Workshop on

“The Costs and Benefits of International Banking”
Eltville, 18 October 2011

Bettina Brüggemann
University of Frankfurt

Jörn Kleinert
University of Graz

Esteban Prieto
University of Tübingen

“A gravity equation of bank loans“

A Gravity Equation for Bank Loans*

Comments are welcome

Bettina Brüggemann
University of Frankfurt

Jörn Kleinert
University of Graz

Esteban Prieto
University of Tübingen

First Version: January 2011
This Version: May 2011

Abstract

We present a gravity equation for bank loans derived from a firm's search for the "ideal" loan offered by a bank. A loan offer can have various dimensions (maturity, amount, timing, collateral, disclosure requirements) that involve costs that go beyond the mere interest rate. Taking all the costs into account, a firm chooses the cost minimizing bank loan. Based on this decision criterion, we derive the probability of a firm from country i choosing any bank from country j for the desired loan contract. We then use this probability to model the total volume of bank loans issued in country j to firms from country i. The resulting equation resembles a gravity equation for cross-border bank loans. The derived specification requires fixed-effects for the lending country and the borrowing country. Based on the theoretical model, we estimate the gravity equation using different methods while controlling for the unobserved heterogeneity proposed by our theoretical model.

JEL classification: L14, F34, G21

Keywords: Product differentiation, Gravity Equation, Cross-border bank lending

*E-Mails of the authors: bettina_brueggemann@hof.uni-frankfurt.de, joern.kleinert@uni-graz.at, esteban.prieto@wiwi.uni-tuebingen.de.

1 Introduction

International bank lending data can be explained surprisingly well by the gravity forces which also explain international trade so well (see among others Buch 2005). The success of these gravity forces is puzzling given that transport costs, which are used to explain the negative effect of distance on trade, should not be important for cross-border loans. While the gravity model has often be applied in the context of international finance in general and banking activities in particular, there are only very few attempts to explain the underlying source of the success of the gravity equation in explaining cross-border finance activities (Martin & Rey 2004, Okawa & van Wincoop 2010). These approaches focus on explaining cross-border equity positions. We provide a theoretical foundation for the gravity equation in cross-border bank lending in order to (i) study the underlying economic relationships which allows for a meaningful interpretation of the results and (ii) to search for the proper specification of the gravity equation. Our theory gives an explanation for the success of the gravity equation to explain cross-border bank lending data. We pay particular attention to the role of distance, because the negative effect of distance in the transfer of "weightless" loans is often seen as puzzling. We see the role of distance between the firm and the envisaged banking market in the fact that distance raises the firm's costs when it searches for the best partner in a particular banking market. In addition distance increases the bank's costs when monitoring the firm.

Our theoretical model starts from the observation that a loan offer has various criteria in addition to the interest rate (e.g. maturity, amount, timing, collateral, disclosure requirements). Comparing two offers is therefore not trivial and the "ideal" offer does not necessarily feature the lowest interest rate. Moreover, a loan contract involves two parties and both must agree on the deal. Since we rely on aggregate data and cannot observe all the relevant characteristics, the offers in our study include a stochastic (i.e. unobserved) part which includes relationship-specific elements. Firms choose their preferred offer depending on the observed part and the part unobserved by the researcher. We assume that the unobserved parts are unrelated. Assuming further that the cost minima are Gumbel extreme value distributed, the probability that a firm chooses the offer of a particular bank can be derived. Summing over all banks from a particular country and the total need for external finance at home, we compute total cross-border loans from a particular lending country to the borrowing country. The resulting equation resembles a gravity equation for bank loans.

We then estimate the gravity equation using confidential locational cross-border bank lending data from the *Bank of International Settlement*. This data allow a direct empirical test of the theory. The results support the predictions of the theoretical model. We demonstrate the large differences that result if we deviate from the specification we derived from the theory. Moreover, we confirm that cross-border bank lending decreases with the distance between two countries. The negative effect of the distance coefficient survives various robustness tests. The distance effect prevails even after controlling for determinants which capture the ability of country pairs to reduce contracting, search and monitoring costs, and the effectiveness/quality of the lending country's banking system.

The gravity equation has a long tradition in the context of international trade but has been applied for other cross-border activities too. As mentioned above, there are only few attempts to find a theoretical justification for the use of the gravity equation in finance. Martin & Rey (2004) use a general equilibrium framework to model assets as imperfect substitutes that assure against different risks and account for transaction costs in cross-border asset trade. Portes & Rey (2005) use their structural gravity equation to explain cross-border asset holding. Okawa & van Wincoop (2010) use a portfolio approach to derive a gravity equation for equity. While resulting in a very similar gravity equation, their approach is very different from ours. They start from a representative investor holding a portfolio of different assets. In contrast to their approach, we focus on cross-border bank loans only. We have more to say about debt markets than about cross-border asset markets and derive a gravity equation based on an analysis of the relationship between a firm demanding a bank loan and the potential lending bank.

Our paper is most closely related to Head & Ries (2008), who build on Anderson et al. (1992) to derive a gravity representation for FDI. FDI is thereby seen as an outcome of the market for corporate control. Head & Ries (2008) model a matching between two firms in an auction for the specific assets of the acquisition target. The firm with the highest ability to use the assets profitably can afford to pay the highest price and therefore controls the assets. We have a similar auction in mind, when we model a loan as outcome of a firms search for the cheapest supplier of bank credit. Hence, instead of assuming a representative firm with several loans, we are closer to the "ideal loan approach".

2 A theory of aggregate cross-border bank loans

We start with a single firm that searches for a loan with particular characteristics (amount, maturity, timing) offered by a bank at the best conditions possible. These characteristics include the interest rate, fixed fees, collateral, reporting requirements and other variables a loan contract might feature. The OECD for instance states in the description of their interest rates in their Financial Indicator (MEI): "...rates vary not only because of inflation ... but also because of a number of other influences, including the amount, purpose and period of the transaction, the credit-worthiness of the borrower, the collateral offered and/or guarantees/guarantors available, the competition for the transaction, government policy. As a consequence, there will be numerous rates applying to a large number number of transactions that are in effect at any one time in any one country." (OECD2011)

Some of these features are observable to the researcher, others are not. Using the observable part, we predict the probability that a particular bank k from a foreign country j provides a representative firm g in country i with a loan. Using only country-level variables, this probability is not too informative with respect to an individual bank but they can be used to calculate the probability that a firm in country i receives a loan from any bank from country j. We use this probability to compute the stock of loans from banks from country j to firms in i.

2.1 Firm-bank relationship

Assume that a firm g from country i searches for a bank loan in N relevant countries including its home country. This firm chooses bank k in country j if this very bank offers the loan for the lowest overall cost c_{igjk}. The costs to the borrowing firm g are specific to its loan contract with bank k. They depend on the interest rate r_{igjk} for the loan on which the firm and the bank have agreed, but also on other factors a_{jk} that capture loan- and bank-specific cost factors. We might think of these additional factors captured by a_{jk} as features like the amount the firm borrows, the collateral, non-monetary variables such as timing and maturity, and on factors such as the efficiency of the bank involved. Let's assume that all these characteristics as well as the interest rate can be collapsed in one variable, overall costs c_{igjk}, and compared between any two loan offers.

