2014) empirically examines whether geography affects CEO compensation and finds statistically and economically significant effects. Kedia, Koh, and Rajgopal (2015) discover that firms are more likely to initiate accounting policies after the public announcement of a restatement by another firm in their industry or neighborhood. In a recent paper, Parsons, Sulaeman, and Titman (2015) even show that financial misconduct is contagious. The general idea is that the misbehavior of one actor can change perceptions of acceptable behavior of others, which is again a typical sociological interpretation of the effect.

Half of the studies find *irrational* herding behavior in corporate financial policy and interpret this effect via psychological or sociological effects, whereby these studies exclude other geographical factors, such as regional industry clustering by geographic idiosyncrasies in labor markets, etc. The other half interpret the results in terms of *rational* herding. In almost all cases, the geographically close firms serve as a peer group connected via a social network or other network partners, like e.g. geographically close sports stars in robustness checks. Geographical closeness is designated often by a location in the same district, area, etc., or by using data on distance relative to real headquarters. Some studies include differentiation of leaders and followers in the peer group analysis. In particular, in the first stage, a few leaders adopt a practice, which is then adopted by followers in contact with the leaders.

In some sense, our study is also close to the study of Leary and Roberts (2014) showing that peer firms influence the determination of corporate capital and financial decisions and policies. In their analysis, the authors examined which firms and CEOs mimic their respective peers, which firms are mimicked, and which factors influence the mimicking behavior. Leary and Roberts (2014) interpret their results with help of several neo-institutional models on corporate financial policy (Chevalier and Scharfstein 1996, Bikhchandani, Hirshleifer, and Welch 1998). Augmenting this field of research, we find further evidence for behavioral finance as we include the cultural dimension in our study. The general supporting idea is that financial decision makers adopt the behavior of other financial decision makers, when they are culturally close to each other. The latter could – but do not have to – be geographically close. Reasons can be e.g. rational, through fewer costs due to information advantages, or sharing of similar institutions and legal structures (Barkema and Vermeulen 1997). The behavioristic argument is based less on a familiarity bias from a psychological viewpoint but rather on the conformity effect in the psych-sociological viewpoint (e.g. Shiller 1998). Both arguments lead to the fact that cultural differences may influence the definition of the peer group. However, there are no studies on the peer group definition via direct cultural differences. Firstly, studies explain the foreign or domestic bias influenced by cultural distance on portfolio holdings, and partly in cases of mergers and acquisitions. Secondly, there are studies on cross listing as a financing decision influenced by cultural distance.

Firstly, focusing on capital market-related literature, Chan, Covrig, and Ng (2005) examine mutual funds of 26 countries and found robust evidence to support the statement that a disproportionately larger fraction of investment in domestic stocks can be explained by familiarity variables (distance between countries as well as common language), while controlling for other variables show significant effects on the domestic bias and foreign bias. They attribute this significant effect to investment barriers, the extent of informational asymmetry between foreign and domestic investors, and in association the different deadweight costs for foreign investors. Recent studies refine the results by applying different measures for cultural distance. Whereas Beugelsdijk and Frijns (2010) and Anderson et al. (2010) largely rely on Hofstede's (2001) measures of individualism and uncertainty avoidance, Aggarwal, Kearney, and Lucey (2012) show how different dimensions of cultural distance interact with geographic distance in shaping cross-country foreign investment patterns, including both aggregate and disaggregated Kogut-Singh measures of cultural distance. This coincides with results from Ahern, Daminelli, and Fracassi (2015), who find that the greater the distance is between two countries along each of the Hofstedes cultural dimensions, the smaller the volume of cross-border mergers between the countries is. Additionally, Grinblatt and Keloharju (2001) determine that investors in Finland (a country where both Finnish as well as Swedish are official languages) are more likely to hold, buy, and sell the stocks of Finnish firms located near to the investor, that communicate in the investor's native tongue, and that have chief executives of the same cultural background. The effect is less prominent among institutional investors and for large firms. Comparable to our study, the work by Grinblatt and Keloharju (2001) reflects this effect within Finland, where the information and barrier arguments are not as striking as in the international context. Consequently, the familiarity bias seems to be a more reasonable explanation, based on the empirical observations. Secondly, we examine the management decision side of corporate finance, where a connection between the results of home bias and those of international cross listings can be observed. Whereas in the first case diversification might play a central role, in the second case cross-listing activity is more common across markets in which diversification gains play a minor role. Studies like that of Sarkissian and Schill (2004) have examined the market preference of firms listing their stocks abroad. Factors that influence overseas-listing decisions are those related to geographical proximity of the foreign market along with other variables that proxy for the degree of familiarity (language-based measure). The idea is that a culturally-different shareholder base can expose the company to shareholders with different cultural values and beliefs, which can then influence corporate financial decisions. Employing Hofstede's cultural dimensions, Dodd, Frijns, and Gilbert (2015) found that firms from developed countries display greater cross-listing propensity towards culturally-similar countries. In summary, effects of cultural proximity are frequently investigated and verified in financial research.

In our case, the situation differs in three aspects. We look at herding among banks in Germany: First, we analyze independent firms from one industry via a full census. As each firm is privileged with territorial exclusivity, our research is based on an intersection-free full coverage of the entire national territory; as such, we have no regional industry clustering to control for. Second, every investor has access to all annual financial statements of all banks in Germany and therefore is able to compare bank risk-taking behavior for all banks. Third and not unimportantly, the same legal system (bank law) and same written language are consistent over the whole of Germany and thus present only low barriers to the flow of information. Consequently, investors as well as each bank's management have access to and can interpret the information from a peer group of other banks. Hence, they could specify the makeup of peer groups by several criteria, e.g. geographic closeness, or "best-in-class". Among others, language is an established proxy for cultural proximity, as Egger and Lassmann (2012) demonstrate in a meta-analysis. Due to the fact that bank management shares the same mother tongue, language is, not suited in our study as a proxy for intra-national cultural proximity. However, the similarity of dialects could fulfill this need. Falck et al. (2012) measure cultural proximity by using a quantification of the co-occurrences of individual dialect features among German regions (Lameli 2013). Hereby the dialect is more than just a measure of similarity in language, it is a measure of culture and transmission of knowledge, values, and other factors that influence behavior from one generation to the next, via teaching and imitation (Boyd and Richerson 1985). In this regard, Germany is a special case in Europe. Due to territorial fragmentation in Germany until 1871, various economic and social networks were established more within the fragments of Germany as among the fragments. The results of this fragmentation are still present in cultural life, economic relations, and the dialects spoken. Consequently, Falck et al. (2012) find that intra-national gross migration flows are positively affected by historical dialect similarity in Germany. Using the same measure, Lameli et al. (2015) find that local dialects have a positive impact on regional trade within Germany, while Bauernschuster et al. (2014) find that risk-loving and better-educated people are more willing than others to cross cultural borders as determined by dialect within Germany. In line with these studies, we hypothesize that bank management chooses its peer group along the cultural dimensions delineated by dialect similarity.

To examine the relationship between bank risk-taking behavior and cultural location, we conduct a full sample survey of all 1,111 German cooperative banks from 2007 to 2010. Via the degree of risk-taking – generally a complex management decision, for which the inability of the bank management to perfectly measure or observe the determinants of the optimal risk-taking behavior comes into play – we expect bank risk-taking, in particular for locally-oriented cooperatives, to be linked to cultural herding. With this sample, we cover 60 percent of all banks residing in Germany; more importantly, we analyze a full coverage of the entire national territory, without intersections. Each of the 1,111 cooperative banks is organizationally independent of one another. In their entirety, they represent a self-contained set appropriate for investigation, because other types of banks (public banks of similar size, major banks, specialized banks or branches of foreign banks) differ substantially in terms of risk-taking. Bank risk-taking can be measured either using a capital market approach (Anderson and Fraser 2000, Akhigbe and Whyte 2003, Flannery and James 1984) or via accounting ratios (e.g. Delis and Kouretas 2011, Jin et al. 2013, Fernández and González 2005). We follow the latter and benefit from the fact that state authorities require publication of a large number of bank risk ratios, and apply the following three measures to comprehensively proxy risk: (I) The solvability ratio ($SOLVA$) to measure the risk appetite of a bank, (II) the Z-Score (Z) to proxy a bank's distance from insolvency (e.g. Jin et al. 2013, Niu 2012), and (III) securities (SEC) to capture trading risk. Distance is measured via dialect similarity between the different banks' headquarters. The idea behind this measure is to quantify the co-occurrences of individual linguistic features among German regions as shown in the historical "Sprachatlas des Deutschen Reichs" (Lameli 2013). Looking at an distinct bank i at time t, $DISTSOLV_{it}, DISTZ_{it},$ and $DISTSEC_{it}$ represent three cultural distance dependent risk variables of the neighboring banks. $DISTSOLV_{it}$ refers to the interrelation between distance and the solvability ratio, while $DISTZ_{it}$ and $DISTSEC_{it}$ depict the interrelation between the Z-Score and securities, respectively. For banks close to more risk-averse banks in the nearest cultural neighborhood, we find a 17.14% higher solvability ratio, a 17.77% higher Z-Score, and a 38.43% higher securities ratio. Similar to other studies (Parsons, Sulaeman, and Titman 2015) analyzing herding behavior, we face the so-called reflection problem (e.g. Manski 1993), which applies when trying to infer whether the average behavior in some group influences the behavior of the individuals comprising the group. Due to its endogenous nature, we solve this problem using a 2SLS approach for a dynamic panel data model. We find that a bank's risk-taking behavior has a highly significant influence on the risk-taking of culturally nearby banks. This effect is persistent while controlling for typical macroeconomic, bank-specific, and – to eliminate unintentional herding – regional determinants. For $DISTSOLV_{it}$, $DISTZ_{it}$, and $DISTSEC_{it}$ we obtain highly significant positive coefficients, indicating that banks within a cultural cluster tend to behave in a more risk-seeking manner: $DISTSOLV_{it}$ = 0.221 (p-value < 0.001), if proximity between banks is measured by the similarity of dialects, $DISTZ_{it}$ = 1.662 (p-value = 0.011) and $DISTSEC_{it}$ = 72.058 (p-value < 0.001). In line with related studies, we use a variety of control variables for the empirical analysis. We apply population density per square km of the state in which a bank is located (POP_{it}), growth rate of gross fixed capital formation in the region ($GFCFGR_{it}$), bank size ($SIZE_{it}$), and capitalization of a bank (CAP_{it}) as regional and bank level control variables. The percentage of the German credit hurdle (CH_t), the 3m EURIBOR on annual average (IRS_t), and the GDP growth rate ($GDPGR_t$) are applied as macroeconomic control variables. Controlling for these characteristics, the influence of the culturally-close peer group is still valid. We proxy cultural closeness both by linguistic distance (measured via dialect proximity) as well as geographically (measured via great-circle distances between banks' headquarters). Linguistic, rather than geographical distances, result in stronger indications of more intense herding. To test the validity of our results, we conduct several robustness tests for the validity of the economic model (14 tests) and the econometric model (10 tests); seventy-two (72) tests are conducted in total. Our results are generally robust to both a series of changes in the empirical methodology and to various model specifications. Our empirical findings strongly support the assertion that the risk-taking behavior of a bank is a function of the risk-taking behavior of its cultural peer group, even dominating effects from an alternative geographical peer group. This effect clearly underpins the behavioristic theories explaining herding behavior.

We contribute to the existing literature with regards to following the following strands, beginning with the literature of herding in corporate finance. Following Pool, Stoffman, and Yonker (2015), we introduce, in addition to a geographical measurement of distance, a cultural distance measure to disentangle geographical from cultural effects. Comparable to Leary and Roberts (2014), we use their idea of having a peer group based on networks. We extend the literature in clearly advocating for the behavioral reasoning behind herding. Secondly, we contribute to the literature of cultural closeness in relation to economic decision-making by measuring cultural closeness via a linguistic measure within a country. We build on the work of Sarkissian and Schill (2004) on cultural closeness in corporate financial policy, and extend it by assessing cultural closeness within a country following Grinblatt and Keloharju (2001). The choice of a peer group follows cultural closeness within a country, where information asymmetry and barriers are less prominent. We further find bank risk-taking to be influenced by cultural differences following Shiller (1998), who proposed that people pay much more attention to ideas or facts reinforced by conversation, ritual, and symbols. This enables us to connect irrational herding, resulting from psychosociological phenomena like mere exposure and conformity effects, with cultural distance. Thirdly, we contribute to literature on bank risk-taking. Bank risk-taking studies are mainly bank-specific studies on e.g. efficiency (Fiordelisi, Marques-Ibanez, and Molyneux 2011) and capitalization (Dell'Ariccia, Laeven, and Marquez 2014), micro studies on e.g. competition (e.g. Boyd and Nicoló 2005), or macro studies related to e.g. regulation (Jin et al. 2013). We enlarge these studies by relating cultural herding with bank risk-taking behavior, and interpreting risk-taking in light of cultural peer group behavior. We further examine the solvability ratio as an additional bank risk-taking factor, which to date hasn't been analyzed in other studies. Fourthly, we contribute to literature about economic geography, by analyzing fragmentation in human behavior. The results of this fragmentation are still present in cultural life, various dialects, and economic relations (Falck et al. 2012, Bauernschuster et al. 2014, Lameli et al. 2015). It has been shown that culture seems to be a driving force in economic exchange. In their study on bilateral trust, Guiso, Sapienza, and Zingales (2009) show that, e.g., the history of conflicts, as well as religious, genetic, and somatic similarities of populations clearly influence trade between two regions. The empirical problem is how to define a good proxy for cultural proximity. We find evidence that this fragmentation is also present in bank behavior, when measured via linguistic proximity.

The remainder of the paper is organized as follows. We discuss the econometric model, including the definition of variables and our estimation methodology in section 2. In section 3, data is presented. Empirical results emerge from section 4, in which we focus on both geographical and linguistic distance measures. A variety of robustness tests are conducted in section 5. We discuss our findings, especially the superiority of linguistic measures over geographical ones, in section 6, and conclude in section 7.

2. Econometric model and research design

In the following section, we show that a fixed-effects model is most suitable to analyze underlying panel data. To provide valid results, both distance and risk are proxied by a variety of appropriate measures. Applying an instrumental variable approach, we also solve problems related to endogeneity (reflection problems).

Before our empirical model can be expanded to its final form in equation (2.1d), the general model to be estimated is given by:

$$r_{it} = \alpha_1 r_{i(t-1)} + \zeta_1 g_{it} + \delta_1{}' b + \gamma_1{}' ma + u_{it}, \tag{2.1a}$$

where the bank risk variable r_{it} of bank i, $i \in N, N = \{1, \dots, n\}$ in time t, $t \in \{1, \dots, T\}$ with $T = 4$ and time in years is $r_{it} \in \{SOLVA_{it}, Z_{it}, SEC_{it}\}$ and is written as a function of lagged bank risk $r_{i(t-1)}$, our distance variable, $g_{it} \in \{DISTSOLV_{it}, DISTZ_{it}, DISTSEC_{it}\}$, a set of regional and bank level control variables $b = (POP_{it}, GFCFGR_{it} SIZE_{it}, CAP_{it})'$, and macroeconomic control variables, $ma = (CH_t, IRS_t, GDPGR_t)'$, are common to all banks (see later in this section for operationalization and discussion of variables). Accordingly, we impose time independent effects for each entity possibly correlated with the regressors, which allows us to control for unobserved heterogeneity. As our model in (2.1a) is a panel data model on bank risk-taking, the residual is given by $u_{it} = \mu_i + \varepsilon_{it} + \lambda_t$, where μ_i is the individual bank-dependent effect but not time-dependent effect called individual effect or heterogeneity, and ε_{it} is the error term with $iid \sim (0, \sigma^2)$. We use a one-way-error component model, i.e. the time dependent error component $\lambda_t = 0$. Year dummies yr are included in the further robustness analysis as well.

Concerning panel data, two assumptions can be made: (1) the random-effects assumption states that the individual specific effects are uncorrelated with the independent variables. (2) The fixed-effect assumption states that the individual specific effects are correlated with the independent variables (e.g. Hausman and Taylor 1981). To determine whether the data contains fixed or random effects, we test the panel regressions against an unweighted ordinary least squares (OLS) estimation. For the random effects model, we conduct a Lagrange Multiplier test (Green 2012). For the fixed effects model, we conduct a simple F-test. Subsequently, we use the Hausman-Wu test to verify that the fixed effects model dominates the random effects model. This is not surprising, because the structure of our data implies that the bank specific effects are most probably correlated with our banks' specific as well as regional variables.

Moreover, we test the residuals for autocorrelation within the fixed effects model. Therefore, we use a modified Durbin-Watson test according to Bhargava, Franzini, and Narendranathan (1982) in association with Baltagi, Song, and Koh (2001). To check for heteroscedasticity, we apply an adjusted Breusch-Pagan-test according to Juhl and Sosa-Escudero (2014) in the fixed-effects model. To address autocorrelation as well as heteroscedasticity, we use so-called Rogers robust estimators for the standard errors. According to Petersen (2008), this estimator is particularly appropriate in the case of firm effects shown here. Additionally, we try the Stock/Watson correction in a further analysis (Stock and Watson 2008). However, our major concern in model (2.1a) is an endogeneity problem, as follows (Bouwman 2014, Leary and Roberts 2014): We propose that the distinct bank follows neighboring banks in its risk-taking behavior. Reversing the argument, the neighboring banks also imitate the behavior of the distinct bank. Mathematically spoken this means we add to equation (2.1a) the equation

$$g_{it} = \alpha_2 g_{i(t-1)} + \zeta_2 r_{it} + \delta_2{}' bg + \gamma_2{}' ma + v_{it}, \qquad (2.1b)$$

leading to a simultaneous equation model containing (2.1a) and (2.1b) to measure the influence of g_{it} on r_{it} without the influence of r_{it} on g_{it}. The equation (2.1b), which can also be used to construct the instrumented variable $\hat{g}_{it}$, comprises the lagged variable $g_{i(t-1)}$, the risk variable of the distinct bank r_{it} (which serves in this equation modelling as our instrumental variable), the macroeconomic variables ma, and the regional and distance dependent bank specific control variables $bg = (gPOP_{it}, gGFCFGR_{it}, gSIZE_{it}, gCAP_{it})'$, which pertain to the neighboring bank. Thus, to measure the risk-taking behavior of the neighboring bank, the equation (2.1b) is constructed in the same way as measuring the risk-taking behavior of the distinct bank. By introducing a second equation, we "killed two birds with one stone" due to the fact that an instrumental variable approach is also inevitable to measure the possible endogeneity and to re-estimate the equation after observing endogeneity. In order to test for endogeneity, we follow Green (2012) by applying an instrumental variable approach. To construct our instrumented variable we solve the equation system leading to the reduced form, whereby the second equation of the reduced form reads as

$$g_{it} = \beta_1 g_{i(t-1)} + \beta_2' bg + \beta_3' ma + \beta_4' b + v_{it}. \qquad (2.1c)$$

The resulting univariate version of the Durbin-Hausman-Wu specification (Green 2012) shows the conjectured endogeneity in most cases. The only exceptions are the models in which we use differencing in the time series (see robustness tests). Finally, yet most importantly, we have to use a two stage least squares approach. Firstly, we estimate equation (2.1c), yielding the instruments $\hat{g}_{it}$, and secondly, we estimate (2.1a) using this instrumented variable (including also lagged bg). Our final model reads as

$$r_{it} = \alpha_1 r_{i(t-1)} + \zeta_1 \hat{g}_{it} + \delta_1{}' b + \delta_2{}' bg + \gamma_1{}' ma + u_{it}. \qquad (2.1d)$$

We use three different variables to proxy bank risk-taking behavior and to capture different aspects of bank risk. The first one is the solvency ratio $SOLVA_{it}$, which measures the risk appetite of a bank. As of 2008 it is known as "solvency ratio" or "solvability ratio", and before as "Grundsatz I Kennziffer" (in German language). The higher the ratio, the more risk-averse a bank's risk-taking is, and vice versa, as it is calculated with the ratio of regulatory capital to weighted risk assets. The second variable is the Z-Score Z_{it}, which indicates with a higher value more bank stability and risk-averse behavior of the bank, and vice versa. It is a common proxy for insolvency risk or a bank's distance from insolvency and widely used in bank risk-taking literature (Agoraki, Delis, and Pasiouras 2009, Bourgain, Pieretti, and Zanaj 2012, Delis and Kouretas 2011, Hadad et al. 2011, Jin et al. 2013, Nguyen 2013, Niu 2012, Pathan 2009, Zhang, Wang, and Qu 2012). The score is calculated by $Z_{it} = \ln\left(\frac{ROA_{it}+EA_{it}}{\sigma(ROA_{it})}\right)$. We use the natural logarithm to address the natural skewness of the Z-Score, e.g. Houston *et al.* (2010). The first component is the return on assets of bank i in t, the second component is equity on assets of bank i in t, and the last component is the volatility of ROA_{it}, where we use a "high-minus-low" proxy for $\sigma(ROA_{it})$ due to our four years panel, given by $\sigma(ROA_{it})_i = \max_{t\in\{1,\dots,T\}}(ROA_{it}) - \min_{t\in\{1,\dots,T\}}(ROA_{it})$. The third bank risk-taking measure is securities SEC_{it} of a bank i measured by the asset amounts of bills, bonds, and shares. This variable is our proxy for trading risk, and is often mentioned in the annual reports of cooperatives as an instrument of risk adjustment. A reduction in securities shows more risk-averse behavior of a bank, and vice versa. However, we want to use a consistent result interpretation of all bank risk-taking proxies, where a higher risk proxy indicates lower bank risk-taking behavior (as for $SOLVA_{it}$ and Z_{it}). Therefore, we scale the value of securities with the total assets of a bank and use the inverse as final ratio. Concluding, all three proxies of bank risk-taking behavior are higher if the behavior of a bank is more risk-averse, and lower in the case of more risk-seeking behavior. With a time-lag of one year, we also use the risk proxy variables as indicators of the speed of adjustment of a bank's risk-taking behavior. We expect the sign to be positive and the coefficient to be between zero and one, as an indication of how fast a bank adopts its own risk-taking. A dynamic model is characterized by these time-lagged variables and is well-pronounced in related studies (Delis and Kouretas 2011, Agoraki, Delis, and Pasiouras 2009) [6].

