

Rudolph, Stephan

**Working Paper**

## Estimating gravity equations with endogeneous trade costs

Dresden Discussion Paper Series in Economics, No. 01/10

**Provided in Cooperation with:**

Technische Universität Dresden, Faculty of Business and Economics

*Suggested Citation:* Rudolph, Stephan (2010) : Estimating gravity equations with endogeneous trade costs, Dresden Discussion Paper Series in Economics, No. 01/10, Technische Universität Dresden, Fakultät Wirtschaftswissenschaften, Dresden

This Version is available at:

<https://hdl.handle.net/10419/39800>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

*Dresden Discussion Paper Series  
in Economics*



---

**Estimating Gravity Equations with Endogeneous  
Trade Costs**

STEPHAN RUDOLPH

*Dresden Discussion Paper in Economics No. 01/10*

ISSN 0945-4829

Address of the author(s):

Stephan Rudolph  
Technische Universität Dresden  
Helmholtzstr. 10  
01069 Dresden  
Germany

e-mail : [stephan.rudolph@tu-dresden.de](mailto:stephan.rudolph@tu-dresden.de)

Editors:

Faculty of Business and Economics, Department of Economics

Internet:

An electronic version of the paper may be downloaded from the homepage:

<http://rcswww.urz.tu-dresden.de/wpeconomics/index.htm>

English papers are also available from the SSRN website:

<http://www.ssrn.com>

Working paper coordinator:

Andreas Buehn

e-mail: [wpeconomics@mailbox.tu-dresden.de](mailto:wpeconomics@mailbox.tu-dresden.de)

# Estimating Gravity Equations with Endogeneous Trade Costs

*Stephan Rudolph*  
*Technische Universität Dresden*  
*01069 Dresden*  
[stephan.rudolph@tu-dresden.de](mailto:stephan.rudolph@tu-dresden.de)

## Abstract:

A basic assumption of the gravity equation of international trade is that increasing trade costs lower exports. But intuition and theory imply that a high export volume lowers bilateral trade costs as well, because a fixed cost intensive trade sector probably bears lower average costs with more trade. In this case, standard gravity estimation might be biased due to simultaneity. This paper finds an empirical interdependency between exports and trade costs. Using a simultaneous equation model to face this problem improves the estimates compared to the standard gravity specification.

JEL-Classification: F13, F17, C33, C5

Keywords: Gravity Equation, Trade Policy, Simultaneity Problem

# 1 Introduction

Gravity Equations are probably the most favorite tool among economists to explore the determinants of international trade. The idea is to regress bilateral trade flows on country sizes and distance. Country size is usually measured by the respective countries' GDPs, geographic distance and a set of policy variables. These policy variables can control for membership in special country groups or blocks, currency unions, custom unions and so on. The outcomes are often used to consult policymakers.

During the last decade, critique was raised that the magnitudes of the estimated coefficients from gravity equations are overestimated or implausibly high. One reason for this critique was a study by McCallum (1995), who concluded that trade between two Canadian provinces is higher by 22% compared to a Canadian province and a U.S. state of the same size and distance: Removing the border would have an astronomically high impact on trade between Canada and the U.S.

This and other results have motivated researchers to find theoretically and econometrically proper solutions to improve the results of estimating gravity equations. This is also the motivation of this paper. The theory behind this paper is that trade costs do not only determine trade flows, but trade flows also determine trade costs if there are economies of scale (falling average costs) in the trade sector, since the trade sector is likely to be fixed cost intensive. This can be shown by combining the gravity derivation from a general equilibrium model (Anderson and van Wincoop, 2003) with a simple model of a bilateral trade cost sector that minimizes trade costs with a given

trade volume to derive an endogenous trade cost tariff equivalent (Rudolph, 2009). The consequence is a mutual causality between trade flows and trade costs which might bias gravity results and thus has to be checked. The econometrics of this paper confront the theoretical considerations by introducing a simultaneous equation model.

Endogeneity in gravity equations has already been faced by some studies. Literature on trade and growth argues that countries' GDPs depend on exports. For instance, Frankel and Romer (1999) run a gravity function and use the estimated values to compute "constructed trade shares" for each country. These constructed trade shares are used as instruments in a further step to regress per capita income on trade shares, population and area. However, they cannot find evidence that controlling for endogeneity improves the results gorgeously.<sup>1</sup> One reason might be that GDP is a function of net rather than gross exports and net exports (exports minus imports) are normally quite small compared to a country's GDP. Apart from that there should not be much simultaneity between GDP and *bilateral* exports since GDP rather depends on multilateral (net) exports. These considerations possibly help to explain, why taking potential endogeneity of GDPs into account has not prevailed so far in the gravity literature. Another literature branch, represented by Baier and Bergstrand (2007) amongst others, concentrates on the question if there is a reverse causality between bilateral trade flows and free trade agreements (FTA). This literature argues that signing up a FTA is in many cases motivated by the fact that the signing countries have considerable trade flows between each other, a priori. Studies addressing this

---

<sup>1</sup> "As a result, the hypothesis that the IV and OLS estimates are equal cannot be rejected" (Frankel and Romer, 1999, p. 388).

problem basically find evidence for simultaneity between FTAs and exports. But this literature does not focus on the endogeneity of overall trade costs. In my study, I use a recently developed index for comprehensive trade costs to analyze the simultaneity between exports and trade costs as a whole.

The paper is structured as follows. Section 2 overviews a theoretical foundation of the gravity equation with economies of scale in trade costs. Building up on these theoretical suggestions, Section 3 introduces the econometric estimation strategy. Section 4 describes the data used in the estimation, section 5 presents the results of the simultaneous equation estimation and gives a comparison to the standard estimation strategies. Section 6 concludes.

## **2 Theory**

The theoretical literature about the gravity equation takes trade costs between two countries as exogenously given: trade costs affect the volume of bilateral trade. In all these models, trade costs enter in terms of iceberg costs which can be interpreted as trade costs per unit trade volume. If trade is interpreted as a fixed cost intensive service between two countries, these per unit trade costs should fall in the bilateral trade volume. Building up on this idea, Rudolph (2009) augmented the theoretical derivation of the gravity equation by Anderson and van Wincoop (2003) with bilateral trade sectors and shows that the existence of economies of scale in these trade sectors leads to falling average trade costs. This model is introduced in the remainder of this section.