While in our view the firm has the part of approaching banks and ask for loan offers, banks are be no means passive or non-optimizing here. We think of the whole procedure as three steps: First the firms ask a bank for the conditions for a loan for a specific project. Second, the firm and the bank negotiate the conditions and the bank makes a make-it or leave-it offer. The firm accepts the offer or not depending on the offered conditions by the other banks it had applied for a loan. Hence, the offered conditions depend also on the characteristics of the bank such as its evaluation of firm and project, and its alternative lending opportunities. Usually, overall costs c_{igjk} are not observable to the researcher. However, some elements of the costs can be observed. The interest rate r_{igjk} on which the two partners agree on is certainly affected by the prevailing interest rate r_j in the country of the bank. Think of the prevailing interest rate in j, r_j, as the average of the interest rates of all loan contracts between firms and banks in this country. The interest rate is also affected by the contracting costs and the costs of monitoring t_{igjk} incurred by the bank. Although the relationship-specific monitoring and contracting costs t_{igjk} are also not observable, they depend systematically on observable variables such as distance (and distance related variables) between the bank and the firm. This observable part of the monitoring costs can have several dimensions like physical and cultural distance, familiarity with the business, institutional features in the country of the debtor or the specific financing needs. The variable τ_{ij} denotes the part of the monitoring cost that is determined by the average distance between the two countries involved. The interest rate that lender and borrower agree on, r_{igjk}, can now be decomposed into an observable part and an unobservable part ϵ with zero mean:

$$r_{igjk} = r_j + \tau_{ij} + \epsilon_{igjk}.$$

Using the same logic, we rewrite the overall costs c_{igjk} as a function of observable variables at the country level and firm-bank-relationship specific unobservable parts that are collapsed in the error term ϵ_{igjk} which has zero mean. The borrowing firm's costs for a loan from bank k can then be described as

$$c_{igjk} = \beta r_j + \gamma \tau_{ij} + \delta a_j + \epsilon_{igjk},$$

where a_j stands for average bank characteristics in country j that might affect the cost of the loan. β, γ, and δ weigh the cost components. Note that our analysis does not depend on the assumption of a representative firm. In addition, the model is flexible enough to take into account any number of firms n in country i. Moreover, since the observable part of the cost function does not depend on firm-specific factors, we can interpret this observable part as the average costs of borrowing from a foreign bank k in country i.[1] We will stick to this interpretation of the above cost function from now on.

Average costs $\bar{c}_{ij}$ include only country averages and bilateral distance-related variables. Thus, the bank-specific deviations from these country-specific variables are also random from the perspective of the researcher, with ϵ_{igjk} collecting all random elements. The cost function can then be written as $c_{igjk} = \bar{c}_{ij} + \epsilon_{igjk}$. Firm g from country i choose bank k from country j to take out a loan if this loan minimizes their costs. The probability that firm g from country i chooses bank k from country j is then given by

$$\begin{aligned}
\mathbf{P}_{igjk} &= \Pr\left(c_{igjk} = \min\{c_{l1} \cdots c_{ln_l}; l = 1 \cdots N\}\right) \\
&= \Pr\left(\bar{c}_{ij} + \epsilon_{igjk} < \bar{c}_{il} + \epsilon_{igl1}; \cdots; \bar{c}_{ij} + \epsilon_{igjk} < \bar{c}_{il} + \epsilon_{igln_l};\;\; l = 1 \cdots N, jk \neq lh\right) \\
&= \Pr\left(\bar{c}_{ij} - \bar{c}_{il} + \epsilon_{igjk} < \epsilon_{igl1}; \cdots; \bar{c}_{ij} - \bar{c}_{il} + \epsilon_{igjk} < \epsilon_{igln_l};\;\; l = 1 \cdots N, jk \neq lh\right) \\
&= 1 - \Pr\left(\bar{c}_{ij} - \bar{c}_{il} + \epsilon_{igjk} \geq \epsilon_{igl1}; \cdots; \bar{c}_{ij} - \bar{c}_{il} + \epsilon_{igjk} \geq \epsilon_{igln_l};\;\; l = 1 \cdots N, jk \neq lh\right) \\
&= \underbrace{\prod_{l=1}^{N} \prod_{h=1}^{n_l}}_{lh \neq jk} \left[1 - F(\bar{c}_{ij} - \bar{c}_{il} + x)\right], l = 1 \cdots N,
\end{aligned}$$

where $l = 1 \cdots N$ is the country index and the index $h = 1 \cdots n_l$ denotes the different banks in country l that are considered by a firm in country i. The random terms ϵ are i.i.d.

[1] Firm heterogeneity only enters the cost function through the error term ϵ_{igjk}, which has zero mean by assumption. When averaging over all firms it would therefore simply cancel out, which renders our model very flexible with regard to the number of firms included.

Their cumulative distribution function is denoted by F, the corresponding density f. For any given realization x of ϵ_{igjk}, bank loan from k in country j are chosen with probability density

$$\underbrace{\prod_{l=1}^{N}\prod_{h=1}^{n_l}}_{lh\neq jk}\left[1-F(\bar{c}_{ij}-\bar{c}_{il}+x)\right],$$

where we multiply over all eligible bank loan variants h in all countries l. None of the deterministic variables in this equation varies from bank to bank; the variation is between countries only. We therefore simply rewrite it as $\prod_{l=1}^{N}\left[1-F(\bar{c}_{ij}-\bar{c}_{il}+x)\right]^{n_l}$, keeping in mind that n_j excludes the firm-bank relationship in question, jk. Accounting for all possible realizations of x, the probability $\mathbf{P}_{ijk}$ of choosing bank k in country j can be calculated using:

$$\mathbf{P}_{ijk}=\int_{-\infty}^{\infty}f(x)\prod_{l=1}^{N}\left[1-F(\bar{c}_{ij}-\bar{c}_{il}+x)\right]^{n_l}dx. \tag{1}$$

The probability that a firm g from country i choose a bank k from country j depends on the average cost characteristics in country j, on the bilateral characteristics, and on the unobservable relationship-specific cost components ϵ.

2.2 Cross-border bank loans at the country level

We are interested in the minimum values of the random component ϵ_{igjk}. For a bank, the probability to be selected equals the probability to offer the ideal package. Thus, we need an extreme value statistic to approximate the distribution of the minimum values of the firm-bank-relationship specific cost component ϵ. The Gumbel distribution is a natural approximation. Thus, we assume the minima of the random components ϵ are Gumbel distributed. The Gumbel distribution is the limiting distribution of several common distributions such as the normal, the log-normal, the exponential or the logistic distribution. As limiting distribution it is an asymptotic distribution that does not depend on the sampled distribution function which we don't know. As the normal distribution arises as limiting distribution for the mean the Gumbel distribution arises as limiting distribution for extreme values (maxima and minima). It is obtained for the greatest or smallest value in random samples of increasing size for continuous random variables that are i.i.d. However, it has to be unbounded in the direction of the extreme value. In the case of minima, this holds only true for the normal distribution. We therefore assume that ϵ is normally distributed, which is also the natural way to account for the many unobserved characteristics that we argued for above which's unobserved parts are summed up to establish ϵ.

The Gumbel distribution has a double exponential form. For minimum values of the residuals with zero mean, it is given by

$$F(x) = 1 - \exp\left[-\exp\left(\frac{x}{\sigma} - \gamma\right)\right] \tag{2}$$

where σ is a constant scale parameter describing the "horizontal stretching" of the distribution, and γ is Euler's constant ($\gamma \approx 0.5772$). From the cumulative distribution function $F(x)$, the density function $f(x)$ can be derived as

$$f(x) = \frac{1}{\sigma}\exp\left(\frac{x}{\sigma} - \gamma\right)\left\{\exp\left[-\exp\left(\frac{x}{\sigma} - \gamma\right)\right]\right\}.$$

Applying the Gumbel distribution, the probability that a bank k from country j provides loans to a firm from country i is thus given by

$$\mathbf{P}_{ijk} = \int_{-\infty}^{\infty} \frac{1}{\sigma}\exp\left(\frac{x}{\sigma} - \gamma\right)\left\{\exp\left[-\exp\left(\frac{x}{\sigma} - \gamma\right)\right]\right\} \prod_{l \neq j}^{N} \exp\left[-\exp\left(\frac{\bar{c}_{ij} - \bar{c}_{il} + x}{\sigma} - \gamma\right)\right]^{n_l} dx. \tag{3}$$

Solving this integral (see Appendix) yields the following expression for the probability of choosing bank loan variant jk:

$$\mathbf{P}_{ijk} = \frac{\exp\left(-\frac{\bar{c}_{ij}}{\sigma}\right)}{\sum_{l=1}^{N} n_l \exp\left(-\frac{\bar{c}_{il}}{\sigma}\right)}. \tag{4}$$

