MODELLING PRODUCTION-CONSUMPTION FLOWS OF GOODS IN EUROPE: THE TRADE MODEL WITHIN TRANSTOOLS3

Gerard de Jong,
Institute for Transport Studies, University of Leeds, UK, and Significance, The Netherlands
Reto Tanner, Institute for Transport Studies, University of Leeds, UK, and Delft University of Technology, The Netherlands
Jeppe Rich
Mikkel Thorhauge
Otto Anker Nielsen
DTU Technical University of Denmark
John Bates
John Bates Services, UK

1. INTRODUCTION

The Transtools3 model is a new forecasting model system for passenger and freight transport in Europe, developed by a consortium led by DTU from Denmark for DG MOVE of the European Commission. It consists of three main blocks: the passenger transport model, the freight and logistics model and the network assignment model. This paper focuses on the trade model, a specific sub-model within the freight and logistics model. This is the top-right box in Figure 1, which presents the overall freight and logistics model. The trade model produces growth between base year and future year in the goods flows between production and consumption zones (PC flows, measured in tonnes). Together with the base PC matrix, the trade model produces future year aggregate PC matrices. The freight and logistics model as a whole is a form of an aggregate-disaggregate-aggregate or ADA model (see Ben-Akiva and de Jong, 2013).

In the literature, there are basically three different approaches (Ivanova, 2014) for modelling PC flows (that is for modelling the transport distribution):

- Gravity-models
- Input-output (IO) models
- Spatial Computable General Equilibrium (SCGE) Models.

The most commonly used method is the gravity model (e.g. used for the Netherlands in Significance et al., 2010; or for Sweden in Edwards, 2008). In such models the flow between zone i and zone j is a function of the product of production and attraction measures of zone i and zone j respectively divided by a some measure of distance or (generalised) transport cost. This makes transport distribution sensitive to changes in transport cost and time (a form of induced demand).

Gravity models to explain trade or transport in money or weight units have long been regarded as models without a theoretical justification, borrowed from physics because they turned out to work well in transport too. In more recent economic literature, a theoretical basis for the gravity model has been derived from the factor proportions model (Deardorff, 1998), Ricardo’s trade model (Eaton and Kortum, 2001, 2002) or monopolistic competition with differentiated products (Anderson and van Wincoop, 2003; Bergstrand et al., 2013).

Input-output tables describe, in money units, what each sector of the economy (e.g. textile manufacturing) delivers to the other sectors, also including the final demand by consumers, import and export. National input-output tables have been developed for many countries. A special form of input-output table, which for many countries does not exist, is a multi-regional or spatial input-output table. This not only includes deliveries between sectors, but also between regions (trade flows). The input-output model (e.g. Marzano and Papola, 2004) assumes that for forecasting, the multi-regional input-output table can be scaled up on the basis of predicted sectoral growth. The new input-output table can then give the future trade flows between regions, using either:

- Fixed technical and trade coefficients: the present production and trade patterns are extrapolated into the future.
- Elastic technical and trade coefficients: functions are estimated (e.g. multinomial logit) in which the fraction that is consumed in region i of the production of sector s in region j depends on the total production of region j in sector s and the (generalised) transport cost, in relation to other regions.

The third option for production and attraction is the computable general equilibrium (CGE) model that establishes equilibrium in several related markets (not only transport, but also goods markets, labour markets, land markets). CGE models in economics (not focussing on transport) often include economic issues that are not handled in transport models, such as type of competition and economies of scale. SCGE models have at least the same data requirements as IO models (e.g. multiregional IO tables or make and use tables). Examples of spatial CGE models that are used for transport distribution are Bröcker et al. (2010), Ivanova et al. (2006) and Ivanova et al. (2007).

For this application we have selected the gravity model instead of the IO and the SCGE models because in Transtools3 we had no ambitions to model other markets than transport (such as the labour or the land market) and up-to-date information on regional input-output relations for our study area was missing for many regions. Conversely, we did have a base PC matrix as input from the ETISplus project. This matrix was also already expressed in tonnes, so this way we could also avoid the conversion from money to tonnes, that needs to be part of IO and SCGE models applied to transport.

The trade model explains the transport flows, either between NUTS3 zones or between countries, by NST/R 1 commodity type, based on the (unconstrained) gravity formulation, from characteristics of the zones and their transport resistance. The dependent variable is the base PC matrix, as delivered by another project for DGMOVE, the ETIS+ project. The independent variables are GDP, GDP per capita (both obtained from the World Bank) and dummies for common trade zone (EU, EFTA), common currency zone (EURO), common language, zones being neighbours and zones being in the same country. The resistance variable in the current estimates is geographic distance, modelled using a spline function, but we plan to replace this later by transport costs to make the trade model sensitive to transport policies that alter transport costs.

To take into account the influence of relative trade resistance between countries instead of absolute resistance (in line with trade theory), we estimated not only least squares regression models but also fixed effects and random effect models.

A key problem of standard regression models is that they cannot handle the case where transport flows are zero, which is very relevant here since many countries do