## 2.1 Trade costs with economies of scale

In theoretical foundations of the gravity equation, trade costs (or trade impediments, respectively) are usually modeled as iceberg trade costs. Iceberg-costs are a constant exogenously given markup (“iceberg-factor”)  $t_{ij}$  onto the factory price  $p_i$ , so that the price of the (composite) commodity of country  $i$  paid in country  $j$  is  $p_{ij} = t_{ij} \cdot p_i$ . The price of the commodity from  $i$  is in country  $j$  higher by the factor  $t_{ij}$  than in the country of origin  $i$  due to trade costs. It is assumed that  $t_{ij} > 1$  for all countries  $j \neq i$  and that the domestic trade cost factor  $t_{ii} = 1$ . This is to ensure that commodities are more expensive abroad than on the domestic market. Modeling trade costs in this way lead to three properties. First, since the trade volume including transport costs (gross trade volume) is  $X_{ij} = t_{ij} \cdot p_i \cdot c_{ij}$  with quantity  $c_{ij}$  sent from  $i$  to  $j$ , the trade volume can be decomposed into total trade costs  $(t_{ij} - 1) \cdot p_i \cdot c_{ij}$  plus transport cost exclusive (net) trade volume  $p_i \cdot c_{ij}$ .<sup>2</sup> Second, it can be shown that a fraction  $(t_{ij} - 1)/t_{ij}$  of the amount of goods shipped from  $i$  to  $j$  is lost in transport.<sup>3</sup> Finally, iceberg-costs are a measure for average trade costs and not for total trade costs, because, following its definition, the iceberg-factor can be denoted as gross trade volume divided by net trade volume:

$$t_{ij} = \frac{p_{ij} \cdot c_{ij}}{p_i \cdot c_{ij}}. \quad (1)$$

---

<sup>2</sup> To bring this mathematically into one line:  $X_{ij} = p_{ij} \cdot c_{ij} = t_{ij} \cdot p_i \cdot c_{ij} = (t_{ij} - 1) \cdot p_i \cdot c_{ij} + p_i \cdot c_{ij}$ . The last expression shows that  $X_{ij}$  equals total trade costs (first summand) plus the net trade volume (second summand).

<sup>3</sup> Assume for simplicity that  $p_i = 1$  and  $c_{ij} = 1$  and e.g.  $t_{ij} = 1.25$ . This means, country  $i$  must send 1.25 units to  $j$  so that one unit arrives. In this case a fraction  $0.25/1.25 = 0.2$  of the trade volume sent by country  $i$  would be lost.



This implies that  $t_{ij}$  is nothing else than the tariff-equivalent factor for bringing a value of \$ 1.00 of country  $i$ 's composite export good to country  $j$ . Therefore, iceberg-cost-factor  $t_{ij}$  is nothing else than a *per-unit-cost* of trade.

Rudolph (2009) models a comprehensive profit-maximizing trade sector between each pair of countries  $i, j$  and derives an average cost function which is equal to the iceberg-factor  $t_{ij}$ . The output of this bilateral trade sector is the net trade volume  $X_{ij}^0$  between the two countries. The result of profit maximization in this sector yields an average trade cost function,

$$t_{ij} = t_{ij}(w_{ij}^k, X_{ij}^0). \quad (2)$$

These average trade costs  $t_{ij}$  describe the costs of bringing a value of \$ 1.00 from country  $i$ 's composite trade volume to country  $j$  as a function of net trade volume  $X_{ij}^0$  and a bilateral input cost vector  $w_{ij}^k = (w_{ij}^1, \dots, w_{ij}^K)$ . The index  $k \in (1, \dots, K)$  denotes respective input factors like shipping a good via ocean or air, paying for tariffs, translating contracts etc. As long as there are economies of scale in the trading sector, e.g. caused by the presence of fixed costs of infrastructure, average cost function (2) will decline with the bilateral trade volume: The more two countries trade with each other, the lower are the average bilateral trade costs. The result is the presumption that  $\partial t_{ij} / \partial X_{ij}^0 < 0$ . If there were no scale effects of trade volume on average trade costs,  $\partial t_{ij} / \partial X_{ij}^0 = 0$  would be the case.

In a critique of modeling trade costs as iceberg costs, Grossman (1998) suggests a logarithmic form to concretize the trade cost function. Following this suggestion, a logarithmic form is applied to the average trade cost function (2) and, according to the outlined hypotheses, augmented by the net trade

volume  $X_{ij}^0$ :

$$t_{ij}(w_{ij}^k, X_{ij}^0) = e^{\beta_0} (w_{ij}^k)^{\beta_k} (X_{ij}^0)^{\beta_X} . \quad (3)$$

If there are economies of scale in the trade sector, the elasticity of the average trade costs with respect to the net trade volume,  $\beta_X$ , is expected to be lower than 0. The empirical question on these scale effects will be checked in sections 3 to 5 of this paper.

## 2.2 Deriving the Gravity Equation

In a general equilibrium framework with many countries trading composite goods that are differentiated by country of origin, Anderson and van Wincoop (2003) derive the following gravity equation:

$$X_{ij} = \frac{Y_i \cdot Y_j}{Y_w} \cdot \left( \frac{t_{ij}}{\Pi_i \cdot P_j} \right)^{1-\sigma} . \quad (4)$$

Here,  $Y_i$  and  $Y_j$  are the exogenously given GDP's of the countries,  $Y_w$  the GDP of the whole world, and  $\sigma$  is the elasticity of substitution. It is assumed that  $\sigma > 1$ , which is supported by empirical evidence (see Anderson and van Wincoop, 2004).  $\Pi_i$  and  $P_j$  denote the exogenously given multilateral resistances of the exporting or the importing country, respectively. This is an index for the overall accessibility to trade of a certain country. In the second multiplier of this equation, bilateral average trade costs  $t_{ij}$  appear in relation to the respective countries multilateral resistances. For illustration, imagine two countries lying isolated from the rest of the world on one island in the ocean, far away from the next continent. Bilateral average trade costs measured by iceberg-factor  $t_{ij}$  might be low and this should guarantee for a higher trade volume between both countries. But the relatively high trade

costs to the complete rest of the world have an additional positive effect on the bilateral trade volume. If the same two island-countries were two small countries in the middle of a huge continent with many huge countries, the multilateral resistances were probably much lower and thus, trade volume between the two countries were lower, even if for the GDP's and  $t_{ij}$  the same levels are chosen.