This equation gives us the same probability for every bank k from each country j that a firm from country i takes into consideration for borrowing. Thus, because we use only aggregated data, bank-level probabilities are not very informative. We are much more interested in the probability that a firm in country i chooses any one of the banks from country j to take out a loan. We can obtain an expression for this probability by simply aggregating the probabilities over all n_j eligible banks in country j. The probability of a firm choosing any bank in country j can then be written as

$$\mathbf{P}_{ij} = \frac{n_j \exp\left(-\frac{\bar{c}_{ij}}{\sigma}\right)}{\sum_{l=1}^{N} n_l \exp\left(-\frac{\bar{c}_{il}}{\sigma}\right)}. \tag{5}$$

To arrive at an expression for the total amount of bank loans floating from country j to country i, BA_{ji}, we simply multiply the above probability by the total amount of bank loans in country i, BL_i:

$$BA_{ji} = \frac{n_j \exp\left(-\frac{\beta r_j + \gamma \tau_{ij} + \delta a_j}{\sigma}\right)}{\sum_{l=1}^{N} n_l \exp\left(-\frac{\beta r_l + \gamma \tau_{il} + \delta a_l}{\sigma}\right)} BL_i. \tag{6}$$

In this equation, we also replace the country-specific cost variables by their different components which we defined in the beginning of this section. Total loans are affected by the prevailing interest rate in country j, r_j, by the monitoring costs τ_{ij} that a bank in country j has to incur when lending to a firm in country i, and the average quality of country j's banks, captured by the average bank characteristics a_j. The effects of these variables are discounted by the sum of the effect of these variables in all the other countries a firm from country i could have chosen a loan from.

2.3 Comparison to gravity in equity

Before we turn to estimating the derived gravity equation, we pause to compare our approach to two other attempts to give the gravity equation a theoretical foundation. The main difference is that we look particularly at the cross-border loans segment of banking activities. Thereby, we model a firm-bank relationship, where the firm takes an active part while the bank is also optimizing: the firm approaches different banks to obtain a loan offer, respectively, and compares the different loan offers it receives. We model the relationship like this, because we believe that in the bank lending market the loan demanding firm has its active role. There are other segments of the financial markets such as the bond market and the equity market where the bank itself is searching for the optimal portfolio. Particularly the equity segments has been in the focus of research on cross-border bank activities. Although the bank might not decide about the alternatives, loans and equity, independently, we do not have an omitted variable bias in our approach because the bank's loan offer includes alternative opportunities to use the financial means.

Martin & Rey (2004) propose a two-country model with two periods to study portfolio (equity) investments. Individuals invest in differentiated assets in order to hedge against idiosyncratic risk in a stochastic world where each assets yields a payment in one out of L stages of nature that can occur in period two. Individuals are risk-averse with constant relative risk aversion. Together with the structure of the payoffs in the second period, risk aversion creates a love of diversity similar to the love of variety in trade models which yields a CES-like structure of asset holdings. Based on this structure Martin & Rey (2004) derive a gravity equation which is in its structure very close to gravity equations from trade models but proposes different "mass" and "distance cost" variables. In addition, the expected return on assets in the investment country shows up in the equation. There are no fixed effects proposed which stems from the two-country structure. Coeurdacier & Martin (2009) extend the model to n countries and yield a gravity equation with a multilateral resistance term for the home and the host country.

In contrast to Martin & Rey (2004), who basically derive a gravity equation for financial holdings when countries trade claims on Arrow Debreu securities, Okawa & van Wincoop (2010) derive their gravity equation from standard static portfolio theory. In their model, investors can hold claims on risky assets from many different countries, which are affected by a global and by country-specific risk factors. Okawa & van Wincoop (2010) also allow for trade in a risk-free asset as well as an asset whose return depends only on global risk. In addition, they introduce financial frictions that are mainly due to information asymmetries. Assuming asset market clearing, they derive a gravity equation where economic masses are given by the supply of equity in one country and total equity holdings in the other country, and distance enters the equation through a relative friction term that relates bilateral financial frictions to multilateral resistance terms. Okawa & van Wincoop (2010) introduce three different, easily applicable methods of estimating these bilateral frictions based on a number of observables, such as distance, common language, same legal system.

Thus, although all three theoretical approaches motivate a gravity equation, the proper specifications and the interpretation of the variables are rather different. Moreover, the theories apply to different segments of the banking sector, a feature that we use in the empirical part to assess our approach. We can break down banks' claims on foreign borrowers by their sector, i.e. other banks, private non-banks and public. We focus on the bank-private non-bank relationship.

3 Data and Empirical Specification

Equation (6) is a gravity equation. The dependent variable, BA_{ij}, is the total amount of bank loans in the debtor country i from lending country j. We construct BA_{ji} using year-end data on cross-border bank claims from the unpublished locational bilateral banking data provided by the Bank of International Settlement (BIS). The sample that we use covers the period 2000 to 2006.

The BIS provides two different sets of banking statistics: a locational banking statistic and a consolidated banking statistic. The consolidated banking data provide information on international claims of domestic bank head offices on a *consolidated* basis, i.e. the data consolidate financial positions of the bank head office and their foreign affiliates. This data thus provides a detailed picture of the country exposure risk of a reporting countries banking system. Besides the aggregate volume of international claims, the consolidated banking statistic also provides disaggregated data based on the sector of the borrower:

banks; public sector; and non-bank private sector. In contrast to the consolidated banking statistics, the BIS's locational data is *not* consolidated but based on the residence principle (balance-of-payment principle), i.e. they include information on the *gross* on-balance sheet asset and liability *vis-à-vis* non-resident entities (banks and non-banks). The BIS locational banking statistic thus provides information on the actual amount outstanding of cross-border claims held by the domestic banking system. In the empirical setup we will resort to the locational banking data. The reason is that the theoretical model laid out in the previous section makes predictions concerning the structure of cross-border lending from country A to country B. The information on international financial claims provided in the locational banking statistics allows discriminating between actual cross-border lending and lending from foreign affiliates. This feature of the data makes the locational statistics more appropriate for a structural estimation of the gravity equation derived above.

There are two variables accounting for economic masses: total loans in the loan receiving country BL_i and the number of banks in the loan sending country, n_j. The two mass variables have a positive effect on bilateral bank loans. Our theory predicts that the coefficients of these variables is one. We use the *Financial Structure Database* collected by Beck et al. (2009) to find data for these variables. We measure total loans in the debtor country i using private credit by deposit money banks and other financial institutions. In the original dataset this variable is defined as claims on the private sector by deposit money banks and other financial institutions divided by GDP. In order to arrive at the level of total loans in the borrowing country BL_i we multiply this variable with GDP. While total loans in the host country are available for most countries, the number of foreign banks n_j is harder to come by. The problem is that n_j refers to all foreign banks over which a firm searches for the "ideal loan" which is necessarily unknown. We proxy n_j by the inverse of the share of the top three banks in the market, as a measure of the effective number of banks in the loan granting country. Since this measure is independent of the total number of banks in country j, we also include fixed effects for country j. The number of banks over which a firm searches depends, however, also on other determinants like the intensity of contact between the two countries or the degree to which the banks in country j are known to firms in country i. In the theoretical framework this feature is captured by the dependence of the number of banks a firm screens on the distance between the firm and the bank. In the empirical framework these characteristics are accounted for by the physical distance and other distance related variables between the two countries $dist_{ij}$. Distance and distance related variable are from the *CEPII*.