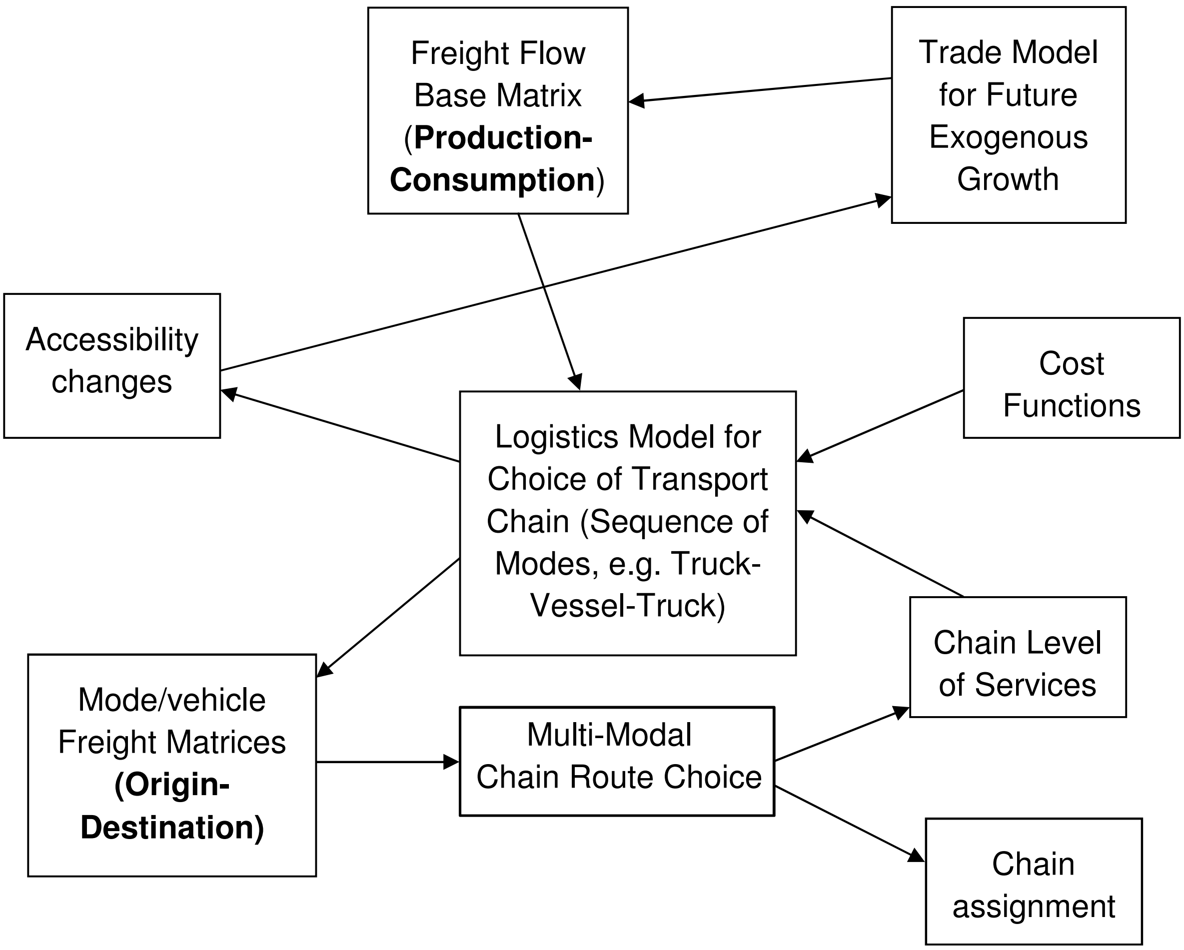

Figure 1. General structure of the Transtools3 freight and logistics model.

not trade (within a commodity group) with each other. In the ordinary least squares regression, the estimation takes place on the positive observations only, omitting the zero-trade data. As a result there can be a sample selection bias. In order to capture the decision whether or not to trade, we estimated Heckman sample selection models with a simultaneous trade selection equation and a positive demand equation to correct for this effect.

In the current model implementation, we only use the predicted changes over time (scenario-based) in GDP by zone and in the population to compute changes in the transport flows by commodity type and zone pair. All other variables are assumed to remain constant.

In Section 2 of this paper, we present the gravity model for trade between zones. The data are described in Section 3. Section 4 discusses the main estimation results. The model implementation is handled in Section 5 and Section 6 gives the conclusions of the paper.

2. THE GRAVITY MODEL FOR TRADE

2.1 Accounting for relative trade cost

The modern theoretical literature on gravity-based trade models (Anderson and van Wincoop, 2003, 2004; Kepaptsoglou et al. 2010; Plummer et al., 2010) has emphasised that trade between two countries is not simply determined by the absolute trade costs between the two countries, but by the relative trade cost (the trade cost of country *i* from importer *j* relative to its overall trade cost for all the countries from which it imports). In an empirical gravity model, this can be taken into account by adding multilateral resistance terms. However a simpler method is to use importer or exporter fixed effects (dummy variables), which are meant to capture effects that are specific to a country including its overall level of imports or exports.

We therefore now first present the model with fixed effects γ_j, which are defined at the level of destination countries *j* (we also estimated a model with fixed effects at the origin country level *i*):

$$\ln\left(x_{ij}\right) = \gamma_j + \sum_{k=1}^{K_D} \alpha_k \cdot \ln\left(d_{ij(k)}\right) + \beta_1 \cdot \ln\left(gdp_i\right) + \beta_3 \cdot \ln\left(\frac{gdp_i}{pop_i}\right) + \ldots$$
$$\ldots + \beta_5 \cdot euefta_{ij} + \beta_6 \cdot euro_{ij} + \beta_7 \cdot neig_{ij} + \beta_8 \cdot lang_{ij} + \varepsilon_{ij} \qquad (1)$$

where:
x_{ij}: flow of goods between country i and j, in tonnes;
d_{ijk}: distance splines, for distance bands k, with distance measured as crow-fly kilometres;
gdp: gross domestic product in euro of 2010;
pop: population;
euefta: dummy that equals 1 if both countries are member of EU or EFTA; 0 otherwise;
euro: dummy that equals 1 if both countries have the EURO as a currency; 0 otherwise;
neig: dummy that equals 1 if both countries are neighbours; 0 otherwise;
lang: dummy that equals 1 if both countries have the same language; 0 otherwise;

In general Greek symbols indicate parameters to be estimated.

The model is double logarithmic in its continuous variables, which is in line with the multiplicative gravity model formulation and also usually works better (here as well) than linear models and yields coefficients which can be directly interpreted as elasticities.

Note, in this model we cannot estimate parameters for the variables gdp_j and gdp_j / pop_j because these are specific to the destination countries j and would be perfectly correlated with the destination country dummy. But, using one constant per destination country may more accurately explain the effects specific to the destination country that influence the imports and reduce thus the variance of the error terms ε. From this follows that in this model we cannot calculate the total impact of a rise of GDP in the total economy. This is a big disadvantage in application.