Note that trade volume in equation 4 is measured as a gross trade volume. Dividing both sides by  $t_{ij}$  yields the gravity equation in terms of the net trade volume:

$$X_{ij}^0 = \frac{Y_i \cdot Y_j}{Y_w} \cdot t_{ij}^{-\sigma} \cdot (\Pi_i \cdot P_j)^{\sigma-1}. \quad (5)$$

Gravity equation (5) can be treated as interdependent with average trade cost function (3), since net trade volume can be seen as a function of average trade costs and average trade costs can be seen as a function of net trade volume.

## 2.3 The Interaction between Trade Volume and Trade Costs

Most researchers applying the gravity equation are interested in the changes of policy variables: what happens if a country signs a trade agreement, enters a monetary union or anything similar? These effects are captured by  $w_{ij}^k$  in the trade cost function (3). How does the interaction between trade volume and trade costs work?

Logarithmizing equations (3) and (5) yield:

$$\ln X_{ij}^0 = K + FE_{ij} + \ln Y_i + \ln Y_j - \sigma \ln t_{ij}, \quad (6)$$

$$\ln t_{ij} = \beta_0 + \beta_k \ln w_{ij}^k + \beta_X \ln X_{ij}^0, \quad (7)$$

with a constant  $K = \left[ \ln \frac{1}{Y_w} \right]$  and the fixed effects  $FE_{ij} = (\sigma - 1) \ln (\Pi_i \cdot P_j)$ . Now consider a change in a component  $k$  of the input cost vector  $w_{ij}^k$ , ceteris paribus. This could be a change of a certain tariff rate, an energy price shock or anything else. Such a change will have a direct impact on the average trade costs measured as a tariff equivalent. The size of this change is quantified by the elasticity  $\beta_k$ . In a next step, the changed average trade costs will affect the bilateral net trade volume quantified by its elasticity  $\sigma$ . Inserting equation (7) into (6) and solving for  $X_{ij}^0$  yields

$$\ln X_{ij}^0 = \frac{1}{1 + \sigma \beta_X} (K' + FE_{ij} + \ln Y_i + \ln Y_j - \sigma \beta_k \ln w_{ij}^k), \quad (8)$$

with  $K' = K + \sigma \beta_0$ . The expression in parentheses is the gravity equation as it is usually estimated. The fraction  $\frac{1}{1 + \sigma \beta_X}$  is a bias term which is likely to enlarge the estimates of the gravity equation (see Rudolph, 2009).

### 3 Econometric Model

The traditional strategy to estimate a gravity equation is: take trade as the endogenous variable and regress it on country sizes and a set of trade cost proxies. But our theoretical considerations suggest (i) that these trade cost proxies affect trade costs rather than exports and (ii) that trade costs could be affected by the exports, reversely, due to economies of scale in international trade. If this reverse causality between exports and trade costs consists,

estimating a simultaneous equation model (SEM) should be the appropriate strategy.<sup>4</sup>

Consider equations (7) and (6) and augment them by the residual terms  $u_{ij}$  and  $v_{ij}$  to get the structural equations:<sup>5</sup>

$$\ln X_{ij}^0 = \alpha_0 + \alpha_{Y_i} \ln Y_i + \alpha_{Y_j} \ln Y_j + \alpha_t \ln t_{ij} + u_{ij}, \quad (9)$$

$$\ln t_{ij} = \beta_0 + \beta_k \ln w_{ij}^k + \beta_X \ln X_{ij}^0 + v_{ij}, \quad (10)$$

Since equation (9) depends on  $t_{ij}$  and equation (10) depends on  $X_{ij}^0$ , the gravity equation becomes a system of interdependent or simultaneous equations. The adequate estimator is the two-stages-least-square (2SLS) or three-stages-least-square (3SLS) estimator. 2SLS means that in a first step, all endogenous variables of the equation system ( $t_{ij}$  and  $X_{ij}^0$ ) are regressed on all exogenous variables of the equation system ( $Y_i$ ,  $Y_j$  and  $w_{ij}^k$ ):

$$\ln X_{ij}^0 = \pi_1 + \pi_2 \ln Y_i + \pi_3 \ln Y_j - \pi_4 w_{ij}^k + \eta_{ij}, \quad (11)$$

$$\ln t_{ij} = \pi_5 + \pi_6 \ln Y_i + \pi_7 \ln Y_j - \pi_8 w_{ij}^k + \epsilon_{ij}, \quad (12)$$

with parameters  $\pi_{(\cdot)}$  and residual terms  $\eta_{ij}$  and  $\epsilon_{ij}$ . These two equations are called the reduced form equations. Note that the reduced form equation (11) is identical with the traditional specification of the gravity equation. In a second step, the estimated values for the endogenous variables ( $\hat{t}_{ij}$  and  $\hat{X}_{ij}^0$ ) are used as instruments to estimate the initial structural equations (9) and (10).

This procedure is necessary, because theory implies that both structural

<sup>4</sup> For an conceptional overview see e.g. Greene (2000, chapter 16).

<sup>5</sup> The fixed effects are omitted here for simplicity.

equations (9) and (10) contain endogenous variables. From an econometric point of view, endogeneity of variables means that these variables are correlated with the error terms  $u_{ij}$  and  $v_{ij}$ . The consequence are inconsistent estimates of the parameters. Since there is a correlation between the error terms  $u_{ij}$  and  $v_{ij}$ , using a feasible generalized least squares (FGLS) estimator, where the estimators are weighted by the variance-covariance-matrix, helps to improve the results. This procedure is known as 3SLS.

Since a panel data set will be used, certain techniques must be used to achieve consistent results.<sup>6</sup> As a baseline case, a pooled regression model is estimated where the panel structure of the data is not considered. Anderson and van Wincoop (2003) postulate to control for the countries' multilateral resistances. To do so, I estimate a least square dummy variable (LSDV) model with dummy variables controlling for exporting and importing countries (unilateral or country fixed effects) as well as for the respective year. This strategy was first suggested by Mátyás (1997). Most recent studies use the two-way fixed-effects model (or country-pair fixed effects model), which controls for country-pairs (bilateral fixed effects) and time. This specification appears in most studies as the most appropriate one.

There are two central questions from an analytical point of view:

**Question 1** *Are there economies of scale in the trade sector?*

This should be the case if in a SEM specification the estimated value of  $\beta_X$  in equation (10) is significantly lower than 0.