The different cost components have a negative effect on the volume of cross-border bank loans. The country-specific cost function $c_{igjk} = \beta r_j + \gamma \tau_{ij} + \delta a_j + \epsilon_{igjk}$ specified above explains the costs by the prevailing interest rate r_j, the monitoring costs τ_{ij}, and the quality of the banking sector a_j. We proxy the monitoring costs by physical distance (in logs) and other distance related variables. Note that in the theoretical set-up distance enters the equation at two different points: in the monitoring costs which strongly depend on the distance between the lending bank and the borrowing firm *and* in determining the actual number of banks per country j that a firm considers for borrowing (see above). Since both effects enter linearly they can be added and estimated together. We approximate the loan sending country's lending rate r_j by the average (implicit) bank lending rate in country j. We construct r_j using the *OECD Banking Statistic on income statement and balance sheet*. We calculate an implicitly bank lending rate in country j by dividing the banking system's interest income by its total assets. Since the quality of the banking sector in country j a_j is largely unobserved we introduce country-j specific fixed-effects D_j to appropriately control for this unobserved heterogeneity.

The sum in the denominator in equation (6) is the same for every bilateral relationship of a particular receiving country i. This sum differs however between any two receiving countries. In this sense the denominator is closely related to what has become known in the trade literature as multilateral (price) resistance terms (Anderson & van Wincoop 2003). In order to account for these "multilateral (cost) resistance terms" we introduce country i-(receiving country) specific fixed effects D_i. Equation (6) shows that neglecting to account for the sum in the denominator when estimating a gravity equation for bilateral bank lending introduces an omitted variable bias. The results would be biased because the multilateral resistance terms are correlated with r_j, τ_{ij} and a_j. Applying these considerations to equation (6) and adding an error term, we can express the empirical analogue to the gravity equation for cross-border bank lending as:

$$BA_{ji} = \exp\left[(\beta_1 r_j + \beta_2 \tau_{ij} + D_j - D_i)\right] n_j^{\beta_3} BL_i^{\beta_4} \epsilon_{ij}. \tag{7}$$

The common procedure for estimating a gravity equation as given in equation 7 is to take logs of both sides, which yields

$$ln(BA_{ji}) = \beta_1 r_j + \beta_2 \tau_{ij} + \beta_3 ln(n_j) + \beta_4 ln(BL_i) + D_i + D_j + ln(\epsilon_{ij}) \tag{8}$$

The parameters in equation (8) can be estimated using linear regression with country i-(receiving country) specific-fixed effects (D_i) to account for multilateral (cost) resistances, and country j-(sending country) specific-fixed effects to control for its size and the unobserved banking market characteristics in country j. Fixed-effects ordinary least squares (FE-OLS) estimation produces consistent parameters estimates given that the variance of the error term ϵ_{ij} in equation (7) is independent of the explanatory variables. If this condition is violated, linear least squares estimates are inconsistent. To cope with the inconsistency of the FE-OLS estimator Silva & Tenreyro (2006) propose using a Poisson pseudo-maximum likelihood estimator which is robust to different patterns of heteroskedasticity. Poisson estimation allows estimating the gravity equation from its multiplicative form directly. The gravity equation in multiplicative from can be written as:

$$BA_{ji} = exp\left[\beta_1 r_j + \beta_2 \tau_{ij} + \beta_3 ln(n_j) + \beta_4 ln(BL_i) + D_i + D_j\right] \epsilon_{ij}. \tag{9}$$

This specification follows directly from the multiplicative form in equation 7.

According to the simulation results in Silva & Tenreyro (2006) and Westerlund & Wilhelmsson (2011), the bias introduced by log-linearizing a gravity equation can be substantial. In order to get a feeling about the importance of this issue in the context of international bank lending, we compare the results of both estimation procedures.

An issue to note is that our theoretical model is derived in a static framework. In its standard form, the gravity equation deduced from the theoretical framework is therefore derived for a cross-section estimation. However, it is straight forward to extent the empirical model to a panel analysis. Panel data and the related panel econometric methods have some well known advantages over cross-sectional econometric methods. One main advantage of using a panel econometric framework is the reduction of the "risk of obtaining biased results" by controlling for time-invariant specific effects (Baltagi 2001). Furthermore, using panel data models is advantageous because such data give more information, more variability, less collinearity amongst the variables, more degrees of freedom and more efficiency (Baltagi 2001). We therefore want to exploit the panel nature of our dataset. The economic content of the parameter estimates from a panel estimation is however different from their cross-sectional counterpart. The outcome from a cross-section estimation represent long-term relationships, whereas the parameter estimates of a panel regression might better be interpreted as short-term, within group effects (Egger & Pfaffermayr 2003). Both interpretation go along with our theoretical model. Therefore, when we apply our gravity equation to panel data, the general set-up remains unchanged, but we need to take into account the time dimension in the sums and in the unobserved characteristics of the banking sector in country j in order to get the structural relationships between the variables. The correct empirical specification in a panel setup thus reads as follows:

$$ln(BA_{jit}) = \beta_1 r_{jt} + \beta_2 \tau_{ij} + \beta_3 ln(n_{jt}) + \beta_4 ln(BL_{it}) + D_{it} + D_{jt} + ln(\epsilon_{ijt}). \tag{10}$$

The analogue multiplicative panel-data econometric specification is given by

$$BA_{jit} = exp\left[\beta_1 r_{jt} + \beta_2 \tau_{ij} + \beta_3 ln(n_{jt}) + \beta_4 ln(BL_{it}) + D_{it} + D_{jt}\right] \epsilon_{ijt}. \tag{11}$$

In a panel setting, the multilateral resistance variables are time varying. Consequently, ignoring the time variation in the fixed effects can produce biased estimates because of an omitted variables bias. We therefore apply year-country fixed effects (i.e. time-varying country-specific fixed effects) for both lending and receiving countries in our empirical specification.[2]

Finally, the gravity representation is very flexible in including additional variables, because the explanatory variables specify monitoring costs τ_{ij} and the effective number of banks n_j, both depending on distance, and the quality a_j of the lending country's banking sector. Buch (2002) for instance uses *(geographical) Distance*, *Language*, *EU membership*, *Capital controls* and the *Legal system* to catch the different dimensions of distance between the lender and the borrower. The quality of the banking sector is assumed to be reflected by *GDP per capita*, the *Index of human development*, and *Restrictions in banking and finance* in her empirical approach. Papaioannou (2009) in contrast concentrates on political and institutional factors which he shows are also important determinants in explaining international bank lending. Our theoretical model gives testable predictions only for the variables included in equation (7), but as stated above there is some flexibility in specifying the monitoring costs τ_{ij} and the quality of lending country's banking sector. Since our aim is not to differentiate between different theoretical models, we are as parsimonious as possible when choosing our variables. Instead, our main goal in the empirical part is to document the importance of including theoretically motivated fixed effects, at the receiving and host country level, in order to properly identify the effects of the main variables of interest.

4 Results

In Table 1 we present the results of the estimation of the gravity equation. Column (1) in Table 2 shows the results using the Poisson estimator which estimates the gravity equation in its multiplicative form. Column (2) and column (3) present results using the conventional fixed effects OLS approach. In column (2) the dependent variable is the logarithm of the stock of cross-border bank loans while in column (3) the dependent variable is constructed as the logarithm of one plus the stock of cross-border bank lending ($ln(1+BA_{ij})$). The latter approach is a common procedure to deal with zeros in international trade or financial holding data when estimating log-linearized gravity equations.

The Poisson estimator produces coefficient estimates on the economic mass variables, total bank loans in the debtor country BL_{it} and the effective number of banks in the lending country n_{jt}, which are in line with the predictions of the theoretical model. Both mass variables have the expected positive sign and are significant at the 1% level. This indicates that bilateral lending activity is significantly larger between countries with larger financial systems. It is worth noting that the coefficients on total bank loans in the

[2] See Baier & Bergstrand (2007) and Baldwin & Taglioni (2006) for similar approaches to deal with country effects in a panel setup

receiving country and the coefficient on the number of banks in the lending country are significantly different from one.[3] According to Silva & Tenreyro (2006), coefficients on mass variables which are different from unity might be explained by the empirical fact that smaller countries tend to be more open to international markets.