A third way to take account of relative trade costs, which does not have this disadvantage, is the random effects model. In this case, we try to explain the effects specific to the destination country by a number of observed variables that are specific to the destination country. We also add a destination-country-specific error term ϑ, since we believe we cannot fully explain the destination-country-specific effect on the basis of the variables that we have. Thus, we believe there are unobserved effects that are specific to the destination country. The model was estimated using generalised least squares (GLS), which also yields the variances of both error terms.

$$\ln\left(x_{ij}\right) = \sum_{k=1}^{K_D} \alpha_k \cdot \ln\left(d_{ij(k)}\right) + \beta_1 \cdot \ln\left(gdp_i\right) + \beta_2 \cdot \ln\left(gdp_j\right) + \beta_3 \cdot \ln\left(\frac{gdp_i}{pop_i}\right) + \beta_4 \cdot \ln\left(\frac{gdp_j}{pop_j}\right) + \ldots$$

$$\ldots + \beta_5 \cdot euefta_{ij} + \beta_6 \cdot euro_{ij} + \beta_7 \cdot neig_{ij} + \beta_8 \cdot lang_{ij} + \vartheta_j + \varepsilon_{ij} \qquad (2)$$

2.2 Accounting for relations with zero trade

In fact, companies do not only decide how much they want to export to a specific country, but also whether they will export to a country at all. Often, there are barriers to export to a country. Some are there for purely economic reasons. One such example is a car manufacturer who needs a certain density of sales representatives in a region, so that the customers have a service point within an acceptable distance from where they live. Therefore, if a car manufacture expects not to sell at least a certain amount of cars at a given price in a country, he will not enter the market at all. Other type of barriers are costs due to differences in the language spoken, or differences in regulation, difficulties in culture, uncertainties and risks with respect to the regulation and tariffs. All these can prevent one company exporting to another country. Of course, all these factors may affect some companies more than others and thus one could expect that in each country, there are at least some companies that export goods to any other country. However, the data shows that for origin-destinations combinations at the country level, depending on the commodity type, 15-85% of the observations are missing values (which can best be interpreted as: no trade).

This situation with two related choices, one discrete choice to participate in trade ("selection") and one continuous choice on the amount of trade (when positive;

“demand”) can be modelled using the model that Heckman originally developed to explain labour market participation and wages (Heckman, 1979). Applications of this model to trade between countries can be found in Linders and de Groot (2006) and in Gomez Herrera (2010).

The demand equation is:

$$\ln\left(x_{ij}\right) = \alpha_1 \cdot \ln\left(d_{ij}\right) + \beta_1 \cdot \ln\left(gdp_i\right) + \beta_2 \cdot \ln\left(gdp_j\right) + \beta_3 \cdot \ln\left(\frac{gdp_i}{pop_i}\right) + \beta_4 \cdot \ln\left(\frac{gdp_j}{pop_j}\right) + \varepsilon_{ij} \qquad (3)$$

The selection equation is:

$$s_{ij} = \gamma_1 \cdot \ln\left(d_{ij}\right) + \delta_1 \cdot \ln\left(gdp_i\right) + \delta_2 \cdot \ln\left(gdp_j\right) + \delta_3 \cdot \ln\left(\frac{gdp_i}{pop_i}\right) + \delta_4 \cdot \ln\left(\frac{gdp_j}{pop_j}\right) + \delta_5 \cdot lang_{ij} + \xi_{ij} \qquad (4)$$

where, $s_{ij} = 0$ if trade not observed or zero, and $s_{ij} = 1$ otherwise.

Note that we need at least one instrumental variable. This is a variable that only has an impact on the selection choice but not on the demand (amount) of imports or exports. In eq. (4) we include the dummy for the same language for this, as an example. In the estimation in Section 4, we use all the dummies from eq. (2) also in (4), but not in (3).

The above Heckman model was estimated simultaneously with the Maximum Likelihood method. It is not possible in the software used (Stata) to estimate a Heckman model that also has random effects, so the Heckman models here do not have dummies or an error term for origin- or destination- specific effects.

3. THE DATA USED

Our preferred model was estimated on data at the country level, and implemented (see Section 5) at the level of TT3 zones (which is the NUTS3 level, or subdivisions of those). However, we also received from the ETISplus project transport flows in tonnes, by NST/R 2 at the NUTS3 level for 2010. We also estimated gravity-based trade models on these data (de Jong et al., 2015). We prefer a model estimated on the country data, since at this level the flows are observed data, obtained from international organisations. At the zonal level, ETISplus made a synthetic split using GDP and population data, so that to some degree estimating a trade model at this level is remodelling the model used for imputing the trade flow data. This is not the case when estimating at the country level, where we also have GDP and other explanatory data directly from international organisations. At the country level, there are also good reasons to believe that zero observations really indicate the absence of trade. At the zonal level, zero (or missing observations) might indicate other things.

The basis of our data is a production-consumption matrix (PC matrix) for ETISplus level 3 (EZ2006_3) zones. In the following we refer to this as the ETIS-3 zones. Each observation covers the flow of a specific type of goods following the NST/R level 2 (NST/R 2) classifications from an origin zone to a destination zone (ETIS-3?). This flow data was produced by the ETISplus project (2014, 2015). They used a PC matrix of observed data at country level and then imputed trade flows for each individual pair of zones using methods that take into account the GDP and other variables of the ETIS-3 zones. We estimated models explaining this PC matrix, as well as models explaining the matrix of flows aggregated to the country to country level (using 214 countries in total). As explained above we prefer the latter models, and these will be presented in this paper. As a check, we will compare the results of estimation at the country level with those at the zonal level.

Our main explanatory variables are GDP, GDP per capita and country-country distance. As data source for the GDP (and population) we use the World Bank database “World development indicators (WDI)”, GDP at current prices in USD, which we converted to EURO of 2010 using a factor of 1.32414. For distance, we use crow- fly distance between the points defined by the longitude and the latitude of each pair of countries. We also defined a number of dummy explanatory variables (prepared largely manually), see list of variables in Section 2 below eq. (1).