---

<sup>6</sup> See e.g. Cheng and Wall (1999) or Baltagi, Egger, and Pfaffermayr (2003) for a useful overview of panel data estimation strategies for gravity equations.

**Question 2** *How are estimates biased if economies of scale in the trade sector consist, but there is not controlled for?*

To analyze this question, I estimate a restricted version of the equation system (9) and (10) using an instrumental variable (IV) regression where  $\alpha_t = 0$  for both the gravity and the trade cost equation. If the theoretical suggestions on the bias term are right, the estimates for the parameters of the exogenous trade variables  $w_{ij}^k$  should be systematically higher than in the SEM specification.

## 4 Data

The data set covers all 30 OECD countries over the years 1995 to 2006. Many studies use broader data bases like the IMF's directions of trade statistics, where more than one hundred countries over more than fifty years are covered. In these data sets, zero trade flows occur, meaning that very small and remote countries might have no trade relations with each other (see Helpman, Melitz, and Rubinstein, 2008). The persistence of zero trade flows postulates alternative non-linear estimation strategies. In this paper, a data set with globalized developed OECD countries is used to make sure that the proposed estimation of a linear SEM model remains consistent.

The data for bilateral exports is taken from the OECD Database for Structural Analysis 2008 (OECD STAN) and converted into logs ( $\text{exp}_{ij}$ ). GDP data is taken from the OECD national account statistics and converted into logs ( $\text{gdpi}$  and  $\text{gdpj}$ ). The  $w_{ij}^k$ -variables are performed as follows. Distance is calculated using the great circle formula between the capitals or economic

centers<sup>7</sup> of two countries in kilometers, converted into logs. For the measure of common language (lang), common border (bor), commonwealth of nations (cwn), former east block (ebl), island (isl), access to the sea (sea) and EU-membership (eu), dummy variables are used. These dummy variables take the value 1 if the condition that is controlled by the dummy applies, and 0 else. Furthermore, for the dummies cwn, ebl, isl, sea and eu, a distinction is made if only one country or both of them fall under the fact controlled by the dummy. For example sea1 indicates that only one country of the country pair has access to the sea and the other country is landlocked while sea2 indicates that both countries have access to the sea. Note that all of these dummy variables as well as the distance variable are constant over time except the variable for EU-membership since several countries became EU-members during the period. The variables trfi and trfj are the logs of the “Freedom of Trade Index” published by the Bell Foundation. The exchange rate volatility is calculated as the monthly standard deviation from the annual mean relative to the annual mean for each bilateral exchange rate and converted into logs. The monthly data for the US Dollar exchange rates of the respective countries were taken from the OECD Financial Indicators database and recalculated into bilateral exchange rates.<sup>8</sup>

---

<sup>7</sup> In the case of Canada (Toronto), Germany (Frankfurt), Turkey (Istanbul) and United States (Chicago).

<sup>8</sup> Because the Euro-countries are taken as one in this database, exchange rate changes between Euro-countries before the introduction of the Euro were calculated from historical data taken from EUROSTAT. Because this EUROSTAT database does not cover Greek Drachma and the Slovak Korun, since these countries introduced the Euro later, monthly data for the Greek Drachma exchange rate was taken from the US Federal Bank, for the Slovak Korun from the Slovak National Bank.



So far, the data set for a standard gravity framework is explained:  $\text{exp}_{ij}$ ,  $\text{gdp}_i$ ,  $\text{gdp}_j$  and a set of trade cost proxies. Since these trade cost proxies influence exports indirectly via trade costs, a measure for overall average trade costs is needed. Novy (2007) derives an index for the geometric mean of the overall trade costs (measured in iceberg costs) between two countries from the theoretical gravity equation derived by Anderson and van Wincoop (2003), which is also a starting point for the theoretical considerations of this paper. This trade cost index can be computed by the formula:

$$t_{ij} = \left( \frac{X_{ii}X_{jj}}{X_{ij}X_{ji}} \right)^{\frac{1}{2(\sigma-1)}}. \quad (13)$$

The higher the trade volume *inside* the respective countries relative to the trade volume *between* the two countries, the higher is the bilateral trade cost, and vice versa. The trade volume inside a country, or the domestic trade, can be interpreted as the country's production minus the sum of the exports into all countries. Since export data is measured in gross shipments while GDP data is based on value added and, additionally, contains services that are not considered in the export data, GDP is not suitable to calculate this index. Instead, following Wei (1996) and Novy (2007), production data for goods extracted from the OECD STAN Database is used and converted in US Dollars using the OECD Financial Indicators annual exchange rates. Unfortunately, production data is not available for some countries. Furthermore, Mexico and Turkey are not considered in this data set at all. Therefore, missing values of production were constructed over the following three steps. In a first step, I assume that in countries with higher productivity (measured by per-capita-income, source: World Development Indicators, WDI, 2008) the relation value added to production is higher. Thus, I calculate the elasticity of the value added/production-ratio with respect to per-capita-income using ordinary least squares. In a second step, I compute the missing data

points from the estimated values of this regression if there is no data for production, but data for value added in the OECD data.

There are still many missing data points and Mexico and Turkey are still omitted from the data. Hence, in a third step, I take the value added data from the World Development Indicators 2008 and, using an adjusted regression (intercept = 0) between OECD and WDI data, I find that OECD data systematically is 95% of the WDI values. Consequently, I multiply WDI data for value added by factor 0.95 and pursue the same procedure as in the first and the second step to compute missing production estimates for the case that there is no value added data available in the OECD STAN database, but in the WDI database.

Another crucial issue is the elasticity of substitution between the countries' composite goods,  $\sigma$ . In a survey of the empirical literature, Anderson and van Wincoop (2004) find that this elasticity takes values between 5 and 10. Thus, following Novy (2007), the elasticity of substitution is set  $\sigma = 8$ .<sup>9</sup> With the data and the assumption about  $\sigma$ , the logs of trade cost index,  $t_{ij}$ , can be computed.