The two variables representing the cost components of firms' funding have negative effects on the volume of bilateral bank lending. The negative distance coefficient is consistent with the theoretical prior that firms tend to have more problems with appropriately screening remote foreign banking markets. Simultaneously, the negative distance effect might also reflect that monitoring costs incurred by the lending bank increase with distance. The point estimate of the distance coefficient is -0.368 and significantly smaller than the common, a-theoretical prior of -1 (Baldwin & Taglioni 2006). The other cost variable in the gravity equation is the bank lending rate in the loan sending country. The coefficient on r_{jt} is negative and significant at the 5% level. This reflects the theoretical rational, that pure (interest rate) cost considerations play an important role for firms when accepting a loan offer. Hence, higher average lending rates in the loan sending country act as a barrier for cross-border lending.

The R^2 of the gravity equation is 0.819, which shows that this simple and parsimonious model is a reasonable good description of cross-border bank lending data. Only about 18% of the variation in the data is left unexplained.

We now compare the results obtained using the Poisson estimator with the results using the common FE-OLS procedure. We consider this to be an important task, since we are not aware of any systematic evaluation of the different estimators applied to bank lending data. It is therefore unclear, whether the concerns raised by Silva & Tenreyro (2006) are quantitatively or qualitatively important.

First, in contrast to the Poisson estimator, the distance coefficient in both OLS specifications is significantly larger than the corresponding coefficient produced by the Poisson estimator (point estimates of -0.852 and -0.881). Also, the proxy for the effective number of banks exerts a significantly larger effect in the OLS framework. The coefficients on the stock of bank loans in the debtor country produced by the OLS regressions are comparable to the results from the Poisson regression. Most strikingly however, the coefficient on the lending rate in the loan sending country j turns insignificant in the OLS specification using the logarithm of the stock of cross-border bank lending as dependent variable. This is in contrast to the predictions of the theoretical model that the interest rate is an important characteristic in the firm's decision to accept a loan offer. In the model with $ln(1+BA_{ij})$

[3] Note that our measure of the number of banks in the receiving country a firm applies to is only a, possibly crude, approximation to the correct theoretical construct. The fact that the coefficient on this variable is significantly different form the theoretical prior might therefore just reflect the fact that our approximation captures only parts of the "true" underlying factor. We will explore this issue further in the next version of the paper.

as dependent variable the lending rate has a similar point estimate is significant. A general picture which emerges from our result is that the point estimates produced by the two OLS models are very similar. This suggests that differences between the Poisson and OLS estimator are not due to truncation (because of zeros in the data) but rather due to log-linearization of the model in the presence of heteroskedasticity which leads to inconsistent estimates.

The last three rows in Table 1 present three test statistics to check the adequacy of the estimated models, and the two models' assumptions concerning the structure of heteroskedasticity. The RESET test is a test for the correct specification of the conditional mean of the proposed model. In a nutshell, the RESET test checks whether the model is correctly specified by amending the proposed model with the squared fitted value of the estimated model, i.e. one amends the model with $(x'b)^2$, with b being the estimated coefficients. Under the hypothesis that the model is correctly specified, testing the significance of the additional regressor provides a test of the adequacy of the proposed model. In both OLS models, the RESET test rejects the null hypothesis that the parameter on the additional variable is zero. This means that the model specification used in the OLS approach is inappropriate. In contrast, this test provides no evidence for misspecification of the gravity model when estimated using the Poisson estimator.
The Park test (see Park 1966), suggested by Manning & Mullahy (2001), is a test of the hypothesis that the non-linear gravity equation can be consistently estimated based on the log linear model (see also Silva & Tenreyro 2006). The test rejects the Null hypothesis at any conventional significance level. The test statistic labeled GNR was proposed by Silva & Tenreyro (2006) to check the validity of the assumption imposed by the Poisson Pseudo Maximum Likelihood estimator that the conditional variance of the dependent variable is proportional to the conditional mean of the dependent variable. The p-value of the test is given by 0.113 which implies that the Poisson estimator's assumption cannot be rejected.[4]

In sum, using the multiplicative form of the gravity equation to estimate the gravity model delivers results which are in line with the theoretical model. When applying the traditional logarithmic approximation of the non-linear gravity model we observe striking differences, although the empirical gravity equation is specified in a way consistent with the theoretical model.[5] Further, tests to check the validity of the different models consistently favor the non-linear model over the log-linearized version. Insofar, our results support

[4]Note however that even if the GNR test is rejected, the Poisson estimator still produces consistent coefficient estimates. A rejection merely means that there exists a different non-linear Pseudo Maximum Likelihood estimator which is more efficient.

[5]Some authors have argued that a correct specification of the gravity equation has to control for omitted heterogeneity at the country pair level. Such effects enter the regression, and if not controlled for the error term, via the multilateral resistance variables, but are not controlled for by (time-varying) home and receiving country dummies. Note however, that our theory does not suggest introducing such time-varying country pair effects in the empirical model. Another disadvantage of country-pair effects is that time-constant country-pair variables, like the distance effect, are no longer identified since their effects are

the concern stated by Silva & Tenreyro (2006) that the linearization of a gravity equation might produce biased estimates. Given these results we proceed in the following sections with the Poisson estimator as our preferred estimation method and omit to present the results obtained from the log-linear model.

Effects of theory derived fixed effects

Our theoretical model specifies the empirical model. Particularly, our theory states that it is crucial to control for unobserved heterogeneity at the loan receiving and the loan sending country. In addition, in a panel framework we must account for the time variation in the loan receiving and the loan sending country fixed effects. It is natural to ask whether these features have important qualitative or quantitative implications for the results. We therefore compare our result from the theory consistent model specification to the results from specifications which ignore to control for one or more levels of unobserved heterogeneity.

In the columns (2) to (5) in Table 2 we show the results of models in which we ignore the time variation in the multilateral resistance variables (column (2)), the borrowing country fixed effects (column (3)), the lending country fixed effects (column (4)), and any form of multilateral resistance effects (column (5)). For ease of comparison, we present in column (1) the results from the benchmark regression once again.

Ignoring the time variation in the resistance variables seems to introduces a bias in the estimate of the effect of the number of banks n_{jt}. The coefficient drops significantly from 0.369 in the benchmark specification to 0.043. Additionally, the coefficient is rendered insignificant. The coefficient on the lending rate in the loan sending country r_{jt} becomes virtually zeros and is also insignificant. The coefficients on total loans and on the distance between the country pair are hardly affected. The estimates from the model without time variation and receiving country fixed effects (column 3) produces similar results than the model with fixed effects at the sending country and the receiving country level. Only the distance coefficient and the coefficient on total loans are slightly larger. Interestingly however, ignoring only unobserved effects from the source country (column 4) produces point estimates which are similar to those obtained from the correct specification. In fact, only the distance coefficient is significantly different from the results from the correct specification (at the 5% level). This illustrates, that it is possible that different biases compensate each other (Baldwin & Taglioni 2006). If we ignore any kind of unobserved heterogeneity and time variation, but only control for time effect which are common to all country pairs (column 5), we observe significantly different estimates on aggregate bank

completely wiped out by the bilateral dummies. Since these variable are of crucial importance in our theoretical model we restrain from including country-pair fixed effects.

loans in the borrowing country (at the 5% level) and on the distance coefficients (at the 1% level).

These results bear two immediate implications. First, using a misspecified model for the geographical properties of international bank lending might lead to erroneous conclusions. Second, without a proper theoretical framework which guides the empirical model specification it is difficult to discriminate between different empirical model proposals. As shown in our example, it is possible that a wrong model produces results which are in line with economic intuition and indistinguishable from the theory-consistent model specification.[6]

Tracing the effects of source banking market characteristics

According to our theoretical model, the quality of the foreign banking market from which firms borrow is an important determinant of overall credit costs. To understand how banking system specific characteristics affect international bank lending, we add to our baseline specification a set of control variables which characterize the profitability and the efficiency the lending country's banking market.[7] Both, the efficiency and the profitability of the banking system should have a direct impact on the borrowing costs of firms. Additionally, these specifications can reveal whether our two source country variables in the baseline regression, the effective number of banks n_{jt} and the lending rate r_{jt}, unintentionally capture effects of other banking market specific variables. Results of the regressions when augmenting our baseline specification with the net interest margin in the lending country ($margin_{jt}$), the return on assets (roa_{jt}) and the cost-income ratio ($cost - income_{jt}$) can be found in Table 3.