Regarding the distance analysis, we expect that the risk-taking behavior of nearby banks has an influence on the risk-taking behavior of bank i in t if the proxy of risk-taking is similar between these nearby banks. We follow Addoum et al. (2015), Bouwman (2014), and Rathgeber and Wallmeier (2011), who find evidence for this effect in the payout policy of German savings banks. We adapt their methodology in measuring regional effects by calculating distance functions $distsol_{it}(r_{distvar})$, $distz_{it}(r_{distvar})$, and $distsec_{it}(r_{distvar})$, in which each function refers to one of

[6] Following these studies, we estimate the model (2.1d) using a System Generalized Method of Moments estimator following Arellano and Bover (1995) and Blundell and Bond (1998), instrumenting $r_{i(t-1)}$ to address endogeneity concerns related to the lagged dependent variable. However, due to our specific research design, the main results are based on a 2SLS approach without instrumenting $r_{i(t-1)}$.

our three risk-taking proxies. For use as variables, the function is converted to the distance variables $DISTSOLV_{it}, DISTZ_{it},$ and $DISTSEC_{it}$. The distance function indicates the distance-weighted risk measure of the neighboring banks. Consequently, to generate the variable of the distance function, we take the average risk proxies of the banks within a radius, aggregate them with respect to the radius, and to obtain scaled values between 0 and 1, normalize this figure by dividing it by the length of the domain $maxdist\text{-}\ d(1{,}2)$ (Rathgeber and Wallmeier 2011)

$$g_{it}(SOLVA) = \sum_{k=1}^{N-1} ar_{kt}(SOLVA)\frac{dist(k, k+1)}{maxdist - d(1{,}2)} \tag{2.2}$$

where $ar_{kt}(SOLVA)$ is defined by the series $ar_{kt}(SOLVA) = \frac{(ar_{k-1t}(SOLVA)(k-1)+r_{kt}(SOLVA))}{k}$ and the $dist(k, k+1)$ is the distance between two banks. Furthermore, we state $dist(N-1, N) = maxdist - dist(N-2, N)$ whereby the maximum distance is set to 1000, i.e. the approximate maximum diameter of Germany. In case the distance between two banks is zero $dist(k-1, k) = 0$, the average risk measure $0.5r_{k-1t}(SOLVA) + 0.5r_{kt}(SOLVA)$ is applied and the number of banks is reduced accordingly. We include all "neighboring banks" *N-1* to calculate the measure. However due to construction, banks at great distance are assigned only a low weight.

[Figure 1]

To illustrate the methodology, figure 1 shows an example bank. The distinct bank is surrounded by ten banks with $SOLVA$ of 0.11-0.22. In a radius of 14 km the average $SOLVA$ of the neighboring banks is 0.136. In a radius of 15 km the average $SOLVA$ $ar_{kt}(SOLVA)$ increases to average 0.145. However this results in only a small increase in the distance weighted $OLVA$ $g_{it}(SOLVA)$, which is dominated by the two neighboring banks and their high-distance weights, leading to 0.135 for $g_{it}(SOLVA)$. Insofar, the $g_{it}(SOLVA)$ is the simple L^1 −norm of the unweighted solvabilty ratios $ar_{kt}(SOLVA)$. These calculations exemplarily underpin the huge influence of the nearest banks on the value of the measure.

As the measure in (2.2) is the crucial point of our analysis, we provide various alterations (cp. Supplementary Information SI). Furthermore, we have two different alternatives to measure the distance: A linguistic and a geographical one. Our primary measure is the linguistic distance measure, interpreted as cultural distance. The idea behind this measure is to quantify the co-occurrences of individual linguistic features between German regions (Lameli 2013), resulting in an index of linguistic similarity – proven to be a very good proxy for culturally similarity according to economics literature (Falck et al. 2012, Bauernschuster et al. 2014, Lameli et al. 2015). The data is taken from the historical "Sprachatlas des Deutschen Reichs" (Linguistic Atlas of the German Empire), created during the second half of the 19th century. Via questionnaires, approximately 45,000 locations of the German Empire were surveyed by translating Standard German sentences into local dialects. The realizations of 66 prototypical features related to pronunciation and grammar are quantified within each region and compared between the regions (Lameli 2013) leading to a measure of 66 for absolute identical dialects and 0 for unfamiliar dialects (see similar Chan, Covrig, and Ng (2005). For compatibility reasons we transform the similarity measure into a distance measure by taking the difference of 1,000 and the-quotient-of-1,000-and-66 times the original measure (Germany diameter ≈ 1,000 km). From a historical perspective, dialects store cultural information typical to the separation or assimilation of social groups (e.g. religion, territorial associations, local or rural customs), and reflect historical interactions between people of different locations. In the time of the data exploration primarily dialects were used in everyday language throughout the German Empire (Standard German as a spoken language was established only in the 20th century) this data captures the most detailed differentiation of cultural space in Germany. In particular, dialects are a measure of culture and transmission of knowledge, values, and other factors influencing behavior from one generation to the next one (Boyd and Richerson 1985).

Our second measure is strictly a geographical one. The distance of one bank to another one is measured via the great circle distance (Addoum et al. 2015, Bouwman 2014)[7]. To account for differences in the distance measures (geographical vs. linguistic) we included both measures in our regression (2.1d). Due to the high collinearity of both distance measures, we orthogonalize the measures by regressing the linguistic measures over the geographical distances. Including the regression residuals $\widehat{res}_{it}$, which capture the effects of the linguistic distances beyond the geographical ones in the model, yields

$$r_{it} = \alpha_3 r_{i(t-1)} + \zeta_3 \hat{g}_{it} + \varsigma_3 \widehat{res}_{it} + \delta_3{}' b + \gamma_3{}' ma + u_{it}. \qquad (2.3)$$

In case of both distance measures, we expect a positive sign for the variables $DISTSOLV_{it}, DISTZ_{it},$ and $DISTSEC_{it}$ of the orthogonalized vector $\widehat{res}_{it}$ (linguistic distances) and of the instrumented vector $\hat{g}_{it}$ (geographical distances). A higher value of the distance variables shows a higher percentage of culturally or geographically close banks (peer group banks) with a similar risk proxy, and vice versa. We assume that this fact may give banks some confidence in continuing with current business strategies, and taking on more risk, due to a kind of "regional confidence". Less risk-seeking behavior is due to lower bank risk-taking proxies. In addition, we expect to see some nonlinearity in the imitating. Banks with risk measures deviating by only a small amount from the risk measures of the neighboring banks have little or no tendency to adjust their risk measure. In contrast, banks with risk measures deviating by a huge amount from the risk measures of the neighboring banks need to adjust their risk measure immediately. To control for these possible nonlinearities we substitute the distance variable with its natural logarithm and further address some construction biases

[7] More detailed calculation steps and equations are provided in the Appendix.

in the robustness tests. To acquire an impression of the distance variables, we plot the variables ar_{kt} for different risk proxies depending on the radius measured in the linguistic distance in figure 2.

[Figure 2]

Thereby, the average ar_{kt} is divided into 5 quintiles for each risk proxy of the distinct bank. For all cases and proxies, we can derive from figure 2 that a higher distinct risk proxy coincides with higher variables ar_{kt} representing the risk proxy of the neighboring banks. However, the effect is much more pronounced when the neighboring banks are located within a radius of less than 200 km.

Our regional controls POP_{it} and $GFCFGR_{it}$ of the vector g are all related to the state of a bank. POP_{it} is the variable of population per square kilometer of the 16 German states, where the respective banks are located. It is calculated for each year. $GFCFGR_{it}$ gives the growth rate of the gross fixed capital formation of the state a bank is located in, calculated per year. In some robustness tests EWG_i is a dummy variable of the former East and West German states during the time of the Iron Curtain from 1961 to 1989. We expect some influences from history on today's banking business – especially as cooperative banks have been integrated into the socialist planning system, and banking in former East and West Germany had been separated until the fall of the wall in 1989. The East and West German dummy variable is valued at one for a bank in a former West German state and zero for a bank in a former East German state. It is calculated for each individual and constant over time. The majority of banks are in former West German states (see figure 3)[8]. Furthermore, we use bank level controls: a bank's asset size $SIZE_{it}$, the natural logarithm of the EUR value of its total assets, is a widely used control for size effects of a bank (Fiordelisi, Marques-Ibanez, and Molyneux 2011). We expect a negative relation to our dependent variables; the bigger a bank, the more sensitive is its reaction to market conditions (Niu 2012, Saunders, Storck, and Travlos 1990) and the greater the possibilities of taking on higher risks, and vice versa. The capitalization structure CAP_{it} of a bank is measured as equity plus 0.5 times special items, with an equity portion plus participation certificates, to total assets. We therewith follow the literature (Pathan 2009, Delis and Kouretas 2011) and account for the strong influence of capital structure on bank risk-taking. We expect a positive relation with our bank risk-taking proxies. A higher capitalization indicates good soundness and structure of the bank, leading to higher risk proxies (i.e., more risk-averse behavior), and vice versa.

The vector ma gives the macroeconomic control variables, which we include to control for global economic effects and business conditions within the German economy during the years 2007 to 2010. In this time economic conditions were very challenging due to the global financial crisis. We control

[8] A short note should be made on the assignment of the cooperative bank in Berlin (bank code 100 900 00). The assignment of the bank is unclear, as Berlin was split by the wall through the middle of the city. As the amount of banks in the former East states is negligibly small, we decide to assign Berlin to the former West German states to reduce its bias.

for the credit hurdle CH_t, which reports the percentage of companies in Germany that feel credit lending is restrictive. We expect a negative relation to our risk proxies as a higher value indicates worse economic conditions, and vice versa, which is why risk-seeking bank behavior and therefore lower risk-taking proxies are assumed. The short-term lending rate IRS_t is the 3month EURIBOR on annual average in percent. Following e.g. Agoraki, Delis, and Pasiouras (2009), we expect a positive relation as a low interest rate environment increases bank risk-taking, and vice versa. The last control is the GDP growth rate $GDPGR_t$, to account for economic growth during the panel periods (Hadad et al. 2011, Fiordelisi, Marques-Ibanez, and Molyneux 2011). We expect a negative relation to our risk proxies as more growth bears the chance of more business, which also comes with more risks, and vice versa.

We use alternative macroeconomic controls in the robustness tests. NPL_t is a control for non-performing loans (NPL) in Germany, representing the NPL market potential in billion EUR relative to the total credit volume in Germany of credit banks without credits to banks in billion EUR. We include this control in the robustness tests, as one of the main risk proxies is the ratio of non-performing-loans to total loans NPL/TA (Stojanovic, Vaughan, and Yeager 2008). Moreover, it is an indicator of backward-looking realized credit risk. We expect a negative relation as bank risk-taking increases with NPLs. The last control is the inflation rate $INFLR_t$ (Houston et al. 2010). We expect a negative relation to our bank risk-taking proxies as the risk-taking of a bank might increase with business conditions under higher inflation rates, and vice versa. Table 1 summarizes all variables in use.

[Table 1]

3. Data

Banks in Germany are regulated and controlled by the Federal Financial Supervisory Authority (BaFin), the German Federal Bank and the European Central Bank (ECB). Based on §26 of the German Banking Act (Kreditwesengesetz, KWG) each bank is periodically obliged to submit its financial statements, management reports, as well as audit reports to supervisory authorities. The study is based on a panel dataset of 1,138 legally independent German cooperative banks over the period of 2007 to 2010. All banks are located in one single legal area, within which all residents speak the same language. The specificity that each cooperative bank is privileged with territorial exclusivity is of particular relevance to our study. Our research is thus based on intersection-free full coverage of the entire national territory. Using financial statements, we analyze each balance sheet and income statement of each German cooperative bank over this period. All data is free and publicly available, albeit mostly hand collected. The completeness of the data, a direct consequence of the disclosure requirements in German Banking Act, makes it a new data set. Furthermore, the complete coverage of one banking industry and the resulting comparability among these banks gives it a unique setting.

Cooperative banks differ from other financial institutions mainly in their ownership structure. All stakeholders with an account at a cooperative bank are potential members and owners of the bank who elect their board of directors in a "one-person-one-vote system", regardless of amount invested. However, it misses controlling stakes or the stock market as an instrument to discipline the management. Cooperative banks are characterized by low absolute firm sizes, rapid structural changes due to a high number of mergers (also an instrument to avoid bankruptcy), relatively high numbers of branch offices, their high geographic coverage with branches, and by volumes of loans which are often significantly lower than the volumes of deposits (Lang and Welzel 1996). Based the specific organizational form, the amount of liable capital (tier 1 and tier 2 capital) of German cooperative banks is relatively fixed. Associated therewith, each cooperative bank can (and is constrained to) make its own risk policy – independently from other cooperative banks within the legal framework. In detail, a bank's risk policy is regulated by (a) its top management, (b) its middle management, (c) its supervisory board, and (d) its supervisory authority. While the former is recruited from the entire federal territory, both middle management, supervisory board and authority stem from regional states. We plot the geographical distribution of German cooperative banks over the period 2007-2010 in figure 3.

[Figure 3]

We see a concentration in the regions of former West Germany, compared to just 39 banks in former East German states. However, cooperative banks represent nearly 60% of all German banks – by far the majority of the German banking sector. Other German banks are not easily comparable to German cooperative banks. The second largest group of banks (more than 20%) is public banks, governmentally owned, with different capital structures and other legal restrictions on risk-taking behavior. The remainders are specialized banks like mortgage banks, investment banks, major banks, or branches of foreign banks also not directly comparable in their risk-taking behavior. After cleaning the data for mergers and acquisitions, we obtain a total of 4,444 observations, resulting in 1,111 observations per year over the period 2007-2010. Descriptives for the main variables and altered distance measures are reported in Table 2.

[Table 2]

The solvency ratio $SOLVA_{it}$ differs between 10.4% and 25% – already high values, as the required value of bank regulation is eight percent. The data on the securities SEC_{it} indicates that the total assets of the banks ranges between 2.12 and 16.39 times more than the absolute value of securities. The big difference may be explained with structural characteristics of cooperative banks as they focus on small business with wide regional distribution. Some banks may not have any kind of securities in their balance sheets. Although negative values of Z_{it} occur due to logarithmic scaling, the original minimal value is still a very small positive. For the distance variables $DISTSOLV_{it}$, $DISTZ_{it}$, and $DISTSEC_{it}$, we find values in a range similar to that of the corresponding risk proxies. Values differ between linguistic and geographical distance measures, which is a result of the different ways distance is defined. Concerning our controls, we reckon structural differences in the bank-level controls $SIZE_{it}$ and CAP_{it}, as the data report high individual fluctuations. Further, all geographic and macroeconomic control variables show high variation, e.g. the GDP growth rate ($GDPGR_t$), the credit hurdle (CH_t), or the short-term interest rate (IRS_t). We expect the variation to occur due to our analyzed crisis period and because the general economic status was highly influenced and quite volatile. Further, we reckon a high negative growth rate in $GFCFGR_{it}$, which is already -24.97% for one state (that of Hesse (German "Hessen") in 2009). The big difference together with the enormous standard deviation indicates tough economic times. The standard deviation for the credit hurdle CH_{it} is remarkably high as well, and points to tough lending conditions during our panel period. Regarding our correlation analysis[9], we find high correlation between CAP_{it} and Z_{it}. The variable CAP_{it} shows the capitalization of the bank while Z_{it} is a proxy for insolvency risk. As they are naturally dependent from each other, we accept the correlation. However, we do orthogonalize the variables with high correlations occurring between the variables CH_t and $GFCFGR_{it}$, CAP_{it}, IRS_t, and $GDPGR_t$ and between the variables $GFCFGR_{it}$ and $GDPGR_t$.

4. Empirical results

Our results of estimating the models in (2.1c), (2.1d) and (2.3) with the two-step estimator are provided together with the conducted model fit tests in table 3.

[Table 3]

Concerning the geographical matter, we find that a bank's risk-taking behavior has a highly significant influence on the risk-taking of nearby banks within the same cluster, as reported by the variables $DISTSOLV_{it}$,$DISTZ_{it}$, and $DISTSEC_{it}$. Affiliation to the cluster is given in cases of similar risk-taking proxies of the banks, calculated with the distance function and variables. The highly significant results of the distance variables on all risk proxies emphasize the result. Additionally, the findings suggest that banks within a regional cluster tend to behave in more risk-seeking behavior if the neighboring banks are risk seeking, too (and vice versa), in line with our expectations. In all models, the instrumental variable approach for the distance variables is applied because the endogeneity test clearly indicates serious endogeneity. The validity of the instruments is tested via a Sargan-test (Sargan 1958), presented in table 4. Altogether we could not find a significant correlation among the residuals of our 2SLS and the tested instrumental variables.

[9] The correlation matrix is available upon request.

[Table 4]

The highly significant influence of a specific bank's risk-taking behavior on the risk-taking of its nearby banks is also e.g. documented by the regional distribution of $SOLVA$ in Germany in figure 4. Here some clusters of high $SOLVA$ e.g. in (a) Northern Hesse, (b) North-Thuringia / Western Saxony, (c) Northern Baden-Wuerttemberg, and (d) Central Lower Saxony can be observed, whereas clusters of low $SOLVA$ can be found in (e) Baden or at the (f) coastal area.

[Figure 4]

The logarithmic model performs (see reset test) better than the linear model, which might be due to the nonlinearity in the nature of the proposed risk-taking behavior of neighboring banks. Lastly yet importantly, the variables based on dialects have higher explanatory power than the measures based on geographical distance. On the one hand, the geographical based measures show significant regression coefficients. On the other hand, the significance vanishes after inclusion of the orthogonalized dialect-based measures. To analyze the difference in more detail, we perform several tests. First, we look at the major differences between the linguistic and geographical distances, which are (first decile) generated by banks situated on borderlines between the major dialect regions in Germany.

[Figure 5]

In this regard, figure 5 shows a classification of German dialects representing the distribution around the year 1900. Although various migratory movements have taken place since then, the linguistic picture in terms of dialects in the German domain has remained largely intact. First, we find, for instance, that especially on the border between the Rhine Franconian area ("Rheinfränkisch") and Swabia ("Schwäbisch") – one of the major borders of the whole language area – a strong differentiation of the banks' risk-taking can be seen, also indicating a border for this phenomenon. Second, we find the smallest differences between the two measures (last decile) for banks located more in the centers of the dialect areas (e.g. Cologne, South east Bavaria) indicating processes of convergence for bank risk-taking within these areas. We interpret these results as an effect of cultural familiarity. Consequently, we construe this result as the previously mentioned bank managers' and board members' policy effort to optimize their relative performance to their cultural peer group. However, the choice of the peer group follows a sociologic phenomenon. Third, we analyze the differences between the two (i.e. geographic and linguistic distance) measures with respect to the neighboring banks. Due to the construction of the distance variable the 10 neighboring banks have a big influence on the value of the measure. Hence, we compare the set of the 10 neighboring banks in each measure and construct a measure of similarity with a domain of 0 to 10 leading to the variable sim_{it}. Hereby, 10 is the maximum value in case of two equal sets of neighboring banks, whereas 0 stands for an empty intersection. Due to the discrete structure, we apply Spearman's correlation and correlate year-wise and separately – for each risk variable $SOLVA_{it}, SEC_{it}$, and Z_{it} –

the absolute differences between the geographic and linguistic distance measures on the measure of similarity in table 5. We find in each year and for each variable clearly negative and significant correlation coefficients.

[Table 5]

Altogether, we conclude that the differences in both measures are generated by the differences in the neighboring banks. Due to the nature of our model our variables might contain a unit root and consequently our results might be spurious. However, we account for this phenomenon to some degree by including yearly dummy variables in a robustness test. Alternatively and more notable, we conduct a panel unit root test, namely the ADF-tests according to Maddala and Wu (1999) and Choi (2001). Both tests are based on an aggregation of the individual F-statistic using different weighting schemes. Whereas Maddala and Wu's test statistic follows the chi²-distribution, Choi's statistics follows the central limit theorem. According to table 6, we can reject a unit root for all time-series, except for the distance variables for SEC_{it}. This is true for both test statistics. As we cannot reject the unit root hypothesis, we can also not conclude the existence of a unit root, especially when looking at the relatively high test statistics for the Maddala/Wu test. In addition, a potential unit root in the distance variables would not lead to spurious results due to the stationarity of our dependent variable. All in all, this shows the robustness of our results.

[Table 6]

In the previous part of this section, we focus on empirical results related to distance variables (as reported in the uppermost third of table 3). In the following, we refer to the control-related results, gathered in the middle third of table 3. Regarding the control variables, we observe mixed results. The lagged dependent variables $SOLVA_{i(t-1)}, Z_{i(t-1)}$, and $SEC_{i(t-1)}$ are significant in rare cases. This indicates that the importance of the speed of adjustment in risk-taking measured via $SOLVA_{it}$ and Z_{it} and SEC_{it} is negligible, whereby the result is not driven by our instrumental variable approach. This can be shown comparing the results with the results of not using the instrumental variable. However, $SOLVA_{i(t-1)}, Z_{i(t-1)}$, and $SEC_{i(t-1)}$ are excluded from the estimation of the instrumental variable. Common control variables of regional, macroeconomic, and bank level factors indicate strong influence on risk-taking. We often find different directions on the regional controls POP_{it}, and $GFCFGR_{it}$, coinciding our expectations. However, we can only report significant coefficients on POP_{it} (positive) and $GFCFGR_{it}$ (negative) in case of SEC_{it}. This slightly changes after orthogonalization of the growth rate of the state gross fixed capital formation. For Z_{it} and SEC_{it}, coefficients are sometimes significant. The two bank level controls $SIZE_{it}$ and CAP_{it} often report the expected relations and also significant coefficients in most instances; however, sometimes they show the opposite sign. The results on $SIZE_{it}$ suggest our expected negative relation, which is that the risk proxies decrease with increasing $SIZE_{it}$ and indicate higher bank risk-taking behavior and vice versa, for $SOLVA_{it}$ and SEC_{it}. However, the relation is significant for the solvency ratio only while the other significant relation is positive and reported for Z_{it}. The positive relation of the capitalization structure variable CAP_{it} is highly significant only for the Z-Score and in line with our expectations that a strong capitalization diminishes risk-taking, and vice versa. The last set of control variables contains the macroeconomic controls CH_t, IRS_t, and $GDPGR_t$. The credit hurdle CH_t reports mostly negative but insignificant coefficients for all three risk proxies. The insignificance is due to some collinearity. After orthogonalization the result changes slightly and we find some negative significant coefficients, in line with our expectations. This indicates that a more restrictive credit lending increases risk-taking behavior of cooperative banks, and vice versa. The short-term interest rate IRS_t reports often insignificant and expected relations for $SOLVA_{it}$ and SEC_{it}. We expect a positive relation where low interest rates increase bank risk-taking, and vice versa. Only for Z_{it} is the relation negative and significant. The same applies to $GDPGR_t$ suggesting the expected negative relation with $SOLVA_{it}$ and SEC_{it}; here, risk-taking behavior increases with GDP growth and vice versa. Concluding our final results, we find strong impact of the distance variables and most control variables on bank risk-taking behavior. Further, our priority for risk proxies is with SEC_{it} and $SOLVA_{it}$ rather than with Z_{it} as they are more powerful regarding results and specifications test (see table 4).

5. Robustness tests

To test the quality of our results, we conduct several robustness tests (each for all three dependent variables, with 72 tests in total) structured in two parts. Part A tests the validity of the economic model (14 tests) while Part B tests the validity of the econometric model (10 tests). Each test is listed with the alternated specification or explanation in table 7. Our major concern is the robustness regarding the measurement of distance variables.