NST/R	(0) ln_tonnes_0	(1) ln_tonnes_1	(2) ln_tonnes_2	(3) ln_tonnes_3	(4) ln_tonnes_4	(5) ln_tonnes_5	(6) ln_tonnes_6	(7) ln_tonnes_7	(8) ln_tonnes_8	(9) ln_tonnes_9
Distance 0-20km	0	0	-0.701*	-0.533*	-0.398	-0.986*	-0.890*	-0.524*	0	0
	(.)	(.)	(-2.38)	(-2.14)	(-1.45)	(-4.86)	(-4.73)	(-2.17)	(.)	(.)
Distance 20-50km	0	0	0	0	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Distance 50-100km	0	0	0	0	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Distance 100-300km	1.327	0.657	2.068	1.827	-1.257	-0.975	-1.779	1.429	-0.991	0.321
	(0.76)	(0.41)	(0.81)	(0.80)	(-0.55)	(-0.50)	(-1.00)	(0.64)	(-0.69)	(0.22)
Distance 300-500km	-2.881*	-3.236*	-3.169+	-2.125	-3.409*	-1.358	-2.667*	0.834	-3.191*	-3.548*
	(-2.59)	(-3.21)	(-1.81)	(-1.45)	(-2.29)	(-1.15)	(-2.34)	(0.59)	(-3.51)	(-3.82)
Distance 500-1000km	-2.025*	-1.985*	-0.481	-2.048*	-2.150*	-2.041*	-3.325*	-1.450*	-2.141*	-1.965*
	(-4.62)	(-4.99)	(-0.60)	(-3.43)	(-3.48)	(-4.39)	(-7.17)	(-2.34)	(-5.95)	(-5.41)
Distance 1000-2000km	-2.200*	-2.182*	-0.314	-0.655+	-0.558	-2.903*	-2.449*	-0.520	-2.639*	-2.902*
	(-8.90)	(-10.01)	(-0.57)	(-1.80)	(-1.38)	(-10.77)	(-8.90)	(-1.30)	(-13.15)	(-14.83)
Distance 2000+km	-0.339*	-0.0479	0.399+	-2.141*	-0.0960	-1.005*	-1.177*	-0.547*	-1.075*	-0.696*
	(-4.64)	(-0.80)	(1.89)	(-15.00)	(-0.65)	(-11.17)	(-12.41)	(-3.78)	(-17.63)	(-12.64)
Ln(origin gdp)	**0.824***	**0.899***	**0.474***	**0.735***	**0.587***	**0.921***	**1.054***	**-0.00223**	**1.123***	**1.175***
	(39.12)	**(56.08)**	**(8.42)**	**(20.92)**	**(14.93)**	**(35.88)**	**(39.03)**	**(-0.05)**	**(66.81)**	**(80.25)**
Ln(destination gdp)	**0.598***	**0.618***	**0.430***	**0.499***	**0.532***	**0.781***	**0.625***	**0.619***	**0.931***	**0.814***
	(12.43)	**(20.14)**	**(5.46)**	**(7.80)**	**(5.73)**	**(19.70)**	**(14.80)**	**(11.65)**	**(30.71)**	**(28.93)**
Ln(origingdp/cap)	**-0.274***	**-0.211***	**-1.051***	**-0.425***	**-0.352***	**-0.393***	**-0.556***	**-0.627***	**0.119***	**-0.0177**
	(-8.37)	**(-8.39)**	**(-9.78)**	**(-7.00)**	**(-5.22)**	**(-9.95)**	**(-13.22)**	**(-8.89)**	**(4.32)**	**(-0.78)**
Ln(dest. gdp/cap)	**-0.111**	**-0.0938***	**0.422***	**0.316***	**-0.0390**	**-0.0936+**	**-0.163***	**-0.390***	**-0.137***	**-0.104***
	(-1.64)	**(-2.10)**	**(3.06)**	**(3.41)**	**(-0.30)**	**(-1.70)**	**(-2.73)**	**(-5.30)**	**(-3.30)**	**(-2.59)**
Both member of EU or EFTA	0.743*	1.103*	0.264	-1.204*	0.357	0.435*	0.357*	-0.0620	0.215+	1.183*
	(5.07)	(8.65)	(0.77)	(-5.44)	(1.45)	(2.72)	(2.15)	(-0.26)	(1.83)	(10.31)
Both Euro as currency	0.596*	-0.0153	0.245	-0.298	0.562*	0.124	0.423*	1.191*	0.324*	0.181
	(3.36)	(-0.10)	(0.76)	(-1.20)	(2.13)	(0.65)	(2.22)	(4.63)	(2.25)	(1.24)
Neighbour countries	1.734*	1.287*	0.949*	1.933*	1.570*	1.184*	1.720*	1.730*	1.037*	0.659*
	(7.03)	(5.77)	(2.36)	(5.85)	(4.60)	(4.54)	(6.76)	(5.38)	(5.12)	(3.21)
Both same language	0.742*	1.020*	0.899*	0.968*	0.903*	0.671*	0.777*	0.106	0.945*	0.963*
	(5.64)	(9.06)	(2.81)	(4.76)	(4.01)	(4.30)	(5.06)	(0.49)	(8.56)	(9.37)
constant	-9.103*	-8.608*	0	0	0	0	0	0	-10.86*	-11.49*
	(-4.88)	(-5.35)	(.)	(.)	(.)	(.)	(.)	(.)	(-7.40)	(-7.73)
N	6388	7905	1379	4039	2619	5063	4442	2380	6465	8686

t statistics in parentheses
+ p<0.10, * p<0.05

Table 1: Estimation results for a model at the country level with country-specific random effects at the destination.

Many trade models explain trade measured in money units. This is then followed by a conversion step to go from money units to tonnes (needed because subsequent sub-models, such as modal spit, are in tonnes). Since we are using data on goods flows in tonnes, we do not require this additional conversion step (one could say that we are explaining transport rather than trade). The downside of this is that we cannot easily link the model to economic statistics of trade in money units.

4. ESTIMATION RESULTS

All models are estimated per NST/R 1 commodity type (10 models). The ETISplus data were aggregated from NST/R 2 to this level. We found that the GDP elasticities of trade flows in tonnes were rather similar for models estimated on zonal and country data. We started the estimation of the trade models on the country data (positive flows only) by estimating a gravity model without fixed or random effects, using ordinary least squares estimation, and then moved on to estimate fixed effects models (both models not reported here for the sake of space) with fixed effects referring to the destination countries. The results show that the origin GDP elasticities do not change much compared to the models without fixed effects.

The estimation results of the random effects model are in Table 1. The distance coefficients indicate that trade is strongest between countries that are geographically close (for some products, transport flows only start decreasing after a distance of 300 km). As expected, GDP has a positive impact at both the origin and destination end (except for commodity 7, fertilisers, at the origin end, but this coefficient is not significant). On top of this, GDP per capita usually has a negative effect: this confirms our hypothesis that richer countries will have a stronger focus on less material-intensive sectors and therefore require less transport. The dummies standing for trade facilities/easements usually have the expected positive influence on trade. The impact of adding random components to the model has a limited impact on the GDP elasticities (compared to the least squares model on the same data). When compared to the fixed effects model, the random effect model has the advantage that we can estimate all GDP elasticities and therefore derive the total GDP effect on trade from the model (while still taking account of destination-specific effects). The total elasticities of a rise in GDP at fixed population (sum of all GDP and GDP/capita elasticities) are shown in Table 2. These are very similar to those of the least squares model and the variation by type of commodity is quite plausible. Note that the value for fertilisers decrease, which is counterintuitive. The reason for this is an insignificant negative parameter estimate for NSTR7. The fact that it is insignificant means that it is not significantly different from zero (and thus could be any value around zero), and therefore should be fixed to zero. It should be noted that when calculating the elasticities of the model it is important to avoid double-counting for trade within countries. Hence, if we increase all GDP variables for all origins and destinations, this would imply a double-counting within countries. We have adjusted for this and the GDP elasticities reflect a general flat-rate GDP increase for all countries.