The net interest margin captures characteristics related to cost and income efficiency in the banking market (Beck et al. 2009). Higher values of this variable indicate lower levels of banking efficiency as there is a higher wedge between lending and deposit rates. The net interest margin has a negative and significant effect on cross-border bank lending indicating that less efficient banking market engage less in international lending. Using our theory, the economic interpretation is that banking market characterized by high inefficiency have a relative disadvantage in providing competitive loan offers to foreign firm. In fact, the negative, and still significant effect of the lending rate indicates that dealing with inefficient banks (banking markets) incurs cost over and beyond pure price (interest rate) effects.

[6]Note that we label an empirical model theory-consistent if it is specified according to the model presented in section 2.

[7]The variables used in these regressions are also taken from the *Financial Structure Database* from Beck et al. (2009).

Return on assets is a widely used indicator for bank profitability. Since more profitable banks should be able to provide more competitive contract terms we expect the coefficient of this variable to be positive. The estimated coefficient has the wrong sign and is not significantly different from zero.

Finally, the cost-income ratio, measured as the ratio of overhead cost to gross revenue, is a measure of non-operating cost efficiency. Higher values of the cost-income ratio indicate lower levels of cost efficiency. Similar to net interest margin, we expect that banking systems with lower ratios of overhead cost to revenues should be able to offer better loan terms. This should lead to higher stocks of outstanding cross-border loans for these banking systems. The coefficient on the cost-income ratio has the expected negative sign indicating that less (non-operating) cost efficient banking systems have lower cross-border loans outstanding. Concerning the four standard variables derived from the theory we observe that the main results are not changed due to the inclusion of these additional control variables. None of the coefficient estimates is significantly different from the results obtained from the benchmark regression presented in Table 1 and Table 2.

What conclusions would have been drawn if we would have used a misspecified model? Inspecting column 2 to 5 in Table 3 reveals some striking differences. First, all bank market specific variable turn insignificant if we ignoring time variation in the resistance terms or if we omit receiving country fixed effects (column(2) and column (3)). Second, omitting sending country fixed effects or any kind of unobserved heterogeneity renders the coefficient on return on assets significant (and still negative) while the lending rate in country j turns insignificant.

Search, contracting and monitoring costs

So far the distance between two countries has been the only variable intended to capture the effects related to contracting, search and monitoring costs. As already indicated, this cost component has various dimensions. In order to understand more deeply what is behind the various costs of searching, contracting, and monitoring we augment our gravity specification with variables that have been identified to be important in this context by other authors. In Table 4 we introduce dummy variables indicating the presence of a common border ($contig_{ij}$), the same official language ($comlang_{ij}$) and a common legal origin ($comlegor_{ij}$).[8]

First, the same language might reduces the cost of writing loan contracts which might be able to reduce transaction costs. Also, a common language should ex ante increase the probability that a firm applies to a certain bank. Both effects should lead to a larger

[8]The common legal origin variable is taken from Porta et al. (1998)

aggregate lending volume between two countries. The coefficient on the dummy variable indicating the same language has the wrong sign but is insignificant.

Whether two countries have a common border (*contig*) with one another has large, independent effects on *trade* between these countries even if one is already controlling for distance. The border dummy might thus control for the fact that center to center distances might overestimate the effective distance between neighboring countries. In fact, neighboring countries generally trade relatively more. According to the results in Table 4, the presence of a common border does not seem to have an independent effect on the volume of bilateral bank lending.

Finally, the dummy variable indicating whether a country pair have the same legal origin exerts a positive and significant effect. The same legal origin proxies a similar legal system or a comparable business conduct. Insofar as a common legal system or the similarity of the business conduct is able to reduce the time and resources devoted to writing a loan contract, transaction costs between these countries are smaller; the reduction in transaction cost in turn increases the cross-border lending between these country pairs.

In Table 5 we add to the baseline model variables that are specific to the borrowing country and are related to the costs a bank incurs when lending to a foreign firm. Specifically, we add to our baseline specification six composite measures of governance which we retrieve from the *Worldwide Governance Indicator* (WGI) collected by the World Bank: *Voice and Accountability*, *Political Stability and Absence of Violence*, *Government Effectiveness*, *Regulatory Quality*, *Rule of Law*, and *Control of Corruption*.[9]

We find that most of the indicators measuring governance quality are insignificant. Only two measures have a significantly positive effect on bilateral lending volume. First, better regulatory quality increases international bank lending. According to the WGI data documentation, this measure captures the perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development (see also Kaufmann et al. 2009). To the extend that policies and regulations promoting private sector development also include issues like contract enforcement, promoting disclosure rules etc., it is not surprising that this measure has a positive effect on international bank lending. Cross-border banking, and banking activities in general, are promoted by increasing the effectiveness of such regulatory rules since these rules increase the adaptability of political institutions and courts in an environment of incomplete contracts. Our finding of a positive effect of better regulatory quality on bilateral bank lending is also in line with the results presented in Aggarwal & Goodell (2009): countries with better regulatory quality are in general less market based but rely

[9]We add each of the variables one at the time but do not control for each of these characteristics simultaneously. The reason is that the variables from the WGI database are highly correlated with each other (see Table 6). Introducing two or more of these variables at the same time would run the risk of introducing high multicollinearity into the model.

more on bank financing. Second, the variable Voice and Accountability (*voice*) also affects cross-border bank lending positively. The variable *voice* captures the perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media (see Kaufmann et al. 2009). Variables like control of corruption, rule of law or political stability do not affect the amount of foreign bank lending in a country.

Concerning the effects of our variables of interest, we observe that these robustness checks do not change the previous results. The two mass variables still exert a positive and significant effect. Also, the coefficients on the distance variable and the lending rate remain negative and significant. Furthermore, the point estimates are statistically indistinguishable from those obtained in the benchmark regression.

5 Conclusion

This paper gives a theoretical foundation for the gravity equation for cross-border bank lending. Starting from the relationship between the borrowing firm and the lending bank, we derive the volume of bank credit in country i from country j, based on a problem of cost minimization. The cost function not only consists of the direct costs represented by the interest rate but is also on cost factors rooted in the specificity of the respective firms, banks and financial systems. The theoretical framework delivers a formulation for the amount of cross-border bank lending which resembles a gravity equation. The gravity equation of bank loans features multilateral (cost) resistance terms and unobserved lending country characteristics which need to be accounted for when applying a gravity framework to bank lending data.

Using the theoretical gravity formulation, we estimate a gravity equation for international bank lending. The results of the empirical implementation of our structural gravity equation provide strong evidence in favor of the predictions derived from the theoretical model. Consistent with the theory, we find that aggregate bank loans in the receiving country and the effective number of banks in the host country are relevant constructs to proxy economic masses. We also find strong evidence that the cost for foreign firms to obtain cross-border loans is an important determinant. Furthermore, our results suggest that it is important to control for the unobserved multilateral resistance terms. Ignoring the resistance variables leads to biased results and erroneous conclusion about the effects of some covariates.

Our results are robust to a number of different model specification. In a series of robustness tests, we find that host country specific banking market variable are only marginally relevant in explaining international bank lending if we control for unobserved host and receiving market characteristics in a way as suggested by our theoretical model. Similar to the results from the trade literature, we find that international bank lending is also affected by distance-related variables such as the same language and colonial relationships. In addition, better regulatory quality in the debtor country increases bilateral lending activity. These findings give support to the importance of non-interest rate costs in the theory, like the cost of setting up and enforcing a loan contract. Searching and monitoring costs are also important characteristics for the geographical distribution of cross-border bank lending. Insofar, our theory gives a coherent framework in which one can think about the determinants of international bank lending.

References

Aggarwal, R. & Goodell, J. W. (2009), 'Markets and institutions in financial intermediation: National characteristics as determinants', *Journal of Banking & Finance* **33**(10).