[Table 7]

In Part A, we focus on economic tests concerning our model. In the test A1 we include year dummies instead of macroeconomic variables due to collinearity between macroeconomic time series and yearly dummies. In other words, we perform an econometric model with fixed bank and year effects. Jin et al. (2013) address nonlinearity in bank size and therewith a possible bias by including a square and cube size of the variable in the model. We follow lead of Jin et al. (2013) in accounting for size effects in bank risk measures, by examining in test A2 if the variable $SIZE$ has a nonlinear influence on bank risk-taking. Additionally, we still expect some influences from German history in today's banking business (especially as cooperative banks in the Eastern part of Germany had been integrated into the socialist planning system until 1990). Consequently and due to the significant differences in the data coverage, we exclude the eastern banks from our sample. Lastly, the inspected banks are audited by regionally-organized auditors. These auditors are responsible for all banks in one German state or in a German region consisting of several small states, and might consequently influence the risk-taking behavior of banks in a certain region. Consequently we analyze one of these states and separately focus on the state of Bavaria (A4). This state appears to be suitable as it home of the most cooperative banks is the by far largest state in Germany and also has the largest number of different dialects. For the tests A5 to A7, we use the orthogonalized values of CH_t and $GFCFGR_{it}$ and address the correlation differently. We start with the original data set but with orthogonalized values. Additionally, we replace the macroeconomic controls (where most correlation occurs) with a new vector of macroeconomic controls ma^* given by $ma^* = (NPL_t, INFLR_t)'$. The last alternative is to drop CH_t, CAP_{it}, and $GDPGR_t$ and insert ma^* instead. The last tests A8-A14 address the definition of our distance measure. In A8-A11, we use alternative measures for the risk-taking behavior of the neighboring banks. The simplest one is given by equally weighted risk proxy of the neighboring banks. In case of $SOLVA_{it}$ this leads to

$$g2_{it}(SOLVA) = \frac{1}{k}\sum_{k=1}^{k} r_{kt}(SOLVA). \tag{5.1}$$

As part of our calculations, ee completely varied the number of banks k by 3 to 10 banks. As the discriminative capacity of the measure decreases, and only present the case of 3 and 5 banks in the following. Additionally we define k in A12-A13 based on distance, as $k_{max} = max\left(\underset{k\in\{1,2,\dots N-1\}}{\text{argmax}}(distsort(i,k) < dmax), 2\right)$, where $distsort$ is the distances between the distinct bank i and the neighboring banks in ascending order. The maximum distance $dmax$ was set to different values. However, we remain with 30 km as a typical driving distance for short business trips. To sum up, k_{max} is the number of banks in a 30 km radius around the distinct bank. In case zero banks are in this radius (13 cases for the geographic measure and 117 for the linguistic measure) the value was set to 2. All in all, these measures are coarsening of the measure in our base case as information is lost. In addition to this simple measure, we construct in A14 a measure accounting for the possible nonlinear behavior in risk-taking with respect to the neighboring banks. The idea is, as described above, that a bank only has a tendency to change its behavior in case it deviates largely from the behavior of the neighboring banks. Consequently, the measure indicates zero in a certain range and ±1 outside the range, leading to the following indicator function, which fully measures the distance, but makes a coarsening of the risk measure:

$$I_{kt}(SOLVA) = \begin{Bmatrix} 1 & r_{kt}(SOLVA) > \mu(r) + 0.5\sigma(r) \\ 0 & |r_{kt}(SOLVA) - \mu(r)| \leq 0.5\sigma(r) \\ -1 & r_{kt}(SOLVA) < \mu(r) - 0.5\sigma(r) \end{Bmatrix} \tag{5.2}$$

where $\mu(r)$ and $\sigma(r)$ are the mean and standard deviation of the risk proxy $SOLVA$ over the full sample in t. We choose a range of one standard deviation as a normal risk proxy to have about 50% of the banks with normal deviation and 50% of the banks with abnormal deviations from the mean. The measure itself is defined as the distance-weighted indictor function of the neighboring banks:

$$g3_{it}(SOLVA) = \sum_{k=1}^{N-1} aI_{kt}(SOLVA) \cdot dist(k,k+1) \Big/ (maxdist - d(1,2)) \tag{5.3}$$

where $aI_{kt}(SOLVA)$ is defined by the sequence $aI_{kt}(SOLVA) = \frac{(aI_{k-1t}(SOLVA)(k-1)+I_{kt}(SOLVA))}{k}$. A high measure implies that the neighboring banks deviate largely positively from the mean, and a low measure vice versa. All measures are again estimated according to (2.1d) using the instrumental variables approach. Table 8 illustrates the coarsening of the different measures.

[Table 8]

In Part B, we vary our model from the original specification in two ways: Modifications of the panel model as well as using year-wise cross sections. Instead of using instrumental variables according to (2.1c), using first differences is an alternative methodology to overcome the specification, respectively endogeneity problems. Hence we estimate the model using first differences; however, we drop the lagged risk proxy in order to obtain a comprehensive number for our firm-year-observation. The two-stage version of the Durbin-Hausman-Wu test, applied due to the short time series observed, shows no endogeneity. Hence, we can use also this approach to correct for endogeneity. Due to the fact that the lagged variable sometimes shows surprising behavior, we estimate the base case model without the lagged risk proxy, leading to a static version of equation (2.1d). Additionally, we alter the variance estimator according to Stock and Watson (2008), leading to a slightly different standard error estimator. Lastly, we change the estimation methodology, leading to a simple cross section equation

$$r_{it} = \alpha_3 r_{i(t-1)} + \zeta_3 \hat{g}_{it} + \varsigma_3 \widehat{res}_{it} + {\delta_3}' b + {\gamma_3}' ma + u_i, \tag{5.4}$$

applied year-wise to our dataset, insofar as we could separate years of the financial crisis based on intensity. For our distance variable, we use the instrumented variable $\hat{g}_{it}$ as well as the instrumented variable $\hat{g}2_{it}$. In the latter case we set k=3. However, we try also different numbers of neighboring banks k, leading to similar results. Again, we estimate the cross section regression in first differences instead of applying an instrumental variable approach

$$\Delta r_{it} = \zeta_3 \widehat{\Delta g}_{it} + \varsigma_3 \widehat{res}_{it} + {\delta_3}' \Delta b + {\gamma_3}' \Delta ma + u_i. \tag{5.5}$$

Hereby, the residuals $\widehat{res}_{it}$ are generated from a regression where the dependent and independent variables are defined in first differences. Durbin-Hauman-Wu-test again shows the applicability of this approach to circumvent the endogeneity problems. In table 9, we present the basic results of the robustness tests to discover whether there is a difference to the base case[10].

[10] Corresponding tables to the robustness tests are available upon request. In addition to the presented robustness test, we conducted a System Generalized Method of Moments (GMM) estimation of Arellano and Bover (1995) and Blundell and Bond (1998). We follow related literature (e.g. (Delis and Kouretas 2011) in which it is used to address endogeneity problems. Overall, our results remain robust. Estimation details are available upon request.

[Table 9]

Regarding the significance and signs of the results in Part A, we have no deviations at all for $SOLVA_{it}$ and SEC_{it} and only one for Z_{it}, which strengthens the economic rationale behind our analysis. The only differences are in the controls and lags, which deteriorate especially in case of the Z_{it}. This relation coincides largely with the findings in Part B. The results in test A2 with different types of $SIZE_{it}$ stay the same for our main variables; however, the significances in the controls differ slightly. We cannot indicate a clear nonlinear relation of $SIZE_{it}$ to bank risk-taking. Moreover, the results of the tests A4 to A7, in which we apply the orthogonalization and use other ways to address the correlation in the control variables, are robust. We can therewith show that the correlation in these variables does not affect the consistency of the estimator, as we still receive the same results for our main explanatory variables as with other ways of addressing correlation. However, the results on the alternative macroeconomic variables $NPL_t, INFLR_t$ in ma^* appear neither to be important nor to report a clear direction. We further find mixed results, although we expect two negative relations with bank risk-taking behavior. It strengthens our original choice of macroeconomic variables, which report significant estimates mostly according to our expectations. This coincides with the fact that we cannot observe any difference after including yearly dummies instead of macroeconomic variables. To control for possible effects from risk-taking behavior in Eastern and Western Germany, we run regressions solely for Western Germany and observe the same results as in West Germany. Also restricting the group of observed banks to Bavarian banks alters the results only in terms of significance due to the smaller sample size.

One of the strongest proofs from the robustness tests are the overall (in Part A and B), constant, and highly significant coefficients of our main cultural regressors. The distance variables indicate strong and high robustness of the cultural findings. Most importantly, the different distance variables show similar results. We especially observe that the linguistic distance measure is superior to the measure based on geographic distances, underpinned by the majority of significant residuals (orthogonalized linguistic) (see e.g. A1-3 or B1-4). Significance remains also in the cases of simpler measures. The results are more pronounced in case of the 3 neighboring banks instead of 5 neighboring banks. This could be interpreted as banks imitating smaller peer groups. Again, the nonlinear measure performs (especially for SEC_{it}, and $SOLVA_{it}$) better than the linear, which can be proved by a reset test. However, it is not as significant as in the base case. This coincides with the fact that the model in A14 performs well in the linear version. With this definite test results, we can prove the reported evidence for the regional effects and the matter of culture for different model specifications, regarding econometric as well as economic variations.

In Part B, we conduct sensitivity analyses concerning the panel model (B1-B4; whereby distance variables are measured directly in case B1 and B2, instead of the instrumental variable approach used

otherwise) and some tests on cross section (B5-B10). Using these techniques we try to avoid the generation of “apparently valid results” that are actually invalid. This may lead to false-positive findings – a common danger of the applied estimation method. In tests B5 to B10 we receive quite robust results for different model specifications. Using a different distance measure (here 3 neighboring banks) (B9 and B10), applying a linear model (B6), and first differencing instead of using instrumental variables (B7-B8) affects only the linguistic measures in rare cases. However, the geographical based measures are more unstable, and the model seems highly sensitive with respect to the control variables. Especially for the robustness test of bank variables, expected signs change. The opposite holds for the lagged dependent variable. In addition to that, the nonlinear model performs slightly better than the linear. Besides, the East and West German control is sometimes significant for all three dependent variables with a negative relation. This suggests that the location of a bank in former West Germany has an increasing effect on risk proxies, indicating a lower bank risk-taking behavior, and vice versa. Finally, the models coincide more with our theoretical derivations in the year 2010 than in the year 2008, which might be a result of the financial crisis.
Another issue with potentially false-positive results in our instrumental variable approach is that our instruments are misspecified. Hence, as alternative we use the first differencing method to stress our instrumental variable, by testing the endogeneity by a Durbin-Wu-Hausman-test. The test yields no endogeneity in cases of $SOLVA_{it}$ and SEC_{it}. However, in some cases for Z_{it} the test statistic implies a p-Value at the borderline of significance. The values indicate a valid model specification strongest for SEC_{it}, again followed by $SOLVA_{it}$, and acceptable validity for Z_{it}. Again we find significant results for the linguistic distance variables and the orthogonalized version. All in all, these results underpin the robustness and strengthen validity of our analysis. Lastly, we test the base case also in the non-dynamical version without lagged dependent variables and alter the standard error estimator. Again, there is no fundamental change in the results. Concluding the robustness tests, we can prove robust results to different model specifications, economic alternatives ,and a high and noticeable robustness in the distance variables.

6. Discussion

Regarding our first research question, if cooperative banks vary in risk-taking behavior following culturally similar peer groups, we find support in results of the distance variables and find empirical evidence for cultural herding effects in German cooperative banking, even while controlling for geographic variables. With regards to our second research question, we find support that cultural distance measures show an even stronger herding effect compared to geographical distance measures. Our results contribute to several different strands in literature.

Firstly, we contribute to the literature in bank risk-taking by finding a new determinant influencing bank risk-taking behavior. However, concerning control variables, evidence is rather mixed, e.g. we find coinciding results for IRS_t with Delis and Kouretas (2011) or for $SIZE_{it}$ with Niu (2012). We can partly confirm the influence of the lagged risk proxies as found by Delis and Kouretas (2011) and Agoraki, Delis, and Pasiouras (2009) if results are estimated in differences in the instrumental variable regressions. We assume that different results occur due to methodological deviations, e.g. in the instrumental variables techniques.

Regarding the herding of corporate finance managers we confirm the results of several authors (Bouwman 2014, Kedia and Rajgopal 2008, Kedia, Koh, and Rajgopal 2015, Addoum et al. 2015, Parsons, Sulaeman, and Titman 2015). We especially see similarities to the paper analyzing financial policy (Leary and Roberts 2014, Dougal, Parsons, and Titman 2015, Gao, Ng, and Wang 2011). That the choice of capital structure is comparable to our results is due to the equity ratio character of $SOLVA$ as a direct consequence of the leverage effect. Hence, a bank managers' strategy is oriented to the behavior of the neighboring banks, by assuming they hold superior information – as managers in the industrial sector do in the case of capital structure. These results may be interpreted in a neo-institutional context, some in a behavioral context. In the first context, it is a cost-saving method to choose a capital structure or a risk exposure in line those of superior market participants or even to share blame, if the allegedly-superior market participant also fails. In the second context, managers emulate the capital structure or risk-taking of other market participants due to psychological or sociological effects like anchoring, mere exposure, and conformity effect. More specifically, our results on the distance variables align with the study of Rathgeber and Wallmeier (2011), as an example of regional clustering in savings banks as a critical determinant for dividend payout policy. Along this line, we find regional clustering as an important determinant for the risk-taking behavior of cooperative banks and develop the findings of Rathgeber and Wallmeier (2011) regarding three aspects: instead of savings banks, we analyze cooperative banks, and instead of the payout policy we examine risk-taking behavior. Most importantly, we find a dominant cultural peer group effect. The use of this variable is also the central distinctive feature in comparison to literature on geographic herding in corporate finance and capital markets (see in the latter case Pool, Stoffman, and Yonker (2015), who disentangle social connection from community effect). Here, peer groups are most of the time chosen by geographic closeness or by the largest companies.

Contrarily, our peer group choice seems to be inspired by cultural closeness. While Felbermayr and Toubal (2010) use, e.g., voting data for the European Song Contest to construct a measurement of cultural proximity, we contribute to the literature analyzing financial decisions by cultural distance variables.

As Chan, Covrig, and Ng (2005) point out, the investment decision may be driven by a familiarity bias. According to Sarkissian and Schill (2004) this is also true for cross-listing decisions. Several studies show that cultural distances can explain cross-listing decisions and foreign bias in portfolio choice. We observe such a cultural closeness also in the choice of the investigated banks, which are choosing their peer on the basis of cultural affinity and confirm the literature's previous findings in another setting as a sort of robustness test.

A rational herding model can explain our results. We could interpret the observed herding with the model of Zwiebel (1995), that bank managers' types are inferred from their relative performance, because managers perceived to be below a cutoff type are fired. However, it is more likely that a behavioristic aspect comes into play. This impression is corroborated by the fact that the peer group is chosen by cultural distance variables. However, we cannot decide if the choice of the peer group is the choice of (a) its top management, (b) its middle management, or (c) its supervisory board. Due to the clear cultural closeness of the peer group we interpret our results more in a behavioristic sense.

Lastly, another interpretation in the vein of economic geography or cultural studies cannot be excluded. That is, that the banks' actors – be they from the region under discussion by origin or not – are oriented along cultural ties formed over many generations and this cultural background still seems to be intact today. Independent of whether the actors are aware of the cultural impact or not, they fall back on geographical differentiation – arguably nothing more than a geographic distinction of aggregated individual traditions of many sorts of verbal and non-verbal interactions. This is, in other terms, an effect of in-group favoritism (Tajfel 1970). Hence, according to this interpretation banks managers or bank owners do not follow their peer group in risk-taking behavior. Their risk preference is determined by cultural factors (e.g. uncertainty avoidance index of Hofstede). Therefore, the measured effect is lastly the result of the fact that different risk-taking behavior is present in different cultures, and therefore a clustering can be observed. Although we cannot rule out that cultural distance itself influences the risk-taking behavior, we do not believe that this aspect is relevant to our investigation: The cultural distance between the regions studied in our paper is relatively small in relation to the magnitude of the observed effect.

7. Conclusion

Banks' locations and economics can potentially influence their risk-taking behavior in a number of ways. While banking literature emphasized the importance of capital, regional, and global economic conditions, more recent literature in corporate finance emphasizes influence from neighbors, which motivates our analysis. In particular, we conjecture that bank managers follow a culturally close peer group in risk-taking behavior. We conduct a full survey sample among all German cooperative banks from 2007-2010 using the solvability ratio, the Z-Score, and securities to proxy bank risk-taking behavior. We find banks' risk-taking behavior to be significantly influenced by the risk-taking of neighboring banks, i.e. the peer group. Therein, we focus more on cultural than geographical closeness. This setting enables us to further support the behavioristic arguments of irrational herding in finance. The aspect of proximity is measured as *cultural* proximity, i.e. the similarity of dialects, and, additionally, as *geographical* distance (based on kilometers).

With the help of a two stage least squares approach we find that cultural closeness can even better explain a bank's risk-taking behavior than possible when only referring to geographical distances. Compared to geographical variables, homogeneity within a cluster is even greater if linguistic variables are considered; although the cultural basis (66 dialects) is coarser than in the case of our continuous geographical distance variable. We control for macroeconomic, bank level, and regional factors. After conducting two different sets of robustness tests, we can report robust results to different model specifications, economic alternatives, and noticeably robustness in the distance variables. We contribute to existing literature by providing the first empirical evidence for the influence of cultural proximity, measured by linguistic closeness of risk-taking behavior of banks. By linking geographical herding and cultural influences in finance we find support for irrational herding as a cause of cultural peer group effects. Due to choice of the peer group, our findings underline the importance of considering behavioral tendencies, especially cultural identity, in herding behavior and might enable targeted control of these propensities in decision situations.

Due to our specific study design, we face some limitations and options for future research: Firstly, a bank's balance sheet ratios as predictors to analyze risk-taking behavior might be biased due to the specific crisis period of our study. Future research can work on subsample tests before and after the crisis, and can further alter risk-taking measurements for measures regarding the capital market. Secondly, the scope of our study is limited to the German banking system and the particular setting of cooperatives therein. We recommend future studies on banks in other countries or in an international setting. An investigation with similar geographical granularity would e.g. be possible in Italy, since its present national state has developed from a variety of small states relatively recently, similar to the case for Germany.

References

Addoum, Jawad M., Alok Kumar, Nhan Le, and Alexandra Niessen-Ruenzi, 2015, Geographic Proximity and Financial Contagion: Spill-over Effects from Local Bankruptcies, *Working Paper*.

Aggarwal, Raj, Colm Kearney, and Brian Lucey, 2012, Gravity and culture in foreign portfolio investment, *Journal of Banking & Finance* 36, 525–538.

Agoraki, Maria-Eleni K., Manthos D. Delis, and Fotios Pasiouras, 2009, Regulations, competition and bank risk-taking in transition countries, *Journal of Financial Stability* 7, 38–48.

Ahern, Kenneth R., Daniele Daminelli, and Cesare Fracassi, 2015, Lost in translation? The effect of cultural values on mergers around the world, *Journal of Financial Economics* 117, 165–189.

Akhigbe, Aigbe, and Ann M. Whyte, 2003, Changes in market assessments of bank risk following the Riegle–Neal Act of 1994, *Journal of Banking & Finance* 27, 87–102.

Anderson, Christopher W., Mark Fedenia, Mark Hirschey, and Hilla Skiba, 2010, Cultural influences on home bias and international diversification by institutional investors, *Working Paper*.

Anderson, Ronald C., and Donald R. Fraser, 2000, Corporate control, bank risk taking, and the health of the banking industry, *Journal of Banking & Finance* 24, 1383–1398.

Arellano, Manuel, and Olympia Bover, 1995, Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics* 68, 29–51.

Baltagi, Badi H., Seuck H. Song, and Won Koh, 2001, Testing Panel Data Regression Models with Spatial Error Correlation, *Working Paper*.

Barkema, Harry G., and Freek Vermeulen, 1997, What Differences in the Cultural Backgrounds of Partners Are Detrimental for International Joint Ventures?, *Journal of International Business Studies* 28, 845–864.

Bauernschuster, Stefan, Oliver Falck, Stephan Heblich, Jens Suedekum, and Alfred Lameli, 2014, Why are educated and risk-loving persons more mobile across regions?, *Journal of Economic Behavior & Organization* 98, 56–69.

Beugelsdijk, Sjoerd, and Bart Frijns, 2010, A Cultural Explanation of the Foreign Bias in International Asset Allocation, *Working Paper*.

Bhargava, A., L. Franzini, and W. Narendranathan, 1982, Serial Correlation and the Fixed Effects Model, *Review of Economic Studies* 49, 533–549.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, 1998, Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades, *Journal of Economic Perspectives* 12, 151–170.

Blundell, Richard, and Stephen Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* 87, 115–143.

Bosch, Karl, 1998, *Mathematik-Taschenbuch* (R. Oldenbourg Verlag, Munich).

Bourgain, Arnaud, Patrice Pieretti, and Skerdilajda Zanaj, 2012, Financial openness, disclosure and bank risk-taking in MENA countries, *Emerging Markets Review* 13, 283–300.

Bouwman, Christa H., 2014, The Geography of Executive Compensation, *Working Paper*.

Boyd, John H., and Gianni de Nicoló, 2005, The Theory of Bank Risk Taking and Competition Revisited, *The Journal of Finance* 60, 1329–1343.

Boyd, R., and P. J. Richerson, 1985, *Culture and the Evolutionary Process* (The University of Chicago Press, Chicago).

Brown, Jeffrey R., Zoran Ivkovich, Paul A. Smith, and Scott Weisbrenner, 2008, Neighbors Matter: Causal Community Effects and Stock Market Participation, *The Journal of Finance* 63, 1509–1531.

Chan, Kalok, Vicentiu Covrig, and Lilian Ng, 2005, What Determines the Domestic Bias and Foreign Bias? Evidence from Mutual Fund Equity Allocations Worldwide, *The Journal of Finance* 60, 1495–1534.

Chevalier, Judith A., and Davis S. Scharfstein, 1996, Capital-Market Imperfections and Countercyclical Markups: Theory and Evidence, *American Economic Review* 86, 703–725.

Choi, In, 2001, Unit root tests for panel data, *Journal of International Money and Finance* 20, 249–272.

Delis, Manthos D., and Georgios P. Kouretas, 2011, Interest rates and bank risk-taking, *Journal of Banking & Finance* 35, 840–855.

Dell'Ariccia, Giovanni, Luc Laeven, and Robert Marquez, 2014, Real interest rates, leverage, and bank risk-taking, *Journal of Economic Theory* 149, 65–99.

Devenow, Andrea, and Ivo Welch, 1996, Rational herding in financial economics, *European Economic Review* 40, 603–615.

Dodd, Olga, Bart Frijns, and Aaron Gilbert, 2015, On the Role of Cultural Distance in the Decision to Cross-List, *European Financial Management* 21, 706–741.

Dougal, Casey, Christopher A. Parsons, and Sheridan Titman, 2015, Urban Vibrancy and Corporate Growth, *The Journal of Finance* 70, 163–210.

Egger, Peter H., and Andrea Lassmann, 2012, The language effect in international trade. A meta-analysis, *Economics Letters* 116, 221–224.

Falck, Oliver, Stephan Heblich, Alfred Lameli, and Jens Südekum, 2012, Dialects, cultural identity, and economic exchange, *Journal of Urban Economics* 72, 225–239.