## 5 Results

Table 1 shows the estimates of the traditional gravity estimation strategy, where all determinants of the trade costs appear directly in the estimation equation. In the first column, the results of the pooled regression approach

---

<sup>9</sup> A sensitivity analysis with  $\sigma = 5$  and  $\sigma = 10$  leads to exactly the same results.

are shown. Only two regressors, namely the trade freedom index of the exporting country ( $trfi$ ) and the case that both countries have access to the sea ( $sea2$ ) do not have a significant impact on bilateral trade volume. Note that the trade of freedom index of the importing country ( $trfj$ ) has got a significantly negative impact on bilateral trade, which means that more liberal importers have lower imports. This result is counter-intuitive. The second column represents the results of estimating the LSDV model with unilateral fixed effects. The time-constant dummies for the case that both the importing and the exporting country are subject to the situation controlled by the respective dummy are dropped due to collinearity. In this model, all variables are significant except the importing countries trade of freedom index ( $trfj$ ). Note that exchange rate volatility seems to have a significantly positive effect on trade in this specification: More uncertainty about exchange rates enhances trade. Also note that the signs of some dummies change compared to the pooled regression specification. The third column of table 1 shows the results for the two-way fixed-effects estimator (bilateral fixed effects). Here, all time invariant variables are dropped due to collinearity. All estimated parameters are significant and have got the expected signs: higher trade freedom and membership in the European Union have got positive impacts on trade, exchange rate volatility lowers the bilateral trade relations.

To compare the three kinds of specifications, the residuals are plotted in figure 1. While the residuals of the pooled and the country fixed effects specifications increase with the logarithmized export volume, in the country-pair fixed effect model they are distributed around zero independently from the export volume. This observation indicates that the country-pair fixed effects model should be preferred. Furthermore the standard errors of the estimated

parameters are lower and the results are all significant and intuitive. Therefore, the two-way fixed-effects estimator should be preferred compared to the other two estimators.<sup>10</sup>

Table 2 shows the results after using the instrumental variable and simultaneous equation model techniques. The left three columns show the results under the restriction that there are no economies of scale in the trade sector using the IV estimator, the right three columns comprise the unrestricted case of estimating a SEM specification. Three important outcomes can be seen from this table. First, the impact of exports on trade costs is significantly negative in the three SEM specification. Second, the signs of the trade cost variables ( $w_{ij}^k$ ) have the expected signs at least in the bilateral fixed effects specification. Third, the estimates for trade cost proxies are lower in the SEM specification compared to the IV specification. That means that controlling for economies of scales in the trade sector lowers the estimated direct impacts of the trade cost proxies.

Two tests are performed: (i) the Sargan-test to check for the appropriateness of the chosen instruments in the restricted IV-model and (ii) the Durbin-Wu-Hausman-test to check whether a simultaneity problem consists.

In the Sargan test,  $\text{exp}_{ij}$  is regressed on  $\text{gdpi}$ ,  $\text{gdpj}$  and  $\text{tij}$  instrumented by the trade cost proxies to achieve the residual vectors  $\hat{u}_{ij}^{IV}$ . This is the residual vector from exactly the IV-regression reported in the upper-left part of table

---

<sup>10</sup> It was also tested whether a random effects model is adequate. The Hausman-test rejects the null-hypothesis that there are no systematic differences in the parameters of two-way fixed effects and random effects, which implies that the two-way fixed effects estimator has to be preferred. In the remaining analysis, the random effects estimator is not further discussed.

2. Then, the achieved residual vector  $\hat{u}_{ij}^{IV}$  is regressed on all exogenous variables in the model:  $\text{gdpi}$ ,  $\text{gdpj}$  and all trade cost proxies. From this regression, the Sargan-test-statistic can be computed as the product of the number of observations and the R-squared. A R-squared close to 0 implies that there is less correlation between the instruments and the error term and therefore the instruments tend to be exogenous. This is the null hypothesis. Under the alternative hypothesis, the instruments are correlated with the error term and therefore endogenous. The test statistic is distributed  $\chi_{k-r}^2$ , where  $k - r$  is the difference between the number of instruments (or trade cost proxies, respectively) minus the number of endogenous variables of the right hand side (which is one:  $t_{ij}$ ). Only in the case of the LSDV-model, the null hypothesis of valid instruments cannot be rejected. In the other models, especially in the preferred bilateral fixed effects model, instruments are not valid and the pure instrument variable estimator appears to be an inadequate measure.

The Durbin-Wu-Hausman-test also consists of two steps. In the first step,  $t_{ij}$  is regressed on all exogenous variables in the model to calculate the estimated residual vector  $\hat{\epsilon}_{ij}$ , see reduced form equation (12). In the second step, this residual vector is plugged into the structural equation of interest (9) as an additional regressor. The null hypothesis is that the coefficient of the residuals is 0. If the F-test suggests that the coefficient of  $\hat{\epsilon}_{ij}$  is significantly unequal to 0, an interdependent relationship between equations (10) and (9) is likely. In this case, OLS estimates are biased and inconsistent while the 2SLS/3SLS results are at least consistent. In all three panel specifications the null hypothesis is rejected and consequently the application of the SEM-strategy should be preferred.

To examine which kind of specification of the SEM estimation has the best fit, the residuals of the pooled, country fixed effects and country-pair fixed effects model are plotted in figures 2 (for the gravity equation) and 3 (for the trade cost equation). Like in figure 1, the residuals of the two-way model are distributed around zero which indicates that the country-pair fixed effects specification makes the best fit. Note that in the SLS estimation, the deviation of the residuals from zero is lower than in the OLS estimation.

Two central research questions were proposed in section 3. Summing up, the results suggest the following answers:

**Answer 1** *Are there economies of scale in the trade sector?*

Obviously yes. The results reported by table 2 yield the following three conclusions: First, the effect of exports on trade costs is significantly negative in the SEM-specification. Second, the Sargan-test indicates that the used instruments ( $w_{ij}^k$ -variables without  $\text{exp}_{ij}$ ) are not valid in the pooled regression and the two-way fixed effects model. Third, the Hausman-test indicates simultaneity in all three models.

**Answer 2** *How are estimates biased if economies of scale in the trade sector consist, but there is not controlled for?*

Ignoring the endogeneity of trade costs tends to overestimate the effects of trade cost proxies. This can be seen after comparing the left part of table 2 with the right part. Nearly all of the  $w_{ij}^k$ -variables are considerably lower in the SEM estimation.

## 6 Conclusion

Studies that apply the gravity equation take trade costs as exogenously given. But theoretical considerations and intuition imply that exports between two countries depend on bilateral trade costs and bilateral trade costs depend on exports between the two countries if there are economies of scale in the trade sector. If this interdependence between exports and trade costs consists and is not controlled, estimates might be biased.