Anderson, J. E. & van Wincoop, E. (2003), 'Gravity with gravitas: A solution to the border puzzle', *American Economic Review* **93**(1).

Anderson, S., de Palma, A. & Thisse, J.-F. (1992), *Discrete Choice Theory of Product Differentiation*, MIT Press.

Baier, S. L. & Bergstrand, J. H. (2007), 'Do free trade agreements actually increase members' international trade?', *Journal of International Economics* **71**(1).

Baldwin, R. & Taglioni, D. (2006), Gravity for dummies and dummies for gravity equations, Cepr discussion papers, C.E.P.R. Discussion Papers.

Baltagi, B. (2001), *Econometric Analysis of Panel Data, 2^{nd} edition*, John Wiley & Sons.

Beck, T., Demirguc-Kunt, A. & Levine, R. (2009), Financial institutions and markets across countries and over time - data and analysis, Policy Research Working Paper Series 4943, The World Bank.

Buch, C. M. (2002), 'Are banks different? evidence from international data', *International Finance* **5**(1).

Buch, C. M. (2005), 'Distance and international banking', *Review of International Economics* **13**(4).

Coeurdacier, N. & Martin, P. (2009), 'The geography of asset trade and the euro: Insiders and outsiders', *Journal of the Japanese and International Economies* **23**(2).

Egger, P. & Pfaffermayr, M. (2003), 'The proper panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects', *Empirical Economics* **28**(3).

Head, K. & Ries, J. (2008), 'Fdi as an outcome of the market for corporate control: Theory and evidence', *Journal of International Economics* **74**(1).

Kaufmann, D., Kraay, A. & Mastruzzi, M. (2009), Governance matters viii : aggregate and individual governance indicators 1996-2008, Policy Research Working Paper Series 4978, The World Bank.

Manning, W. G. & Mullahy, J. (2001), 'Estimating log models: to transform or not to transform?', *Journal of Health Economics* **20**(4).

Martin, P. & Rey, H. (2004), 'Financial super-markets: size matters for asset trade', *Journal of International Economics* **64**(2).

Okawa, Y. & van Wincoop, E. (2010), Gravity in international finance, Working Papers 072010, Hong Kong Institute for Monetary Research.

Papaioannou, E. (2009), 'What drives international financial flows? politics, institutions and other determinants', *Journal of Development Economics* **88**(2).

Park, R. E. (1966), 'Estimation with heteroskedastic error terms', *Econometrica* **34**(888).

Porta, R. L., de Silanes, F. L., Shleifer, A. & Vishny, R. W. (1998), 'Law and finance', *Journal of Political Economy* **106**(6).

Portes, R. & Rey, H. (2005), 'The determinants of cross-border equity flows', *Journal of International Economics* **65**(2).

Silva, J. M. C. S. & Tenreyro, S. (2006), 'The log of gravity', *The Review of Economics and Statistics* **88**(4).

Westerlund, J. & Wilhelmsson, F. (2011), 'Estimating the gravity model without gravity using panel data', *Applied Economics* **43**.

6 Tables

Table 1: Panel gravity equation for cross-border bank lending

The dependent variable BA_{ij} are assets of reporting country j in country i. BL_i = total bank loan in receiving country i. $dist_{ij}$ = distance between reporting country j and receiving country i. n_j = inverse of the 3-bank concentration ratio in country j. $r_j = j$ average implicit bank lending rate in country j. The RESET test for detecting neglected nonlinearities. Rejection of the Null hypothesis indicates that the model is misspecified. The Park test is a test of the hypothesis that the model is consistently estimated by OLS. Rejection of the Null indicates that estimating the model in log-linear form is inappropriate. The GNR test checks the assumption of the Poisson estimator that the conditional variance of the dependent variable is proportional to the conditional mean. Rejection of the Null hypothesis indicates that the assumption is violated. Robust standard errors in brackets. ***, **, * indicate significant at the 1, 5, 10 % level.

	PPML	OLS	OLS (1+BA_{ij})
BL_i	0.558***	0.596***	0.624***
	[0.050]	[0.035]	[0.035]
$dist_{ij}$	-0.368***	-0.852***	-0.881***
	[0.030]	[0.035]	[0.037]
n_j	0.369***	0.724***	0.652***
	[0.067]	[0.070]	[0.090]
r_j	-0.145**	-0.041	-0.085**
	[0.069]	[0.039]	[0.039]
N	5209	4895	5209
R^2	0.819	0.728	0.731
RESET Test (p-value)	0.701	0.024	0.010
Park-Test (p-value)	0.000	-	-
GNR (p-value)	0.113	-	-

Table 2: Panel gravity equation for cross-border bank lending

The dependent variable are assets of reporting country j in country i. BL_i = total bank loan in receiving country i. $dist_{ij}$ = distance between reporting country j and receiving country i. n_j = inverse of the 3-bank concentration ratio in country j. $r_j = j$ average implicit bank lending rate in country j. The RESET test for detecting neglected nonlinearities. Rejection of the Null hypothesis indicates that the model is misspecified. Robust standard errors in brackets. ***, **, * indicate significant at the 1, 5, 10 % level.

	Benchmark	both country fixed effects + year controls ignoring time variation	sending country fixed effects + year controls ignoring time variation	receiving country fixed effects + year controls ignoring time variation	year controls
BL_i	0.558***	0.597***	0.694***	0.619**	0.644***
	[0.050]	[0.162]	[0.019]	[0.250]	[0.024]
$dist_{ij}$	-0.368***	-0.328***	-0.446***	-0.427***	-0.505***
	[0.030]	[0.030]	[0.024]	[0.048]	[0.034]
n_j	0.369***	0.043	0.048	0.444***	0.409***
	[0.067]	[0.093]	[0.108]	[0.035]	[0.038]
r_j	-0.145**	-0.071	-0.08	-0.118***	-0.118***
	[0.069]	[0.044]	[0.063]	[0.030]	[0.036]
N	5209	5209	5209	5209	5209
R^2	0.819	0.668	0.375	0.287	0.268
RESET Test (p-value)	0.701	0.133	0.006	0.147	0.096

Table 3: Gravity equation for cross-border bank lending and banking characteristics

The dependent variable are assets of reporting country j in country i. BL_i = total bank loan in receiving country i. $dist_{ij}$ = distance between reporting country j and receiving country i. n_j = inverse of the 3-bank concentration ratio in country j. r_j = average implicit bank lending rate in country j. $margin_j$ = the net interest rate margin in country j. roe_j = average bank return on equity in country j. $cost-income_j$ = bank cost to income ratio in country j . The RESET test for detecting neglected nonlinearities. Rejection of the Null hypothesis indicates that the model is misspecified. Robust standard errors in brackets. ***, **, * indicate significant at the 1, 5, 10 % level.

	Benchmark	both country fixed effects + year controls ignoring time variation	sending country fixed effects + year controls ignoring time variation	receiving country fixed effects + year controls ignoring time variation	year controls
	Poisson Pseudo Maximum Likelihood				
BL_i	0.566***	0.607***	0.694***	0.646***	0.653***
	[0.050]	[0.162]	[0.019]	[0.243]	[0.024]
$dist_{ij}$	-0.365***	-0.328***	-0.446***	-0.431***	-0.506***
	[0.030]	[0.030]	[0.024]	[0.049]	[0.034]
n_j	0.445***	-0.005	0.001	0.533***	0.484***
	[0.072]	[0.099]	[0.117]	[0.040]	[0.041]
r_j	-0.105*	-0.066	-0.074	-0.019	-0.026
	[0.054]	[0.044]	[0.063]	[0.023]	[0.027]
$netintmargin$	-0.227***	-0.082	-0.081	-0.257***	-0.220***
	[0.080]	[0.060]	[0.074]	[0.035]	[0.038]
roa	-0.086	0.019	0.011	-0.402***	-0.421***
	[0.139]	[0.048]	[0.063]	[0.052]	[0.052]
$cost-income$	-0.025***	-0.002	-0.002	-0.012***	-0.014***
	[0.009]	[0.003]	[0.004]	[0.003]	[0.003]
N	5209	5209	5209	5209	5209
R^2	0.821	0.834	0.669	0.411	0.318
RESET Test (p-value)	0.907	0.076	0.004	0.271	0.148

Table 4: Panel gravity equation for cross-border bank lending and additional control variables

The dependent variable are assets of reporting country j in country i. BL_i = total bank loan in receiving country i. $dist_{ij}$ = distance between reporting country j and receiving country i. n_j = inverse of the 3-bank concentration ratio in country j. r_j = average implicit bank lending rate in country j. $contig_{ij}$ = dummy variable which equals one if country i and j share a common border and zero otherwise. $colonoy_{ij}$ = dummy variable which equals one if country i and j have colonial relationship. $comlang_{ij}$ = dummy variable which equals one if country i and j have common language. Robust standard errors in brackets. ***, **, * indicate significant at the 1, 5, 10 % level.