Felbermayr, Gabriel J., and Farid Toubal, 2010, Cultural proximity and trade, *European Economic Review* 54, 279–293.

Fernández, Ana I., and Francisco González, 2005, How accounting and auditing systems can counteract risk-shifting of safety-nets in banking: Some international evidence, *Journal of Financial Stability* 1, 466–500.

Fiordelisi, Franco, David Marques-Ibanez, and Phil Molyneux, 2011, Efficiency and risk in European banking, *Journal of Banking & Finance* 35, 1315–1326.

Flannery, Mark J., and Christopher M. James, 1984, The Effect of Interest Rate Changes on the Common Stock Returns of Financial Institutions, *The Journal of Finance* 39, 1141–1153.

Gao, Wenlian, Lilian Ng, and Qinghai Wang, 2011, Does Corporate Headquarters Location Matter for Firm Capital Structure?, *Financial Management* 40, 113–138.

Green, W., 2012, *Econometric Analysis* (Prentice Hall, Upper Saddle River).

Grinblatt, Mark, and Matti Keloharju, 2001, How Distance, Language, and Culture Influence Stockholdings and Trades, *The Journal of Finance* 56, 1053–1073.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2009, Cultural biases in economic exchange?, *The Quarterly Journal of Economics* 124, 1095–1131.

Hadad, Muliaman D., Agusman Agusman, Gary S. Monroe, Dominic Gasbarro, and James K. Zumwalt, 2011, Market discipline, financial crisis and regulatory changes: Evidence from Indonesian banks, *Journal of Banking & Finance* 35, 1552–1562.

Hausman, Jerry A., and William E. Taylor, 1981, Panel Data and Unobservable Individual Effects, *Econometrica* 49, 1377–1398.

Heath, Chip, and Amos Tversky, 1991, Preference and belief: Ambiguity and competence in choice under uncertainty, *Journal of Risk and Uncertainty* 4, 5–28.

Hirshleifer, David, 2001, Investor Psychology and Asset Pricing, *The Journal of Finance* 56, 1533–1597.

Hofstede, Geert, 2001, *Culture's Consequences*. *Comparing Values, Behaviors, Institutions and Organizations Across Nations*, Tilburg University (SAGE Publications, Inc.).

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2005, Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Traders of Money Managers, *The Journal of Finance* 60, 2801–2824.

Houston, Joel F., Chen Lin, Ping Lin, and Yue Ma, 2010, Creditor rights, information sharing, and bank risk taking, *Journal of Financial Economics* 96, 485–512.

Huberman, Gur, 2001, Familiarity Breeds Investment, *Review of Financial Studies* 14, 659–680.

Ivkovich, Zoran, and Scott Weisbrenner, 2007, Information Diffusion Effects in Individual Investors' Common Stock Purchases: Covet Thy Neighbors' Investment Choices, *Review of Financial Studies* 20, 1327–1357.

Jin, Justin Y., Kiridaran Kanagaretnam, Gerald J. Lobo, and Robert Mathieu, 2013, Impact of FDICIA internal controls on bank risk taking, *Journal of Banking & Finance* 37, 614–624.

Juhl, Ted, and Walter Sosa-Escudero, 2014, Testing for heteroskedasticity in fixed effects models, *Journal of Econometrics* 178, 484–494.

Kedia, Simi, Kevin Koh, and Shivaram Rajgopal, 2015, Evidence on Contagion in Earnings Management, *The Accounting Review* 90, 2337–2373.

Kedia, Simi, and Shiva Rajgopal, 2008, Neighborhood Matters: The Impact of Location on Broad Based Stock Option Plans, *Journal of Financial Economics*.

Lameli, A., 2013, *Strukturen im Sprachraum*. *Analysen zur arealtypologischen Komplexität der Dialekte in Deutschland* (De Gruyter, Berlin, Boston).

Lameli, Alfred, Volker Nitsch, Jens Südekum, and Nikolaus Wolf, 2015, Same Same But Different. Dialects and Trade, *German Economic Review* 16, 290–306.

Lang, Günter, and Peter Welzel, 1996, Efficiency and technical progress in banking Empirical results for a panel of German cooperative banks, *Journal of Banking & Finance* 20, 1003–1023.

Leary, Mark T., and Michael R. Roberts, 2014, Do Peer Firms Affect Corporate Financial Policy?, *The Journal of Finance* 69, 139–178.

Maddala, G. S., and Shaowen Wu, 1999, A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics* 61, 631–652.

Malloy, Christopher J., 2005, The Geography of Equity Analysis, *The Journal of Finance* 60, 719–755.

Manski, Charles F., 1993, Indentification of Endogenous Social Effects: The Reflection Problem, *Review of Economic Studies* 60, 531–542.

Marquis, Christopher, Mary A. Glynn, and Gerald F. Davis, 2007, Community Isomorphism and corporate social action, *Academy of Management Review* 32, 925–945.

Mizruchi, Mark S., 1989, Similarity of Political Behavior Among Large American Corporations, *American Journal of Sociology* 95, 401–424.

Nguyen, Tu, 2013, The disciplinary effect of subordinated debt on bank risk taking, *Journal of Empirical Finance* 23, 117–141.

Niu, Jijun, 2012, An empirical analysis of the relation between bank charter value and risk taking, *The Quarterly Review of Economics and Finance* 52, 298–304.

Parsons, Christopher A., Johan Sulaeman, and Sheridan Titman, 2015, The Geography of Financial Misconduct, *Working Paper*.

Pathan, Shams, 2009, Strong boards, CEO power and bank risk-taking, *Journal of Banking & Finance* 33, 1340–1350.

Petersen, M. A., 2008, Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, *Review of Financial Studies* 22, 435–480.

Pool, Veronika, Noah Stoffman, and Scott Yonker, 2015, The People in Your Neighborhood: Social Interactions and Mutual Fund Portfolios, *The Journal of Finance* 70, 2678–2732.

Rathgeber, Andreas, and Martin Wallmeier, 2011, Regionales Clustering im Ausschüttungsverhalten von Sparkassen, *Zeitung für Betriebswirtschaft, ZfB* 81, 1341–1377.

Sargan, J. D., 1958, The Estimation of Economic Relationships using Instrumental Variables, *Econometrica* 26, 393–415.

Sarkissian, Sergei, and Michael J. Schill, 2004, The Overseas Listing Decision. New Evidence of Proximity Preference, *Review of Financial Studies* 17, 769–809.

Saunders, Anthony, Elizabeth Storck, and Nickolaos G. Travlos, 1990, Ownership Structure, Deregulation, and Bank Risk Taking, *The Journal of Finance* 45, 643–654.

Scharfstein, Davis S., and Jeremy C. Stein, 1990, Herd Behavior and Investment, *American Economic Review* 80, 465–479.

Shiller, Robert J., 1998, Human Behavior and the Efficiency of the Financial System, *NBER Working Paper*.

Stock, James H., and Mark W. Watson, 2008, Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression, *Econometrica* 76, 155–174.

Stojanovic, Dusan, Mark D. Vaughan, and Timothy J. Yeager, 2008, Do Federal Home Loan Bank membership and advances increase bank risk-taking?, *Journal of Banking & Finance* 32, 680–698.

Tajfel, H., 1970, Experiments in Intergroup Discrimination, *Scientific American* 223, 96–102.

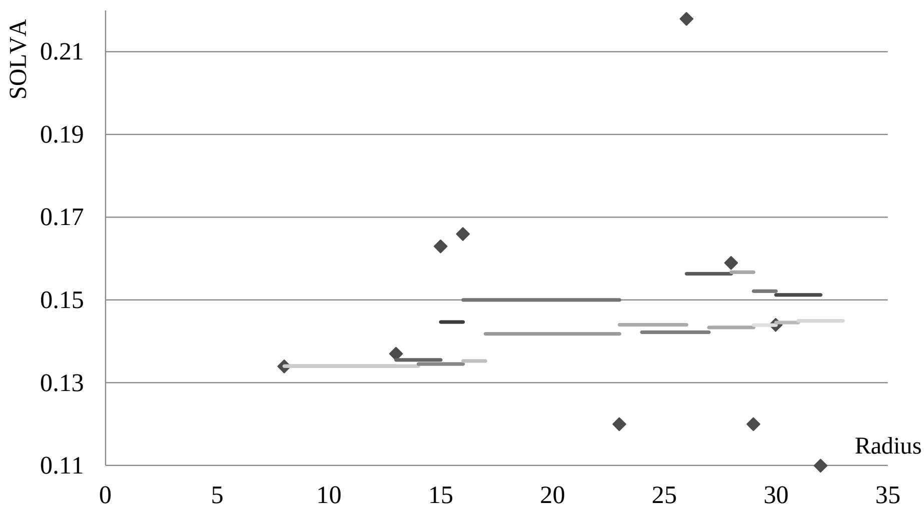
Wiesinger, P., 1983, *Die Einteilung der deutschen Dialekte*. In: Besch, W. et al. (Eds.), Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung. de Gruyter, 2nd half-volume, Berlin, New York, pp. 807–900.

Zeckhauser, Richard, Jayendu Patel, and Darryll Hendricks, 1991, Nonrational actors and financial market behavior, *Theory and Decision* 31, 257–287.

Zhang, Jianhua, Peng Wang, and Baozhi Qu, 2012, Bank risk taking, efficiency, and law enforcement: Evidence from Chinese city commercial banks, *China Economic Review* 23, 284–295.

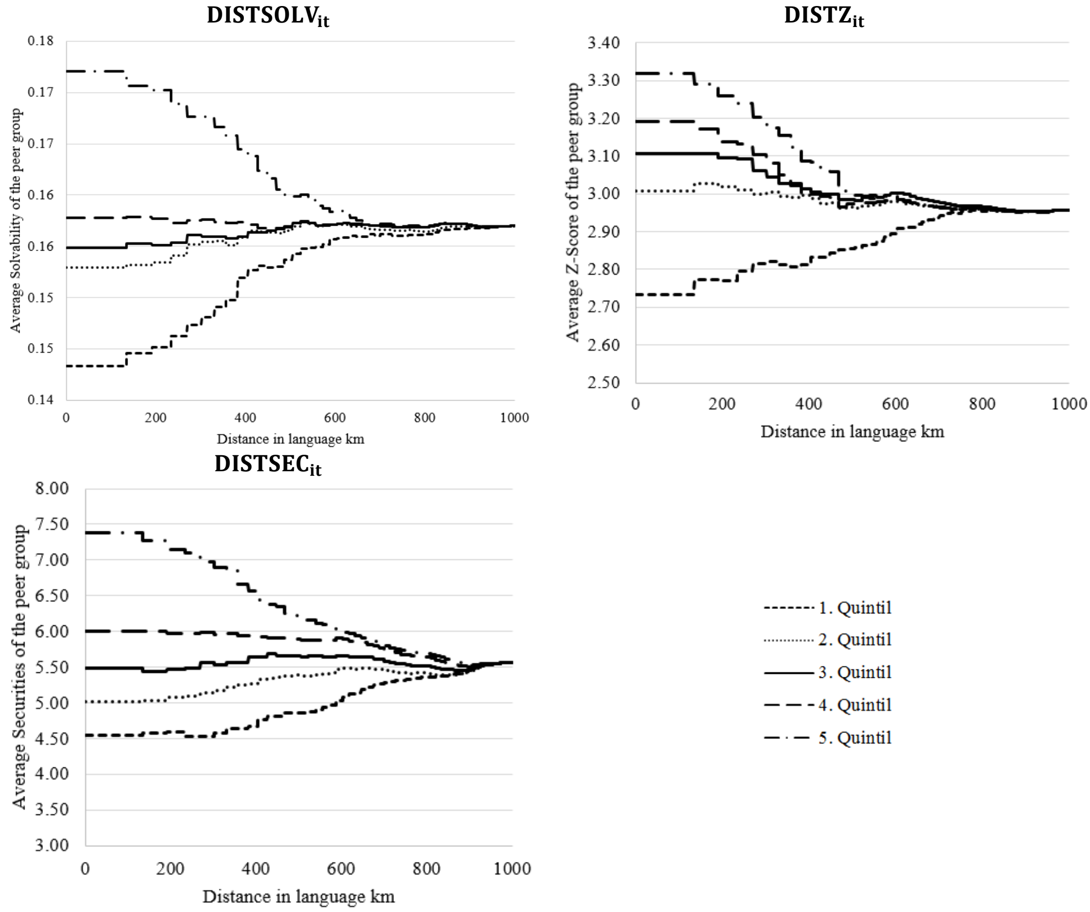
Zwiebel, Jeffrey, 1995, Corporate Conservatism and relative compensation, *Journal of Political Economy* 105, 1–25.

Figure 1: Calculation and radius sensitivity of the distance variable (Solvability Ratio)



Note: This figure is a plot of a distance variable calculated for 1-10 neighboring banks of an distinct bank. The solvability ratio of each neighboring bank depending on the geographical distance is depicted by dots. The average solvability ratio within a radius $ar_{kt}(SOLVA)$ is depicted by the dark grey line. The light grey line represents the distance weighted measure $g_{it}(SOLVA)$.

Figure 2: Radius sensitivity and the distance variables



Note: This figure is a plot of each linguistic variable ar_{kt} calculated for 1000 different radii each. The sample (over the period 2007–2010) is drawn from databases described in section 3 and the values are aggregated over all 4 years. The data set for the distance variables ar_{kt} determined by the neighboring banks' solvency and solvability is separated into quintiles of the risk proxies ($SOLVA_{it}$, Z_{it}, SEC_{it}) of the distinct banks.

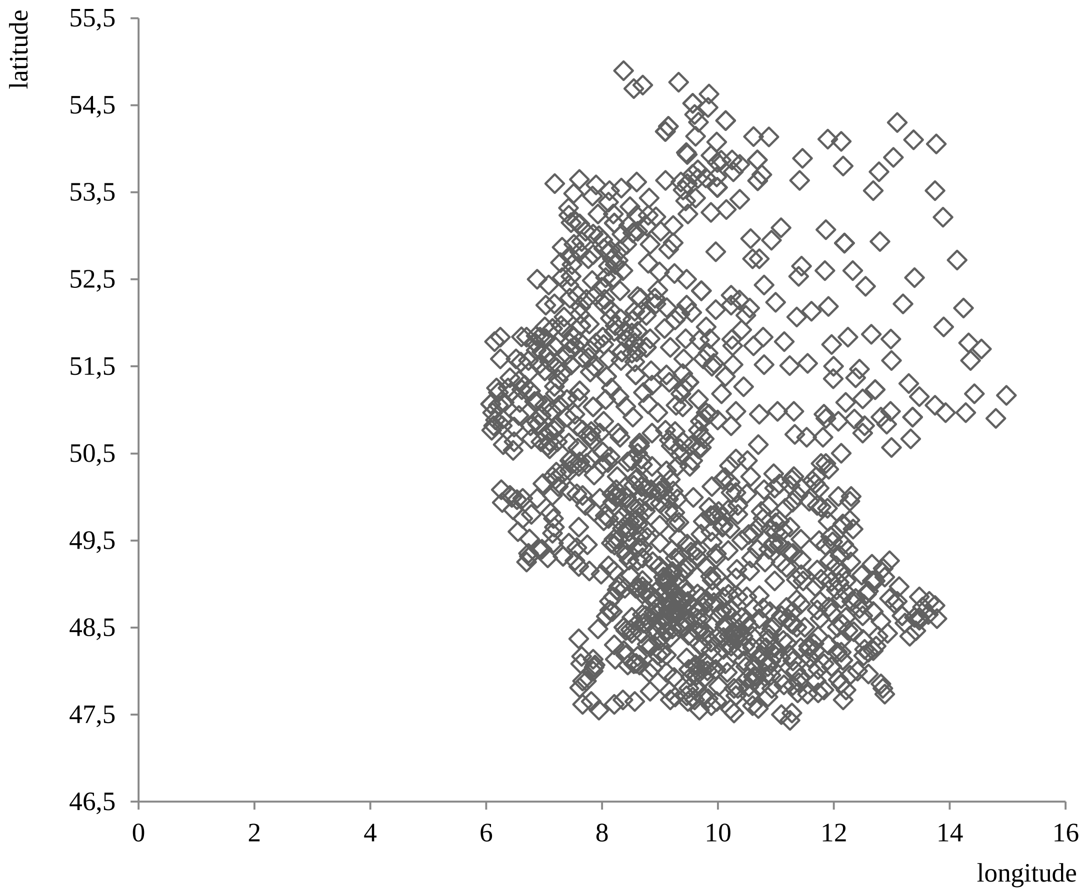
Table 1: Description of variables

Descriptions		***Data Sources 2007 - 2010***	
Panel 1: dependent variables for bank risk-taking behavior			
$SOLVA_{it}$	Solvency ratio: proxy of the general risk appetite of a bank	*Electronic Federal Gazette/ Annual Reports of cooperative banks/ Notes*	
Z_{it}	Z-Score: proxy of a bank's distance to insolvency	*Electronic Federal Gazette/ Annual Reports of cooperative banks/ Balance sheet items*	
SEC_{it}	Securities: proxy of a bank's trading risk	*Electronic Federal Gazette/ Annual Reports of cooperative banks/ Balance sheet items*	
Panel 2: variables for geographic and linguistic analysis		***Geographical***	***Linguistic***
$DISTSOLV_{it}$	Distance variable based on the risk proxy $SOLVA_{it}$	*DeStatis Regional data base for postal codes, odgb data base and latlong.net for latitude and longitude data*	*66 dialects of the "Sprachatlas des deutschen Reichs"*
$DISTZ_{it}$	Distance variable based on the risk proxy Z_{it}		
$DISTSEC_{it}$	Distance variable based on the risk proxy SEC_{it}		
Panel 3: control variables for bank level and regional controls			
POP_{it}	Population per square km of the state in which a	*DeStatis Regional data base*	

	bank is located	
$GFCFGR_{it}$	Growth rate of gross fixed capital formation of the state in which a bank is located	*DeStatis Regional data base*
EWG_i	Dummy of former East and West Germany of bank i, 1 if located in former West, 0 if located in former East Germany	*DeStatis Regional data base*
$SIZE_{it}$	Natural logarithm of bank size in EUR values	*Electronic Federal Gazette/ Annual Reports of cooperative banks/ Balance sheet items*
CAP_{it}	Capitalization of a bank via equity to total assets	*Electronic Federal Gazette/ Annual Reports of cooperative banks/ Balance sheet items/*
g	Set of control variables for bank level and regional controls pertaining to a specific bank	-
bg	Set of control variables for bank level and regional controls pertaining to the neighboring bank	-
Panel 4: control variables for macroeconomic controls		
CH_t	Credit hurdle, percentage of restrictive credit lending	*Ifo Instiut Germany*
IRS_t	3m EURIBOR on annual average in percent, interest rate short term	*Deutsche Bundesbank*
$GDPGR_t$	GDP growth rate	*DeStatis via GENESIS online*
NPL_t	NPL market potential in billion EUR to the total credit volume in Germany of credit banks without credits to banks in billion EUR	*BaFin*
$INFLR_t$	Inflation rate	*DeStatis via GENESIS online*

Note: This table shows all variables used in the regression analysis, clustered in four panels for (1) the dependent variables, (2) variables for the geographic and linguistic analysis including the geographic/cultural distance measures, (3) control variables for bank level and regional controls, and (4) control variables for macroeconomic controls. We provide the variable's name, its description and the corresponding data source within one line.

Figure 3: Geographical distribution of all cooperative banks in Germans, 2007-2010



Note: This figure is a scatter plots of the latitudinal and longitudinal data of all German cooperative banks included in the sample from 2007-2010. It depicts the banks' distribution over Germany.

Table 2: Summary statistics

Variable	Mean	sd	min	max
$SOLVA_{it}$	0.157	0.034	0.104	0.250
Z_{it}	2.958	1.889	-5.759	6.585
SEC_{it}	5.563	3.208	2.118	16.391
$DISTSOLV_{it}$ (linguistic)	0.152	0.018	0.072	0.197
$DISTZ_{it}$ (linguistic)	2.875	1.385	-2.943	5.789
$DISTSEC_{it}$ (linguistic)	5.450	1.355	1.887	10.862
$DISTSOLV_{it}$ (geographic)	0.156	0.007	0.139	0.169
$DISTZ_{it}$ (geographic)	2.935	1.335	1.399	5.278
$DISTSEC_{it}$ (geographic)	5.503	0.793	3.871	7.320
POP_{it}	1.189	0.547	0.726	2.294
$GFCFGR_{it}$	1.168	8.035	-24.974	14.093
EWG_{i}	0.931	0.254	0.000	1.000
$SIZE_{it}$	19.352	1.023	17.487	21.265
CAP_{it}	0.314	0.479	0.000	2.248
CH_{t}	34.496	6.107	26.17	43.419
IRS_{t}	2.738	1.729	0.811	4.634
$GDPGR_{t}$	0.825	3.585	-5.100	4.000
Distance measures used in robustness tests				
$DISTSOLV_{it}$ (linguistic, g2, 3 banks)	0.159	0.023	0.105	0.242
$DISTZ_{it}$ (linguistic, g2, 3 banks)	2.981	1.590	-5.759	6.410
$DISTSEC_{it}$ (linguistic, g2, 3 banks)	5.577	2.232	2.118	15.874
$DISTSOLV_{it}$ (geographic, g2, 3 banks)	0.157	0.021	0.107	0.243
$DISTZ_{it}$ (geographic, g2, 3 banks)	2.970	1.565	-5.759	6.502
$DISTSEC_{it}$ (geographic, g2, 3 banks)	5.583	2.169	2.118	16.391
$DISTSOLV_{it}$ (linguistic, g2, 5 banks)	0.158	0.018	0.110	0.217
$DISTZ_{it}$ (linguistic, g2, 5 banks)	2.968	1.508	-2.638	6.191
$DISTSEC_{it}$ (linguistic, g2, 5 banks)	5.613	1.857	2.153	12.207
$DISTSOLV_{it}$ (geographic, g2, 5 banks)	0.157	0.018	0.109	0.218
$DISTZ_{it}$ (geographic, g2, 5 banks)	2.974	1.481	-4.142	6.295
$DISTSEC_{it}$ (geographic, g2, 5 banks)	5.579	1.910	2.238	13.826
$DISTSOLV_{it}$ (linguistic, g2, 30 km)	0.157	0.023	0.104	0.250
$DISTZ_{it}$ (linguistic, g2, 30 km)	2.966	1.592	-5.759	6.585
$DISTSEC_{it}$ (linguistic, g2, 30 km)	5.631	2.322	2.118	16.391
$DISTSOLV_{it}$ (geographic, g2, 30 km)	0.157	0.016	0.104	0.246
$DISTZ_{it}$ (geographic, g2, 30 km)	2.968	1.447	-5.735	6.231
$DISTSEC_{it}$ (geographic, g2, 30 km)	5.593	1.733	2.118	16.391
$DISTSOLV_{it}$ (linguistic, g3)	-0.003	0.217	-0.688	0.661
$DISTZ_{it}$ (linguistic, g3)	0.007	0.201	-0.714	0.572
$DISTSEC_{it}$ (linguistic, g3)	0.002	0.262	-0.714	0.701
$DISTSOLV_{it}$ (geographic, g3)	0.001	0.019	-0.134	0.150
$DISTZ_{it}$ (geographic, g3)	-0.001	0.017	-0.150	0.063
$DISTSEC_{it}$ (geographic, g3)	-0.008	0.069	-0.357	0.300
No. of observations	**4,444**			

Note: This table provides summary statistics of the main variables, as well as statistics for various distance measures used in the robustness tests. We show the mean, standard deviation, minimum and maximum values of each variable.