In the Heckman model some of the GDP elasticities of the demand equation (see Appendix A) are higher now than before, but many are also rather similar to the model in Tables 1 and 2 that was estimated on the positive observations only. The impact of GDP on the selection equation (trading or not) is on the other hand usually much smaller.

Product type	Elasticity
0 Agricultural prod. & live animals	0.79
1 Foodstuffs and animal fodder	0.86
2 Solid mineral fuels	0.27
3 Petroleum products	1.10
4 Ores and metal waste	0.46
5 Metal products	1.09
6 Crude and manufactured minerals	0.78
7 Fertilisers	-0.24
8 Chemicals	1.69
9 Machinery	1.45

Table 2: Elasticity of trade flow in tonnes in random effects model if the GDP increases (by 1%) and population remains constant.

During the estimation of the trade-model it has been considered whether we should apply a disaggregated version at the level of the NUTS3 zones or an aggregated version at the level of the countries. Although we have attained encouraging results using a disaggregated approach we have decided to use a two-stage estimation approach instead. In a first stage, we estimate a reference model at the aggregation level of the countries. At this stage we estimate the impact of GDP effects but also the impact of EFTA and EURO dummies on the overall trade pattern. As previously discussed, this is preferable as this reflect the level for which the trade data has been collected. In the first stage we absorb effects related to all countries. More specifically, we used the parameters for the random effect model at country level (the results presented in Table 1) and fixed these in a subsequent estimation at the zone level. In the second-stage estimation, we apply the parameters from the first stage as regard GDP and EFTA/EURO dummies in a NUTS3 version of the model where we estimate logsum parameters. It is not possible to estimate logsum parameters at the level of the countries as this virtually “destroy” the variation in the logsum variables from the logistic model. As for the first model we include fixed effects related to the countries in order to prevent the logsum parameters from explaining unwanted variation. Hence, the approach involves a double absorption in order to estimate the correct main effects at each level. The estimated logsum variables appear from Table 3 below.

All the parameters are significant and with the correct sign. However an exception is the logsum parameter for NSTR2 (Solid mineral fuels) which was found to be negative and only slightly significant. This is not surprising given the earlier results and it has been decided to fix this value to 0.

	(1) ln_tonnes_0	(2) ln_tonnes_1	(3) ln_tonnes_2	(4) ln_tonnes_3	(5) ln_tonnes_4	(6) ln_tonnes_5	(7) ln_tonnes_6	(8) ln_tonnes_7	(9) ln_tonnes_8	(10) ln_tonnes_9
ln_o_gdp	0.824	0.899	0.474	0.735	0.587	0.921	1.054	-0.002	1.123	1.175
ln_d_gdp	0.598	0.618	0.43	0.499	0.532	0.781	0.625	0.619	0.931	0.814
ln_o_gdp_cap	-0.274	-0.211	-1.051	-0.425	-0.352	-0.393	-0.556	-0.627	0.119	-0.018
ln_d_gdp_cap	-0.111	-0.094	0.422	0.316	-0.039	-0.094	-0.163	-0.39	-0.137	-0.104
cty_both_EU_EFTA	0.743	1.103	0.264	-1.204	0.357	0.435	0.357	-0.062	0.215	1.183
cty_both_EURO	0.596	-0.015	0.245	-0.298	0.562	0.124	0.423	1.191	0.324	0.181
cty_neigh	1.734	1.287	0.949	1.933	1.57	1.184	1.72	1.73	1.037	0.659
lang_same	0.742	1.02	0.899	0.968	0.903	0.671	0.777	0.106	0.945	0.963
cntry_same	2.202	2.428	3.059	2.097	3.486	1.597	2.833	3.432	1.017	1.409
	(278.871)	(341.09)	(236.361)	(390.437)	(375.926)	(224.792)	(416.866)	(322.995)	(159.475)	(224.089)
LogSum_NSTR_0	0.219									
	(168.795)									
LogSum_NSTR_1		0.223								
		(223.05)								
LogSum_NSTR_2			0							
			-							
LogSum_NSTR_3				0.04						
				(91.606)						
LogSum_NSTR_4					0.056					
					(41.201)					
LogSum_NSTR_5						0.301				
						(305.503)				
LogSum_NSTR_6							0.382			
							(373.995)			
LogSum_NSTR_7								0.397		
								(261.085)		
LogSum_NSTR_8									0.41	
									(470.569)	
LogSum_NSTR_9										0.29
										(327.156)
Intercept	-9.538	-9.504	-14.393	-12.59	-11.417	-12.199	-10.299	-0.816	-8.78	-12.604
	(-105.409)	(-118.787)	(-126.565)	(-185.125)	(-64.125)	(-135.462)	(-94.105)	(-8.148)	(-83.572)	(-161.173)
Observations	1144042	1281877	356789	1249786	636420	1005574	1150606	491506	1197438	1472537
Adjusted R^2	0.2047	0.3655	0.3094	0.4088	0.3061	0.3155	0.3341	0.3293	0.2556	0.3505

t statistics in parentheses

Table 3: Trade model with fixed effects estimated at the NUTS3 level using a 2-step procedure to include logsums from the logistic model.

The trade-model in Transtools3 represents a relative advance model in which the choice between multiple logistical chains is dealt with in a nested logit representation. The Transtools3 model is described in more details in Jensen et al. (2016) and extent previous regional and national models (Rich et al. 2011; de Jong and Ben-Akiva, 2007) to a wider geographic scope.

5. MODEL IMPLEMENTATION

In the application of the trade model, we only use the GDP and GDP per capita elasticities, assuming that distances and the dummies do not change (the model however can also be used however to calculate the trade effects of changes in the composition of the European Union, such as Brexit, or the EURO zone).

The trade model then reads in the 2010 base PC matrix and income and population changes per zone. The base PC matrix comes from ETIS+. It has the variables as described below in Table 4.