	Benchmark	both country fixed effects + year controls ignoring time variation	sending country fixed effects + year controls ignoring time variation	receiving country fixed effects + year controls ignoring time variation	year controls
			Poisson Pseudo Maximum Likelihood		
BL_i	0.565***	0.593***	0.695***	0.623**	0.648***
	[0.051]	[0.155]	[0.019]	[0.248]	[0.025]
$dist_{ij}$	-0.412***	-0.390***	-0.516***	-0.315***	-0.504***
	[0.039]	[0.039]	[0.035]	[0.057]	[0.045]
n_j	0.382***	0.033	0.044	0.415***	0.405***
	[0.066]	[0.085]	[0.104]	[0.035]	[0.036]
r_j	-0.156**	-0.064	-0.074	-0.112***	-0.119***
	[0.067]	[0.040]	[0.059]	[0.030]	[0.036]
contig	-0.113	-0.113	-0.429***	0.516***	-0.027
	[0.106]	[0.103]	[0.107]	[0.134]	[0.152]
comlang	-0.058	-0.073	0.115	-0.14	-0.021
	[0.098]	[0.107]	[0.135]	[0.123]	[0.138]
comlegor	0.505***	0.554***	0.260***	0.143	0.131
	[0.056]	[0.055]	[0.083]	[0.104]	[0.090]
N	5209	5209	5209	5209	5209
R^2	0.84	0.853	0.663	0.379	0.287
RESET Test (p-value)	0.529	0.505	0.116	0.107	0.160

Table 5: Panel gravity equation for cross-border bank lending and additional control variables

The dependent variable are assets of reporting country j in country i. BL_i = total bank loan in receiving country i. $dist_{ij}$ = distance between reporting country j and receiving country i. n_j = inverse of the 3-bank concentration ratio in country j. r_j = average implicit bank lending rate in country j. $contig_{ij}$ = dummy variable which equals one if country i and j share a common border and zero otherwise. $colonoy_{ij}$ = dummy variable which equals one if country i and j have colonial relationship. $comlang_{ij}$ = dummy variable which equals one if country i and j have common language. Robust standard errors in brackets. ***, **, * indicate significant at the 1, 5, 10 % level.

	Poisson Pseudo Maximum Likelihood					
BL_i	0.529***	0.551***	0.515***	0.540***	0.527***	0.541***
	[0.053]	[0.058]	[0.053]	[0.054]	[0.055]	[0.055]
$dist_{ij}$	-0.357***	-0.366***	-0.353***	-0.363***	-0.357***	-0.363***
	[0.030]	[0.030]	[0.030]	[0.030]	[0.031]	[0.030]
n_j	0.370***	0.370***	0.371***	0.371***	0.371***	0.370***
	[0.066]	[0.067]	[0.063]	[0.067]	[0.065]	[0.066]
r_j	-0.146**	-0.145**	-0.144**	-0.146**	-0.144**	-0.145**
	[0.068]	[0.069]	[0.066]	[0.069]	[0.067]	[0.068]
voice	0.263**					
	[0.118]					
ruleoflaw		0.034				
		[0.134]				
regul			0.380**			
			[0.148]			
polstab				0.143		
				[0.121]		
gov					0.189	
					[0.133]	
corrupt						0.093
						[0.111]
N	5187	5187	5187	5187	5187	5187
R^2	0.821	0.82	0.822	0.82	0.821	0.82

A Appendix

The probability of a firm in country i choosing bank k in country j was given by the following integral:

$$\mathbf{P}_{ijk} = \int_{-\infty}^{\infty} \frac{1}{\sigma} \exp\left(\frac{x}{\sigma} - \gamma\right) \left\{ \exp\left[-exp\left(\frac{x}{\sigma} - \gamma\right)\right]\right\} \prod_{l \neq j}^{N} \exp\left[- \exp\left(\frac{\bar{c}_{ij} - \bar{c}_{il} + x}{\sigma} - \gamma\right)\right]^{n_l} dx$$

Simplifying the equation by setting $y_{ij} = \exp(\bar{c}_{ij}/\sigma)$ and $\delta = exp(x/\sigma - \gamma)$ and adjusting the integral's range, we arrive at the following, simplified integral:

$$\mathbf{P}_{ijk} = \int_{0}^{\infty} \exp\left[-\delta \left(\sum_{l=1}^{N} n_l \frac{(y_l)^{-1}}{(y_j)^{-1}} \right)\right] d\delta..$$

Integrating yields

$$\begin{aligned}
\mathbf{P}_{jk} &= - \frac{(y_j)^{-1}}{\sum_{l=1}^{N} n_l (y_l)^{-1}} \left\{ \exp\left[-\delta \left(\sum_{l=1}^{N} n_l \frac{(y_l)^{-1}}{(y_j)^{-1}} \right)\right]\right\}_{0}^{\infty} \\
&= \frac{(y_j)^{-1}}{\sum_{l=1}^{N} n_l (y_l)^{-1}} \\
&= \frac{\exp\left(-\frac{\bar{c}_{ij}}{\sigma}\right)}{\sum_{l=1}^{N} n_l \exp\left(-\frac{\bar{c}_{il}}{\sigma}\right)}.
\end{aligned}$$

Table 6: Correlation Matrix

	BL_i	$dist_{ij}$	n_j	r_j	$margin$	roa	$z-score$	contig	comlang	colony	voice	ruleoflaw	regul	polstab	gov	corrupt
BL_i	1.000															
$dist_{ij}$	-0.109	1.000														
n_j	0.069	0.228	1.000													
r_j	-0.011	-0.126	-0.115	1.000												
$margin$	0.105	0.009	0.218	0.257	1.000											
roa	0.005	-0.079	-0.057	0.136	-0.113	1.000										
$z-score$	-0.040	-0.009	-0.009	-0.010	0.194	-0.757	1.000									
contig	0.185	-0.428	-0.026	0.053	0.040	0.005	0.020	1.000								
comlang	-0.034	-0.106	0.068	0.122	0.137	0.001	0.074	0.334	1.000							
colony	-0.070	-0.012	-0.004	0.208	0.039	-0.024	0.040	0.127	0.343	1.000						
voice	0.410	-0.256	0.026	0.015	0.073	0.012	-0.035	0.175	0.017	-0.129	1.000					
ruleoflaw	0.522	-0.281	0.007	0.001	0.044	0.009	-0.034	0.169	-0.012	-0.138	0.804	1.000				
regul	0.523	-0.252	0.028	-0.003	0.054	0.010	-0.037	0.154	-0.022	-0.180	0.818	0.914	1.000			
polstab	0.242	-0.260	-0.012	0.022	0.025	0.004	-0.015	0.134	-0.018	-0.119	0.743	0.840	0.774	1.000		
gov	0.586	-0.283	0.021	-0.004	0.059	0.007	-0.035	0.186	0.000	-0.153	0.815	0.957	0.937	0.798	1.000	
corrupt	0.530	-0.243	0.011	0.005	0.052	0.005	-0.031	0.170	0.012	-0.121	0.777	0.959	0.903	0.791	0.959	1.000