Table 3: Main results (Two stage least square)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Distance		Linguistic			Geographic/Linguistic			Geographic			Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables												
LnDistSolva(geo)				0.316			0.791***					
LnDistZ(geo)					13.139			21.496***				
LnDistsec (geo)						-9.753*			52.426***			
DistSolva(lin/ort)	0.221***			0.214***						1.376***		
DistZ(lin/ort)		1.662***			1.412**						0.25**	
Distsec(lin/ort)			72.058***			50.82***						1.168***
Independent Variables: Controls												
Lag dep Var	-0.088***	-0.041	0.052	-0.087***	-0.039	0.052	-0.081***	-0.051	0.05	-0.086***	-0.049	0.054
(u)POP	-0.081	5.498	107.64***	-0.055	-3.958	109.39***	0.054	-7.953	95.706***	-0.15	7.478	8.271***
GFCFGR	<0.001	0.002	-0.019***	<0.001	-0.003	-0.018***	<0.001	-0.006	-0.015***	<0.001	0.003	-0.014**
Size	-0.003	0.412***	-0.904**	-0.004	0.375***	-0.892*	-0.005	0.368***	-0.849*	-0.004	0.424***	-0.95**
Cap	0.038	52.855***	-2.897	0.038	51.688***	-2.92	0.043*	52.748***	-2.776	0.039*	53.791***	-3.226
CH	0.013	-0.302	-1.564	0.012	-0.356	-1.228	0.028	-0.574	-1.116	0.004	-0.311	0.276
IRS	0.01	-0.215	-0.49	0.01	-0.536	0.003	0.025	-0.909	0.554	0.003	-0.207	0.569
GDPGR	0.013	-0.318	-1.563	0.012	-0.192	-1.203	0.026	-0.285	-1.07	0.004	-0.335	0.334
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.7%	80.2%	86.0%	90.7%	80.2%	86.0%	90.6%	80.0%	85.9%	90.7%	80.0%	85.9%
Durbin Watson	2.95***	1.783***	2.801***	2.855***	1.801***	2.645***	2.7***	1.864***	2.681***	2.876***	1.811***	2.568***
LM Hetero	0.397	0.449	0.139	0.017	0.003	0.063	1.112	0.107	0.675	0.269	0.199	0.106
HausmannWu	10975***	461.2***	5351***	10222***	131.6***	3999***	2656***	109***	1234***	129090***	108.1***	4480***
Reset Test										2.59***	19.80***	12.28***

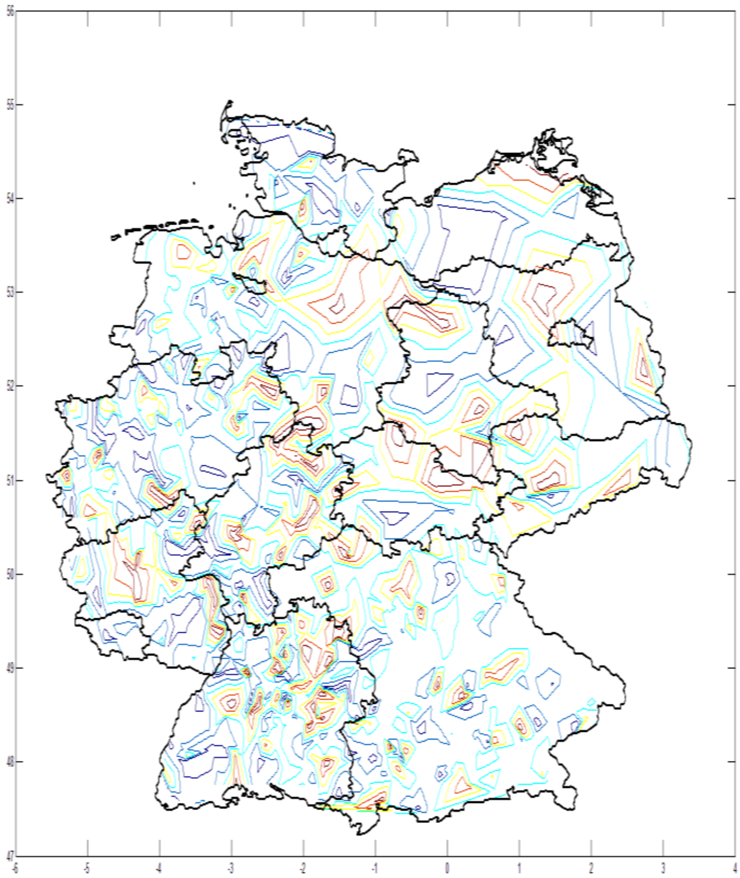
Note: The estimated models are $r_{it} = \alpha_1 r_{i(t-1)} + \zeta_1 g_{it} + \delta_1{}'b + \gamma_1{}'ma + u_{it}$ (2.1d) and $g_{it} = \beta_1 g_{i(t-1)} + \beta_2' bg + \beta_3' ma + \beta_4' b + v_{it}$ (2.1c). The dependent variable in columns 1-12 is as defined in table 1. The sample (over the period 2007–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns (1)-(3) and (10)-(12) the variables are based on linguistic distances (columns (7)-(9) on geographic and (4)-(6) on geographic as well as on linguistic variables). The distance variables in column (1)-(9) are log transformed to address nonlinearity in imitating. Model (10) to (12) use linear ("lna") distance variables. Columns (4)-(6) also contains the residuals after orthogonalizing of $r_{it} = \alpha_3 r_{i(t-1)} + \zeta_3 \hat{g}_{it} + \varsigma_3 \widehat{res}_{it} + \delta_3{}'b + \gamma_3{}'ma + u_{it}$ (2.3) and are the main results of our estimations. The independent control variables of all models with regression coefficients are defined as in table 1. In the fourth to last line the modified Durbin Watson statistic (Panel data) is depicted. In the third to last line the value of the LM-Test for heteroscedasticity and in the second to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. In case of the last three columns a reset test was performed, of which the results are depicted in the last line. All specifications include firm fixed effects as well as a test of whether the data contained fixed effects (positive for all models). Rogers robust standard errors clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table 4: Sargan test final results

Model	(2.1d) log	(2.1d) log	(2.1d) log	(2.3) log	(2.3) log	(2.3) log	(2.1d) log	(2.1d) log	(2.1d) log	(2.1d) lna	(2.1d) lna	(2.1d) lna
Dep. Var.	Solva	Z	Sec	Solva	Z	Sec	Solva	Z	Sec	Solva	Z	Sec
Sargan test	9.57	15.02	5.24	13.65	3.70	10.97	9.81	14.68	4.40	14.42	13.51	3.82

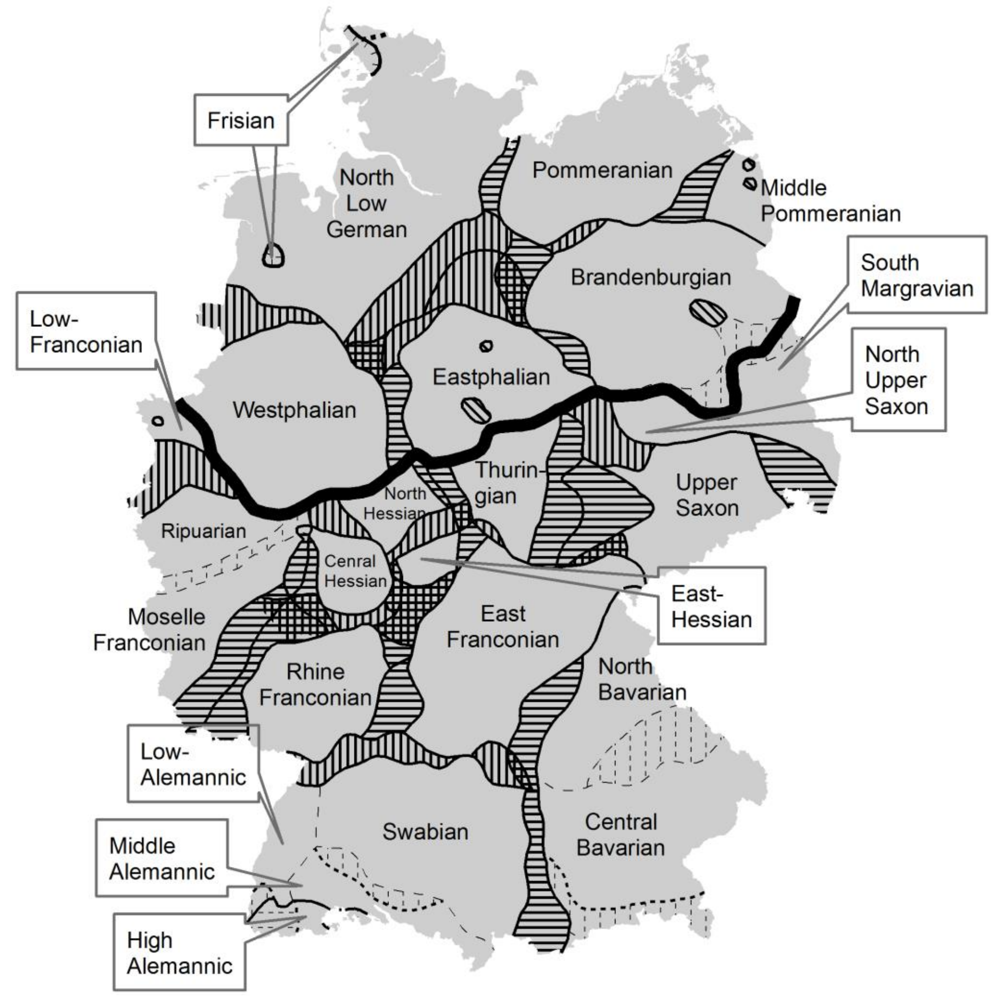
Note: The dependent variable in columns 1-12 is as defined in table 1. The sample (over the period 2007–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent instrumental variables in the 1-3 and 10-12 columns are based on linguistic distances. In addition to that, these instrumental variables in column 1-9 are log-transformed. Columns 4-6 also contain the residuals after orthogonalizing (model 2.3). (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively based on the Chi^2 distribution with 14 degrees of freedom.)

Figure 4: Heat map of the risk proxy distribution of all cooperative banks in Germany, 2009



Note: This graph depicts the regional distribution of $SOLVA$ in Germany, 2009. Warm colors indicate high $SOLVA$, cold a low $SOLVA$.

Figure 5: Language distribution in Germany



Note: The map shows the distribution of German dialects following Wiesinger (1983). The major dialect groups are the Low German dialects in the north and the High German dialects in the southern part.

Table 5: Comparison of linguistic and geographic distance measure

	2007	2008	2009	2010
DISTSOLV	-0.068**	-0.174***	-0.192***	-0.129***
DISTZ	-0.180***	-0.101***	-0.138***	-0.104***
DISTSEC	-0.194***	-0.143***	-0.129***	-0.144***

Note: The table contains the Spearman correlation for $|g_{it}(GeoSOLVA) - g_{it}(LingSOLVA)|$ and sim_{it}. (*, **, and *** denote significance at the 10%, 5%, and 1% levels).

Table 6: Panel unit root test for main variables

	SOLVA	Z	SEC	DIST-SOLV (geo)	DISTZ (geo)	DIST-SEC (geo)	DIST-SOLV (lin)	DISTZ (lin)	DIST-SEC (lin)
Maddala/ Wu Chi²-test	4458***	8619***	3604***	4229***	8324***	1638	5410***	7976***	1584
Choi normal dist-test	34***	96***	21***	30***	92***	-9	48***	86***	-10

Note: The dependent variable in columns 1-9 is as defined in table 1. The sample (over the period 2007–2010) is drawn from databases described in section 3 and is restricted to firm years (4,444 observations) where values of all variables are available. The first line is the Chi²-test statistic distributed test statistic according to Maddala and Wu (1999), the second line the normally distributed-test statistic according to Choi (2001), both with a null hypothesis of a unit root in the panel (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table 7: Structure and specifications of robustness tests

Part A: validity of the economic model	
Test A1	**Include Interaction:** YEAR effects instead of macroeconomic variables (collinearity of macro and year)
Test A2	**Nonlinear SIZE relationship:** Add SIZEsq and SIZEcube
Test A3	**Without East Germany:** Only use banks situated in West Germany
Test A4	**Bavaria only**: Including only Bavarian banks (as Bavaria is the largest German state with most banks and the largest number of different dialects)
Alternative correlation purging of CH and GFCFGR	
Test A5	**CH and GFCFGR**: Use variables with orthogonalization
Test A6	**CH and GFCFGR**: Drop CH IRS GDPGR and use instead INFLR NPL (ma^*)
Test A7	**CH and GFCFGR**: Drop CH GDPGR CAP and insert INFLR NPL (ma^*)
Validity of the distance variables with a new data set for each test: All estimated as instruments:	
Test A8	**Distance variables (1a):** Alternative measure definition of the risk proxy of the neighboring banks: Equally weighted average risk proxy of the 3 next banks
Test A9	**Distance variables (1b):** Logarithm of equally weighted average risk proxy of the nearest (geographic/linguistic) 3 banks
Test A10	**Distance variables (1c):** Equally weighted average risk proxy of the nearest (linguistic/geographic) 5 banks
Test A11	**Distance variables (1d):** Log of equally weighted average risk proxy of the nearest (linguistic/geographic) 5 banks
Test A12	**Distance variables (1e):** Equally weighted average risk proxy of the (linguistic/geographic) banks in a range of max. 30 km
Test A13	**Distance variables (1f):** Log of equally weighted average risk proxy of the (linguistic/geographic) banks in a range of max. 30 km
Test A14	**Distance variables (2): "**0" in case of the risk proxy deviates max 0.5 standard deviations from the risk proxy of all banks. In case of a risk proxy higher than 0.5 standard deviations "1", and in case it is less than -0.5 standard deviation of risk proxies of all banks "-1".
Part B: Validity of the econometric model	
IV Approach and Static Model	
Test B1	**First-Differencing (1):** All variables are used in time differences (Distance variable is measured directly, no instrumental variable approach)
Test B2	**First-Differencing (2):** All variables are used in time differences (logarithmic Distance variable is measured directly, no instrumental variable approach)
Test B3	**Static Model:** Lagged dependent variables are dropped as independent variables
Test B4	**Stock Watson Correction:** Alternative covariance estimator
Other econometric approach cross section	
Test B5	**Cross Section (1a):** Estimation is performed for each year separately (Instrumental variable for distance/logarithm).
Test B6	**Cross Section (1b):** Instrumental variable for distance/linear
Test B7	**Cross Section (1c):** Estimation is performed in first differences (log No instrumental variable approach).
Test B8	**Cross Section (1b):** Estimation is performed in first differences (linear No instrumental variable approach).
Test B9	**Cross Section (2a):** Instrumental variable for distance/linear: here equally weighted average risk proxy of the 3 next banks
Test B10	**Cross Section (2b):** Instrumental variable for distance/logarithm: here equally weighted average risk proxy of the 3 next banks

Note: In this table, we present the structure of our robustness tests. In Part A, we focus on the economic variations. We therefore mention the change compared to the original model, but do not list the specification, as long as it is the original specification for all tests in Part A. As the specification is a major part of the tests in Part B, we list all specifications that vary during the 10 tests. To simplify matters, non-varying specifications are left out of the explanation list in the table.

Table 8: Different distance measures applied

Distance measure		Scale of risk proxy	Scale of distance	Weighting scheme
Weighted average risk proxy	g	metric	metric (0-1000 km)	Distance weighted
Average risk proxy of 3 neighboring banks	g2	metric	ordinal (3 banks)	Unweighted
Average risk proxy of 5 neighboring banks	g2	metric	ordinal (5 banks)	Unweighted
Average risk proxy of banks in radius 30km	g2	metric	metric (0-30 km)	Unweighted
Weighted deviation from mean	g3	ordinal (1,0,-1)	metric (0-1000 km)	Distance weighted

Notes: In this table, we present the different distance measures (column 1 and symbol column 2). They differ in terms of the level of measurement regarding the risk proxy (column 3) as well as the distance (column 4). Additional measures are aggregated using different weighting schemes (column 5).

Table 9: Result comparison of robustness tests

	Dist (geo)		Dist(lin)		Ortho (lin)			Lag	Controls			FE
	+/-	*	+/-	*	+/-	*	geo		reg	bank	mac	
SOLVA												
Part A: Validity of the economic model												
A1	✓	✓	✓	✓	✓	✓	×	×	✓	✓		✓
A2	✓	✓	✓	✓	✓	✓	×	×	✓	✓	×	✓
A3	✓	✓	✓	✓	✓	✓	×	×	×	✓	×	✓
A4	✓	×	✓	✓	✓	✓	×	×	✓	×	×	✓
A5	✓	✓	✓	✓	✓	✓	×	×	×	✓	×	✓
A6	✓	✓	✓	✓	✓	✓	×	×	×	✓	✓	✓
A7	✓	✓	✓	✓	✓	✓	×	×	✓	✓	×	✓
A8	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓
A9	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓
A10	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓
A11	✓	✓	✓	✓	✓	×	✓	×	✓	✓	×	✓
A12	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓
A13	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓
A14	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓
Part B: Validity of the econometric model												
B1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		×
B2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		×
B3	✓	✓	✓	✓	✓	✓	×		×	✓	×	✓
B4	✓	✓	✓	✓	✓	✓	×	×	✓	✓	×	✓
B5	✓	×	✓	×	✓	✓	×	✓	×	✓		
B6	✓	×	✓	✓	✓	✓	✓	✓	✓	×		
B7	✓	×	✓	✓	✓	✓	×		✓	×		
B8	✓	×	✓	✓	✓	✓	×		✓	×		
B9	✓	×	✓	✓	×	×	✓	✓	×	✓		
B10	×	×	✓	✓	✓	×	×	✓	×	✓		
Z-Score												
Part A: Validity of the economic model												
A1	✓	✓	✓	✓	✓	✓	✓	×	×	×		✓
A2	✓	✓	✓	✓	✓	✓	×	×	×	✓	×	✓
A3	✓	✓	✓	✓	✓	✓	×	×	×	×	✓	✓
A4	×	×	✓	×	✓	×	×	✓	×	✓	✓	✓
A5	✓	✓	✓	✓	✓	✓	×	×	✓	✓	✓	✓
A6	✓	✓	✓	✓	✓	✓	×	×	×	×	×	✓
A7	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	×	✓
A8	✓	✓	✓	✓	✓	×	✓	×	×	×	✓	✓
A9	✓	✓	✓	✓	✓	✓	✓	×	×	×	✓	✓
A10	✓	✓	✓	✓	✓	×	✓	×	×	×	✓	✓
A11	✓	✓	✓	✓	✓	✓	✓	×	×	×	✓	✓
A12	✓	✓	✓	✓	✓	×	✓	×	×	×	✓	✓

A13	✓	✓	✓	✓	✓	×	✓	×	×	×	✓	✓
A14	✓	✓	✓	✓	✓	×	✓	×	×	×	✓	✓
Part B: Validity of the econometric model												
B1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
B2	✓	✓	✓	✓	✓	✓	✓	✓	×	✓		✓
B3	✓	✓	✓	✓	✓	✓	×		×	×	✓	✓
B4	✓	✓	✓	✓	✓	✓	×	×	×	×	✓	✓
B5	✓	×	✓	✓	✓	×	×	✓	✓	×		
B6	✓	×	✓	✓	×	×	×	✓	✓	×		
B7	✓	×	✓	✓	✓	✓	×		✓	✓		
B8	✓	×	✓	✓	✓	✓	×		✓	✓		
B9	✓	✓	✓	✓	✓	×	×	✓	✓	×		
B10	✓	✓	✓	✓	✓	×	✓	✓	✓	×		
SEC												
Part A: Validity of the economic model												
A1	✓	✓	✓	✓	✓	✓	×	✓	✓	×		✓
A2	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓
A3	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓
A4	✓	×	✓	✓	✓	×	×	✓	✓	×	×	✓
A5	✓	✓	✓	✓	✓	✓	×	✓	×	×	✓	✓
A6	✓	✓	✓	✓	✓	✓	×	✓	×	×	✓	✓
A7	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓
A8	✓	✓	✓	✓	✓	×	×	✓	✓	×	✓	✓
A9	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓
A10	✓	✓	✓	✓	✓	×	×	✓	✓	×	✓	✓
A11	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓
A12	✓	✓	✓	✓	✓	×	×	✓	✓	×	×	✓
A13	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓
A14	✓	✓	✓	✓	✓	✓	×	✓	✓	×	×	✓
Part B: Validity of the econometric model												
B1	✓	✓	✓	✓	✓	✓	✓	×	×	×		×
B2	✓	✓	✓	✓	✓	✓	✓	×	×	×		×
B3	✓	✓	✓	✓	✓	✓	×		✓	×	✓	✓
B4	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓
B5	✓	✓	✓	✓	✓	✓	✓	✓	×	×		
B6	✓	✓	✓	✓	✓	✓	✓	✓	×	×		
B7	✓	×	✓	✓	✓	✓	×		✓	×		
B8	✓	✓	✓	✓	✓	✓	×		✓	×		
B9	✓	✓	✓	✓	✓	✓	✓	✓	×	×		
B10	✓	✓	✓	✓	✓	×	✓	✓	×	×		

Note: We summarize the robustness tests of our model, in total: 72 tests. The columns (1)-(2) contain the results for the geographic distance variable, whereas columns (3)-(4) are dedicated to the linguistic distance variable. The results for the regression (2.3) and (5.5) are depicted in (column (5)-(7), whereby (5)-(6) contains the orthogonalized linguistic distance and (7) the geographic distance variable. The results for the lagged dependent variable can be found in column (8). Columns (9)-(11) depict the results for the controls (reg: regional controls, bank: bank controls, mac: macroeconomic controls). In case of the distance variables, if hypothesized results for the signs ("+/-") and / or the significance ("*") are obtained, it is marked with ✓, otherwise with ×. Turning to the geographical variable in case of model (2.4 and 5.4) a ✓means significant and positive results, otherwise described the results by ×. In case of the control variables a wrong sign is indicated by × and a correct sign by ✓. FE in column (12I stands for Hausman specification test on the fixed effects model. Again significant results are indicated by ✓, otherwise ×.

Appendix: Construction of the geographical and linguistic distance variables

To explain the calculation of our distance measure, we introduce the following example for the construction of the geographical and linguistic distance variable. The basis for both types of distance measures is the distance between the headquarters of two banks in terms of geographical and linguistic distances. For geographical distances, the distance is calculated given latitudes and longitudes. For this example, we assume the latitude $\varphi_i = 48°$ and the longitude λ_i =8° of bank i and the latitude $\varphi_j = 51°$ and the longitude $\lambda_j = 10°$ of bank j.