The empirical results of this paper give evidence that economies of scale in international trade do exist. Using a 3SLS/2SLS regression yields that more trade between two countries implies lower bilateral trade costs. Comparing the results of this regression with an assimilable IV approach, where the impact of exports on trade costs is neglected by assumption, shows that ignoring the interaction of exports and trade costs tends to result in higher coefficients. This result might be a contribution to the broad discussion of presumably too high coefficient estimates in studies using gravity equation.

## A Figures and Tables

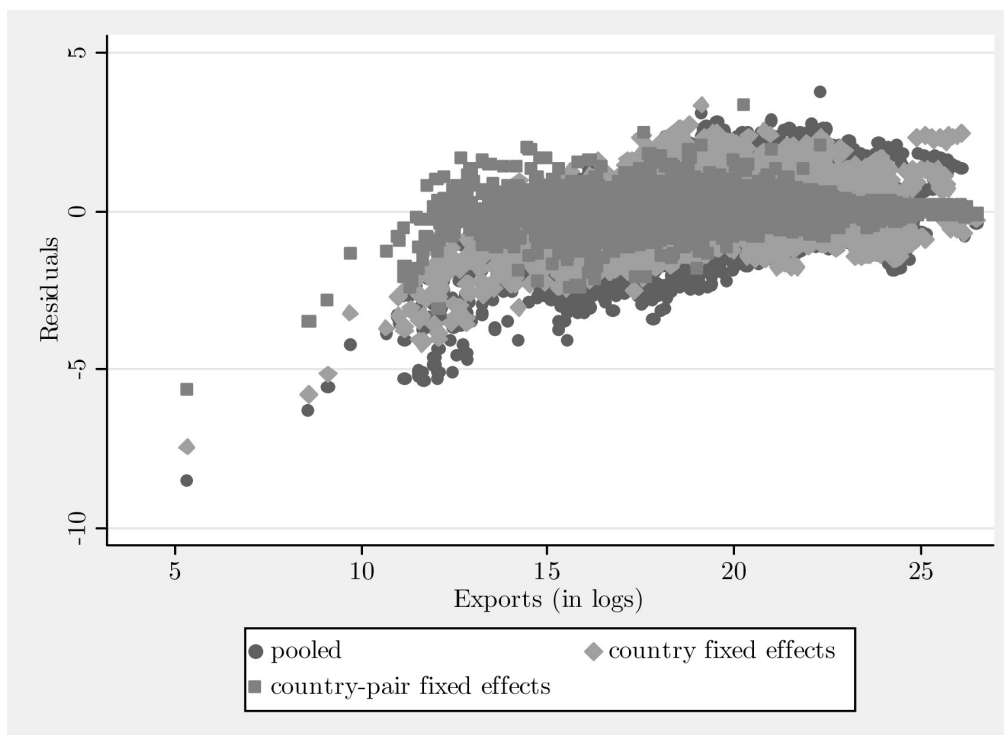


Figure 1: OLS Estimation – Residuals of the three model specifications.



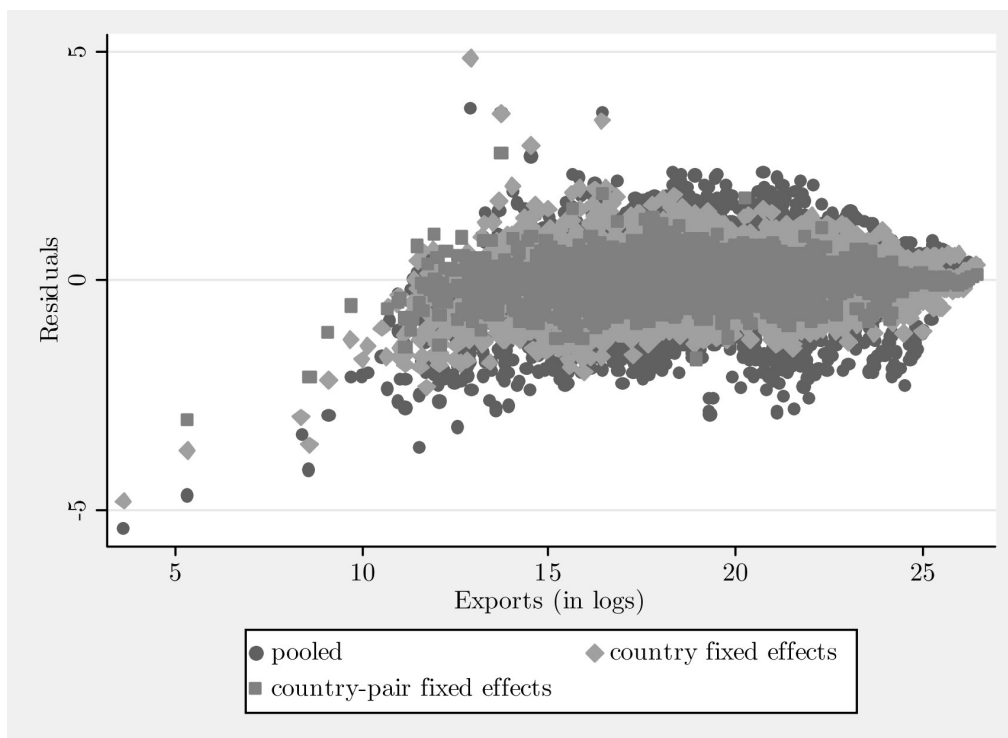


Figure 2: SLS Estimation of the Gravity Equation– Residuals of the three model specifications.