Variable	Description	Transformations
OriginEZ2006	Production zone using the NUTS3 system of 2006	Transfer to TT3 zones
DestinationEZ2006	Consumption zone using the NUTS3 system of 2006	Transfer to TT3 zones
NSTR2	Commodity type using the NST/R classification at 2 digits	Aggregation to NST/R 1 digit
Tonnes	Goods transport flow in tonnes	

Table 4: Data structure of trade matrices.

Within each scenario that is applied in runs with Transtools3, we need a percentage growth (between the future and the base year) in gross domestic product and in population for each zone (growth is the same for a zone, irrespective of whether it serves as a P or a C):

$$\%changeGDP_P = \%changeGDP_C$$

$$\%changePOP_P = \%changePOP_C$$

From these, we can also calculate the percentage growth in GDP per capita. In addition, a change in the level of service between a Production and Consumption zone (PC-pair) also influence the trade level among the zones. In order to account for such an effect, the percentagewise change in LogSums between a Production and Consumption zone.

The trade model explains %changeF, the percentage change (due to income and population change per zone) for each PC matrix cell value F_{PCg}, that can be applied together with the base PC matrix to obtain the future year PC matrix. For each NSTR commodity class g we have:

$$\%changeF_{PCg} = e^{GDP}{}_{Pg} * \%changeGDP_P + e^{GDP}{}_{Cg} * \%changeGDP_C + e^{GDPCAP}{}_{Pg} * \%changeGDPCAP_P + e^{GDPCAP}{}_{Cg} * \%changeGDPCAP_C + e^{LOGSUM}{}_{g} * \%changeLOGSUM_{CPg}$$

Where $e^{y}{}_{zg}$ denotes the estimated elasticity for changes in y (e.g. GDP) for zones at the z end (either P or C) for commodity group g (see Table 1). The elasticities for GDP and GDPCAP for the production and consumption zones can be transferred directly from the random effect model (presented in Table 1) as it is a log-log representation where parameters are essentially identical to the parameters. The elasticities for the LogSums on the other hand, cannot be transferred directly. Thus, the LogSum elasticities are computed using simulation. More specifically, we compute an increase in LN(tonnes) between zonepairs due to a 10% increase of the LogSum, and compute the elasticities for each commodity type g with respect to the base values without a 10% increase in the LogSums. Thus, the elasticity for each observation in the data is computed as:

$$e^{LOGSUM}_{CPg} = (100 * \frac{LOGSUM_{CPg} - LOGSUM_{CPg,base}}{LOGSUM_{CPg,base}})/10$$

The overall elasticity measure for each commodity type g is compute based on a weigthed average with respect to the amount of freight between zonepairs. The elasticities for GPD, GDPCAP, and LogSums are listed in Table 5.

	NSTR0	NSTR1	NSTR2	NSTR3	NSTR4	NSTR5	NSTR6	NSTR7	NSTR8	NSTR9
e^{GDP}_{Pg}	0.824	0.899	0.474	0.735	0.587	0.921	1.054	-0.002	1.123	1.175
e^{GDP}_{Cg}	0.598	0.618	0.430	0.499	0.532	0.781	0.625	0.619	0.931	0.814
e^{GDPCAP}_{Pg}	-0.274	-0.211	-1.051	-0.425	-0.352	-0.393	-0.556	-0.627	0.119	-0.018
e^{GDPCAP}_{Cg}	-0.111	-0.094	0.422	0.316	-0.039	-0.094	-0.163	-0.390	-0.137	-0.104
e^{LOGSUM}_{g}	0.533	0.141	0.000	0.079	0.110	0.740	0.209	0.322	0.174	0.164

Table 5: Elasticities for the trade model.

Obviously, the logsum elasticities are artificial in the sense that we cannot directly link these with underlying LoS variables. Hence, it is necessary to link the sensitivity of the logsum to the sensitivity to underlying LoS variables that can be interpreted. This is done below in Table 6.

Type of elasticity	NSTR0	NSTR1	NSTR2	NSTR3	NSTR4	NSTR5	NSTR6	NSTR7	NSTR8	NSTR9
LS-to-Cost elasticity										
LS-to-Time elasticity										
Inferred trade-to-Cost elasticity										
Inferred trade-to-Time elasticity										

Table 6: Elasticities between trade and LoS.

The future flow can be calculated as:

$F_{PCg}(new) = F_{PCg}(base) * (100 + \%changeF_{PCg})/100$

The output of the trade model consists of a new PC matrix (for a scenario in a future year). It contains similar variables to the base PC matrix:

The trade model is then followed by the logistics model (see Figure 1), which is a disaggregate model for the choice of transport chain, estimated on the French ECHO survey and the Swedish Commodity Flow Survey (CFS). We considered doing the application of the transport chain model on a prototypical sample of shipments. However, given the limited dependency on shipment characteristics, it is computationally much more efficient to apply the model at the level of the number of tonnes per aggregate PC flow.

For this reason we chose to apply the transport chain models to the aggregate number of tonnes per NSTR-1 category from the trade model. Having programmed the transport chain choice model, the alternative-specific constants were recalibrated to reflect the observed aggregate mode shares in Europe for the base year (as in the EU Energy and Transport in Figures Statistical Pocketbook for 2010).

The legs of the chain by mode and commodity are summed over the PC relations to produce aggregate OD matrices by mode and commodity type (in tonnes), which are then (after pivoting) used as input to the network assignment.

6. CONCLUSIONS

The paper discussed the existing literature on gravity-based trade models. It described the data and model structures used and presented the estimation results for various specifications. Elasticities for changes in GDP were provided. The paper also discussed the structure of the overall Transtools3 freight and logistics model and how PC matrices from the trade model are combined with the disaggregate transport chain choice model in model application.

Trade models that include country-specific fixed or random effects are more in line with modern economic theory, in particular with the relative costs hypothesis. Fixed effects models have the practical problem that they cannot give the full effect of an increase in GDP on trade (and that is an important reason why we need a trade model in TT3). Therefore, we prefer the random effects model.