Step 1: Geographical distances

We calculate the angles of banks i and j via great circle (Bosch 1998)

$AN = (angle_{ij})$ with $angle_{ij} = \arccos(sin\varphi_i * sin\varphi_j + cos\varphi_i * cos\varphi_j * cos|\lambda_j - \lambda_i|)$=3.26°.

We calculate the distances of banks i and j via great circle (Bosch 1998)

$D = (distance_{ij})$ with $distance_{ij} = r_{earth} * 2\pi * \frac{angle_{ij}}{360°} = 363\ km$ with $r_{earth} = 6371.221\ km.$

Step 2: Linguistic distances

In case of the linguistic distance the latter is given by the quote of the 66 prototypical features identical in the region where both banks are located. An example is $\frac{45}{66} = 0.6818$. For technical reasons the measure is transformed into a distance measure by taking the differencing and rescaling. We calculate the linguistic distances of banks i and j via $D = 1000 - 1000 \cdot 0.6818 = 318$ (linguistic) km. The next steps are applied in exactly the same manner for both measures.

Step 3: Sorting

All banks excluding bank i are sorted according to their distance to bank i beginning with the nearest bank. We assume following example.

	Distance to bank i	Solvability Ratio
Bank 1	100 km	10%
Bank 2	200 km	12%
Bank 3	300 km	11%
Bank 4	300 km	15%
Bank 5	350 km	12%

Step 4: Distance weighted measure

Given the risk measure of the banks in the sample (In this example: Solvability Ratio) the distance-weighted risk measure $g_{it}(SOLVA)$ is calculated stepwise.

Step 4a: The portion of the measure of bank 1 to the distance measure is

$g_{it}(SOLVA, 1) = 10\% \frac{200km-100km}{1000\ km-100km} = 1.11\%$. We assume a maximum distance of 1000 km.

Step 4b: The average ratio in the next region is the average ratio of bank 1 and bank 2 leading to

$ar_{it}(SOLVA, 2) = \frac{10\%+12\%}{2} = 11\%$

Step 4c: The portion of the measure of bank 1 and bank 2 to the distance measure is

$g_{it}(SOLVA, 2) = 11\% \frac{300km-200km}{1000\ km-100km} = 1.22\%$

Step 4d: The average ratio in the next region is the average ratio of bank 1-4 (ties: bank 3 and 4 have the same distance) leading to

$ar_{it}(SOLVA, 3-4) = \frac{10\%+12\%+11\%+15\%}{4} = 12\%$

Step 4d: The portion of the measure of the first bank 1-4 to the distance measure is

$g_{it}(SOLVA, 3-4) = 12\% \frac{350km-300km}{1000\ km-100km} = 0.67\%$

This procedure is repeated for all neighboring banks.

Step 4e: All amounts have to be aggregated to yield the distance-weighted risk measure

$$g_{it}(SOLVA) = 1.11\% + 1.22\% + 0.67\% \ldots = 10.64\%$$

For the robustness tests we applied three different measures:

Step 4e_robust_2.1:

The simplest one is given by equally-weighted risk proxy of the neighboring banks. In case of 5 neighboring banks it can be calculated in one step:

$g2_{it}(SOLVA, 3\ banks) = \frac{1}{3}(10\% + 12\% + 11\%) = 11\%$

$g2_{it}(SOLVA, 5\ banks) = \frac{1}{5}(10\% + 12\% + 11\% + 15\% + 12\%) = 12\%$

Step 4e_robust_2.2:

When defining k depending on distance as $\underset{k\in\{1,2,\ldots N-1\}}{\text{argmax}} (distsort(i,k) < 150\ km) = 1$, with $k_{max} = max(1,2) = 2$ and $g2_{it}(SOLVA) = \frac{1}{2}(10\% + 12\%) = 11\%$

In case of the next measure, which accounts for the possible nonlinear behavior in risk-taking with respect to the neighboring banks, we must first define $\mu(r) = 12\%$ and $\sigma(r) = 2\%$ (the mean and standard deviation of the risk proxy $SOLVA$ over the full sample in t).

In the next step we calculate the indicator function

Step 4a_robust_3:

$$I_{kt}(SOLVA, 1) = \left\{\begin{matrix} 1 & r_{kt}(SOLVA) > 13\% \\ 0 & |r_{kt}(SOLVA) - 12\%| \leq 1\% \\ -1 & r_{kt}(SOLVA) < 11\% \end{matrix}\right\} = -1$$

Step 4b_robust_3: The portion of the measure of the first bank 1 to the distance measure is

$g3_{it}(SOLVA, 1) = -1 \frac{200km-100km}{1000\ km-100km} = -0.11$

Step 4c_robust_3:

$$I_{kt}(SOLVA, 1) = \left\{\begin{matrix} 1 & r_{kt}(SOLVA) > 13\% \\ 0 & |r_{kt}(SOLVA) - 12\%| \leq 1\% \\ -1 & r_{kt}(SOLVA) < 11\% \end{matrix}\right\} = 0$$

Step 4d_robust_3: The portion of the measure of the first bank 1 and bank 2 to the distance measure is $g_{3it}(SOLVA, 2) = \frac{-1+0}{2} \frac{300km-200km}{1000\ km-100km} = -0.056$

Step 4e_robust_3: All portions have to be aggregated to yield the distance weighted risk measure

$$g3_{it}(SOLVA) = -0.11 - 0.056 + \cdots = -0.56$$

Supplementary Information SI

Table SI-1: Results (Yearly dummies)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance		Geographic			Linguistic			Geographic/Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.791***						0.316		
LnDistZ(geo)		10.056***						8.204***	
LnDistsec (geo)			52.426***						-9.753*
DistSolva(lin/ort)				0.221***			0.214***		
DistZ(lin/ort)					1.342**			1.05**	
Distsec(lin/ort)						72.058***			50.82***
Independent Variables: Controls									
Lag dep Var	-0.081***	-0.05	0.05	-0.088***	-0.039	0.052	-0.087***	-0.037	0.052
(u)POP	-0.054	-9.816	95.706***	-0.081	4.831	107.644***	-0.055	-6.876	109.393***
GFCFGR	0	-0.007**	-0.015***	0	0.002	-0.019***	0	-0.005	-0.018***
Size	-0.005	0.351***	-0.849**	-0.003	0.409***	-0.904**	-0.004	0.358***	-0.892*
Cap	0.043*	52.105***	-2.776	0.038	52.68***	-2.897	0.038	51.089***	-2.92
Y2008	0.013**	-3.014***	5.441***	0	0.15	2.968***	0.002	-1.965*	3.752***
Y2009	0.027***	-3.532***	-0.206*	0.002	0.051	-0.26**	0.006	-2.321*	-0.251**
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.6%	80.1%	85.9%	90.7%	80.2%	86.0%	90.7%	80.3%	86.0%
Durbin Watson	2.664***	1.778***	2.667***	2.723***	1.777***	2.649***	2.708***	1.773***	2.64***
LM Hetero	0.057	0.048	0.1	0.11	0.1	0.028	1.056	0.071	0.18
HausmannWu	8991***	1016.5***	3822***	9416***	373***	3838***	9377***	176.6***	3841***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-2: Results (Nonlinear size)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance	Geographic			Linguistic			Geographic/Linguistic		
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.821***						0.344		
LnDistZ(geo)		22.326***						14.168*	
LnDistsec (geo)			52.209***						-10.013*
DistSolva(lin/ort)				0.227***			0.217***		
DistZ(lin/ort)					1.645***			1.372**	
Distsec(lin/ort)						72.17***			50.944***
Independent Variables: Controls									
Lag dep Var	-0.079***	-0.056	0.048	-0.086***	-0.046	0.05	-0.086***	-0.044	0.049
(u)POP	0.1	-8.085	95.9***	-0.04	6.011	107.968***	-0.006	-4.208	109.806***
GFCFGR	0	-0.006*	-0.014***	0	0.002	-0.018***	0	-0.004	-0.018***
Size	-3.192	-165.742**	205.396	-3.59	-145.617*	210.375	-3.601	-157.862**	209.935
$Size^2$	0.163	8.802**	-10.969	0.184	7.754*	-11.236	0.184	8.387**	-11.214
$Size^3$	-0.003	-0.155**	0.194	-0.003	-0.137*	0.199	-0.003	-0.148**	0.198
Cap	0.04*	52.876***	-3.124	0.035	53.114***	-3.253	0.035	51.843***	-3.278
CH	0.287	13.08	-18.169	0.305	11.865	-18.717	0.303	12.467	-18.253
IRS	0.243	10.573	-13.784	0.256	10.035	-14.924	0.255	10.242	-14.313
GDPGR	0.296	13.945	-18.825	0.317	12.349	-19.42	0.315	13.178	-18.929
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.6%	80.2%	86.0%	90.8%	80.3%	86.1%	90.8%	80.3%	86.1%
Durbin Watson	2.841***	1.924**	3.683***	3.114***	1.798***	3.69***	3.782***	2.132***	2.813***
LM Hetero	0.031	5.951**	0.16	27.096***	1.026	0.123	0.868	0.087	0.197
HausmannWu	10120***	920.1***	3867***	183***	940.1***	3907***	137***	285.7***	8268***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-3: Results (West Germany)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance		Geographic			Linguistic			Geographic/Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.869***						0.443*		
LnDistZ(geo)		15.334**						7.536	
LnDistsec (geo)			67.07***						-0.198
DistSolva(lin/ort)				0.219***			0.191***		
DistZ(lin/ort)					1.604***			1.439**	
Distsec(lin/ort)						83.234***			43.5***
Independent Variables: Controls									
Lag dep Var	-0.09***	-0.057	0.04	-0.094***	-0.046	0.042	-0.094***	-0.044	0.042
(u)POP	0.139	-7.477	100.509***	0.018	5.997	120.736***	0.07	-1.346	114.815***
GFCFGR	0	-0.003	-0.015***	0	0.002	-0.019***	0	-0.001	-0.019***
Size	-0.004	0.364***	-1.015**	-0.003	0.374***	-1.064**	-0.003	0.362***	-1.085**
Cap	0.038	53.368***	-3.078	0.034	52.609***	-3.287	0.034	52.121***	-3.205
CH	0.029	-0.423	-1.498	0.011	-0.307	-1.922	0.013	-0.264	-1.464
IRS	0.026	-0.635	-0.407	0.009	-0.213	-0.852	0.011	-0.322	-0.825
GDPGR	0.027	-0.224	-1.498	0.011	-0.323	-1.941	0.012	-0.182	-1.479
NumBanks	1034	1034	1034	1034	1034	1034	1034	1034	1034
NumObsv	3102	3102	3102	3102	3102	3102	3102	3102	3102
R^2	90.6%	79.9%	85.8%	90.7%	80.1%	85.9%	90.7%	80.1%	85.9%
Durbin Watson	2.786***	1.775***	2.706***	3.488***	1.776***	2.733***	2.732***	1.773***	2.68***
LM Hetero	0.556	0.762	1.876	4.177**	2.465	0.174	16.992***	3.089*	0.413
HausmannWu	10004***	1216.4***	4099***	26960***	12127.4***	7136***	3226***	2651.8***	21956***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-4: Results (Bavaria)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance		Geographic			Linguistic			Geographic/Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	1.274						1.139		
LnDistZ(geo)		-1.983						-1.89	
LnDistsec (geo)			31.923						9.873
DistSolva(lin/ort)				0.358***			0.344***		
DistZ(lin/ort)					0.14			0.224	
Distsec(lin/ort)						16.168***			-34.599*
Independent Variables: Controls									
Lag dep Var	-0.282***	0.003	0.013	-0.278***	0.002	0.017	-0.28***	0.003	0.015
(u)POP	0.574	13.2**	-69.727***	-1.595	13.915**	-15.29**	-1.071	13.272*	323.523
GFCFGR	0	0.012	0.175	0.001**	0.012	0.113	0.001	0.011	-0.239
Size	-0.002	-0.161***	-1.048	0	-0.161***	-1.177	-0.001	-0.162***	-1.131
Cap	0.012	23.601***	-1.789	0.003	23.575***	-2.66	-0.004	23.543***	-1.625
CH	0.039	-0.044	1.447	0.038	-0.114	0.926	0.045	-0.034	-3.427
IRS	0.035	-0.011	-0.094	0.03	-0.108	1.339	0.038	-0.003	-2.985
GDPGR	0.035	-0.1	1.111	0.036	-0.149	0.759	0.04	-0.089	-3.17
NumBanks	298	298	298	298	298	298	298	298	298
NumObsv	894	894	894	894	894	894	894	894	894
R^2	77.1%	93.0%	87.7%	77.7%	92.9%	88.0%	77.8%	93.0%	87.8%
Durbin Watson	3.566***	3.701***	3.88***	3.628***	3.869***	3.987***	3.218***	3.47***	3.903***
LM Hetero	0.829	0.237	0.01	0.868	0.33	0.959	0.918	0.33	0.061
HausmannWu	5	718.5***	1068***	24***	78.7***	12	20**	14.1	1622***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-5: Results (including orthogonalized values for CH and GFCFGR)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance	Geographic			Linguistic			Geographic/Linguistic		
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.791***						0.316		
LnDistZ(geo)		21.496***						13.139	
LnDistsec (geo)			52.426***						-9.753*
DistSolva(lin/ort)				0.221***			0.214***		
DistZ(lin/ort)					1.662***			1.412**	
Distsec(lin/ort)						72.058***			50.82***
Independent Variables: Controls									
Lag dep Var	-0.081***	-0.051	0.05	-0.088***	-0.041	0.052	-0.087***	-0.039	0.052
(u)POP	0.054	-7.953	95.706***	-0.081	5.498	107.644***	-0.055	-3.958	109.393***
(uu)GFCFGR	-0.016	0.309	0.683	-0.008	0.194	1.094	-0.009	0.253	0.854
Size	-0.005	0.368***	-0.849*	-0.003	0.412***	-0.904**	-0.004	0.375***	-0.892**
Cap	18.771	-309.125	-797.313	9.371	-164.25	-1268.375	9.955	-240	-997.625
(uu)CH	0.038	-0.743	-1.632	0.019	-0.442	-2.603	0.02	-0.605	-2.045
IRS	0.104	-2.425	-2.87	0.051	-1.138	-6.131	0.054	-1.848	-4.431
GDPGR	0.01	0.06	-0.467	0.006	-0.146	-0.816	0.006	-0.023	-0.623
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.6%	80.0%	85.9%	90.7%	80.2%	86.0%	90.7%	80.2%	86.0%
Durbin Watson	2.854***	1.831***	2.859***	3.329***	1.696***	2.679***	3.322***	1.856***	2.677***
LM Hetero	0.51	0.246	0.039	0.552	0.007	0.06	0.233	0.068	0.1
HausmannWu	6119***	444.8***	3412***	659***	439.6***	4447***	14535***	3344.7***	4390***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-6: Results (including new independent variables CH and GFCFGR: Drop CH IRS GDPGR and use instead INFLR NPL (ma*))

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance		Geographic			Linguistic			Geographic/Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.791***						0.316		
LnDistZ(geo)		21.496***						13.139	
LnDistsec (geo)			52.426***						-9.753*
DistSolva(lin/ort)				0.221***			0.214***		
DistZ(lin/ort)					1.662***			1.412**	
Distsec(lin/ort)						72.058**			50.82***
Independent Variables: Controls									
Lag dep Var	-0.081***	-0.051	0.05	-0.088***	-0.041	0.052	-0.087***	-0.039	0.052
(u)POP	0.054	-7.953	95.706***	-0.081	5.498	107.644***	-0.055	-3.958	109.393***
GFCFGR	0	-0.006	-0.015***	0	0.002	-0.019***	0	-0.003	-0.018***
Size	-0.005	0.368***	-0.849*	-0.003	0.412***	-0.904**	-0.004	0.375***	-0.892*
Cap	0.043*	52.748***	-2.776	0.038	52.855***	-2.897	0.038	51.688***	-2.92
CH	-22.442***	2853***	-1758.9***	-1.45	-92.726	-863.0***	-4.865	1624.4	-1141.7**
IRS	-0.053***	6.139***	-1.198***	-0.004	-0.147	-0.389	-0.012	3.51	-0.631
GDPGR	-0.081***	-0.051	0.05	-0.088***	-0.041	0.052	-0.087***	-0.039	0.052
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.6%	80.0%	85.9%	90.7%	80.2%	86.0%	90.7%	80.2%	86.0%
Durbin Watson	2.664***	1.778***	2.667***	2.723***	1.776***	2.649***	2.708***	1.773***	2.64***
LM Hetero	0.055	0.083	0.01	0.054	0.078	0.025	0.073	1.348	0.172
HausmannWu	8991***	1073.9***	3822***	9416***	583.1***	3838***	9377***	492.5***	3841***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-7: Results (including new independent variables CH and GFCFGR: Drop CH GDPGR CAP and insert INFLR NPL (ma*))

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance	Geographic			Linguistic			Geographic/Linguistic		
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.819***						0.317		
LnDistZ(geo)		51.238***						-0.074	
LnDistsec (geo)			52.668***						-9.46
DistSolva(lin/ort)				0.231***			0.224***		
DistZ(lin/ort)					7.857***			8.026***	
Distsec(lin/ort)						71.922***			50.676***
Independent Variables: Controls									
Lag dep Var	-0.081***	0.002	0.05	-0.088***	0.062	0.052	-0.087***	0.059	0.051
(u)POP	0.057	-41.413*	95.471***	-0.082	-20.854	107.274***	-0.058	-14.482	108.925***
GFCFGR	0	-0.011	-0.015***	0	0.001	-0.019***	0	0.005	-0.019***
Size	-0.006	-0.178	-0.825*	-0.004	-0.093	-0.877*	-0.004	-0.075	-0.866*
IRS	0.058	-4.839	-0.724	0.021	0.102	-2.107	0.024	0.874	-1.601
NPL	69.766	-916.375	-2858	31.504	683.25	-4198	33.293	733.375	-3646
INFLR	0.052	5.577	-2.449	0.033	1.194	-4.213	0.031	0.096	-3.502
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.6%	50.8%	85.9%	90.7%	58.4%	86.0%	90.7%	58.4%	86.0%
Durbin Watson	3.117***	2.914***	2.608***	2.766***	2.315***	3.187***	2.344***	1.662***	2.785***
LM Hetero	3.451*	1.319	2.84*	8.33***	0.503	149.998***	0.949	0.241	1.117
HausmannWu	510***	12.30	51***	46***	23.4***	41***	577***	1222.4***	1127***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-8: Results (Distance variables (1a))

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance		Geographic			Linguistic			Geographic/Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.745***						0.864***		
LnDistZ(geo)		0.129***						0.156***	
LnDistsec (geo)			1.056***						0.343
DistSolva(lin/ort)				0.679***			0.533**		
DistZ(lin/ort)					0.095**			0.054	
Distsec(lin/ort)						1.506***			5.918
Independent Variables: Controls									
Lag dep Var	-0.084***	-0.048	0.052	-0.087***	-0.053	0.049	-0.087***	-0.045	0.052
(u)POP	-0.067	9.661	4.25***	-0.139	10.131	5.295***	-0.095	9.469	49.065
GFCFGR	0	0.003	-0.01	0	0.003	-0.01*	0	0.002	-0.012*
Size	-0.004	0.422***	-0.993**	-0.003	0.429***	-0.933**	-0.003	0.418***	-0.899*
Cap	0.045**	54.71***	-2.389	0.045**	54.43***	-2.597	0.045**	54.18***	-2.344
CH	0.004	-0.347	0.307	0.005	-0.36	0.252	0.003	-0.276	-0.491
IRS	0.003	-0.223	0.429	0.004	-0.228	0.283	0.002	-0.168	-0.197
GDPGR	0.004	-0.381	0.353	0.006	-0.4	0.291	0.003	-0.305	-0.477
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.6%	80.1%	85.9%	90.7%	80.0%	86.1%	90.7%	80.1%	85.8%
Durbin Watson	2.745***	1.9690	2.698***	2.636***	1.791***	2.653***	2.637***	1.792***	2.669***
LM Hetero	0.336	0.339	0.115	1.029	0.031	0.065	0.047	0.052	0.128
HausmannWu	362***	138.1***	4375***	3547***	2419.9***	3959***	27763***	1656.7***	1158***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-9: Results (Distance variables (1b))

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance	Geographic			Linguistic			Geographic/Linguistic		
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.108***						0.127***		
LnDistZ(geo)		0.825***						1.108***	
LnDistsec (geo)			68.2***						-3.419***
DistSolva(lin/ort)				0.092***			0.066**		
DistZ(lin/ort)					0.751***			0.578**	
Distsec(lin/ort)						65.435***			39.076***
Independent Variables: Controls									
Lag dep Var	-0.085***	-0.049	0.052	-0.087***	-0.048	0.048	-0.087***	-0.044	0.049
(u)POP	-0.056	9.45	98.052***	-0.124	9.422	95.037***	-0.066	8.855	100.595***
GFCFGR	0	0.002	-0.017***	0	0.003	-0.016***	0	0.002	-0.017***
Size	-0.004	0.422***	-0.866*	-0.004	0.426***	-0.905**	-0.003	0.416***	-0.895*
Cap	0.045**	54.677***	-3.407	0.046**	53.876***	-2.918	0.045**	53.654***	-3.149
CH	0.009	-0.362	-1.527	0.009	-0.364	-1.368	0.007	-0.283	-1.201
IRS	0.007	-0.236	-0.741	0.007	-0.237	-0.553	0.006	-0.178	-0.388
GDPGR	0.009	-0.396	-1.537	0.01	-0.4	-1.37	0.008	-0.311	-1.195
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.6%	80.1%	85.9%	90.6%	80.2%	86.0%	90.7%	80.3%	86.0%
Durbin Watson	2.988***	1.81***	2.717***	3.12***	1.811***	2.815***	2.686***	1.786***	2.706***
LM Hetero	0.106	0.208	0.177	0.853	0.153	0.021	1.001	0.196	0.107
HausmannWu	1043***	366.9***	61478***	7104***	1428.7***	4341***	10914***	980.6***	7895***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-10: Results (Distance variables (1c))

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance	Geographic			Linguistic			Geographic/Linguistic		
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.825***						0.931***		
LnDistZ(geo)		0.155***						0.19***	
LnDistsec (geo)			9.619						0.96
DistSolva(lin/ort)				0.787***			0.428**		
DistZ(lin/ort)					0.14***			0.079	
Distsec(lin/ort)						5.169			2.193
Independent Variables: Controls									
Lag dep Var	-0.084***	-0.048	0.052	-0.087***	-0.049	0.051	-0.086***	-0.045	0.051
(u)POP	-0.032	9.721	44.495	-0.075	11.334	20.66	-0.03	10.3	11.139
GFCFGR	0	0.002	-0.011	0	0.003	-0.011*	0	0.002	-0.01
Size	-0.005	0.426***	-0.918**	-0.004	0.421***	-0.866*	-0.004	0.418***	-0.878*
Cap	0.046**	54.721***	-2.202	0.043*	54.108***	-1.961	0.044**	54.099***	-1.86
CH	0.003	-0.35	-0.524	0.004	-0.391	-0.106	0.002	-0.284	0.049
IRS	0.002	-0.229	-0.212	0.002	-0.262	-0.064	0.001	-0.179	0.016
GDPGR	0.003	-0.382	-0.51	0.004	-0.428	-0.079	0.002	-0.311	0.078
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.7%	80.1%	85.8%	90.7%	80.0%	85.8%	90.7%	80.1%	85.8%
Durbin Watson	2.64***	1.789***	2.724***	2.727***	1.799***	2.677***	2.988***	1.798***	2.649***
LM Hetero	1.158	0.538	0.184	0.39	0.088	0.005	0.317	0.084	0.308
HausmannWu	233***	374.5***	6198***	13805***	122.5***	31449***	395***	827.8***	8094***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-11: Results (Distance variables (1d))