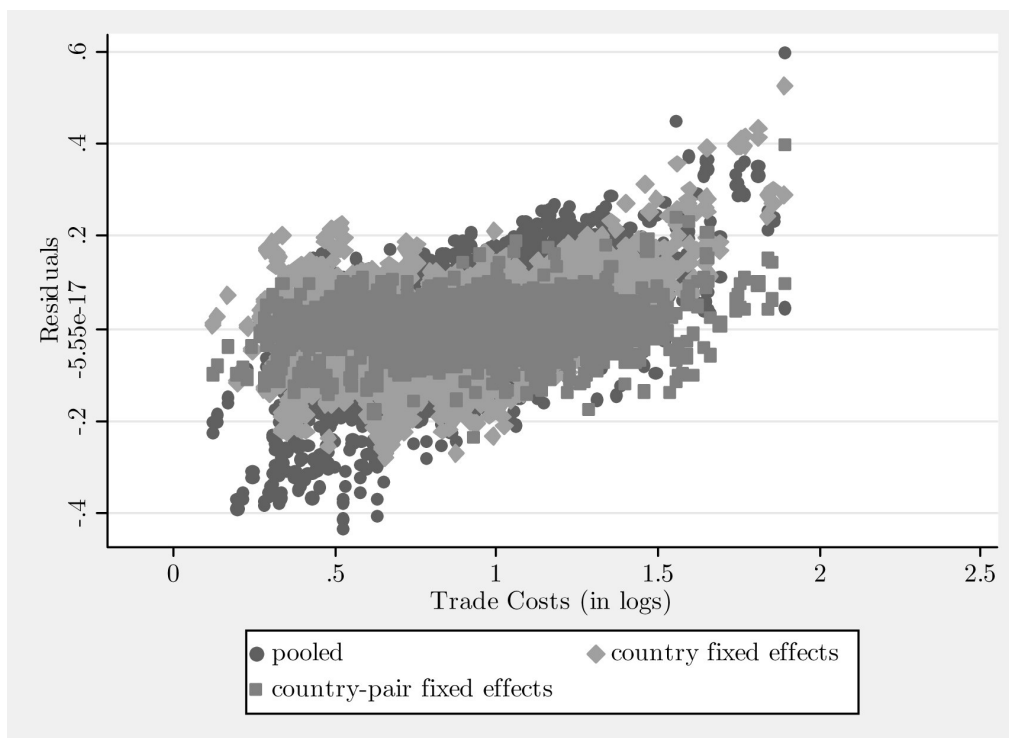


Figure 3: OLS Estimation of the Transport Cost Equation – Residuals of the three model specifications.

expij	Pooled Regression (1)	Unilateral FE (2)	Bilateral FE (3)
gdpi	0.915 (0.008)***	0.441 (0.082)***	0.483 (0.033)***
gdpij	0.896 (0.008)***	0.621 (0.082)***	0.659 (0.032)***
trfi	-0.182 (0.182)	0.351 (0.188)*	0.573 (0.108)***
trfj	-0.444 (0.164)***	-0.050 (0.171)	0.180 (0.084)**
dist	-1.017 (0.013)***	-1.155 (0.018)***	—
lang	0.405 (0.042)***	0.315 (0.046)***	—
bor	0.328 (0.041)***	0.205 (0.041)***	—
cwn1	0.098 (0.030)***	-0.338 (0.027)***	—
cwn2	0.839 (0.077)***	—	—
eb11	-0.266 (0.032)***	-0.835 (0.034)***	—
eb12	1.129 (0.065)***	—	—
isl1	0.094 (0.028)***	-0.147 (0.028)***	—
isl2	0.525 (0.056)***	—	—
sea1	-0.167 (0.046)***	-0.173 (0.024)***	—
sea2	0.047 (0.053)	—	—
eu1	0.209 (0.032)***	0.491 (0.045)***	0.472 (0.036)***
eu2	0.268 (0.038)***	0.840 (0.072)***	0.736 (0.039)***
exvol	-0.050 (0.010)***	0.019 (0.008)**	-0.015 (0.005)***
Constant	-17.122 (1.068)***	0.606 (3.259)	-13.876 (0.621)***
Obs.	8492	8492	8492
Adjusted $R^2$	0.83	0.90	—
$R^2$	—	—	0.98

Note: Robust standard errors in parentheses.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

Table 1: Basic Case – Estimates of the Traditional Gravity Approach.

	Instrumental Variable Estimation (IV)			Simultaneous Equation Model (SEM)		
	Pooled (2SLS)	Unilateral FE (2SLS)	Bilateral FE (2SLS)	Pooled (3SLS)	Unilateral FE (3SLS)	Bilateral FE (2SLS)
(1) expij						
tij	-6.275***	-6.999***	-6.750***	-6.275***	-6.999***	-6.750***
gdpj	0.539***	0.344***	0.300***	0.535***	0.348***	0.300***
gdpj	0.512***	0.527***	0.438***	0.516***	0.524***	0.438***
cons	-2.539***	—	6.116***	-2.542***	—	6.116***
$R^2$	0.922	0.968	0.899	0.923	0.968	0.899
(2) expij						
expij	—	—	—	-0.066***	-0.032***	-0.052***
trfj	-0.154***	-0.037*	-0.124***	-0.019	-0.026	-0.052***
trfj	-0.157***	-0.038**	-0.124***	-0.023*	-0.021	-0.066***
dist	0.139***	0.166***	—	0.095***	0.128***	—
lang	-0.127***	-0.044***	—	-0.046***	-0.034***	—
bor	-0.074**	-0.027***	—	-0.009**	-0.021***	—
cwn1	-0.030***	0.049***	—	-0.009***	0.038***	—
cwn2	-0.102***	—	—	-0.057***	—	—
eb11	0.103***	0.117***	—	0.033***	0.091***	—
eb12	-0.006	—	—	-0.072***	—	—
isl1	0.044***	0.022***	—	-0.007***	0.017***	—
isl2	0.053***	—	—	-0.040***	—	—
sea1	-0.045***	0.024***	—	0.014***	0.019***	—
sea2	-0.128***	—	—	0.000	—	—
eu1	-0.049***	-0.077***	-0.085***	-0.026***	-0.057***	-0.037***
eu2	-0.090***	-0.135***	-0.148***	-0.033***	-0.099***	-0.062***
exvol	0.002	-0.003***	0.005***	0.005***	-0.002***	0.002***
cons	1.186***	—	1.984***	1.576***	—	2.410***
$R^2$	0.580	0.879	0.209	0.880	0.923	0.752
Obs.	8,092	8,092	8,092	8,092	8,092	8,092
Sargan	$\Pr(\chi^2_{15} > 313.11)$ 0.000	$\Pr(\chi^2_{11} > 11.51)$ 0.402	$\Pr(\chi^2_4 > 52.56)$ 0.000	—	—	—
DWH	—	—	—	$\Pr(F_{1,8087} > 47.98)$ 0.000	$\Pr(F_{1,7256} > 14.58)$ 0.000	$\Pr(F_{1,7256} > 17.31)$ 0.000

Table 2: Theory-based estimates without economies of scale in trade (IV) and with economies of scale in trade (SEM).