It would be good to include the Heckman model instead of a one equation model on the positive flows only. However, it is not possible to have random effects and the Heckman specification at the same time. Linders and de Groot (2006) concluded that the Heckman model gave the best treatment of the zero flows, but that simply deleting the zero flows and estimating a model on the positive observations only (as we did in all models except the Heckman model) was acceptable. Therefore, our preferred model for implementation is the random effects model from Table 1. This is the model used in the implementation of the Trantools3 freight and logistics model. However we plan to replace this model with a trade model that will also include the influence on transport cost from the transport chain choice model, so that there will also be an influence of transport costs on the pattern of PC flows, and not only on the choice of transport chain for each given PC flow.

REFERENCES

Anderson, J.E. and E. van Wincoop (2003) Gravity with gravitas: a solution to the border puzzle, *American Economic Review*, 93-3, 171-192.

Ben-Akiva, M.E. and G.C. de Jong (2013) The aggregate-disaggregate-aggregate (ADA) freight model system, in *Freight Transport Modelling* (Eds.: M.E. Ben-Akiva, H. Meersman and E. van de Voorde), Emerald.

Bergstrand, J.H., P. Egger and M. Larch 92013) Gravity redux: estimation of gravity equation coefficients, elasticities of substitution, and general equilibrium comparative statics under asymetric bilateral trades costs, *Journal of International Economics*, 89, 110-121.

Bröcker, J., A. Korzhenevych and C. Schuermann (2010) Assessing spatial equity and efficiency impacts of transport infrastructure projects, Transportation research B, 44, 795-811.

Deardorff, A. (1998) Determinants of bilateral trade: does gravity work in a neoclassic world? in J. A. Frankel (Ed.): *The regionalization of the world economy*, University of Chicago Press, Chicago.

Eaton, J. and S. Kortum (2001) Trade in Capital Goods, *European Economic Review*, 45, 1195–1235. (2002):

Eaton, J. and S. Kortum (2001) Technology, Geography, and Trade*, Econometrica*, 70, 1741–1779.

Edwards, H. with assistance from J, Bates and H. Swahn (2008) Swedish base matrices report – estimates for 2004, methodology, data and procedures, SIKA PM 2008.

ETISplus (2014) D8 ETISplus Final Report, Panteia/NEA: Sean Newton, ISIS: Riccardo Enei, TRT:Claudia de Stasio, KIT: Eckhard Szimba, NTU: Michael Stie Laugesen, TIS: Daniela Carvalho, NTUA: Athanassios Ballis, ITC: Kristiana Chakarova, Zoetermeer.

ETISplus (2015) deliverables, Url: http://www.ETISplus.eu/deliverables/default.aspx

Gomez Herrera, E. (2010) Comparing alternative methods to estimate gravity models of bilateral trade, Department on Economic Theory, University of Granada.

Heckman, J.J. (1979) Sample selection error as a specification bias, *Econometrica*, 47, 153-161.

Ivanova, O. A. Vold and V. Jean-Hansen (2006) PINGO – A model for prediction of regional and interregional freight transport, TØI report 578/2002 (revised), Institute of Transport Economics, Oslo.

Ivanova, O., C. Heyndrickx, K. Spitaels, L. Tavasszy, W. Manshanden, M. Snelder and O. Koops (2007). RAEM version 3.0, Transport Mobility Leuven, Leuven

Ivanovo, 0. (2014) Modelling inter-regional freight demand with input-output, gravity and SCGE methodologies, in L.A. Tavasszy and G.C. de Jong (Eds.): Modelling Freight Transport, Elsevier Insights Series, Elsevier, Amsterdam.

Jensen, A.F., de Jong, G., Thorhauge, M., Rich, J., Dekker, T., Johnson, D., Cabral, M.C., Bates, J., Nielsen, O.A. (2016). A model for freight transport chain choice in Europe. *Transportation Research Part E*. Under review.

Jong, G.C. de and M.E. Ben-Akiva (2007) A micro-simulation model of shipment size and transport chain choice, *Transportation Research B*, 41, pp. 950-965.

Jong, G.C. de, O.A. Nielsen, M. Henriques, R. Tanner, D. Johnson, V. Sowa, I. Vierth and J. Bates (2015) Deliverable 7.1 - Draft report on the freight and logistics model. Seventh Framework Programme theme [SST.2010.1.3-1.], MOVE/FP7/266182/TRANSTOOLS 3, DTU Transport, Copenhagen.

Kepaptsoglou, K., M.G. Karlaftis and D. Tsamboulas (2010) The gravity model specification for modeling international trade flows and free trade agreement effects: a 10-year review of empirical studies, *The Open Economics Journal*, 3, 1-13.

Linders, G.-J.M. and H.L.F. de Groot (2006) Estimation of the gravity equation in the presence of zero flow, Tinbergen Institute Discussion Paper TI 2006-072/3, VU University Amsterdam.

Plummer, M.G., D. Cheong and S. Hamanaka (2010) Methodology for impact assessment for free trade agreements, Asian Development Bank, Mandaluyong City.

Marzano, V. and A. Papola (2004) Modeling freight demand at a national level: theoretical developments and application to Italian demand, Paper presented at the European Transport Conference 2004, Strasbourg.

Rich J., Kveiborg, O., Hansen C.O. 2011, On structural inelasticity of modal substitution in freight transport, *Journal of Transport Geography*, 19(1), 134-146.

Significance, NEA and Demis (2010) Schatting BasGoed. Rapport for DVS, Significance, The Hague.

Appendix: Estimation results based on the Heckman-model at the country level.