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance		Geographic			Linguistic			Geographic/Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.129***						0.145***		
LnDistZ(geo)		1.003***						1.409***	
LnDistsec (geo)			69.825***						-3.558***
DistSolva(lin/ort)				0.119***			0.058		
DistZ(lin/ort)					1.158***			0.863**	
Distsec(lin/ort)						68.883***			35.436***
Independent Variables: Controls									
Lag dep Var	-0.085***	-0.049	0.052	-0.087***	-0.045	0.05	-0.086***	-0.042	0.05
(u)POP	-0.007	9.653	100.411***	-0.06	11.971	100.56***	-0.003	11.035	106.169***
GFCFGR	0	0.002	-0.017***	0	0.003	-0.015***	0	0.002	-0.016***
Size	-0.005	0.428***	-0.898*	-0.004	0.416***	-0.936**	-0.004	0.414***	-0.928**
Cap	0.046**	54.717***	-3.207	0.044*	53.618***	-3.139	0.045**	53.624***	-3.286
CH	0.009	-0.376	-1.538	0.009	-0.427	-1.515	0.007	-0.326	-1.277
IRS	0.007	-0.249	-0.738	0.007	-0.299	-0.682	0.005	-0.219	-0.452
GDPGR	0.009	-0.408	-1.55	0.01	-0.465	-1.528	0.007	-0.353	-1.276
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.7%	80.1%	85.9%	90.7%	80.1%	86.0%	90.7%	80.2%	86.0%
Durbin Watson	3.196***	1.79***	2.765***	2.698***	1.848***	2.65***	2.638***	1.821***	2.756***
LM Hetero	0.354	0.367	0.12	1.576	0.736	0.01	1.094	0.195	0.221
HausmannWu	3243***	76.2***	5458***	948***	297.1***	3858***	247096***	3133***	3096***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-12: Results (Distance variables (1e))

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance		Geographic			Linguistic			Geographic/Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.609***						0.815***		
LnDistZ(geo)		0.249***						0.254***	
LnDistsec (geo)			1.245***						0.456
DistSolva(lin/ort)				0.661***			0.533**		
DistZ(lin/ort)					0.095*			0.011	
Distsec(lin/ort)						1.292***			2.004
Independent Variables: Controls									
Lag dep Var	-0.085***	-0.044	0.052	-0.086***	-0.052	0.05	-0.086***	-0.044	0.052
(u)POP	-0.106	12.453	4.358***	-0.183	9.02	4.802***	-0.145	12.194	34.471
GFCFGR	0	0.001	-0.009	0	0.003	-0.008	0	0.001	-0.009
Size	-0.006	0.416***	-1.01**	-0.003	0.424***	-0.857*	-0.003	0.414***	-0.887*
Cap	0.045**	54.256***	-2.691	0.043*	53.956***	-3.099	0.043*	54.088***	-2.48
CH	0.005	-0.417	0.293	0.006	-0.328	0.243	0.004	-0.316	-0.274
IRS	0.004	-0.297	0.344	0.004	-0.199	0.318	0.003	-0.214	-0.059
GDPGR	0.006	-0.443	0.332	0.006	-0.364	0.277	0.004	-0.339	-0.258
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.6%	80.3%	85.9%	90.7%	80.0%	86.1%	90.7%	80.3%	85.8%
Durbin Watson	2.894***	1.854***	2.682***	2.538***	1.798***	2.582***	2.845***	1.889***	2.495***
LM Hetero	0.245	0.09	0.214	0.362	0.121	0.096	0.257	0.114	0.297
HausmannWu	1706***	749***	4300***	680***	2643.9***	5036***	4274***	1010.1***	104894***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-13: Results (Distance variables (1f))

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance	Geographic			Linguistic			Geographic/Linguistic		
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.1***						0.128***		
LnDistZ(geo)		1.762***						2.035***	
LnDistsec (geo)			70.249***						-3.244***
DistSolva(lin/ort)				0.096***			0.072**		
DistZ(lin/ort)					0.966***			0.626	
Distsec(lin/ort)						73.894***			21.888***
Independent Variables: Controls									
Lag dep Var	-0.085***	-0.044	0.053	-0.086***	-0.046	0.049	-0.087***	-0.037	0.05
(u)POP	-0.079	12.598	101.315***	-0.167	6.283	107.256***	-0.11	9.296	109.144***
GFCFGR	0	0.001	-0.017***	0	0.002	-0.02***	0	0	-0.02***
Size	-0.006	0.411***	-0.829*	-0.003	0.403***	-0.96**	-0.004	0.391***	-0.949**
Cap	0.045**	54.049***	-2.937	0.044*	52.621***	-2.779	0.043*	52.4***	-2.819
CH	0.009	-0.46	-1.556	0.01	-0.302	-1.658	0.008	-0.306	-1.33
IRS	0.007	-0.335	-0.748	0.008	-0.188	-0.793	0.006	-0.216	-0.505
GDPGR	0.01	-0.487	-1.566	0.011	-0.331	-1.668	0.009	-0.322	-1.326
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.6%	80.4%	85.9%	90.7%	80.2%	86.1%	90.7%	80.6%	86.1%
Durbin Watson	2.732***	1.812***	2.734***	3.263***	2.079***	2.637***	2.798***	1.8***	2.747***
LM Hetero	0.301	1.449	0.062	0.346	0.02	0.122	0.383	0.108	0.015
HausmannWu	9657***	1332.8***	5318***	17856***	9089.6***	4841***	2082***	2655.7***	14916***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-14: Results (Distance variables (2))

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance		Geographic			Linguistic			Geographic/Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.543***						0.464***		
LnDistZ(geo)		9.777***						9.785***	
LnDistsec (geo)			17.1***						6.76
DistSolva(lin/ort)				0.045***			0.035***		
DistZ(lin/ort)					0.53			0.196	
Distsec(lin/ort)						12.459***			90.148***
Independent Variables: Controls									
Lag dep Var	-0.083***	-0.06	0.053	-0.084***	-0.056	0.052	-0.084***	-0.058	0.053
(u)POP	-0.173	6.993	67.422***	-0.425**	7.577	73.822***	-0.366*	6.235	67.57***
GFCFGR	0	0.001	-0.014***	0	0.003	-0.012**	0	0.001	-0.014***
Size	-0.004	0.429***	-0.926**	-0.005	0.424***	-0.892*	-0.004	0.423***	-0.928**
Cap	0.035	54.837***	-2.09	0.035	54.365***	-2.282	0.031	54.455***	-2.106
CH	0.008	-0.294	0.382	0.013	-0.302	0.366	0.009	-0.218	-0.758
IRS	0.006	-0.16	0.681	0.01	-0.168	0.66	0.007	-0.098	-0.278
GDPGR	0.009	-0.331	0.442	0.014	-0.343	0.421	0.01	-0.25	-0.745
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.6%	80.0%	85.8%	90.7%	79.9%	85.8%	90.7%	80.0%	85.8%
Durbin Watson	2.909***	1.807***	2.701***	3.034***	1.778***	2.551***	2.872***	1.776***	2.523***
LM Hetero	0.065	0.188	0.062	2.378	0.081	0.116	0.119	0.12	0.009
HausmannWu	2548***	1250.9***	5322***	25***	1106.1***	4703***	10165***	819.3***	3635***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-15: Results (Differences linear)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance		Geographic			Linguistic			Geographic/Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.461**						0.61***		
LnDistZ(geo)		0.941***						0.954***	
LnDistsec (geo)			1.192***						0.994***
DistSolva(lin/ort)				1.406***			1.678***		
DistZ(lin/ort)					1.166***			1.501***	
Distsec(lin/ort)						1.392***			1.461***
Independent Variables: Controls									
Lag dep Var	1.847*	378.29***	-13.688	-1.215	240.262***	-1.802	0.012	134.626**	36.484
POP	0	0.008**	-0.018***	0	0.005	-0.007	0	0.002	-0.006
GFCFGR	-0.005	-0.342**	-0.635	-0.004	-0.341**	-0.612	-0.004	-0.364**	-0.597
Size	-0.008**	0.472***	0.078	0.004	-0.11	0.317	-0.007*	0.261*	-0.011
Cap	<0.001	0.032***	-0.012	<0.001	0.013*	0.015	0	0.011	-0.008
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	28.2%	63.7%	29.2%	37.9%	69.6%	37.4%	39.2%	70.1%	37.5%
Durbin Watson	2.098***	1.613***	2.271***	2.12***	1.616***	2.274***	2.124***	1.619***	2.274***
LM Hetero	0.055	0.049	0.004	0.063	0.086	0.005	0.066	0.082	0.006
HausmannWu	8	87.3***	1	8	237.7***	5	15**	169.5***	89***
SarganTest									
WuTest	0.81	1.503	-0.802	0.626	2.783**	0.033	0.452	2.643**	-0.043

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-16: Results (Differences log)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance		Geographic			Linguistic			Geographic/Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.071**						0.089***		
LnDistZ(geo)		7.298***						7.226***	
LnDistsec (geo)			6.354***						5.576***
DistSolva(lin/ort)				0.217***			0.271***		
DistZ(lin/ort)					7.201***			7.144***	
Distsec(lin/ort)						9.022***			9.977***
Independent Variables: Controls									
Lag dep Var	1.847*	377.13***	-17.839	-1.083	134.039**	-72.288	0.222	138.807**	-9.721
POP	0	0.008**	-0.018***	0	0.003	-0.012**	0	0.004	-0.01**
GFCFGR	-0.005	-0.34**	-0.637	-0.004	-0.379***	-0.596	-0.003	-0.378***	-0.566
Size	-0.008**	0.462***	0.07	0.005*	0.359	0.511***	-0.007**	0.336***	-0.031
Cap	0	0.02**	-0.015	0	0.001	0.023	0*	0.001	-0.015
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R²	28.2%	63.8%	29.2%	37.3%	70.1%	36.9%	38.8%	70.1%	37.2%
Durbin Watson	2.098***	1.613***	2.271***	2.119***	1.611***	2.273***	2.122***	1.614***	2.272***
LM Hetero	0.055	0.053	0.003	0.063	0.179	0.007	0.067	0.191	0.008
HausmannWu	8	88.2***	1	7	151.9***	1	13**	193.3***	2
SarganTest									
DurbinWuHauTest	0.704	0.89	-0.695	0.035	2.601**	1.371	-0.161	2.435**	-1.392

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-17: Results (measure 3 static model)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance		Geographic			Linguistic			Geographic/Linguistic	
Model	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables									
LnDistSolva(geo)	0.82***						0.375		
LnDistZ(geo)		21.715***						13.235	
LnDistsec (geo)			55.202***						-10.136*
DistSolva(lin/ort)				0.214***			0.201***		
DistZ(lin/ort)					1.676***			1.424**	
Distsec(lin/ort)						74.958***			53.053***
Independent Variables: Controls									
(u)POP	0.057	-8.124	99.928***	-0.092	5.468	111.624***	-0.045	-4.054	113.537***
GFCFGR	0	-0.006	-0.013***	0	0.002	-0.018***	0	-0.003	-0.017***
Size	-0.005	0.372***	-0.856*	-0.003	0.415***	-0.912**	-0.003	0.378***	-0.899*
Cap	0.042*	52.707***	-2.737	0.038*	52.823***	-2.852	0.037*	51.649***	-2.878
CH	0.023	-0.496	-0.941	0.01	-0.24	-1.295	0.016	-0.425	-1.588
IRS	0.021	-0.886	0.734	0.008	-0.195	-0.247	0.014	-0.624	-0.258
GDPGR	0.02	-0.202	-0.896	0.01	-0.256	-1.291	0.016	-0.263	-1.587
NumBanks	1111	1111	1111	1111	1111	1111	1111	1111	1111
NumObsv	3333	3333	3333	3333	3333	3333	3333	3333	3333
R^2	90.5%	80.0%	85.8%	90.6%	80.2%	85.9%	90.6%	80.2%	86.0%
Durbin Watson	1.723***	1.621***	1.718***	1.913***	1.643***	1.829***	2.97***	1.638***	1.877***
LM Hetero	0.034	0.06	0.148	0.299	0.198	0.047	0.021	0.842	0.354
HausmannWu	1	145.8***	203***	55***	1334.6***	223***	187***	136***	258***

Note: The dependent variable in columns 1-9 is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firm years (3,333 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines, whereby in columns 1-3 the variables are based on geographic distance, in columns 4-6 the variables are based on linguistic distance, and in columns 7-9 the variables are based on geographic as well as on linguistic distance. The independent control variables of all models with regression coefficients are defined as in table 3. In the third to last line the modified Durbin Watson statistic (Panel data) is depicted. In the second to last line the value of the LM-Test for heteroscedasticity and in the first to last line Hausman-Wu (Fixed vs. Random effects) test statistic is presented. All specifications include firm fixed effects as well as a test, whether the data contained fixed effects (positive for all models). Rogers robust standard errors that are clustered at the firm level are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-18: Results (Cross section log measure 3)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Distance	Geographic									Geographic/Linguistic				
Model	**2008**			**2009**			**2010**			**2008**			**2009**	
Dep. Var.	**(2.1d) log SOLVA**	**(2.1d) log Z**	**(2.1d) log SEC**	**(2.3) log SOLVA**	**(2.3) log Z**	**(2.3) log SEC**	**(2.1d) log SOLVA**	**(2.1d) log Z**	**(2.1d) log SEC**	**(2.1d) lna SOLVA**	**(2.1d) lna Z**	**(2.1d) lna SEC**	**(2.1d) log SOLVA**	**(2.1d) log Z**
Independent Variables: Distance Variables														
LnDistSolva(geo)	-0.076			0.239*			0.236***			-0.1			0.249*	
LnDistZ(geo)		2.339			-0.031			14.885***			2.205			-0.04
LnDistsec (geo)			7.161***			8.745***			2.653**			7.371***		
DistSolva(lin/ort)										0.01**			-0.004	
DistZ(lin/ort)											0.173			0.011
Distsec(lin/ort)												-0.335		
Independent Variables: Controls														
Const	0***	0***	0***	0***	0	0***	0***	0***	0***	-0.099	-6.63***	-7.82**	0.52**	-1.873
Lag dep Var	0.948***	0.801***	0.77***	0.74***	0.679***	0.523***	0.965***	0.879***	0.944***	0.947***	0.801***	0.771***	0.74***	0.679***
(u)POP	-0.003**	0.047*	-0.152	0.002**	0.033	-0.338***	0.001	0.107*	-0.127*	-0.003**	0.044*	-0.142	0.002**	0.033
GFCFGR	0	-0.004*	-0.038**	0	-0.001	-0.023***	0	-0.026**	-0.012	0	-0.004*	-0.04**	0	-0.001
Size	-0.004***	0.07***	-0.191**	-0.001*	0.111***	-0.033	0	0.399***	-0.004	-0.004***	0.072***	-0.196***	-0.001**	0.112***
Cap	0.12***	11.813***	-4.955*	0.08**	9.267***	6.516*	0.025*	62.322***	-1.456	0.117***	11.741***	-4.92*	0.081**	9.265***
EWG	0.007***	0.016	0.016	-0.005**	-0.029	-0.32*	-0.006***	-0.77***	-0.118	0.005*	-0.008	0.11	-0.005*	-0.031
R^2	59.9%	80.8%	65.9%	72.7%	71.7%	62.6%	88.1%	62.9%	80.3%	59.9%	80.8%	66.0%	72.7%	71.7%
FStatistic	235***	663***	305***	419.1***	398.9***	264.1***	1171.1***	266.6***	642.4***	206.1***	580.5***	266.8***	366.6***	348.7***
Breusch-pagan	0.00	0.03	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01

Column	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)
Distance	Geographic/Linguistic				Linguistic								
Model	**2009**	**2010**			**2008**			**2009**			**2010**		
	(2.1d) log	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**
Dep. Var.	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables													
LnDistSolva(geo)		0.24***											
LnDistZ(geo)			11.844***										
LnDistsec (geo)	8.402***			2.525**									
DistSolva(lin/ort)		-0.002			0.01**			-0.003			0.001		
DistZ(lin/ort)			0.761***			0.188*			0.011			0.902***	
Distsec(lin/ort)	1.119***			0.384*			0.18			1.415***			0.487**
Independent Variables: Controls													
Const	-10.78***	0.449***	-22.32***	-3.197*	0***	0***	0***	0***	0	0***	0***	0***	0***
Lag dep Var	0.522***	0.965***	0.898***	0.942***	0.947***	0.801***	0.777***	0.74***	0.679***	0.527***	0.966***	0.891***	0.946***
(u)POP	-0.385***	0.001	0.108*	-0.135*	-0.002**	0.064***	0.152	0.001	0.032	-0.151*	0	0.079	-0.083
GFCFGR	-0.027***	0	-0.02*	-0.012	0	-0.005**	-0.058***	0	-0.001	-0.01*	0	0.004	-0.014*
Size	-0.016	0	0.392***	0.003	-0.004***	0.073***	-0.159**	-0.001**	0.112***	0.011	0	0.398***	0.008
Cap	6.202*	0.025*	59.912***	-1.398	0.121***	11.815***	-3.204	0.071*	9.264***	8.151**	0.016	61.031***	-1.403
EWG	-0.681***	-0.005***	-0.844***	-0.241*	0.004*	0.035	0.6**	-0.003	-0.032	-0.198	-0.006***	-0.83***	-0.101
R^2	62.9%	88.1%	63.5%	80.3%	59.9%	80.8%	65.7%	72.6%	71.7%	62.5%	88.0%	63.2%	80.3%
FStatistic	233.5***	1024***	240.1***	562.9***	235.5***	663.5***	301.7***	418.2***	398.9***	263.1***	1153.9***	270.2***	642.2***
Breusch-pagan	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	1.00	0.00	0.01	0.00	0.00

Note: In columns 1-9 the variables are based on geographic distance, in columns 10-18 the variables are based on geographic as well as on linguistic distance, and in columns 19-27 the variables are based on linguistic distance. The dependent variable in all columns is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firms (1,111 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines of each subtable. The independent control variables of all models with regression coefficients are defined as in table 3. In the last line Breusch Pagan test statistic is presented. White robust standard errors are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-19: Results (Cross section linear measure 3)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Distance	Geographic									Geographic/Linguistic				
Model	**2008**			**2009**			**2010**			**2008**			**2009**	
	(2.1d) log	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**
Independent Variables: Distance Variables														
LnDistSolva(geo)	-0.478			1.533*			1.444***			-0.621			1.58*	
LnDistZ(geo)		0.694			-0.031			5.471***			0.636			-0.031
LnDistsec (geo)			1.162***			1.833***			0.559**			1.196***		
DistSolva(lin/ort)										0.086**			-0.028	
DistZ(lin/ort)											0.056			-0.001
Distsec(lin/ort)												-0.042		
Independent Variables: Controls														
Const	0.162	-5.589***	-1.607	-0.183	-1.847	-6.49***	-0.223***	-20.14***	-1.936*	0.182	-5.459***	-1.804	-0.189	-1.848
Lag dep Var	0.948***	0.801***	0.77***	0.74***	0.679***	0.523***	0.965***	0.879***	0.944***	0.947***	0.801***	0.771***	0.74***	0.679***
(u)POP	-0.003**	0.046*	-0.158	0.002**	0.034	-0.342***	0.001	0.107*	-0.128*	-0.003**	0.044*	-0.151	0.002**	0.034
GFCFGR	0	-0.004*	-0.038**	0	-0.001	-0.024***	0	-0.026**	-0.012	0	-0.004*	-0.039**	0	-0.001
Size	-0.004***	0.07***	-0.191**	-0.001*	0.111***	-0.033	0	0.399***	-0.004	-0.004***	0.072***	-0.195***	-0.001**	0.111***
Cap	0.12***	11.812***	-4.969*	0.08**	9.27***	6.466*	0.025*	62.355***	-1.457	0.116***	11.736***	-4.954*	0.081**	9.27***
EWG	0.007***	0.016	0.063	-0.005**	-0.027	-0.281*	-0.006***	-0.766***	-0.106	0.005*	-0.004	0.105	-0.005*	-0.027
R^2	59.9%	80.8%	65.9%	72.7%	71.7%	62.6%	88.1%	62.8%	80.3%	59.9%	80.8%	65.9%	72.7%	71.7%
FStatistic	235***	663***	305***	419.1***	398.9***	264.2***	1171***	266.4***	642.5***	206.1***	580.5***	266.7***	366.6***	348.7***
Breusch-pagan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01