## References

- ANDERSON, J. E., AND E. VAN WINCOOP (2003): “Gravity with Gravitas: A solution to the Border Puzzle,” *American Economic Review*, 93(1), 170–92.
- (2004): “Trade Costs,” *Journal of Economic Literature*, 42, 691–751.
- BAIER, S., AND J. BERGSTRAND (2007): “Do Free Trade Agreements Actually Increase Members’ International Trade?,” *Journal of International Economics*, 71(1), 72–95.
- BALTAGI, B. H., P. EGGER, AND M. PFAFFERMAYR (2003): “A generalized design for bilateral trade flow models,” *Economics Letters*, 80, 391–397.
- CHENG, I.-H., AND H. J. WALL (1999): “Controlling for Heterogeneity in Gravity Models of Trade and Integration,” Working Paper 1999-010E, Federal Reserve Bank of St. Louis.
- FRANKEL, J., AND D. ROMER (1999): “Does Trade Cause Growth?,” *American Economic Review*, 89(3), 379–399.
- GREENE, W. H. (2000): *Econometric Analysis*. Prentice Hall International, 4th edn.
- GROSSMAN, G. M. (1998): “Comment on A. V. Deardorff, Determinants of International Trade: Does Gravity Work in a Neoclassical World?,” in *The Regionalization of the World Economy*, ed. by J. A. Frankel, pp. 29–31. The University of Chicago Press, Chicago.

- HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): “Estimating Trade Flows: Trading Partners and Trading Volumes,” *Quarterly Journal of Economics*, 123(2), 441–487.
- MCCALLUM, J. (1995): “National Borders Matter: Canada-U.S. Regional Trade Patterns,” *American Economic Review*, 85(3), 615–623.
- MÁTYÁS, L. (1997): “Proper Econometric Specification of the Gravity Model,” *The World Economy*, 20(3), 363–368.
- NOVY, D. (2007): “Gravity Redux: Measuring International Trade Costs with Panel Data,” Warwick Economic Research Papers No. 861.
- RUDOLPH, S. (2009): “The Gravity Equation with Micro-founded Trade Costs,” Dresden Discussion Paper in Economics 11/09, Technische Universität Dresden.
- WEI, S.-J. (1996): “Intra-National versus International Trade: How Stubborn are Nations in Global Integration?,” NBER Working Paper 5531.

## Dresden Discussion Paper Series in Economics

- 09/08 **Broll, Udo / Schmidt, Christian W.:** The Effect of Exchange Rate Risk on US Foreign Direct Investment: An Empirical Analysis
- 10/08 **Lehmann-Waffenschmidt, B. Cornelia:** An Evolutionary Agent-Based Simulation Model for the Industry Lifecycle
- 11/08 **Mukherjee, Vivekananda / Ramani, Shyama V.:** R&D Cooperation, Asymmetrical Technological Capabilities and Rationale for Technology Parks
- 12/08 **Lehmann-Waffenschmidt, Marco / Roth, Gerhard / Thießen, Friedrich:** Die (innere) Logik des Entscheidens – Zur neurobiologischen Fundierung ökonomischer Entscheidungen
- 13/08 **Dopfer, Kurt:** Zur Neubestimmung von Geschichte und Theoriegeschichte in den Wirtschaftswissenschaften – Ko-evolution von Wirtschaftsbeobachtung und Wirtschaftswirklichkeit
- 14/08 **Lessmann, Christian / Markwardt, Gunther:** One Size fits all? Decentralization, Corruption, and the Monitoring of Bureaucrats
- 15/08 **Farzanegan, Mohammad Reza / Markwardt, Gunther:** The Effects of Oil Price Shocks on the Iranian Economy
- 16/08 **Fuchs, Michaela / Wohlrabe, Klaus:** Institutions, Trade and Integration: What can be expected within the enlarged EU?
- 01/09 **Broll, Udo / Egozcue, Martin / Wong, Wing-Keung:** Prospect Theory and Two Moment Model: the Firm under Price Uncertainty
- 02/09 **Broll, Udo / Eckwert, Bernhard:** Modelling Information and Hedging: the Exporting Firm
- 03/09 **Binswanger, Hans Christoph:** Die Wachstumsspirale in der Krise – Ansätze zu einem nachhaltigem Wachstum
- 04/09 **Brunow, Stefan / Hirte, Georg:** Regional Age Structure and Economic Growth: An Econometric Study for German Regions
- 05/09 **Broll, Udo / Kemnitz, Alexander / Mukherjee, Vivekananda:** Globalization and a Welfare Program for the Marginalized
- 06/09 **Tscharaktschiew, Stefan / Hirte, Georg:** An Urban General Equilibrium Model with Multiple Household Structures and Travel Mode Choice
- 07/09 **Tscharaktschiew, Stefan / Hirte, Georg:** How does the Household Structure Shape the Urban Economy?
- 08/09 **Lessmann, Christian:** Fiscal Decentralization and Regional Disparity: Evidence from Cross-section and Panel Data
- 09/09 **Lessmann, Christian / Markwardt, Gunther:** Aid, Growth and Decentralization
- 10/09 **Broll, Udo / Wahl, Jack E. / Wessel, Christoph:** Export and Benefits of Hedging in Emerging Economies
- 11/09 **Rudolph, Stephan:** The Gravity Equation with Micro-Founded Trade Costs
- 12/09 **Biswas, Amit K.:** Import Tariff Led Export Under-invoicing: A Paradox
- 13/09 **Broll, Udo / Wahl, Jack E.:** Mitigation of Foreign Direct Investment – Risk and Hedging
- 14/09 **Broll, Udo / Wahl, Jack E.:** Güterwirtschaftliches Risikomanagement: - Ein Entscheidungsmodell zur Lagerpolitik bei Unsicherheit
- 15/09 **Lukas, Daniel:** Efficiency Effects of Cross-Border Medical Demand
- 16/09 **Broll, Udo / Bieta, Volker / Milde, Hellmuth / Siebe, Wilfried:** Strategic Pricing of Financial Options
- 16/09 **Broll, Udo / Bieta, Volker / Milde, Hellmuth / Siebe, Wilfried:** Strategic Pricing of Financial Options
- 17/09 **Broll, Udo / Wahl, Jack E.:** Liquidity Constrained Exporters: Trade and Futures Hedging
- 01/10 **Rudolph, Stephan:** Estimating Gravity Equations with Endogenous Trade Costs