NST/R	(0) ln_tonnes_0	(1) ln_tonnes_1	(2) ln_tonnes_2	(3) ln_tonnes_3	(4) ln_tonnes_4	(5) ln_tonnes_5	(6) ln_tonnes_6	(7) ln_tonnes_7	(8) ln_tonnes_8	(9) ln_tonnes_9
main										
0 - 1000km	-2.971*	-2.875*	-1.768*	-3.369*	-4.139*	-2.592*	-4.237*	-2.089*	-3.003*	-2.596*
	(-12.67)	(-13.72)	(-3.64)	(-10.00)	(-11.82)	(-10.60)	(-17.40)	(-6.18)	(-16.07)	(-13.63)
1000 - 2000km	-2.393*	-2.226*	-0.374	-0.556	-1.162*	-3.156*	-2.497*	-1.547*	-2.773*	-3.079*
	(-10.19)	(-10.81)	(-0.64)	(-1.59)	(-2.87)	(-12.80)	(-9.79)	(-4.12)	(-15.06)	(-17.08)
> 2000km	-0.970*	-0.511*	-0.0223	-2.159*	-0.724*	-1.048*	-1.339*	-0.688*	-1.103*	-0.943*
	(-14.33)	(-9.76)	(-0.10)	(-16.16)	(-5.35)	(-13.54)	(-16.43)	(-5.31)	(-21.16)	(-20.07)
Ln(origin gdp)	0.897*	0.903*	0.538*	0.991*	0.760*	0.984*	1.158*	0.239*	1.130*	1.171*
	(28.96)	(39.97)	(5.20)	(19.00)	(13.90)	(31.05)	(29.75)	(3.76)	(56.82)	(70.38)
Ln(dest. gdp)	0.761*	0.720*	0.565*	0.818*	0.976*	0.901*	0.762*	0.824*	0.981*	0.856*
	(29.98)	(41.05)	(6.72)	(21.20)	(13.81)	(33.40)	(27.51)	(15.73)	(54.91)	(57.37)
Ln(orig. gdp/cap)	-0.301*	-0.180*	-1.217*	-0.483*	-0.395*	-0.261*	-0.396*	-0.455*	0.163*	0.122*
	(-9.48)	(-7.58)	(-12.05)	(-7.69)	(-6.54)	(-6.87)	(-10.08)	(-6.39)	(5.86)	(5.48)
Ln(dest. gdp/cap)	-0.0488	-0.0781*	0.371*	0.276*	-0.100	-0.110*	-0.160*	-0.371*	-0.157*	-0.0943*
	(-1.60)	(-3.13)	(3.26)	(5.53)	(-1.62)	(-3.28)	(-4.59)	(-6.40)	(-6.73)	(-4.61)
Constant	28.83*	28.57*	15.23*	32.63*	40.96*	21.23*	41.89*	19.87*	26.03*	22.24*
	(9.16)	(10.11)	(2.72)	(7.34)	(9.29)	(6.44)	(13.02)	(4.69)	(10.32)	(8.63)
select										
0 - 1000km	-1.348*	-0.911*	-0.751*	-1.801*	-1.510*	-0.991*	-1.751*	-1.470*	-1.413*	-2.426+
	(-3.89)	(-2.74)	(-4.88)	(-6.62)	(-8.42)	(-4.20)	(-6.50)	(-8.96)	(-3.84)	(-1.73)
1000 - 2000km	-1.384*	-1.094*	-1.102*	-0.917*	-1.405*	-1.766*	-1.318*	-0.843*	-1.660*	-1.050*
	(-8.18)	(-6.30)	(-8.65)	(-6.28)	(-10.54)	(-10.84)	(-8.65)	(-6.62)	(-8.42)	(-3.90)
> 2000km	-0.245*	0.0569+	-0.238*	-0.693*	-0.284*	-0.560*	-0.441*	-0.354*	-0.443*	-0.282*
	(-7.94)	(1.77)	(-6.11)	(-21.10)	(-8.11)	(-16.77)	(-13.18)	(-10.20)	(-12.57)	(-7.03)
ln(origin gdp)	0.420*	0.328*	0.259*	0.440*	0.322*	0.526*	0.576*	0.386*	0.491*	0.376*
	(38.70)	(29.97)	(20.62)	(38.45)	(27.82)	(42.12)	(44.86)	(31.70)	(38.98)	(27.42)
Ln(dest. gdp)	0.278*	0.144*	0.183*	0.240*	0.469*	0.364*	0.300*	0.301*	0.339*	0.236*
	(28.95)	(14.89)	(15.13)	(24.28)	(37.57)	(33.26)	(28.16)	(26.84)	(29.74)	(19.27)
Ln(orig. gdp/cap)	-0.208*	-0.0809*	0.0827*	0.221*	-0.0297+	0.152*	0.109*	0.196*	0.338*	0.179*
	(-15.57)	(-5.98)	(4.30)	(14.41)	(-1.86)	(10.38)	(7.20)	(11.23)	(21.99)	(11.45)
Ln(dest. gdp/cap)	0.0248+	0.0834*	0.142*	-0.00516	0.0499*	-0.0346*	0.0251+	-0.118*	-0.124*	-0.0637*
	(1.91)	(6.21)	(7.38)	(-0.37)	(2.98)	(-2.43)	(1.73)	(-7.58)	(-8.37)	(-3.83)
Both member of	0.832*	0.725*	0.0739	0.595*	0.325*	0.611*	0.595*	0.202*	1.359*	7.693
EU or EFTA	(6.95)	(5.48)	(0.96)	(6.34)	(4.06)	(5.16)	(5.89)	(2.68)	(6.10)	(0.00)
Both EURO as	0.199	0.201	0.589*	-0.295*	0.348*	0.198	0.0554	0.125	0.280	0.624
currency	(1.06)	(0.97)	(6.07)	(-2.30)	(3.06)	(1.11)	(0.38)	(1.20)	(0.76)	(0.00)
Neighbour	0.430	0.638	0.861*	0.617*	0.231	0.000327	0.674+	0.334*	-0.173	6.109
countries	(0.99)	(1.45)	(5.78)	(1.97)	(1.30)	(0.00)	(1.89)	(2.04)	(-0.48)	(0.00)
Both same	0.299*	0.268*	0.180*	0.295*	0.263*	0.155+	0.388*	0.367*	0.458*	0.153+
language	(3.92)	(3.26)	(2.24)	(3.73)	(3.60)	(1.93)	(4.84)	(4.94)	(5.23)	(1.66)
constant	11.47*	8.843+	5.977*	19.11*	12.25*	5.978+	16.00*	12.74*	13.35*	29.64
	(2.42)	(1.94)	(2.88)	(5.15)	(5.05)	(1.87)	(4.36)	(5.76)	(2.65)	(1.53)
athrho										
_cons	0.163*	0.0460	-0.0216	0.449*	0.354*	0.179*	0.199*	0.381*	0.0787*	0.109*
	(3.07)	(0.80)	(-0.14)	(7.79)	(4.83)	(4.76)	(4.12)	(5.24)	(2.30)	(2.65)
lnsigma										
_cons	0.947*	0.846*	1.179*	1.264*	1.168*	0.968*	0.942*	1.083*	0.732*	0.767*
	(99.86)	(105.69)	(61.56)	(84.99)	(62.99)	(94.15)	(83.68)	(52.31)	(82.87)	(100.09)
N	10205	10205	10205	10205	10205	10205	10205	10205	10205	10205

t statistics in parentheses
+ p<0.10, * p<0.05