Column	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)
Distance	Geographic/Linguistic				Linguistic								
Model	**2009**	**2010**			**2008**			**2009**			**2010**		
Dep. Var.	**(2.1d) log SEC**	**(2.3) log SOLVA**	**(2.3) log Z**	**(2.3) log SEC**	**(2.1d) log SOLVA**	**(2.1d) log Z**	**(2.1d) log SEC**	**(2.1d) lna SOLVA**	**(2.1d) lna Z**	**(2.1d) lna SEC**	**(2.1d) lna SOLVA**	**(2.1d) lna Z**	**(2.1d) lna SEC**
Independent Variables: Distance Variables													
LnDistSolva(geo)		1.461***											
LnDistZ(geo)			4.254***										
LnDistsec (geo)	1.712***			0.492**									
DistSolva(lin/ort)		-0.008			0.081**			-0.021			0.016		
DistZ(lin/ort)			0.451***			0.062*			-0.001			0.533***	
Distsec(lin/ort)	0.287***			0.103*			0.072			0.372***			0.132**
Independent Variables: Controls													
Const	-5.906***	-0.225***	-17.76***	-1.644	0.069***	-4.168***	3.567**	0.063***	-1.919***	-0.607	0.015*	-11.62***	-0.176
Lag dep Var	0.521***	0.965***	0.906***	0.941***	0.947***	0.801***	0.776***	0.74***	0.679***	0.526***	0.966***	0.901***	0.944***
(u)POP	-0.397***	0.001	0.097	-0.135*	-0.002**	0.063**	0.121	0.001	0.033*	-0.188**	0	0.068	-0.093
GFCFGR	-0.028***	0	-0.024**	-0.011	0	-0.005**	-0.055***	0	-0.001	-0.013**	0	-0.002	-0.013
Size	-0.016	0	0.395***	0.003	-0.004***	0.073***	-0.158**	-0.001**	0.111***	0.009	0	0.401***	0.008
Cap	6.065*	0.025*	59.581***	-1.397	0.12***	11.804***	-3.345	0.072*	9.267***	7.815**	0.016	60.488***	-1.397
EWG	-0.573***	-0.005***	-0.911***	-0.2*	0.004*	0.036	0.535**	-0.003	-0.029	-0.165	-0.006***	-0.915***	-0.095
R^2	62.9%	88.1%	63.6%	80.3%	59.9%	80.8%	65.7%	72.6%	71.7%	62.6%	88.0%	63.3%	80.3%
FStatistic	233.9***	1023.8***	240.6***	563.3***	235.6***	663.4***	301.9***	418.1***	398.9***	264.1***	1154.2***	271.5***	643.1***
Breusch-pagan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: In columns 1-9 the variables are based on geographic distance, in columns 10-18 the variables are based on geographic as well as on linguistic distance, and in columns 19-27 the variables are based on linguistic distance. The dependent variable in all columns is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firms (1,111 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines of each subtable. The independent control variables of all models with regression coefficients are defined as in table 3. In the last line Breusch Pagan test statistic is presented. White robust standard errors are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-20: Results (Cross section diff log measure 3)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Distance	Geographic									Geographic/Linguistic				
Model	**2008-2007**			**2009-2008**			**2010-2009**			**2008-2007**			**2009-2008**	
Dep. Var.	**(2.1d) log SOLVA**	**(2.1d) log Z**	**(2.1d) log SEC**	**(2.3) log SOLVA**	**(2.3) log Z**	**(2.3) log SEC**	**(2.1d) log SOLVA**	**(2.1d) log Z**	**(2.1d) log SEC**	**(2.1d) lna SOLVA**	**(2.1d) lna Z**	**(2.1d) lna SEC**	**(2.1d) log SOLVA**	**(2.1d) log Z**
Independent Variables: Distance Variables														
LnDistSolva(geo)	0.528***			0.307*			0.569			-0.751			-2.583**	
LnDistZ(geo)		2.175			0.233			2.562***			0.703			-1.573**
LnDistsec (geo)			23.711			13.423*			19.218***			-0.495		
DistSolva(lin/ort)										1.806***			1.837***	
DistZ(lin/ort)											0.091			0.574***
Distsec(lin/ort)												1.677***		
Independent Variables: Controls														
Const	-0.046***	-0.874	0.719	0.004	0.076*	2.64*	-0.016	0.8***	0.116	0.021**	-0.686	-0.417	-0.009**	0.212***
(u)POP	-0.477*	1.241	35.4	-0.307	6.385**	86.257*	0.004	22.384	22.616	0.032	1.776	38.02	-0.598**	3.569
GFCFGR	0	-0.007	-0.03	0*	0	-0.018**	0	-0.005	-0.004	0	-0.007	-0.03	0	0
Size	-0.001	-0.866***	0.671	-0.013***	-0.018	-0.915*	0.003	0.31*	-0.941**	0.002	-0.868***	0.503	-0.009**	-0.018
Cap	-0.007***	0.416***	0.082	0.059*	22.104***	-1.671	0.015	67.453***	-2.123*	-0.005***	0.412***	0.073	0.039	21.291***
EWG	0.003	0.195	-0.303	-0.001	-0.027*	-0.884***	-0.003	0.067	-0.046	0	0.288*	-0.021	-0.002	-0.039**
R^2	4.0%	21.4%	1.1%	3.0%	76.3%	5.0%	4.0%	68.1%	3.9%	18.6%	21.7%	13.1%	14.4%	77.0%
FStatistic	7.6***	50.2***	2.1**	5.6***	591.8***	9.6***	7.7***	393.1***	7.4***	36.1***	43.6***	23.8***	26.6***	528.8***
Breusch-pagan	0.00	0.40	0.34	0.00	0.00	0.01	0.84	0.00	0.00	0.00	0.46	0.36	0.00	0.00
DurbinWuHaus				-0.275	-12.358	1.109	0.047	0.947*	-1811.142				-0.074	1.396

Column	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)
Distance	Geographic/Linguistic				Linguistic								
Model	**2009-2008 (2.1d) log**	**2010-2009 (2.3) log**	**(2.3) log**	**(2.3) log**	**2008-2007 (2.1d) log**	**(2.1d) log**	**(2.1d) log**	**2009-2008 (2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**	**2010-2009 (2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**
Dep. Var.	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables													
LnDistSolva(geo)		-0.493											
LnDistZ(geo)			0.235										
LnDistsec (geo)	-0.238			3.479**									
DistSolva(lin/ort)		1.784***			1.597***			1.557***			1.597***		
DistZ(lin/ort)			0.532***			0.108			0.443***			0.497***	
Distsec(lin/ort)	1.501***			1.831***			1.592***			1.366***			1.437***
Independent Variables: Controls													
Const	-1.21*	0.017	-0.028	0.098	-0.017***	-1.706***	0.343	0.005**	0.057***	1.766***	-0.002	-0.02	0.04
(u)POP	80.116*	0.114	23.947	-8.396	0.04	4.794	35.656	-0.014	3.581	55.909	0.055	20.855	1.122
GFCFGR	-0.008	0	-0.004	-0.005*	0	-0.007	-0.028	0	0	-0.011*	0	-0.001	-0.002
Size	-1.011**	0.001	0.296*	-0.668**	0.001	-0.85***	0.591	-0.009**	-0.018	-0.994*	0.002	0.325**	-0.747**
Cap	-1.983	0.02	63.714***	-2.354**	-0.006***	0.406***	0.073	0.037	21.497***	-1.406	0.018	66.111***	-1.761*
EWG	-0.564**	-0.006	0.054	-0.012	0.001	0.302*	-0.321	-0.003	-0.025	-0.9***	-0.003	0.017	0.034
R^2	14.6%	20.4%	69.2%	19.0%	16.8%	23.2%	12.7%	13.2%	76.9%	14.0%	20.3%	68.5%	17.0%
FStatistic	27***	40.4***	353.5***	36.9***	37.1***	55.6***	26.9***	28***	611***	30***	46.9***	400***	37.6***
Breusch-pagan	0.00	0.80	0.00	0.00	0.00	0.30	0.32	0.00	0.00	0.00	0.90	0.00	0.00
DurbinWuHaus	0.765	-0.005	0.059	-21.432				-0.828*	-20.746	1.378	-0.042	1.055*	-3912.23*

Note: In columns 1-9 the variables are based on geographic distance, in columns 10-18 the variables are based on geographic as well as on linguistic distance, and in columns 19-27 the variables are based on linguistic distance. The dependent variable in all columns is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firms (1,111 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines of each subtable. The independent control variables of all models with regression coefficients are defined as in table 3. In the last line Breusch Pagan test statistic is presented. White robust standard errors are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-21: Results (Cross section diff linear measure 3)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Distance	Geographic									Geographic/Linguistic				
Model	**2008-2007**			**2009-2008**			**2010-2009**			**2008-2007**			**2009-2008**	
Dep. Var.	**(2.1d) log SOLVA**	**(2.1d) log Z**	**(2.1d) log SEC**	**(2.3) log SOLVA**	**(2.3) log Z**	**(2.3) log SEC**	**(2.1d) log SOLVA**	**(2.1d) log Z**	**(2.1d) log SEC**	**(2.1d) lna SOLVA**	**(2.1d) lna Z**	**(2.1d) lna SEC**	**(2.1d) log SOLVA**	**(2.1d) log Z**
Independent Variables: Distance Variables														
LnDistSolva(geo)	3.225***			1.888*			3.281			-0.751			-2.583**	
LnDistZ(geo)		0.861			0.045			1.356***			0.703			-1.573**
LnDistsec (geo)			4.071			1.966***			3.843***			-0.495		
DistSolva(lin/ort)										1.806***			1.837***	
DistZ(lin/ort)											0.091			0.574***
Distsec(lin/ort)												1.677***		
Independent Variables: Controls														
Const	-0.043***	-0.195	0.659	0.004	0.08*	1.737**	-0.014	0.851***	0.117	0.021**	-0.686	-0.417	-0.009**	0.212***
(u)POP	-0.583**	0.476	27.399	-0.303	6.426**	67.935	-0.025	19.374	22.122	0.032	1.776	38.02	-0.598**	3.569
GFCFGR	0	-0.007	-0.031	0*	0	-0.016**	0	-0.005	-0.004	0	-0.007	-0.03	0	0
Size	0	-0.87***	0.68	-0.013***	-0.018	-0.916*	0.003	0.313*	-0.943**	0.002	-0.868***	0.503	-0.009**	-0.018
Cap	-0.007***	0.419***	0.085	0.059*	22.104***	-1.64	0.015	67.493***	-2.137*	-0.005***	0.412***	0.073	0.039	21.291***
EWG	0.004	0.246*	-0.186	-0.001	-0.027*	-0.619**	-0.003	0.098	-0.051	0	0.288*	-0.021	-0.002	-0.039**
R^2	3.9%	21.4%	1.2%	3.0%	76.3%	5.2%	4.0%	68.1%	3.9%	18.6%	21.7%	13.1%	14.4%	77.0%
FStatistic	7.5***	50.2***	2.2**	5.6***	591.8***	10.1***	7.6***	392.3***	7.4***	36.1***	43.6***	23.8***	26.6***	528.8***
Breusch-pagan	0.00	0.39	0.36	0.00	0.00	0.00	0.84	0.00	0.00	0.00	0.46	0.36	0.00	0.00
DurbinWuHaus				-1.228	1.409	0.135	0.296	0.486*	-526.065				-0.427	-0.174

Column	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)
Distance	Geographic/Linguistic				Linguistic								
Model	**2009-2008 (2.1d) log**	**2010-2009 (2.3) log**	**(2.3) log**	**(2.3) log**	**2008-2007 (2.1d) log**	**(2.1d) log**	**(2.1d) log**	**2009-2008 (2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**	**2010-2009 (2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**
Dep. Var.	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables													
LnDistSolva(geo)		-0.493											
LnDistZ(geo)			0.235										
LnDistsec (geo)	-0.238			3.479**									
DistSolva(lin/ort)		1.784***			1.597***			1.557***			1.597***		
DistZ(lin/ort)			0.532***			0.108			0.443***			0.497***	
Distsec(lin/ort)	1.501***			1.831***			1.592***			1.366***			1.437***
Independent Variables: Controls													
Const	-1.21*	0.017	-0.028	0.098	-0.012***	-2.368***	0.1	0.005**	0.059***	0.617**	0.001	0.18	0.041
(u)POP	80.116*	0.114	23.947	-8.396	-0.273	0.416	29.197	0.004	3.439	29.176	-0.024	21.617	1.869
GFCFGR	-0.008	0	-0.004	-0.005*	0	-0.008	-0.026	0	0	-0.008	0	-0.005	-0.002
Size	-1.011**	0.001	0.296*	-0.668**	0.002	-0.866***	0.541	-0.009**	-0.018	-1.025**	0.002	0.29*	-0.752**
Cap	-1.983	0.02	63.714***	-2.354**	-0.005***	0.411***	0.074	0.035	21.487***	-1.576	0.019	63.786***	-1.836*
EWG	-0.564**	-0.006	0.054	-0.012	-0.001	0.295*	-0.104	-0.002	-0.028*	-0.193	-0.005	0.078	-0.001
R^2	14.6%	20.4%	69.2%	19.0%	17.9%	21.6%	12.9%	13.2%	76.9%	14.2%	19.7%	69.2%	16.9%
FStatistic	27***	40.4***	353.5***	36.9***	40***	50.6***	27.2***	28***	611.6***	30.5***	45.1***	412.5***	37.5***
Breusch-pagan	0.00	0.80	0.00	0.00	0.00	0.33	0.28	0.00	0.00	0.00	0.91	0.00	0.00
DurbinWuHaus	-0.11	0.021	0.07	1.317				-4.446	-5.424	-0.043	-0.229	0.529*	-48.382

Note: In columns 1-9 the variables are based on geographic distance, in columns 10-18 the variables are based on geographic as well as on linguistic distance, and in columns 19-27 the variables are based on linguistic distance. The dependent variable in all columns is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firms (1,111 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines of each subtable. The independent control variables of all models with regression coefficients are defined as in table 3. In the last line Breusch Pagan test statistic is presented. White robust standard errors are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-22: Results (Cross section log measure 5)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Distance	Geographic									Geographic/Linguistic				
Model	**2008**			**2009**			**2010**			**2008**			**2009**	
	(2.1d) log	**(2.1d) log**	**(2.1d) log**	**(2.3) log**	**(2.3) log**	**(2.3) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) log**	**(2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**	**(2.1d) log**	**(2.1d) log**
Dep. Var.	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**
Independent Variables: Distance Variables														
LnDistSolva(geo)	-0.002			0			0.01**			-0.1			0.249*	
LnDistZ(geo)		0.653**			0.726			1.193***			2.205			-0.04
LnDistsec (geo)			0.702*			1.995***			0.383**			7.371***		
DistSolva(lin/ort)										0.01**			-0.004	
DistZ(lin/ort)											0.173			0.011
Distsec(lin/ort)												-0.335		
Independent Variables: Controls														
Const	0.082***	-5.403***	3.15**	0.059***	-3.437**	-1.183	0.033***	-13.45***	-0.075	-0.099	-6.63***	-7.82**	0.52**	-1.873
Lag dep Var	0.948***	0.801***	0.775***	0.74***	0.678***	0.527***	0.965***	0.884***	0.946***	0.947***	0.801***	0.771***	0.74***	0.679***
(u)POP	-0.002**	0.058**	0.079	0.001	0.019	-0.197**	0	0.064	-0.072	-0.003**	0.044*	-0.142	0.002**	0.033
GFCFGR	0	-0.004*	-0.046**	0	-0.002*	-0.021**	0	0	-0.014*	0	-0.004*	-0.04**	0	-0.001
Size	-0.004***	0.072***	-0.168**	-0.001**	0.111***	-0.016	0	0.401***	0.001	-0.004***	0.072***	-0.196***	-0.001**	0.112***
Cap	0.122***	11.887***	-3.432	0.071*	9.161***	6.474*	0.017	63.404***	-1.167	0.117***	11.741***	-4.92*	0.081**	9.265***
EWG	0.006**	0.056**	0.331*	-0.004*	-0.041	-0.45*	-0.005***	-0.725***	-0.069	0.005*	-0.008	0.11	-0.005*	-0.031
R^2	59.9%	80.8%	65.7%	72.6%	71.7%	62.4%	88.0%	62.8%	80.3%	59.9%	80.8%	66.0%	72.7%	71.7%
FStatistic	234.9***	665.1***	302.4***	418***	399.5***	261.5***	1157.3***	266***	641.4***	206.1***	580.5***	266.8***	366.6***	348.7***
Breusch-pagan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01

Column	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)
Distance	Geographic/Linguistic				Linguistic								
Model	**2009 (2.1d) log**	**2010 (2.3) log**	**(2.3) log**	**(2.3) log**	**2008 (2.1d) log**	**(2.1d) log**	**(2.1d) log**	**2009 (2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**	**2010 (2.1d) lna**	**(2.1d) lna**	**(2.1d) lna**
Dep. Var.	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**	**SOLVA**	**Z**	**SEC**
Independent Variables: Distance Variables													
LnDistSolva(geo)		0.24***											
LnDistZ(geo)			11.844***										
LnDistsec (geo)	8.402***			2.525**									
DistSolva(lin/ort)		-0.002			0.004			0.013**			0.007**		
DistZ(lin/ort)			0.761***			0.688**			0.302			1.207***	
Distsec(lin/ort)	1.119***			0.384*			0.648*			1.025***			0.412**
Independent Variables: Controls													
Const	-10.78***	0.449***	-22.32***	-3.197*	0.091***	-5.467***	3.131**	0.08***	-2.544**	-0.101	0.028***	-13.31***	-0.07
Lag dep Var	0.522***	0.965***	0.898***	0.942***	0.947***	0.802***	0.776***	0.74***	0.679***	0.527***	0.965***	0.883***	0.946***
(u)POP	-0.385***	0.001	0.108*	-0.135*	-0.002**	0.054**	0.072	0.001*	0.024	-0.145*	0	0.016	-0.087
GFCFGR	-0.027***	0	-0.02*	-0.012	0	-0.004*	-0.051***	0	-0.002	-0.01	0	0.001	-0.014*
Size	-0.016	0	0.392***	0.003	-0.004***	0.072***	-0.162**	-0.001*	0.111***	-0.016	0	0.396***	0
Cap	6.202*	0.025*	59.912***	-1.398	0.121***	11.659***	-3.095	0.062*	9.174***	7.686**	0.015	62.016***	-1.214
EWG	-0.681***	-0.005***	-0.844***	-0.241*	0.007***	0.046*	0.333	-0.004*	-0.037	-0.027	-0.005***	-0.667***	-0.078
R^2	62.9%	88.1%	63.5%	80.3%	59.9%	80.8%	65.7%	72.7%	71.7%	62.3%	88.0%	63.2%	80.3%
FStatistic	233.5***	1024***	240.1***	562.9***	235***	664.7***	302.1***	419.5***	399***	259.9***	1156.4***	270.2***	641.5***
Breusch-pagan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: In columns 1-9 the variables are based on geographic distance, in columns 10-18 the variables are based on geographic as well as on linguistic distance, and in columns 19-27 the variables are based on linguistic distance. The dependent variable in all columns is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firms (1,111 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines of each subtable. The independent control variables of all models with regression coefficients are defined as in table 3. In the last line Breusch Pagan test statistic is presented. White robust standard errors are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).

Table SI-23: Results (Cross section linear measure 5)

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Distance					Geographic							Geographic/Linguistic		
Model	**2008**			**2009**			**2010**			**2008**			**2009**	
Dep. Var.	**(2.1d) log SOLVA**	**(2.1d) log Z**	**(2.1d) log SEC**	**(2.3) log SOLVA**	**(2.3) log Z**	**(2.3) log SEC**	**(2.1d) log SOLVA**	**(2.1d) log Z**	**(2.1d) log SEC**	**(2.1d) lna SOLVA**	**(2.1d) lna Z**	**(2.1d) lna SEC**	**(2.1d) log SOLVA**	**(2.1d) log Z**
Independent Variables: Distance Variables														
LnDistSolva(geo)	-0.01			-0.001			0.058**			0.007			0.051	
LnDistZ(geo)		0.079**			0.083			0.169***			0.111***			0.084
LnDistsec (geo)			0.137*			0.464***			0.124**			0.172*		
DistSolva(lin/ort)										0.031			0.102**	
DistZ(lin/ort)											0.052			0.002
Distsec(lin/ort)												0.104		
Independent Variables: Controls														
Const	0.087***	-4.2***	3.596**	0.06***	-2.093***	-0.313	0.006	-11.295 ***	-0.041	0.083***	-4.267***	3.442**	0.048***	-2.094***
Lag dep Var	0.948***	0.801***	0.774***	0.74***	0.678***	0.526***	0.965***	0.885***	0.943***	0.948***	0.802***	0.774***	0.74***	0.678***
(u)POP	-0.002**	0.058**	0.062	0.001	0.019	-0.215**	0	0.062	-0.078	-0.002**	0.05**	0.016	0.001*	0.019
GFCFGR	0	-0.004*	-0.044**	0	-0.002*	-0.024**	0	0.001	-0.012	0	-0.004	-0.043**	0	-0.002
Size	-0.004***	0.072***	-0.17**	-0.001**	0.111***	-0.02	0	0.399***	0	-0.004***	0.072***	-0.167**	-0.001*	0.111***
Cap	0.122***	11.886***	-3.535	0.071*	9.163***	6.012*	0.017	63.415***	-0.983	0.12***	11.734***	-3.314	0.061*	9.159***
EWG	0.006**	0.056**	0.388**	-0.004*	-0.04	-0.34	-0.005***	-0.729***	-0.092	0.006**	0.045*	0.29	-0.004*	-0.04
R^2	59.9%	80.8%	65.7%	72.6%	71.7%	62.5%	88.0%	62.8%	80.3%	59.9%	80.9%	65.8%	72.7%	71.7%
FStatistic	234.9***	665.1***	302.5***	418***	399.5***	262.1***	1157.2***	265.8***	642.7***	205.4***	582.2***	264.6***	367.1***	349.3***
Breusch-pagan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Column	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)
Distance	Geographic/Linguistic				Linguistic								
Model	**2009**	**2010**			**2008**			**2009**			**2010**		
Dep. Var.	**(2.1d) log SEC**	**(2.3) log SOLVA**	**(2.3) log Z**	**(2.3) log SEC**	**(2.1d) log SOLVA**	**(2.1d) log Z**	**(2.1d) log SEC**	**(2.1d) lna SOLVA**	**(2.1d) lna Z**	**(2.1d) lna SEC**	**(2.1d) lna SOLVA**	**(2.1d) lna Z**	**(2.1d) lna SEC**
Independent Variables: Distance Variables													
LnDistSolva(geo)		0.07**											
LnDistZ(geo)			0.23***										
LnDistsec (geo)	0.465***			0.146**									
DistSolva(lin/ort)		0.025			0.022			0.085**			0.043**		
DistZ(lin/ort)			0.128***			0.08**			0.033			0.165***	
Distsec(lin/ort)	0.014			0.079			0.167*			0.244***			0.127**
Independent Variables: Controls													
Const	-0.306	0.004	-11.261 ***	-0.073	0.081***	-4.197***	3.4**	0.042***	-1.982***	0.341	0.008	-11.218 ***	0.004
Lag dep Var	0.526***	0.964***	0.895***	0.942***	0.947***	0.802***	0.774***	0.74***	0.679***	0.527***	0.965***	0.885***	0.944***
(u)POP	-0.217**	0	0.023	-0.096	-0.002**	0.055**	0.012	0.001*	0.025	-0.164*	0	0.014	-0.1
GFCFGR	-0.024**	0	-0.002	-0.012	0	-0.004*	-0.046***	0	-0.002	-0.011*	0	0.001	-0.013
Size	-0.02	0	0.394***	-0.001	-0.004***	0.072***	-0.163**	-0.001*	0.111***	-0.018	0	0.399***	-0.001
Cap	6.002*	0.016	61.986***	-0.956	0.121***	11.665***	-3.093	0.062*	9.18***	7.452**	0.015	62.291***	-1.069
EWG	-0.342	-0.005***	-0.695***	-0.126	0.006***	0.047*	0.291	-0.004*	-0.036	0.026	-0.005***	-0.685***	-0.095
R^2	62.5%	88.0%	63.2%	80.3%	59.9%	80.8%	65.8%	72.7%	71.7%	62.3%	88.0%	62.9%	80.3%
FStatistic	229.1***	1012.3***	236.3***	562.4***	235***	664.6***	302.5***	419.6***	399***	260***	1156.3***	266.9***	642.9***
Breusch-pagan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: In columns 1-9 the variables are based on geographic distance, in columns 10-18 the variables are based on geographic as well as on linguistic distance, and in columns 19-27 the variables are based on linguistic distance. The dependent variable in all columns is as defined in table 3. The sample (over the period 2008–2010) is drawn from databases described in section 3 and is restricted to firms (1,111 observations) where values of all variables are available. The central independent variables with regression coefficients are depicted in the first six lines of each subtable. The independent control variables of all models with regression coefficients are defined as in table 3. In the last line Breusch Pagan test statistic is presented. White robust standard errors are applied to determine significance (*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively).